Optimal Targeting in Fundraising: A Machine Learning Approach

Tobias Cagala, Ulrich Glogowsky,
Johannes Rincke, Anthony Strittmatter*

February 12, 2021

Abstract

This paper studies optimal targeting in fundraising. In a large-scale field experiment, we randomly provide potential donors with a small unconditional gift. We then use causal machine learning methods to derive the optimal targeting of the fundraising instrument based on socio-economic characteristics, donation history, and geo-spatial information. In the warm list, optimal targeting increases the charity's profits significantly, even if the algorithm uses only the publicly available geo-spatial information. In the cold list, optimal targeting does not raise donations sufficiently to cover the costs of the gift. We conclude that without optimal targeting, charities' fundraising efforts may waste significant resources.

JEL codes: C93; D64; D82; H41; L31; C21
Keywords: Fundraising; charitable giving; gift exchange; targeting; optimal policy learning; individualized treatment rules

---

*Cagala: Deutsche Bundesbank (tobias.cagala@bundesbank.de); Glogowsky: Johannes Kepler University Linz and CESifo (ulrich.glogowsky@jku.at); Rincke: University of Erlangen-Nuremberg (johannes.rincke@fau.de); Strittmatter: CREST-ENSAE, Institut Polytechnique Paris and CESifo (anthony.strittmatter@ensae.fr). **Acknowledgments:** We have benefited from comments by Maja Adena, Philipp Dörrenberg, Michael Knaus, Andreas Haufler, Steffen Huck, Michael Lechner, Benjamin M. Marx, Michael P. Price, and Dominik Sachs. We also thank conference and seminar participants at the Causal Data Science Meeting 2020, the Annual Conference of the National Tax Association 2019, the Annual Conference of the German Economic Association 2019, the Annual Conference of the International Institute of Public Finance 2019, and the WZB Workshop on Charitable Giving in 2018. Anthony Strittmatter gratefully acknowledges financial support from the Swiss National Science Foundation (Spark Project 190422) and the French National Research Agency (LabEx Ecodec/ANR-11-LABX-0047). Furthermore, we are grateful for financial support from the Hans-Frisch-Stiftung and the Emerging Field Initiative at University of Erlangen-Nuremberg. The paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.

1 Introduction

Fundraising is a costly activity: The largest 25 US charities spend between 5% and 25% of total donations on fundraising expenses (Andreoni and Payne, 2011). These numbers are a matter of concern for two reasons: First, high fundraising costs imply that a smaller proportion of overall donations can be used to finance charitable projects. This effect can lead to an underprovision of the provided goods and services and may, thus, lower welfare if the donors' utility depends on provision levels (Rose-Ackerman, 1982; Name-Correa and Yildirim, 2013). Second, high fundraising costs also matter from the charities' perspectives: It is well documented that donors are averse to financing overhead costs (Tinkelman and Mankaney, 2007; Gneezy *et al.*, 2014). Hence, charities with excessive fundraising expenses will be less successful in raising donations. In conclusion, reducing disproportional fundraising costs can be crucial, both from a welfare and a charity-management perspective. However, while there is a broad literature evaluating fundraising instruments such as matching grants and fundraising gifts (surveyed by Andreoni and Payne, 2013), previous research has paid less attention to how charities could increase the fundraising cost efficacy.

This paper shifts focus to a novel approach to increase a fundraising campaigns' efficacy: optimal targeting of fundraising activities based on causal machine learning (see Athey and Imbens, 2019, for a review of the causal-machine-learning methods). Optimal targeting exploits that, due to heterogeneity in donors' preferences or characteristics, the effects of any fundraising campaign are likely heterogeneous across individuals.[1] Charities that ignore this heterogeneity likely waste resources, for example, by directing their fundraising activities to notorious non-donors or donors who do not increase their donation enough in response to the fundraising instrument to result in a net gain for the charity. This paper's key contribution is to show that (a) causal machine learning techniques offer a powerful toolkit for optimal targeting, which (b) allows charities to increase their campaigns' profits and, hence, service and goods provision.

