BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Improving the effectiveness of financial education programs. A targeting approach

by Ginevra Buratti and Alessio D'Ignazio

April 2023

Number 765

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Improving the effectiveness of financial education programs. A targeting approach

by Ginevra Buratti and Alessio D'Ignazio

Number 765 – April 2023

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6643 (online)

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

IMPROVING THE EFFECTIVENESS OF FINANCIAL EDUCATION PROGRAMS. A TARGETING APPROACH

by Ginevra Buratti♦ and Alessio D'Ignazio ♦

Abstract

We investigate whether targeting algorithms can improve the effectiveness of financial education programs by identifying the most appropriate recipients in advance. To this end, we use micro-data from approximately 3,800 individuals who recently participated in a financial education campaign conducted in Italy. Firstly, we employ machine learning (ML) tools to devise a targeting rule that identifies the individuals who should be targeted primarily by a financial education campaign based on easily observable characteristics. Secondly, we simulate a policy scenario and show that pairing a financial education campaign with an ML-based targeting rule enhances its effectiveness. Finally, we discuss a number of conditions that must be met for ML-based targeting to be effectively implemented by policymakers.

JEL Classification: C38, I21, G5.
Keywords: financial education, machine learning, policy targeting, randomized controlled trials.
DOI: 10.32057/0.QEF.2023.0765

Contents

♦ Bank of Italy. Directorate General for Consumer Protection and Financial Education.

1 Introduction[1]

In the last years academic research and policy-makers devoted growing attention to financial literacy. The literature showed that individuals, households and entrepreneurs with low financial literacy levels make poor decisions that, in turn, have a negative impact on both their well-being and the financial welfare of the whole society (Jappelli and Padula, 2013; Lusardi, 2019). These issues have become more pressing in recent years due to the increasing complexity of the financial system. In a rapidly changing economic environment, people are indeed required to make more challenging and forward-looking financial decisions, related e.g. to ageing, more fragmented career paths and less generous pension systems. In order to tackle such issue, a growing number of countries developed dedicated national strategies, with the aim to enhance financial education through coordinated and tailored efforts. Together with the introduction of a new set of supervisory tools, several financial education campaigns are now designed and delivered in many countries (OECD, 2015, 2022).

While the effort to promote financial education is significant and costly, the evidence on the effectiveness of financial education programs is mixed (Willis, 2011; Kaiser et al., 2022a). The impact of financial education programs may be undermined by various factors, such as low participation, poor execution, suboptimal design or inappropriate target audience. In this paper, we will focus on the latter factor. According to Willis (2011), for financial education to be effective the contents of a program should be personalised to each participant. On a similar ground, OECD (2015) argues that ensuring that a financial education program is appropriate for its recipients is crucial for effectiveness. Programs that are not properly tailored to their intended audience may have unintended policy outcomes; for instance, Al-Bahrani et al. (2019) find that, on average, financial education programs undertaken in the US increased the financial knowledge gap between whites and minorities.

Tailoring financial education programs to their recipients is particularly relevant when policy-makers face budget constraints, since these programs are often costly: according to Bargagli Stoffi et al. (2021), policy-makers often want to target those who

[1]The authors would like to thank Paolo Angelini, Magda Bianco, Filippo Calabresi, Emanuele Ciani, Guido de Blasio, Riccardo De Bonis, Kristof De Witte, Tim Kaiser, Annamaria Lusardi and Daniela Marconi for useful suggestions and comments. The usual disclaimer applies. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Italy.

are in greatest need of financial education programs, particularly when facing budget constraints. Similar issues would arise for financial education programs that are characterized by a limited scalability. In such contexts, it is desirable that policy-makers identify the ideal recipients of a financial education program on the basis of their level of financial competences. Unfortunately, this approach may not always be feasible, as identifying individuals based on their level of financial competences may be difficult or impossible due to a lack of observable characteristics. In this case, policy-makers should rely on prediction tools to provide a successful program.

In this paper we investigate whether targeting algorithms - devised by using machine learning (ML) tools - could improve the effectiveness of a financial education program, by helping identify ex-ante the most appropriate recipients. In particular, our paper fits within the literature studying how ML techniques can be used to help policy-makers to increase the effectiveness of programs. Such ML tools have been applied to different fields. For instance, Chalfin et al. (2016) devise an ML algorithm to improve teachers hiring, while Kleinberg et al. (2018) focus on judges who have to decide whether or not to grant bail, by exploiting observable information about the accused. Andini et al. (2018) study the use of ML for targeting a tax bonus intended to spur consumption, while Andini et al. (2022) investigate the use of ML to provide public credit guarantees.

The application of ML targeting techniques to financial literacy programs is also starting to be explored. Bargagli Stoffi et al. (2021) exploit the 2015 survey of the OECD's Program for International Student Assessment (PISA), which included both a general and a financial literacy part, with the latter being administered only to a subset of the sample. The authors focus on Belgian students data and devise an ML algorithm able to predict - using a common set of students' characteristics - the financial literacy score of PISA students who did not participate in the financial literacy part, arguing that this could provide a useful instrument to individuate, and hence target with financial education programs, low performing students.

In this paper we extend the above streams of research. Our aim is twofold. Firstly, we devise an ML targeting rule to identify - from a set of easily observable individual characteristics - recipients who should be (primarily) targeted by an education campaign willing to improve knowledge on the following topics: income and expense tracking, children's future, pensions and insurances. Secondly, we simulate a policy scenario. In particular, our simulation exercise consists of two steps: (i) we apply the ML targeting rule to a new group of individuals (i.e., different from the sample used to devise the ML algorithm) and individuate those among them who should be offered a financial education course (ML-targeted individuals) and those who should not (not ML-targeted); (ii) we evaluate whether the effectiveness of the financial education campaign is different among these two groups of individuals.

Our paper utilizes a rich dataset collected during an Italian financial education campaign conducted by the Financial Education Committee between October and December 2021. An important feature of this campaign was its pairing with a randomized controlled trial (RCT) involving approximately 3,800 individuals who were randomly assigned to one of four groups, including three treatment groups and one control group. This RCT design provides an ideal framework for our policy scenario simulation. In the first part of the paper, we randomly divide the 3,800 individuals into two groups. Roughly three-quarters of the individuals are assigned to the training set, which we use to develop our ML targeting rule. The remaining one-quarter are assigned to the holdout set, which is not used to train the algorithm. In the second part of the paper, we apply the ML targeting rule (developed using the training set) to the individuals in the holdout set to identify those who are most likely to benefit from the policy intervention. Finally, we assess whether the use of ML targeting improves the effectiveness of the financial education campaign.

Our results provide evidence that ML-based targeting can greatly enhance the effectiveness of financial education programs. Specifically, we find that the financial education campaign was effective in improving financial literacy among individuals who were identified as targets by our ML algorithm, while no significant impact was observed among non-targeted recipients in the holdout sample. These findings are robust to different impact measures, as well as to two alternative ML predictive algorithms, and are supported by both a falsification test and rigorous robustness checks.

The remaining of the paper is organised as follows. Section 2 describes the financial education campaign that was carried out in Italy between October and December 2021 and the data. Section 3 sets out the empirical strategy. ML targeting algorithms are discussed in Section 4. Section 5 describes the policy scenario simulation, providing a description of ML-targeted individuals; moreover, the impact of ML-targeting on program effectiveness is also estimated. Section 6 concludes and sets out the policy implications.

2 Data: the financial education campaign in Italy

In October 2021 the Committee for Financial Education (established in 2017 in Italy with the aim to increase financial literacy) launched a campaign to bring financial education to the homes of Italian people. The campaign, undertaken between October and December 2021, was based on three pillars: financial education contents were delivered by means of two TV programs by Rai (the national public broadcasting company) and a short advertisement on some digital media. The two TV programs are a soap opera, "Un posto al sole" (on air Monday-Friday between 8:45 pm and 9:05 pm), and a game show, "L'Eredità" (Monday-Friday, between 6:45 pm and 8 pm), while the short advertisement consisted in a cartoon featuring Sofia, the young mascot of the campaign.

The first pillar of the campaign, i.e. "Un posto al sole", was prominent. Financial education topics treated in the soap involved: income and expense tracking; one's own children's future; pension system and savings for the elderly age. Such topics were mostly well-tailored to the stories of the protagonists, who engaged in long dialogues, discussing them carefully.[2] The game show "L'Eredità" included financial education contents in the form of multiple choice questions, asked to the participants. After each question, the TV host would quickly provide the correct answer and a short explanation of it. In eight episodes 17 questions were asked, involving various topics, from the activity of the Committee to risk diversification and pensions. Finally, the Sofia advertisement consisted in a 30 seconds message, centered on the importance of taking care of one's own finances, broadcasted on TV, radio and digital media.

Delivering financial education contents through mass media might effectively reach people that might not otherwise seek out financial education. For instance, Spader et al. (2009) show that "Nuestro Barrio", a soap opera aimed at Latino immigrants in the USA, was successful in delivering financial education contents. Mixed evidence was instead found by Coville et al. (2019), who investigate borrowing and saving decisions of Nigerian micro-entrepreneurs exposed to a movie delivering financial messages.

The Italian campaign was accompanied by a randomised controlled trial (RCT) to evaluate its effectiveness. While the impact evaluation of the full campaign is outside

[2] The episodes were broadcasted on October 25th, November 1st and 24th, December 3rd and 10th.

the scope of this paper, we rely on the existence of the RCT to test the effectiveness of ML-targeting by means of a policy scenario simulation (see Section 5). The RCT involved a sample of about 3,800 individuals, who were randomly allocated to four groups of equal size: three groups were assigned each to a different form of treatment among the soap opera "Un posto al sole" (T1), the game show "L'Eredità" (T2), and the advertisement of the mascot Sofia (T3), while the remaining one was used as control group. The sample was administered two questionnaires, before and after exposure to treatment (if required).

The ex-ante questionnaire consists of about 50 questions. The first part of the questionnaire covers socio-demographic characteristics, such as gender, age, civil and occupational status, level of education (their own and of their parents), income band, geographical area. The main body is devoted to financial literacy, that is knowledge, attitude and behaviour. Individuals have to self-evaluate their level of financial knowledge, and answer the well-known Big Three questions (Lusardi and Mitchell, 2008). They are asked how often, how easily, and from which sources, they get informed on economic and financial matters, and whether they attended (or would like to attend in the future) a financial education course or event. They have to tell how useful they consider some basics financial concepts (e.g. simple and compound interest rate, inflation, etc.) and how important it is in their opinion to take care of their own finances. They are asked whether they keep track of income and expenses, and which financial/social security/insurance instruments they own. On top of this, there are also a few questions about pension literacy and retirement planning, insurance literacy, and knowledge of the Committee, including its website www.quellocheconta.gov.it and its initiatives, namely the Financial Education Month[3]. Finally, the last part of the questionnaire collects information about the use of mass and social media. In the ex-post questionnaire, individuals are asked again a subset of the questions of the ex-ante questionnaire, namely those related to financial, pension and insurance literacy, as well as knowledge of the Committee. Variables that are observed only once (i.e. in the ex-ante questionnaire) and those observed in both questionnaires are listed in Tables 1 and 2, respectively.

Individuals assigned to the three treatment groups were administered additional

[3]The Financial Education Month is an initiative promoted by the Financial Education Committee every year, since 2018, for the whole month of October. Its calendar is filled with appointments and events for both families and businesses, with the aim to improve skills of the Italian population in matters of savings, investments, pensions and insurance.

follow-up questionnaires (one for each TV episode or digital advertisement broadcasted), meant to assess whether they actually watched the relevant episodes of the TV programs or the digital advertisements. In order to incentivize the compliance of individuals assigned to treatment, a monetary incentive was promised. In particular, those who answered correctly to two-thirds of the follow-up questions, would receive a 10 euro prize (provided as a voucher).

Slightly more than 40 per cent of people participating to the RCT are aged between 36 and 55, while a similar share of individuals has more than 55 years; the remainder of the sample consists of young (18-35 years old) people, see Table 3. Female participants count for the majority of the sample (56 per cent). Exactly half of the RCT participants hold an upper secondary school degree, while about one third hold a bachelor degree or a post-graduate degree; less than 2 per cent have a primary school degree or no formal education at all.

3 Empirical strategy

While in this paper we do not evaluate the impact of the whole financial education campaign carried out in Italy, our empirical strategy was partly shaped with the aim of exploiting the RCT to simulate the use of ML tools in a policy scenario. The impact of the whole financial education campaign, on the other hand, is currently being investigated by a separated research project (see Kaiser et al., 2022b).

3.1 Training and holdout sample

We split the about 3,800 individuals involved in the financial education campaign into two groups: a training sample - about 2,800 individuals, used to devise our ML targeting rule - and a holdout sample - about 1,000 individuals, used to simulate a policy scenario. The allocation of individuals between the two groups was performed following a quasi-randomization. In particular, individuals belonging to treatments T2 (i.e., "L'Eredità" tv game show) and T3 (i.e., Sofia advertisements) were entirely assigned to the training sample, while those belonging to treatment T1 (i.e., "Un posto al sole" soap opera, hereinafter abbreviated as UPAS) and the control group were randomly allocated, in equal proportion, both to the training and to the holdout sample (see Figure 1).

Allocating all individuals belonging to treatments T2 and T3 to the training sample was necessary in order to reach a suitable number of observations to train our ML algorithms. This choice, however, leaves us with only one policy simulation to run in the second part of the paper, that is the one involving ML targeting on treatment T1 (administered through UPAS). Our preference for treatment T1 over T2 and T3 was mainly based on the fact that the UPAS treatment was significantly more prominent during the campaign with respect to the other two (see Section 2). Moreover, the topics treated in the UPAS episodes were broader with respect to the other two forms of treatment, and largely covered by the (ex-ante and ex-post) questionnaires.

Since the allocation of individuals participating to the campaign to the four groups is random, our quasi-randomization into two groups, solely based on group allocation, also yields random samples, which only differ in size. As shown in Figure 2 and Table 8, training and holdout samples are indeed perfectly balanced across a large set of observable characteristics.

3.2 Defining target individuals

To individuate those individuals who should be (primarily) targeted by the financial education campaign we devise a binary indicator, measuring whether people lag behind in three key areas to financial well-being: track income and expenses and budget properly, plan the future, save and invest wisely. Crucially for our policy scenario simulation - performed in the second part of the paper -, such topics are all covered by the soap opera UPAS episodes. To individuate individuals in need of financial education programs we exploit the questionnaires administered before the campaign to the RCT participants. In particular, we consider as financial literacy education needy those individuals meeting at least two among the following four conditions: they do not keep track of income and expenses in any written or digital form; concerning their own children's future, they do not think that early savings are the best option; they are not aware of the current pension system in Italy; in view of their needs in the elderly age, they do not think they should start saving as early as possible.

We prefer the above described indicator to alternative measures employed in the literature to describe lack of financial literacy knowledge, such as the well-known Big three questions[4] (Lusardi and Mitchell, 2008), because the latter largely involve topics that were not covered by the UPAS episodes and hence would leave us with no room to perform a policy simulation and to evaluate the effectiveness of targeted financial education programs. As expected, however, our preferred indicator of individuals in need of financial education shows a positive and statistically significant correlation with two alternative dummies based on the Big three questions, controlling for a set of individual characteristics in the training sample (see Table 5).

3.3 Variables employed in the ML models as predictors

While the initial (pre-treatment) questionnaire consists of about 50 questions (see Section 2), we only employ a subset of them to build the predictor space of our ML algorithms. In particular, we select about half of the available variables. The reason is that we want to simulate a realistic policy scenario, where only a small set of characteristics at the individual (or local, i.e. municipality) level are usually observable and hence available to both build an ML targeting algorithm and use it for out-of-sample predictions. The questions employed to devise our ML algorithms involve some general demographic characteristics, the occupational status, the level of education and that of their parents, the income band and a bunch of variables involving the relationship with financial instruments. These variables are described in Table 6. On the other hand, some variables that could easily be observed by a policy-maker for targeting purposes (such as the province of residence) are not available in the campaign questionnaires. As a result, since more information would yield more accurate ML predictions, our exercise provides a lower bound estimate of the impact of ML targeting to improve the effectiveness of financial education programs.

Finally, it is worth remarking that all variables employed to train our ML algorithms, namely both the binary indicator defining whether an individual needs financial education or not and the full set of predictors, are obviously drawn from the initial questionnaires.

[4]The Big three questions are the following. Question 1: Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? Question 2: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account? Question 3: Please tell me whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund."

3.4 Policy simulation exercise

In the simulation exercise we apply our ML targeting algorithms to the holdout sample individuals. The simulation exercises consists of two steps. Firstly, we determine whether, according to our ML algorithms, an individual should receive a financial education course or not. Secondly, we test whether such ML targeting leads to an improvement of the effectiveness of the campaign.

4 ML algorithms

We illustrate here the ML algorithms used for our prediction exercise, and refer to Hastie et al. (2009) for more details about the algorithms. We start with a decision tree, namely a classification tree since the targeting variable we want to predict is a dummy. As compared to more elaborate models, the main advantage of a decision tree is that the classification rule has a simple graphical representation (precisely, a tree) which is very easy to read, and makes it especially suitable for policy targeting applications where transparency of the assignment mechanism is a main concern. Model interpretability, however, often comes at the price of a lower predictive performance. In particular, the tree suffers from a high variance problem: if we split the training data into two parts at random, and fit a decision tree to both halves, it is not guaranteed that we get similar results. A possible way out is to use decision trees as building blocks to construct more powerful prediction models, such as a random forest. The increased prediction accuracy, however, comes at the expense of some loss of interpretability, since a collection of trees is certainly more difficult to interpret than a single tree.

4.1 Decision tree results

The decision tree uses recursive binary splitting to create a partitioning of the predictor space into a number of distinct, non overlapping regions. To each region is associated a predicted outcome, that is given by the most frequent outcome among training sample observations falling in that region. More precisely, starting from the whole training sample, and at each subsequent step, the algorithm identifies the predictor and the corresponding threshold (or subset of values if the variable is categorical) that can be used to split the observations into two regions leading to the highest reduction in the Gini impurity index. Split after split, the tree continues to grow, and by the time the impurity index of the terminal nodes is as low as possible, the tree has typically got very large, and is likely to overfit the data. A solution is provided by cost complexity pruning, that is we use cross-validation to choose the optimal value of a complexity parameter (cp) and use it to prune the tree.

The output of the decision tree is shown in Figure 3. For the sake of illustration, we pause here to describe the upper part of the tree. From the very first split, we see that individuals without a bank account are ML-targeted. As for bank account holders, if they own at least another instrument different from the bank account, they are not ML-targeted. If instead they do not own any other instrument, we look e.g. at level of education and income band. In particular, individuals holding (at least) a bachelor degree with income above 2001 euro, are not ML-targeted, and so on. To each variable is associated a measure of variable importance, e.g. given by the overall decrease in Gini index due to splits in which that variable appears (variables may appear in the tree more than once). Figure 4 is a variable importance plot, where these quantities are scaled to sum to 100.

4.2 Random forest results

Next, we use random forest. Instead of a single decision tree, we grow many different (decorrelated) trees from many different bootstrapped training data sets. Each tree comes with its own predictions, and for a given test observation we obtain a single prediction by taking the so-called majority vote. A variable importance plot is very useful in this case, since it provides us with an indication of which variables are most important to the procedure, based on the total amount that the Gini index is decreased due to splits over a given predictor, averaged over all the trees. An alternative importance measure is based on how much the accuracy decreases when the variable is excluded. Both importance measures are shown in Figure 5.

4.3 Predictive performance and transparency

On top of the two ML algorithms (decision tree, random forest), we also estimate a linear probability model (LPM). To make these predictions comparable with ML ones, we regress over all variables listed in Table 6, and include pair-wise interactions to account for non linearities. The output of LPM is continuous, so we predict the targeting dummy to be 1 if the predicted probability is larger than 0.5. We use the holdout sample to assess and compare the predictive performances of the various models, as shown in Table 7. As expected, LPM predictive performance is outperformed by that of both the decision tree and the random forest. The latter exhibits the largest share of correctly predicted outcomes, and hence is our preferred algorithm in the policy simulation exercise.

A random forest algorithm is less interpretable than a decision tree, with the former being a sort of a black-box. With a tradeoff between interpretability and predictive performance being at work, we claim that the gains from the greater predictive performance associated to the random forest are larger than the losses linked to its lower interpretability. While, in general, the tradeoff between predictions' precision and algorithmic transparency should be assessed case by case, we argue that there is a second dimension of transparency that concerns the accountability of policy-makers to accomplish their mission best. In this respect, random forest algorithm might be preferable, because the gains in terms of effectiveness associated with ML targeting are larger.

5 Policy scenario simulation (holdout sample)

In this section we simulate a policy scenario to investigate whether ML could improve the effectiveness of financial education campaigns. We do that in two steps. Firstly, we employ our ML algorithms (devised using the training set) to individuate and describe, among the holdout sample individuals, those who should be targeted by a financial education program such as the one carried out through the soap opera UPAS. Secondly, we test the effectiveness of the UPAS program in the holdout sample, separately for both the individuals in the ML target group and for those not targeted by ML.