We exemplify the potential of optimal targeting using a field experiment and studying a widely used fundraising instrument: small unconditional gifts that accompany a solicitation letter. Particularly, we cooperated with a charity and randomly assigned almost 20,000 potential donors to a gift treatment (in which potential donors received a solicitation letter with a gift) and a control group (in which donors only received the letter). In theory, such unconditional gifts should increase donations by triggering a reciprocal reaction of potential donors (Falk, 2007). In practice, the evidence on this tool's effectiveness is, however, mixed. Some studies buttress that unconditional gifts are an effective fundraising instrument, while others find that they backfire and even

[1] For example, donation motives (such as altruism, warm glow, and reciprocity) are heterogeneously distributed across individuals (Falk *et al.*, 2018). Consequently, responses to fundraising activities that leverage (some of) these motives are likely heterogeneous as well.

lower donations (see literature section). This type of effect heterogeneity might very well operate through individual heterogeneities (such as heterogeneous social preferences). Thus, fundraising gifts look like a promising context to study the benefits of optimal targeting.

The challenge of optimal targeting lies in directing fundraising instruments (such as gifts) to only a subset of individuals to maximize a fundraising campaign's expected profits. Yet, who should charities target for that purpose? Intuitively, the profit-maximizing targets are the so-called net donors: individuals whose additional donations are higher than the marginal fundraising cost. In an ideal world, the charity would know each individual's donation with and without fundraising, such that it could trace out the set of net donors. In reality, this set is, however, unknown to the charity. The reason is that donations under both conditions (gift vs. no gift) are unknown before the campaign. Moreover, after the campaign, the charity could only learn a donor's behavior under her assigned condition. These observations set the stage for our optimal-targeting machine-learning approach. The algorithm learns the relationship between the individuals' characteristics and their donation behaviors in the treatment and the control group of the experiment (i.e., with and without the gift). From this relationship, we then (out-of-the-sample) predict who will or will not be a net donor. This procedure results in a transparent rule, specifying the set of predicted net donors who should be targeted. Importantly, although we cannot trace out the causal effect of the characteristics that drive the heterogeneity, the policy-relevant increase in net donations achieved by the targeting rule is causally identified.

More specifically, we implement the optimal-policy-learning algorithm of Athey and Wager (2021), which extends the empirical welfare maximization approach of Kitagawa and Tetenov (2018) with machine learning. In particular, in our main specifications, we estimate optimal targeting rules with the exact policy learning tree of Zhou *et al.* (2018) and show the sensitivity of our results to various alternative estimators (Logit, Logit-Lasso, CART, classification forest). Regarding data, we feed our algorithm with data from various sources and of different types, including socio-economic characteristics, donation history, and geo-spatial information. The geo-spatial data comprises publicly available information from Google maps on economic and cultural activities close to the potential donor's residence.

Our analysis yields the following main results. In the warm list of previous donors, optimal targeting increases the charity's profits significantly compared to two benchmarks scenarios without targeting: one in which everyone receives the gift (increase: 13.8%) and one in which nobody receives the gift (14.3%). The positive effect of targeting persists over two years, which speaks against pull-forward effects. Importantly, charities can realize these gains by relying upon readily available information: data on whether an individual donated in the past or not and publicly available geo-spatial data linked to each donor's residence. Instead, socio-economic characteristics, which can be hard to obtain for charities, are redundant under the optimal targeting rule. In contrast to the warm list, targeting in the cold list (i.e., the sample of previous non-givers) has no bite relative to the benchmark that nobody receives the gift. Moreover, even the optimal targeting rule does not broaden the donor base enough to justify the gift's additional fundraising costs. While the details of our findings may well be specific to our context, our study's general conclusion is that charities that do not optimally target their fundraising efforts may waste significant resources.

The paper's organization is as follows. Section 2 outlines our contributions to the literature, and Section 3 describes the institutional background and design of the experiment. In Section 4, we describe our empirical strategy. Section 5 discusses our results, and Section 6 concludes. Online Appendices A-E provide supplementary materials.

2 Contributions to Literature

This study relates and contributes to several literature strands in economics and marketing. The first relevant strand is the *theoretical fundraising literature in public economics*, which provides a theoretical underpinning for why fundraising targeting can be beneficial. The argument is as follows. Many charities have properties similar to privately-provided public goods (Andreoni and Payne, 2013): contributions are voluntary, and the provided goods are non-excludable and non-rivalrous. In such a context, fundraising tools can counteract free-riding and, hence, underprovision problems (see e.g., Andreoni, 1988; Morgan, 2000; Vesterlund, 2003; Andreoni and Payne, 2003, 2013).[2] However, if fundraising is costly and charities need to compete for donors, competition can push the costs to such high levels that the total service provision falls (Rose-Ackerman, 1982; Aldashev and Verdier, 2010; Aldashev *et al.*, 2014). Targeting of fundraising instruments can then be a tool to maximize donations net of fundraising costs and, thus, provision levels. Specifically, Name-Correa and Yildirim (2013) show in an elegant model that a charity's optimal strategy is to target net donors with fundraising (i.e., donors expected to give more than their solicitation costs). Everyone outside this set would be a "net-free-rider" and should not be targeted. Building on these theoretical arguments, we explore how charities can achieve a data-driven optimal targeting rule in practice.