5.1 Characteristics of ML target vs non-target individuals

Here, we look at ML target and non-target individuals in the holdout sample, and briefly describe their characteristics with respect to some relevant variables. To this aim, we employ the decision tree described in Figure 3.[5] The composition of the two groups (i.e., ML target vs non-target) by level of education is shown in Figure 6. Target

[5]When we use the random forest we obtain almost identical evidence (not reported, available upon request).

individuals tend to have lower levels of education: the proportion of non-graduates in the target group (80%) is almost twice than in the non-target group (45%). However, the proportion of individuals with high education (bachelor degree and above) in the target group is surprisingly large (20%). Another key variable is the occupational status. As can be seen from Figure 7, the target group has a notably larger proportion of house workers, students and unemployed individuals, and a smaller proportion of permanent employees and retired people, as compared to the non-target group. Differences between the two groups are much less marked for age composition (Figure 8) and gender (Figure 9), with a slightly larger proportion of young people and women in the target group. We also look at how well the two groups perform on the Big Three questions (see Figures 10 and 11). As expected, individuals in the target group perform way worse, namely 50% of them fails at least two out of three questions, and 20% fails all three. The last variable we consider is the willingness to attend a financial education course. As shown in Figure 12, the proportion of individuals not willing to attend is much higher in the target group (55%) than in the non-target group (34%), suggesting that voluntary education may not be enough and a specific targeting is needed in order to get these people involved.

5.2 Testing the effectiveness of ML-targeted UPAS campaign

In this section we test whether ML targeting improves the effectiveness of the financial education program carried out by means of UPAS. To this aim, we focus on the holdout sample only.

In order to test for the role of ML in enhancing the effectiveness of the UPAS campaign, we firstly devise a performance indicator. To this aim, we exploit the same four financial knowledge questions that were employed to target financial literacy needy individuals, described in Section 3, and augment them with two more questions to take into account for more general effects of the UPAS campaign on the perceptions of individuals about the relevance of taking care of their own finances and of addressing in young age their needs in the elderly age (the latter topic was discussed within two distinct UPAS episodes). In particular, we exploit two questions asking, respectively, whether people think it is important to take care of their own finances and whether people have thought about their needs in their elderly age. Hence, we end up with a score ranging from 0 (all questions wrongly answered) to 6 (all questions correctly answered).

We compute this score for each individual, both before and after the campaign. Finally, we devise two alternative performance (outcome) variables: a binary indicator, taking value 1 if the score improved over the campaign, and a simple pre-post score difference.

5.2.1 Treatment effect estimates

In order to estimate whether targeting improves the effectiveness of the UPAS campaign on the holdout individuals set, we firstly have to identify the set of treated individuals (i.e. individuals who were exposed to the UPAS treatment). A key challenge arises from the possibility that individuals assigned to treatment might fail to comply. In our case, non-compliance arises when individuals assigned to the UPAS treatment do not watch the soap opera episodes with financial literacy contents or, alternatively, when individuals assigned to the control group are exposed to financial education contents. The first source of non-compliance is the most-relevant one. On the other hand, any non-compliance arising from people assigned either to "L'Eredità" or "Sofia" treatment groups is irrelevant to our analysis, since all individuals in both groups are included solely in the training set: we exploit their pre-treatment data only.

Importantly, compliance is not observable here, and, worse, it can be estimated only poorly. Indeed, to measure non-compliance, individuals assigned to treatment were administered a questionnaire in the days following the relevant UPAS episode. By providing a small monetary prize (10 euro overall, at most) in case of correctly answered questions, such questionnaires were meant to incentivize compliance and, at the same time, to measure it.

However, such procedure does not allow to precisely measure individuals' compliance to their assigned treatment: individuals could still cheat and ask someone else help to fill the follow-up questionnaire (providing a small monetary prize) in their stead, or simply guess their answers. Hence, in order to overcome these challenges and estimate the impact of UPAS in the holdout sample, we identify treated people by exploiting additional information provided in the questionnaire. In particular, people participating to the experiment were given a list of tv-shows and asked whether they usually watched them or not. Such list included the soap UPAS as well. This variable is key to us to individuate truly treated people.

Treated individuals are hence defined as those who are assigned to treatment (i.e., were asked to watch some UPAS episodes) and who claimed to usually watch UPAS in the initial questionnaire. We also provide a second, stricter definition of treatment: we consider as treated individuals those who are assigned to treatment, who usually watch UPAS and who answered correctly at least two thirds of the test questions meant to assess whether they actually watched the UPAS episodes dealing with financial literacy topics. Control group individuals are, instead, those individuals assigned to the control group and who do not usually watch UPAS. All in all, we end up with 110 treated individuals according to the baseline definition of treatment, with 70 treated individuals according to the stricter definition of treatment, and 492 control individuals.

Treated and control individuals are perfectly balanced in terms of levels of a large set of observable characteristics, measured before the treatment (see Tables 9 and 10). Although the assumption of the television habit of watching UPAS[6] being random is fundamentally not testable, the above piece of evidence showing a perfect balance between treated and control group suggests it holds. Under this assumption, a simple mean comparison is a valid treatment effect estimate.

While the second definition of treatment involves also a monetary reward being at work, we claim that the additional condition of reaching a high score in the questionnaires does not undermine the assumption of randomness in the treatment since such reward is quite small (10 euro, at most). In other words, we claim that a 10 euro prize would not change the (random) evening plans of an individual who usually watch UPAS (it is broadcasted at 8:45 pm) or not. On the other hand, an individual who was already planning to watch UPAS will pay more attention knowing that there is a reward (even if small). Once again, this assumption is corroborated by balancing statistics, reported in Tables 11 and 12.

In order to assess the impact of ML-based targeting on program effectiveness, we estimate the average treatment effect on the treated (ATT) by comparing average outcomes of treated and control individuals through OLS estimates. Formally, we estimate the following regression model:

$$y_i = \alpha + \beta \cdot Treated_i + \sum_j \gamma_j X_{ji} + \epsilon_i \tag{1}$$

Here, y_i is, alternatively, a binary indicator taking value 1 if the financial literacy score

[6] The soap opera started in 1996 and most of its about 2 million viewers watched it since then.

of individual i improved over the campaign, or a variable measuring a simple pre-post financial literacy score difference of individual i; $Treated$ is a dummy taking value 1 for treated individuals and 0 for control individuals, according to the above definition of treatment and control groups; X_j is a vector of individual characteristics, employed in the robustness exercises, including several variables such as: initial level of financial literacy, gender, age, level of education, income (see Table 9 for the full set of explanatory variables); i indexes individuals.

We estimate the model firstly on the full holdout sample and then by splitting the holdout individuals into two groups: those who were not a target according to ML, and those who were. In the sample split estimates, we provide results with respect to two alternative ML algorithms: decision tree and random forest. OLS estimation results, reported in Table 13, show that the UPAS soap was effective in the ML-target group, providing ground for the claim that ML could significantly enhance the effectiveness of financial education programs. In particular, in the target group the treatment based on UPAS increases the probability of improving financial literacy (with respect to the topics covered by UPAS) by 14-17 percentage points. Coherently with prior arguments about measuring the treatment, a larger effect of the program is detected when a stricter, more precise definition of treatment is employed.

The baseline results are robust to an alternative definition of performance. In particular, we employ a simple pre-post score difference rather than the binary indicator (assuming value one if the score improved, and zero otherwise). Moreover, treated individuals significantly improve their financial literacy score in the ML target group, regardless of the ML method (decision tree vs random forest) employed (see Table 13).

5.2.2 Falsification test and robustness

In this section we provide some robustness and falsification tests to corroborate our policy simulation exercise. Firstly, we run a falsification test, on the same holdout set, using an outcome variable built on financial literacy questions about topics that were not covered by the UPAS campaign. In this case, we expect no impact of the campaign. Secondly, we perform a robustness exercise by repeating our policy simulation on a very large number of different training and holdout samples. In this case, we expect to find results similar to those of the baseline estimates.

In order to perform our falsification test, we devise a binary indicator of performance building on questions that involve the remaining financial literacy topics, i.e. those not covered by the UPAS campaign (insurances, understanding contracts, risk-return tradeoff, knowledge of FL institutions). As in the baseline estimates, we compute both a binary indicator taking value 1 if the score relative to such questions improved over the campaign, and a simple pre-post score difference. Results, displayed in Table 15, reassuringly show that the coefficients of interest are never statistically significant.

As a first robustness exercise, we run our estimates by controlling for a large set of individual characteristics. Results, reported in Table 14, confirm the baseline findings. The same findings hold when we estimate the ATT by means of propensity score matching (Rosenbaum and Rubin, 1983).[7] As a second robustness exercise, we test whether our results remain valid when different random samples are used to both estimate the ML algorithm (training sample) and run the policy simulation (holdout sample). To perform this test we repeat our exercise 100 times, each time using a different training sample (of about 2,800 individuals) and a different holdout sample (the about 1,000 remaining individuals), both randomly selected. Figures 13 and 14 provide the distributions of the 100 ATT estimates. They confirm our baseline findings, showing that ML could significantly enhance the effectiveness of financial education programs.

6 Conclusions

Increasing the effectiveness of financial education programs is crucial for policymakers, especially when budgets or capacity are limited. Targeting algorithms based on machine learning can help policymakers identify in advance the recipients who would benefit the most from such programs. In this study, we use micro-data from a recent financial education campaign in Italy involving approximately 3,800 individuals to explore the potential policy gains of ML-based targeting.

We trained an ML algorithm on a random subset of about 2,800 individuals using a small set of easily observable characteristics. The algorithm identifies those who would benefit the most from a financial education program. We then used the algorithm to select the preferred recipients of the program among a second subset of about 1,000 individuals who were not included in the training set. Finally, we tested the effectiveness

[7]These results are not provided for the sake of brevity and are available from the authors on request.

of the financial education program using a randomized controlled trial associated with the campaign and found that it was significantly higher among the targeted individuals.

Our policy simulation shows that the financial education campaign is more effective for individuals selected as targets by the ML algorithm, while there are no effects on those in the non-target group. These results are further reinforced by a falsification test and two robustness exercises. Furthermore, it is worth noting that our ML algorithm is estimated using a relatively small sample of about 2,800 individuals with a limited number of observable characteristics. As ML tools perform well with larger datasets and variables, our policy simulation may underestimate the potential contribution of ML-based targeting in improving the effectiveness of financial education programs.

A couple of caveat also apply. The sample of individuals participating to the RCT is not representative of the Italian population (see Section 2). As a result, the ML algorithm devised to individuate those people who should be primarily targeted by a financial education campaign would under-perform if it was employed on the entire population. In order to obtain a targeting algorithm valid for the entire Italian population, a new sample - representative of the entire population - would be needed. The limited external validity of the algorithm, however, does not undermine the findings of the paper.

While we provide evidence that ML-based targeting can significantly improve the effectiveness of financial education programs, several conditions should be met for an effective policy targeting approach. Firstly, a large enough sample of individuals, representative of the population of interest and providing detailed information about both individual characteristics and financial literacy, is required to train the ML algorithm. Secondly, the policy-maker should provide full details about how the targeting algorithm was devised, maximising transparency. Finally, in order to maximise the impact of ML targeting, the sample used to devise the algorithm should contain also information about how the target individuals could be reached in the most effective way (i.e., through traditional media, online information, social media).