The second literature strand is especially closely related to our paper: It is a small, but emerging economic literature on *targeting fundraising among heterogeneous donors*.

[2]Which tool is suited to increase provision depends on the context. Solicitation letters (perhaps providing return envelopes) counteract the underprovision problem in settings with transaction costs (Andreoni and Payne, 2003). By contrast, leadership gifts oppose the underprovision of threshold public goods (Andreoni, 1988) and public goods under imperfect information (Vesterlund, 2003). Lotteries, instead, can increase efficiency in standard public goods settings (Morgan, 2000).

To date, there are only two papers. First, Adena and Huck (2019) apply a targeting strategy to "matching gifts," a fundraising tool where pre-campaign collected funds top up donations above a threshold. Their main result is that charities can crowd in donations by conditioning these thresholds on past giving behavior (i.e., the thresholds are targeted). Second, Drouvelis and Marx (2021) focus on belief-based targeting. They conclude that charities can increase net donations by targeting information treatments to potential donors who hold incorrect (low) beliefs about others' donations. Our paper differs from these studies in important dimensions: Adena and Huck (2019) and Drouvelis and Marx (2021) both build on conceptual considerations to identify dimensions of heterogeneity used for targeting. In contrast, our contribution is to take a more agnostic but also more flexible data-driven approach (based on Athey and Wager, 2021). Particularly, our machine-learning algorithms identify the most crucial dimensions of heterogeneity based on input features, which can flexibly approximate preferences or other drivers of response heterogeneity. On top of that, instead of focusing on debiasing or threshold matching, we consider a different tool: fundraising gifts.

Naturally, the third strand of relevance is, hence, the *literature on fundraising gifts* (see, e.g., the review of List and Price, 2012). While Falk (2007) documents that gifts are a cost-effective tool to increase donations in the warm list, other studies paint a more scattered picture: For example, Landry *et al.* (2010) find zero effects of gifts in the warm list, and Alpizar *et al.* (2008) conclude that conditional on giving, gifts even lower contributions. Yin *et al.* (2020) present similar results.[3] We add to the looming discussion on the causes of effect heterogeneity by showing that individual heterogeneity alone is powerful enough to account for negative and positive effects of gifts.[4] In particular, in our setting, some groups of potential donors increase donations in response to gifts, and others lower them heavily. This finding persists if we restrict our sample to the warm list only. While these findings are insightful in themselves, our main contribution to the literature on gifts is, however, different: We demonstrate that charities can effectively exploit heterogeneities in responses for optimal gift targeting.

By focusing on effect heterogeneity, we further contribute to a fourth literature strand on *heterogeneous responses to fundraising*. This literature identifies heterogeneity to (a) characteristics of the charitable organization and the purpose of the charity (e.g., Okten and Weisbrod, 2000; de Vries *et al.*, 2015), (b) characteristics of the donors (e.g., Andreoni *et al.*, 2003; Andreoni and Vesterlund, 2001), (c) donation motives or preferences of donors (e.g., Harbaugh *et al.*, 2007), and (d) past donation behavior (Schlegelmilch and Diamantopoulos, 1997; Hassell and Monson, 2014), and (e) crowding out (Meer,

[3] Thank-you gifts that are handed out after a donation have also been found to lower donations (Newman and Shen, 2012). Eckel *et al.* (2018) directly compare unconditional and thank-you gifts and show that donors are twice as likely to give when they receive an unconditional, high-quality gift.

[4] Of course, differences in the overall setting (such as varying causes of charities) might also explain why different studies come to varying conclusions.