References

A. Al-Bahrani, J. Weathers, and D. Patel. Racial differences in the returns to financial literacy education. *Journal of Consumer Affairs*, 53(2):572–599, 2019.

M. Andini, G. de Blasio, A. D'Ignazio, and V. Salvestrini. Targeting with machine learning: An application to a tax rebate program in italy. *Journal of Economic Behavior & Organization*, 156:86–102, 2018.

M. Andini, M. Boldrini, E. Ciani, G. de Blasio, A. D'Ignazio, and A. Paladini. Machine learning in the service of policy targeting: The case of public credit guarantees. *Journal of Economic Behavior & Organization*, 198:434–475, 2022.

F. J. Bargagli Stoffi, K. De Beckker, J. E. Maldonado, and K. De Witte. Assessing sensitivity of machine learning predictions.a novel toolbox with an application to financial literacy. *mimeo*, 2021.

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

A. Coville, V. Di Maro, F. Dunsch, and S. Zottel. The nollywood nudge: An entertaining approach to saving. *World Bank Policy Research Working Paper*, (8920), 2019.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer series in statistics. Springer, 2009.

T. Jappelli and M. Padula. Investment in financial literacy and saving decisions. *Journal of Banking & Finance*, 37(8):2779–2792, 2013.

T. Kaiser, A. Lusardi, L. Menkhoff, and C. Urban. Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics*, 145(2, Part A):255–272, 2022a.

T. Kaiser, A. Lusardi, A. Sconti, and B. Zia. Evaluating the impact of a national financial education campaign delivered via television and social media: Experimental evidence from Italy. *AEA RCT Registry. January 25, 2022*, 2022b.

J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1): 237–293, 2018.

A. Lusardi. Financial literacy and the need for financial education: evidence and implications. *Swiss Journal of Economics and Statistics*, 155(1), 2019.

A. Lusardi and O. S. Mitchell. Planning and financial literacy: How do women fare? *American Economic Review*, 98(2):413–417, 2008.

OECD. National strategies for financial education. 2015. https://www.oecd.org/daf/fin/financial-education/National-Strategies-Financial-Education-Policy-Handbook.pdf.

OECD. Evaluation of national strategies for financial literacy. 2022. https://www.oecd.org/financial/education/evaluation-of-national-strategies-for-financial-literacy.htm.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

J. Spader, J. Ratcliffe, J. Montoya, and P. Skillern. The bold and the bankable: How the nuestro barrio telenovela reaches latino immigrants with financial education. *Journal of Consumer Affairs*, 43:56–79, 2009.

L. E. Willis. The financial education fallacy. *American Economic Review*, 101(3):429–34, 2011.

Figures

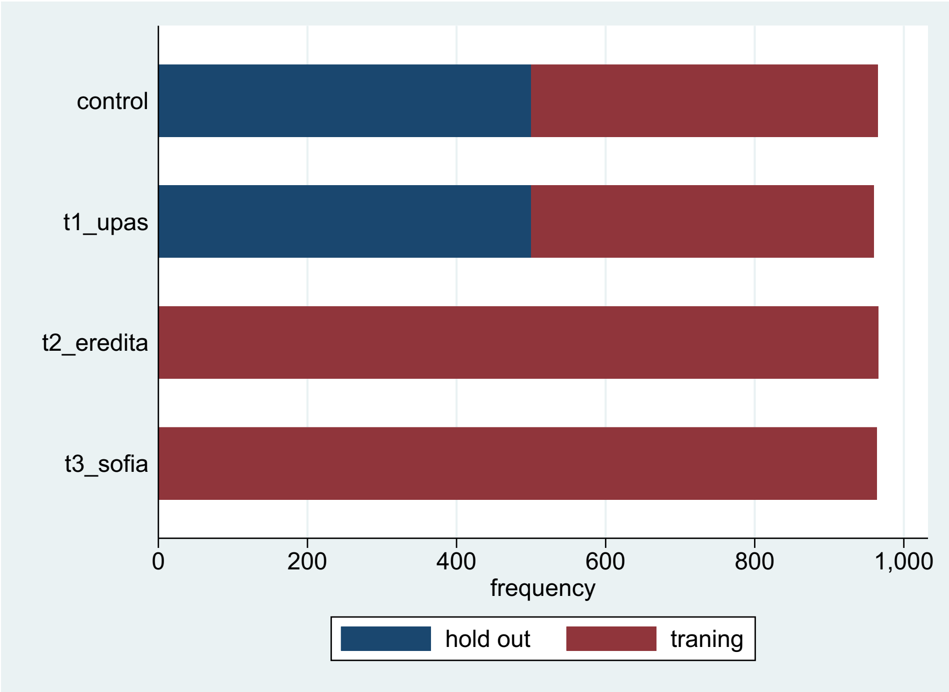

Figure 1: Training and holdout sample for the policy simulation exercise
Notes: The holdout consists of about 1,000 individuals, equally split into UPAS-treated and control.

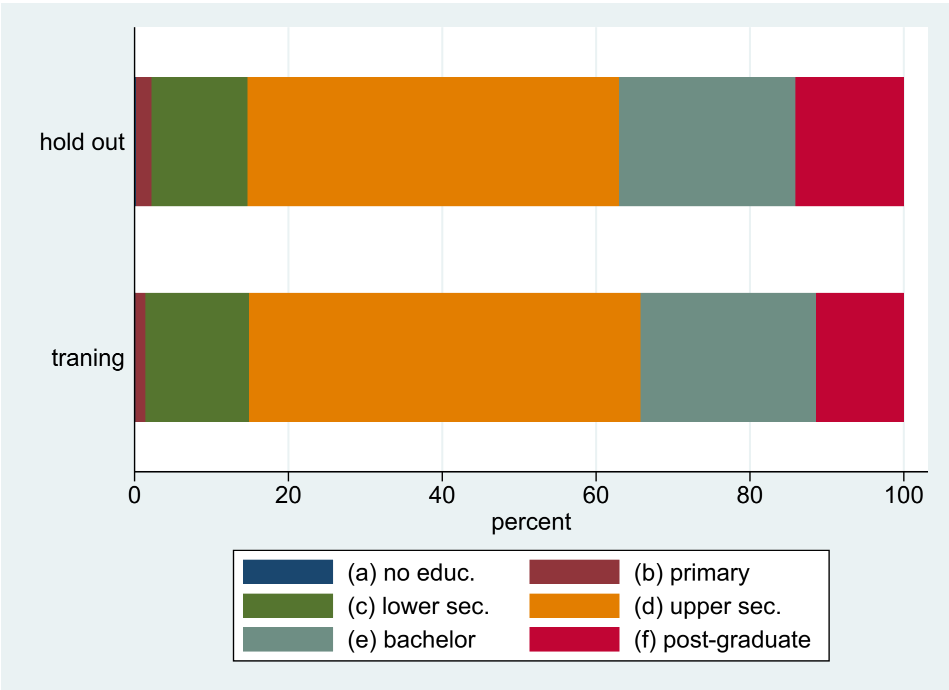

Figure 2: Training and holdout samples: share of individuals by level of education

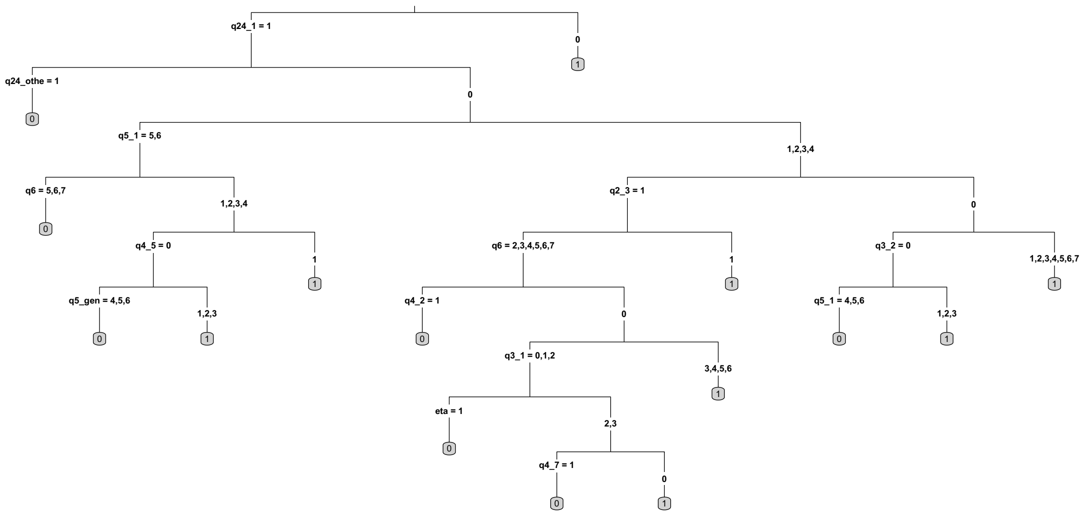

Figure 3: Pruned tree corresponding to minimal cross-validation error

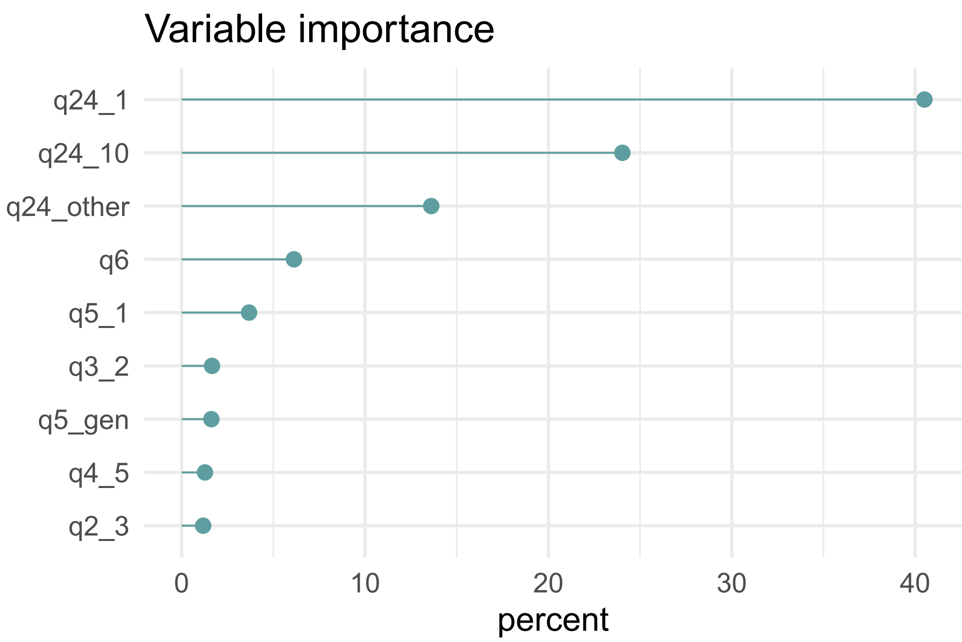

Figure 4: Variable importance plot for the decision tree
Notes: Only the nine most important variables are shown.

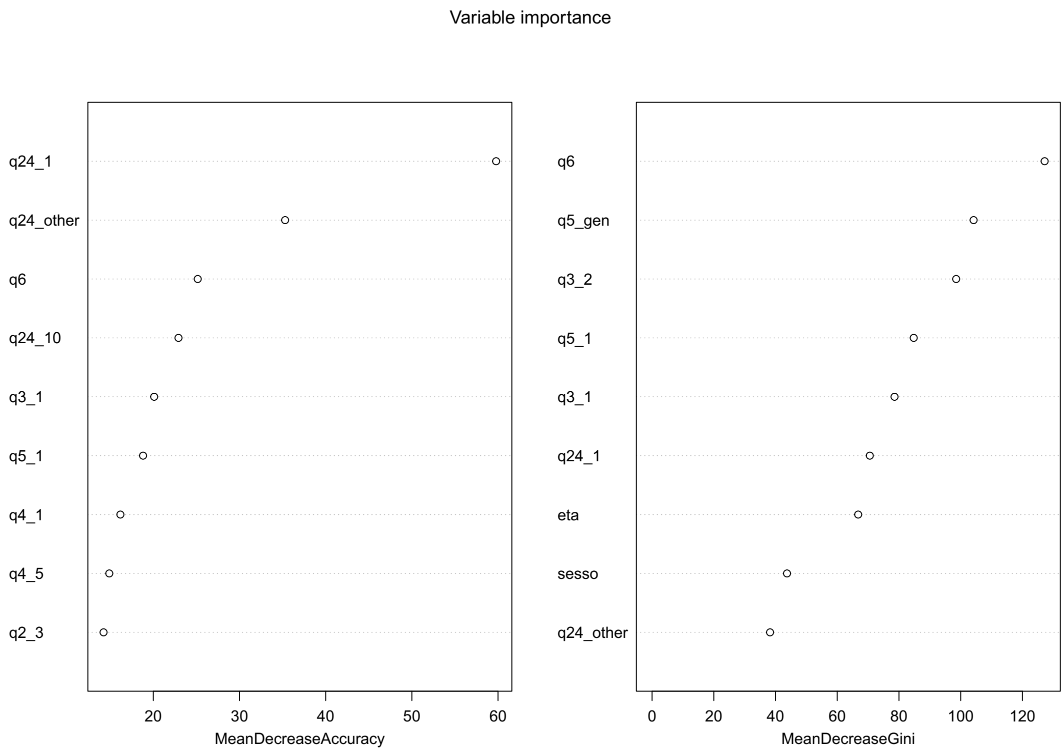

Figure 5: Variable importance plots for the random forest
Notes: Only the nine most important variables are shown. Variable importance is computed using the mean decrease in accuracy (*left*) and the mean decrease in Gini index (*right*).

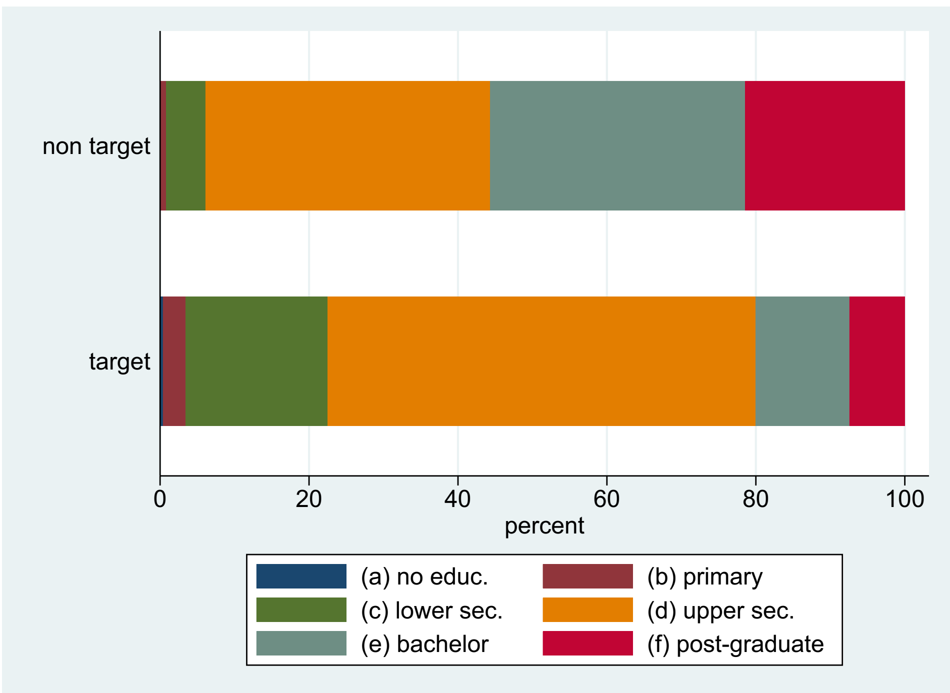

Figure 6: Characteristics of target vs non target individuals: education
Notes: In the target group the share of individuals holding a bachelor degree or post-graduate degree is significantly lower than that of the non target group.

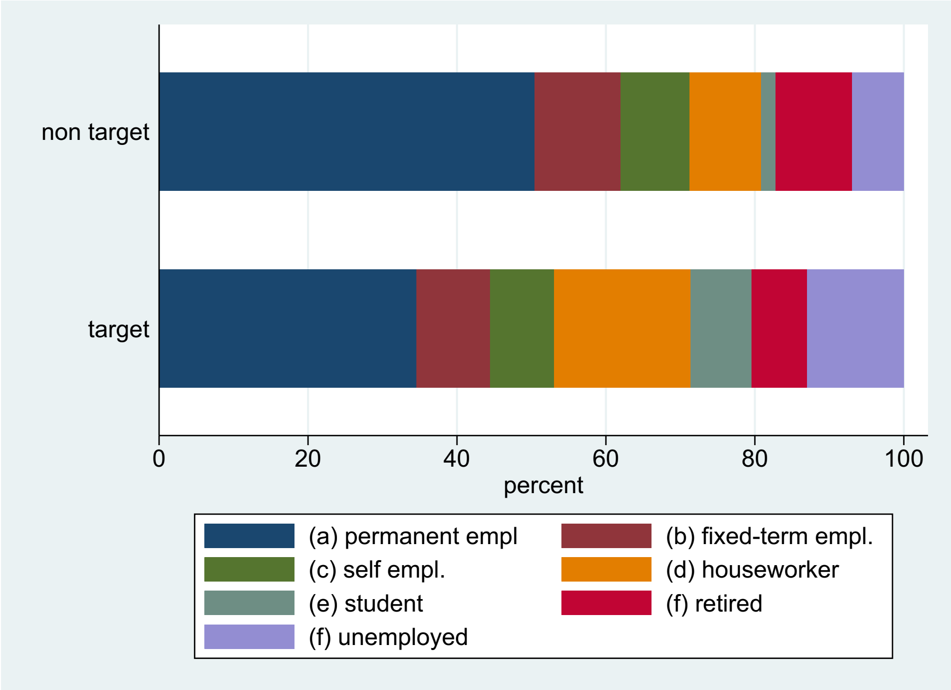

Figure 7: Characteristics of target vs non target individuals: occupational status
Notes: In the target group the share of houseworkers, students and unemployed people is significantly larger than that of the non target group.

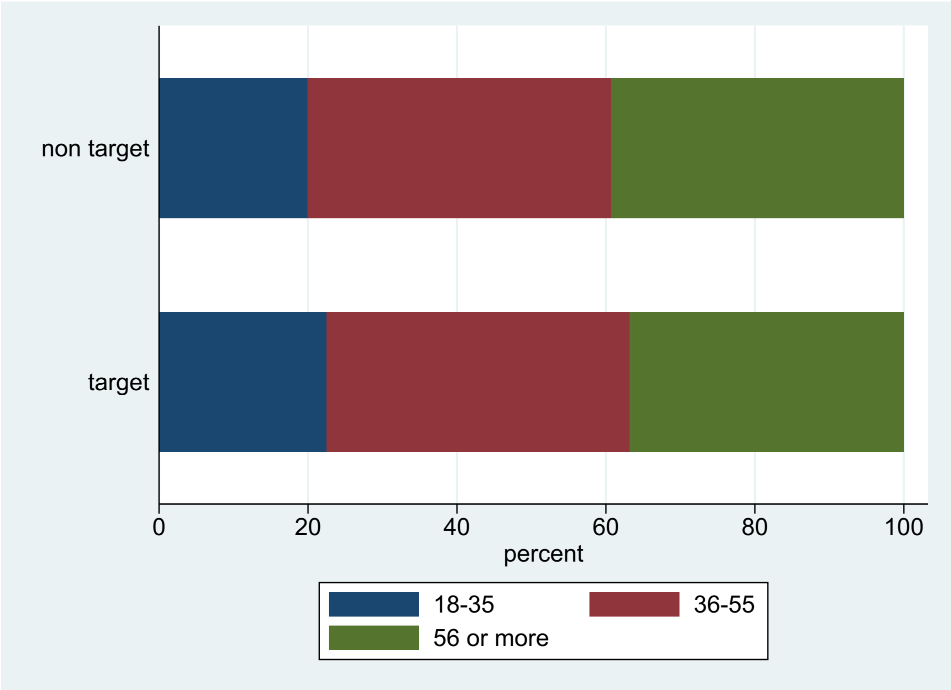

Figure 8: Characteristics of target vs non target individuals: age
Notes: The shares of the different age groups are rather similar across target and non target groups.

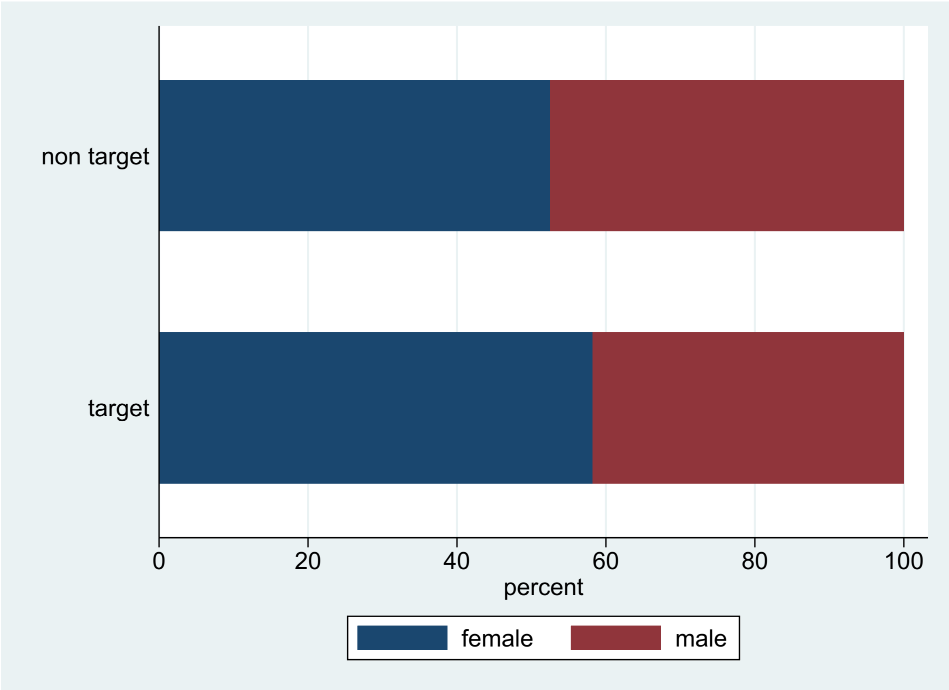

Figure 9: Characteristics of target vs non target individuals: gender
Notes: In the target group the share of female individuals is slightly larger than that of the non target group.

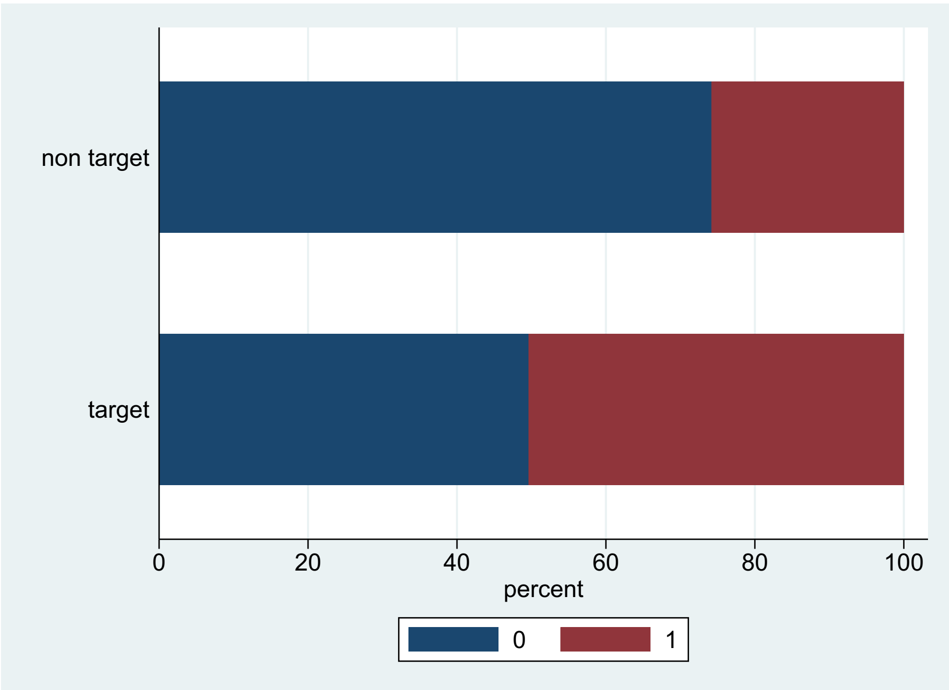

Figure 10: Characteristics of target vs non target individuals: Big Three questions (part 1)
Notes: 1 = at least two of the Big Three questions are answered incorrectly. The target group includes individuals who performs way worse with respect to the non target group.

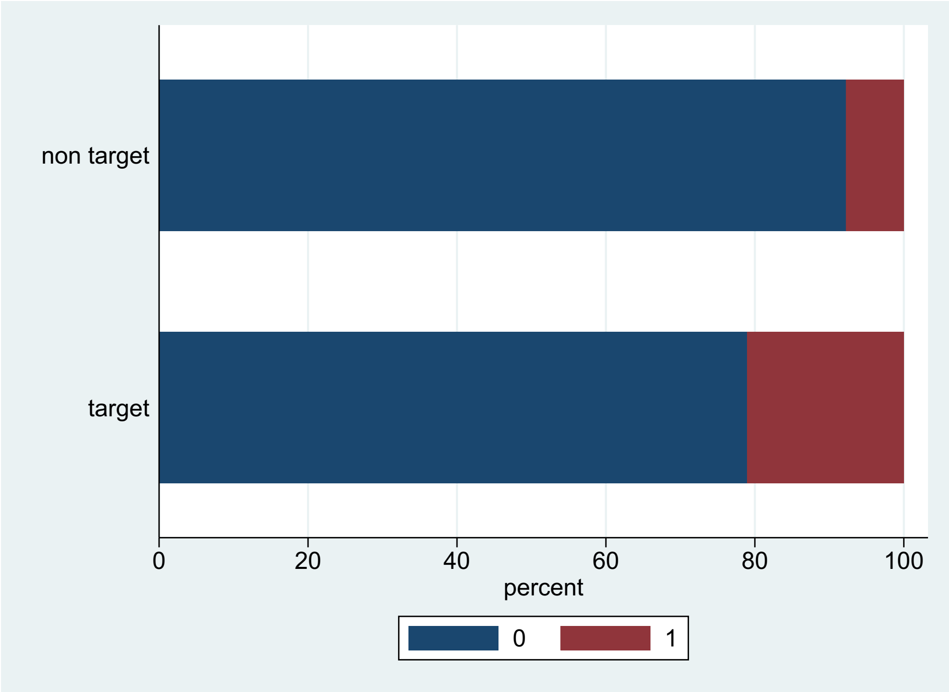

Figure 11: Characteristics of target vs non target individuals: Big Three questions (part 2)
Notes: 1 = all the Big Three questions are answered incorrectly. The target group includes individuals who performs way worse with respect to the non target group.

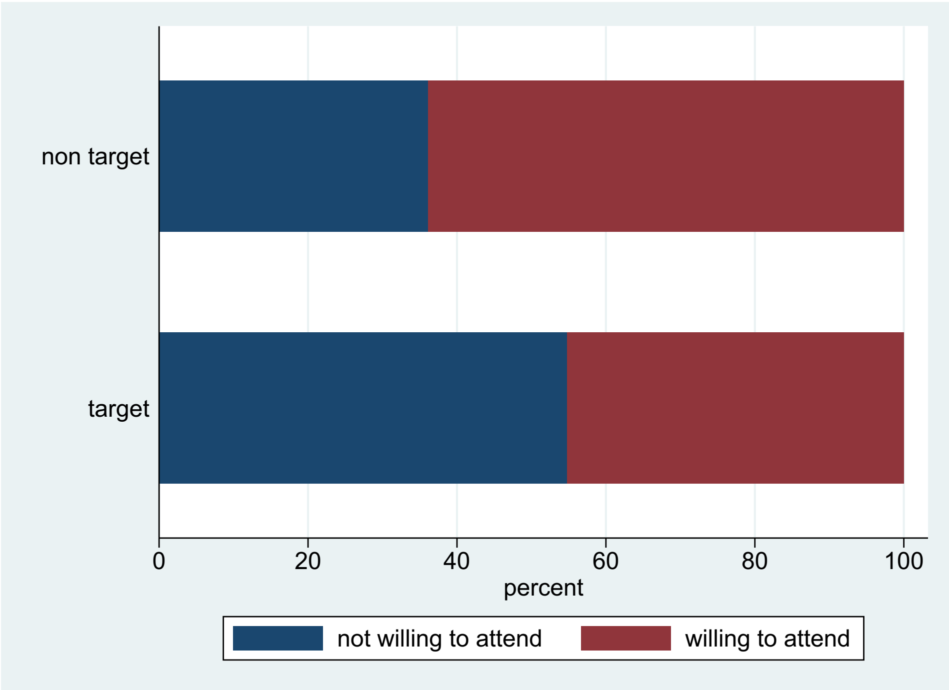

Figure 12: Characteristics of target vs non target individuals: willingness to attend a financial education course
Notes: In the target group the share of individuals not willing to attend FE courses is significantly larger than that of the non target group.

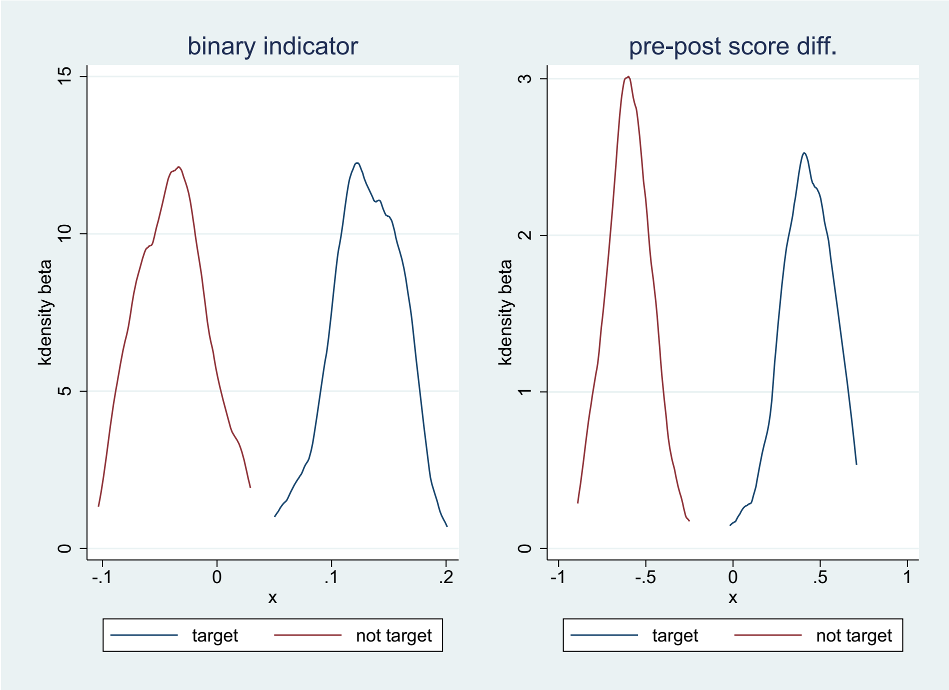

Figure 13: Treatment effect estimates (type I treated) for target and non target groups
Notes: Distribution of beta estimated on 100 random training vs holdout samples.

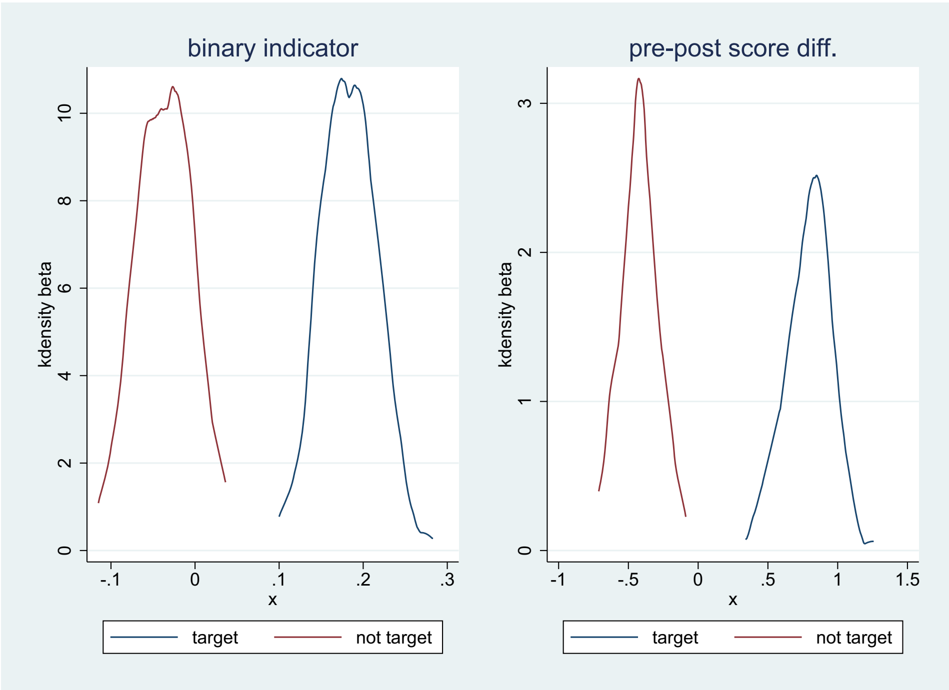

Figure 14: Treatment effect estimates (type II treated) for target and non target groups
Notes: Distribution of beta estimated on 100 random training vs holdout samples.

Tables

Table 1: Complete list and brief description of variables included in ex ante questionnaire only

sesso	Gender
eta	Age
group	RCT treatment group (T1 = UPAS, T2 = L'Eredità, T3 = Sofia, control)
q1	Binary variable indicating whether the individual is the economic decision maker of the household (or the most informed on economic/financial matters)
q2	Civil status
q3_1	Number of minors in the household
q3_2	Number of adults in the household
q4	Occupational status
q5_1	Level of education
q5_2	Mother's level of education
q5_3	Father's level of education
q6	Income band in June 2021
q7	Geographical area (North-East, North-West, Center, South and islands)
q42	Preferred sources of information on general matters: TV, radio, newspapers, the web, other
q43	Preferred sources of information on financial matters: TV, radio, newspapers, the web, other
q44	Preferred (online) sources of information on financial matters: social media, institutional websites, thematic blogs, information sites, online newspapers, podcasts, none, other
q45	Newspapers and magazines bought or read (in printed version) in the last month (multiple choice from a list)
q46	Newspapers and magazines bought or read in the last month (multiple choice from a list)
q47	Radio stations listened to in the last month (multiple choice from a list)
q48	Tv channels watched in the last month (multiple choice from a list)
q49	Frequency of social media use: TikTok, Snapchat, YouTube, Facebook, Instagram, Twitter, Linkedin
q50	Most frequently watched TV fictions (multiple choice from a list, including "Un posto al sole")
q51	Most frequently watched TV game shows (multiple choice from a list, including "L'Eredità")

Table 2: Complete list and brief description of variables included in both questionnaires

q8	Self-reported financial knowledge (1-10 score)
q9	Big Three question: compound interest rate
q10	Big Three question: inflation and purchasing power
q11	Big Three question: risk diversification
q12	Ease of finding information on economics/finance
q13	Frequency of getting information on economics/finance
q14	Having attended or thought about attending a financial education event
q15	Willingness to attend a financial education course or event in the future
q16	Month of the Financial Education Month in Italy (correct answer: October)
q17	Utility of knowing: interest rate (_1), compound interest rate (_2), inflation (_3), risk diversification (_4), risk-return relationship (_4), longevity risk (_5), insurance and risk transfer (_6)
q18	Importance of: finding out about financial issues (_1), taking care of one's own finances (_2), knowing the basics of finance before making an investment (_3)
q19	Risk appetite
q20	Keeping track of income and expenses
q21	Capacity to remedy 2000 euros whithin a month to face an unforeseen need
q22	Frequency of thinking about their future
q23	Thinking about their financial situation makes them anxious
q24_1	Owned financial/social security/insurance instruments (multiple choice from a list)
q25_1	Making ends meet
q25_2	Expenses since the start of the COVID emergency
q26	Case study: tracking income and expenses
q27	Case study: fully understanding before signing a contract
q28	Case study: one's own children's future
q29	Case study: risk-return relationship
q30	Frequency of discussing: daily shopping expenses (_1), extraordinary spending decisions (_2), saving decisions (_3), family budget (_4), economics and finance news (_5)
q31	Self-reported pension knowledge (1-10 score)
q32	Pension system in Italy
q33	Having thought about needs in the elderly age
q34	Saving for the elderly
q35	Self-reported insurance knowledge (1-10 score)
q36	Best policy for loss of self-sufficiency (correct answer: LTC policy)
q37	Overdraft in an insurance contract

Table 3: Sample characteristics

	frequencies	percentages
	panel (a): age	
18-35	817	21.2
36-55	1,591	41.3
56 or more	1,447	37.5
	panel (b): gender	
Female	2,164	56.1
Male	1,691	43.9
	panel (c): education	
No formal education	5	0.1
Primary school	57	1.5
Lower secondary school	511	13.3
Upper secondary school	1,935	50.2
Bachelor degree	881	22.9
Post-graduate degree	466	12.1
	panel (d): area	
North West	948	24.6
North East	549	14.2
Centre	946	24.5
South and Islands	1,412	36.6

Notes: Full sample of individuals participating to the october-december 2021 financial education campaign (training and holdout samples).

Table 4: Description of variables used in the regressions

score_pre	Financial education score measured before the treatment
big_3_allwrong	Dummy variable equal to 1 if all Big Three questions were answered incorrectly, and 0 otherwise
big_3_atlst2wrong	Dummy variable equal to 1 if at least two of the Big Three questions were answered incorrectly, and 0 otherwise
female	Dummy variable equal to 1 for female individuals, and 0 otherwise
older	Dummy variable equal to 1 if the individual is aged 56 or more, and 0 otherwise
graduate	Dummy variable equal to 1 if the individual holds a Bachelor degree or higher, and 0 otherwise
perm_empl	Dummy variable equal to 1 if the individual is a permanent employee in the public or private sector, and 0 otherwise
low_income	Dummy variable equal to 1 if net family monthly income was not greater than EUR 2,000.00 in June 2021, and 0 otherwise
fl_autoperc_low	Dummy variable equal to 1 if self-reported FL score is not greater than 5 (over a 1-10 score range), and 0 otherwise
married	Dummy variable equal to 1 if the individual is married, and 0 otherwise
family_memb_lt18	Dummy variable equal to 1 if all family members are aged 18 or more, and 0 otherwise
south_islands	Dummy variable equal to 1 if the individual resides in Southern Italy or in the islands, and 0 otherwise
part_fe_courses	Dummy variable equal to 1 if the individual partecipated to an FE course or is willing to participate, and 0 otherwise
make_ends_meet_diff	Dummy variable equal to 1 if the individual makes ends meet with difficulty or great difficulty, and 0 otherwise
info_web	Dummy variable equal to 1 if the individual gets informed on general matters on the web, and 0 otherwise
info_fin_web	Dummy variable equal to 1 if the individual gets informed on financial matters on the web, and 0 otherwise
newspapers	Dummy variable equal to 1 if the individual bought or read a newspaper (in printed version) in the last month, and 0 otherwise

Table 5: Training sample: financial literacy needy indicator

	(1)	(2)	(3)	(4)
big_3_allwrong	0.295***	0.200***		
	(0.0259)	(0.0261)		
big_3_atlst2wrong			0.224***	0.144***
			(0.0187)	(0.0192)
female		-0.00569		-0.0106
		(0.0182)		(0.0182)
older		-0.0288		-0.0281
		(0.0196)		(0.0196)
graduate		-0.0296		-0.0230
		(0.0198)		(0.0198)
perm_empl		-0.0236		-0.0247
		(0.0192)		(0.0192)
low_income		0.0895***		0.0777***
		(0.0206)		(0.0208)
fl_autoperc_low		0.0866***		0.0848***
		(0.0197)		(0.0197)
married		-0.0244		-0.0203
		(0.0196)		(0.0196)
family_memb_lt18		-0.0132		-0.0105
		(0.0198)		(0.0198)
south_islands		-0.00268		-0.00430
		(0.0189)		(0.0189)
part_fe_courses		-0.0549**		-0.0547**
		(0.0221)		(0.0221)
make_ends_meet_diff		0.0346*		0.0332*
		(0.0201)		(0.0201)
info_web		-0.0683***		-0.0712***
		(0.0225)		(0.0225)
info_fin_web		-0.111***		-0.107***
		(0.0223)		(0.0224)
newspapers		-0.0262		-0.0386**
		(0.0196)		(0.0195)
Constant	0.475***	0.568***	0.426***	0.548***
	(0.00999)	(0.0367)	(0.0120)	(0.0371)
Observations	2,811	2,811	2,811	2,811
R-squared	0.044	0.110	0.049	0.109

Notes: Training sample. The dependent variable is a dummy taking value 1 if individuals lack knowledge in key financial areas (income and expense tracking; own children's future; pension system; savings for the elderly age). OLS regressions. Independent variables are described in Table 2. Robust standard errors in parentheses. $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 6: Set of variables used as ML predictors

`gender`	age	`q3_1` no. children	`q3_2` no. adults
`q2_1–q2_6` civil status		`q4_1–q4_7` occupational status	
`q5_1` level of education		`q5_gen` parents' level of education	
`q6` income band		`q7_1–q7_4` area	
`q24_1` bank account		`q24_other` other financial instruments	

Table 7: Decision tree, random forest and LPM model performance compared

		Real status Non target	Target	Total
Panel A: decision tree				
	Non target	315	161	476
Predicted status	Target	201	323	524
	Total	516	484	1000
	% Correctly predicted	61,1%	66,7%	63,8%
Panel B: random forest				
	Non target	334	176	510
Predicted status	Target	182	308	490
	Total	516	484	1000
	% Correctly predicted	64,7%	63,6%	64,2%
Panel C: LPM				
	Non target	325	203	528
Predicted status	Target	191	281	472
	Total	516	484	1000
	% Correctly predicted	63,0%	58,1%	60,6%

Notes: Holdout sample. Out-of-sample performance of alternative predictive models.

Table 8: ML training sample and holdout (policy simulation) sample: balancing tests

	training		**holdout**		**diff. of means**	
	mean	sd	mean	sd	b	p
score_pre	3.043	1.545	3.086	1.548	-0.043	(0.452)
big_3_mistakes	0.149	0.356	0.149	0.356	0.000	(0.986)
female	0.558	0.497	0.551	0.498	0.007	(0.704)
older	0.379	0.485	0.384	0.487	-0.005	(0.791)
graduate	0.344	0.475	0.369	0.483	-0.025	(0.158)
perm_empl	0.413	0.492	0.419	0.494	-0.007	(0.710)
low_income	0.205	0.404	0.180	0.385	0.025	(0.089)
fl_autoperc_low	0.408	0.491	0.394	0.489	0.014	(0.454)
married	0.595	0.491	0.585	0.493	0.011	(0.562)
family_memb_lt18	0.555	0.497	0.541	0.499	0.014	(0.462)
south_islands	0.366	0.482	0.375	0.484	-0.009	(0.623)
part_fe_courses	0.266	0.442	0.301	0.459	-0.035*	(0.037)
make_ends_meet_diff	0.377	0.485	0.372	0.484	0.005	(0.784)
info_web	0.631	0.483	0.614	0.487	0.017	(0.341)
info_fin_web	0.521	0.500	0.531	0.499	-0.010	(0.598)
newspapers	0.385	0.487	0.410	0.492	-0.026	(0.156)
Observations	2811		987		3798	

Notes: Holdout sample. t-tests on the equality of means for training and holdout individuals, assuming unequal variances. Variables are described in Table 2. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 9: ML-targeted individuals in the holdout sample: type I treated, treated vs control balancing tests

	treated		**control**		**diff. of means**	
	mean	sd	mean	sd	b	p
score_pre	2.528	1.695	2.531	1.659	0.003	(0.988)
big_3_mistakes	0.167	0.375	0.199	0.400	0.033	(0.523)
female	0.625	0.488	0.551	0.498	-0.074	(0.258)
older	0.472	0.503	0.352	0.478	-0.121	(0.072)
graduate	0.167	0.375	0.180	0.385	0.013	(0.796)
perm_empl	0.375	0.488	0.309	0.463	-0.066	(0.304)
low_income	0.264	0.444	0.270	0.445	0.006	(0.924)
fl_autoperc_low	0.486	0.503	0.516	0.501	0.030	(0.661)
married	0.639	0.484	0.555	0.498	-0.084	(0.197)
family_memb_lt18	0.542	0.502	0.555	0.498	0.013	(0.846)
south_islands	0.403	0.494	0.375	0.485	-0.028	(0.673)
part_fe_courses	0.278	0.451	0.211	0.409	-0.067	(0.260)
make_ends_meet_diff	0.528	0.503	0.488	0.501	-0.039	(0.557)
info_web	0.458	0.502	0.512	0.501	0.053	(0.427)
info_fin_web	0.361	0.484	0.414	0.494	0.053	(0.416)
newspapers	0.431	0.499	0.324	0.469	-0.106	(0.108)
Observations	72		256		328	

Notes: Holdout sample, ML-targeted individuals. t-tests on the equality of means for treated and control individuals, assuming unequal variances. Variables are described in Table 2. Type I treated individuals are considered. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Not ML-targeted individuals in the holdout sample: type I treated, treated vs control balancing tests

	treated		**control**		**diff. of means**	
	mean	sd	mean	sd	b	p
score_pre	3.789	1.094	3.644	1.288	-0.145	(0.462)
big_3_mistakes	0.105	0.311	0.076	0.266	-0.029	(0.589)
female	0.553	0.504	0.534	0.500	-0.019	(0.832)
older	0.421	0.500	0.424	0.495	0.003	(0.976)
graduate	0.579	0.500	0.547	0.499	-0.032	(0.713)
perm_empl	0.474	0.506	0.479	0.501	0.005	(0.954)
low_income	0.053	0.226	0.064	0.244	0.011	(0.786)
fl_autoperc_low	0.211	0.413	0.263	0.441	0.052	(0.477)
married	0.737	0.446	0.623	0.486	-0.114	(0.155)
family_memb_lt18	0.447	0.504	0.534	0.500	0.087	(0.330)
south_islands	0.368	0.489	0.369	0.483	0.000	(0.998)
part_fe_courses	0.395	0.495	0.398	0.491	0.004	(0.967)
make_ends_meet_diff	0.211	0.413	0.25	0.434	0.039	(0.590)
info_web	0.605	0.495	0.746	0.436	0.140	(0.106)
info_fin_web	0.658	0.481	0.657	0.476	-0.001	(0.989)
newspapers	0.605	0.495	0.445	0.498	-0.160	(0.070)
Observations	38		236		274	

Notes: Holdout sample, not ML-targeted individuals. t-tests on the equality of means for treated and control individuals, assuming unequal variances. Variables are described in Table 2. Type I treated individuals are considered. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: ML-targeted individuals in the holdout sample: type II treated, treated vs control balancing tests

	treated		**control**		**diff. of means**	
	mean	sd	mean	sd	b	p
score_pre	2.283	1.772	2.531	1.659	0.249	(0.380)
big_3_mistakes	0.174	0.383	0.199	0.400	0.025	(0.684)
female	0.674	0.474	0.551	0.498	-0.123	(0.112)
older	0.413	0.498	0.352	0.478	-0.061	(0.441)
graduate	0.196	0.401	0.18	0.385	-0.016	(0.803)
perm_empl	0.283	0.455	0.309	0.463	0.026	(0.723)
low_income	0.304	0.465	0.27	0.445	-0.035	(0.640)
fl_autoperc_low	0.522	0.505	0.516	0.501	-0.006	(0.940)
married	0.674	0.474	0.555	0.498	-0.119	(0.124)
family_memb_lt18	0.500	0.506	0.555	0.498	0.055	(0.501)
south_islands	0.326	0.474	0.375	0.485	0.049	(0.523)
part_fe_courses	0.261	0.444	0.211	0.409	-0.05	(0.480)
make_ends_meet_diff	0.543	0.504	0.488	0.501	-0.055	(0.496)
info_web	0.478	0.505	0.512	0.501	0.033	(0.680)
info_fin_web	0.348	0.482	0.414	0.494	0.066	(0.395)
newspapers	0.478	0.505	0.324	0.469	-0.154	(0.059)
Observations	46		256		302	

Notes: Holdout sample, ML-targeted individuals. t-tests on the equality of means for treated and control individuals, assuming unequal variances. Variables are described in Table 2. Type II treated individuals are considered. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12: Not ML-targeted individuals in the holdout sample: type II treated, treated vs control balancing tests

	treated		**control**		**diff. of means**	
	mean	sd	mean	sd	b	p
score_pre	3.708	1.197	3.644	1.288	-0.064	(0.805)
big_3_mistakes	0.125	0.338	0.076	0.266	-0.049	(0.499)
female	0.417	0.504	0.534	0.500	0.117	(0.286)
older	0.417	0.504	0.424	0.495	0.007	(0.948)
graduate	0.625	0.495	0.547	0.499	-0.078	(0.466)
perm_empl	0.500	0.511	0.479	0.501	-0.021	(0.848)
low_income	0.042	0.204	0.064	0.244	0.022	(0.627)
fl_autoperc_low	0.292	0.464	0.263	0.441	-0.029	(0.772)
married	0.708	0.464	0.623	0.486	-0.085	(0.400)
family_memb_lt18	0.417	0.504	0.534	0.500	0.117	(0.286)
south_islands	0.542	0.509	0.369	0.483	-0.173	(0.122)
part_fe_courses	0.417	0.504	0.398	0.491	-0.018	(0.866)
make_ends_meet_diff	0.250	0.442	0.25	0.434	0.000	(1.00)
info_web	0.667	0.482	0.746	0.436	0.079	(0.446)
info_fin_web	0.708	0.464	0.657	0.476	-0.052	(0.609)
newspapers	0.542	0.509	0.445	0.498	-0.097	(0.382)
Observations	24		236		260	

Notes: Holdout sample, not ML-targeted individuals. t-tests on the equality of means for treated and control individuals, assuming unequal variances. Variables are described in Table 2. Type II treated individuals are considered. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 13: Holdout sample policy simulation: estimates.

	Full sample	ML: Decision tree		ML: Random forest	
		target	not target	target	not target
panel (a): y = binary (improvement) indicator					
Type I treated	0.0709	0.118*	-0.0970	0.135*	-0.0206
	(0.0527)	(0.0655)	(0.0833)	(0.0713)	(0.0744)
Observations	602	328	274	302	300
Type II treated	0.0917	0.177**	-0.152	0.170*	-0.00986
	(0.0638)	(0.0784)	(0.102)	(0.0868)	(0.0899)
Observations	562	302	260	280	282
panel (b): y = pre-post score difference					
Type I treated	0.0802	0.160	-0.506	0.551*	-0.545
	(0.244)	(0.312)	(0.367)	(0.333)	(0.338)
Observations	602	328	274	302	300
Type II treated	0.311	0.492	-0.482	0.901**	-0.415
	(0.295)	(0.375)	(0.444)	(0.405)	(0.405)
Observations	562	302	260	280	282

Notes: Holdout sample. OLS regressions. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 14: Holdout sample policy simulation: estimates with additional controls

	Full sample	ML: Decision tree		ML: Random forest	
		target	not target	target	not target
panel (a): y = binary (improvement) indicator					
VARIABLES		target	not target	target	not target
Type I treated	0.0777*	0.132**	-0.0373	0.137**	0.0254
	(0.0456)	(0.0534)	(0.0797)	(0.0613)	(0.0716)
Indiv. ctrls (1)	yes	yes	yes	yes	yes
Observations	602	328	274	302	300
Type II treated	0.0632	0.167**	-0.0848	0.142*	0.0252
	(0.0560)	(0.0671)	(0.0961)	(0.0751)	(0.0877)
Indiv. ctrls (1)	yes	yes	yes	yes	yes
Observations	562	302	260	280	282
panel (b): y = pre-post score difference					
Type I treated	0.141	0.314	-0.215	0.607**	-0.323
	(0.197)	(0.246)	(0.327)	(0.270)	(0.294)
Indiv. ctrls (1)	yes	yes	yes	yes	yes
Observations	602	328	274	302	300
Type II treated	0.184	0.474	-0.170	0.724**	-0.255
	(0.233)	(0.305)	(0.345)	(0.324)	(0.334)
Indiv. ctrls (1)	yes	yes	yes	yes	yes
Observations	562	302	260	280	282

Notes: Holdout sample. OLS regressions. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. (1) control variables included: see Table 9.

Table 15: Holdout sample falsification test: estimates

	Full sample	ML:		Full sample	ML:	
		target	not target		target	not target
	y = binary (improv.) indicator			y = pre-post score difference		
T. I treat.	-0.0147	-0.0142	-0.0317	0.0850	-0.0107	0.106
	(0.0522)	(0.0728)	(0.0719)	(0.235)	(0.293)	(0.365)
Obs.	602	302	300	602	302	300
T. II treat.	0.0528	0.126	-0.0424	0.429	0.578	0.190
	(0.0639)	(0.0851)	(0.0858)	(0.293)	(0.354)	(0.446)
Obs.	562	280	282	562	280	282

Notes: Holdout sample. OLS regressions. Robust standard errors in parentheses. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.