2017). Instead of studying single, selected dimensions of heterogeneity, we combine a range of individual characteristics and past donation behavior. On top of that, our paper adds an additional, rarely-used determinant of heterogeneity to the analysis: geo-spatial characteristics. As this information is easily accessible by charities, it is particularly beneficial in the context of optimal targeting.[5]

Setting the literature in economics aside, our paper has natural connections to the *marketing literature* as well. Kotler and Levy (1969) proposed to "broaden the concept of marketing [with the goal] to combat [organizations] lack of funds." In this spirit, several studies started to apply concepts of marketing to the fundraising context (e.g., Mindak and Bybee, 1971; Schlegelmilch, 1988; Shapiro, 1973). Other studies focused more directly on campaign targeting. These studies' standard approach is to profile or segment donors based on different socio-economic characteristics and previous donation history.[6] In this vein, one of the suggestions is to target fundraising instruments to those with a high baseline donation probability (e.g., Srnka *et al.*, 2003). There is, however, no guarantee that such an approach *per se* maximizes a campaign's profits. For example, Ascarza (2018) reports that a similar strategy backfires in a different context: customer churn. Motivated by this observation, our contribution to the charitable marketing literature is to employ a method that maximizes profits more directly. Particularly, our optimal-policy-learning approach targets gifts to those individuals for whom we predict a positive effect on net donations. To sum up, by using machine learning to identify net donors, we provide a novel way to answer a previously asked, not fully answered, and still vital question in charitable marketing.

Methodologically, we contribute to the small but rapidly growing literature that *applies machine learning methods to target public and private policies* (e.g., Andini *et al.*, 2018; Hitsch and Misra, 2018; Kang *et al.*, 2013; Knaus *et al.*, 2020a; Knittel and Stolper, 2019; Rockoff *et al.*, 2011; Kleinberg *et al.*, 2015). While these study consider contexts such as taxation or labor-market programs, our study is the first that applies optimal-policy-learning algorithms (such as the one of Athey and Wager, 2021) to the context of charitable giving.

[5] Dong *et al.* (2019) and Glaeser *et al.* (2018, 2020) show that geo-spatial characteristics are good proxies for income and other socio-economic characteristics.

[6] For examples, see Bekkers and Wiepking (2011), Casale and Baumann (2015), Farrokhvar *et al.* (2018), Kottasz (2004), Rajan *et al.* (2009), Ranganathan and Henley (2008), Sargeant and Lee (2002), Sargeant and Woodliffe (2007), Schlegelmilch and Diamantopoulos (1997), Shelley and Polonsky (1971), Bekkers and Wiepking (2012), and Wiepking and James (2013).

3 Experimental Design and Data

3.1 The Natural Field Experiment

We implemented a natural field experiment in cooperation with a fundraiser operating within the structure of the Catholic church in an urban area in Germany. The fundraiser sends out annual solicitation letters (per mail) to the population of all members of the Catholic church living in the urban area. All Church members residing in the urban area receive the solicitation letter, irrespective of previous donations. The solicitation letter highlights the fundraiser's cause, which is to support the local church parishes in maintaining their clergy houses, parish centers, and churches, and asks recipients for a donation. Attached to the letter is a bank transfer form pre-filled with the fundraiser's bank account information and the recipient's name. Donations are made exclusively via bank transfer, and the fundraiser does not provide any information about individual donations to the church parishes. Donations are thus made in a private information setting.

Figure 1: The gifts consists of folded cards and envelopes

Notes: This figure represents the gifts. It consists of three different folded cards showing flower motives from paintings of Albrecht Dürer plus three envelopes.

The field experiment was conducted in 2014. All experimental participants received a letter with information about the fundraiser's cause (maintaining clergy houses, parish centers, and churches) and a donation request. A randomly selected treatment group additionally received a small unconditional gift. The gift consisted of three different postcards showing flower motives, together with three envelopes (see Figure 1).[7] The

[7]The specification of the gift relates our design to Falk (2007), who also used postcards and envelopes. The additional cost for the fundraiser to send out a gift letter (relative to sending out a standard letter

letter in the gift treatment group was identical to the control group letter apart from the fact that recipients were informed in a short remark following the two main paragraphs that the included postcards were a gift.

The sampling frame comprised 2,354 individuals from the warm list (individuals who had donated at least once to the fundraiser pre-treatment) and 17,425 individuals from the cold list (people who had never donated to the fundraiser pre-treatment). Based on a stratified randomization scheme, we assigned in the warm [cold] list 1,180 [2,283] individuals to treatment, and 1,174 [15,142] to control.[8]

3.2 Data Description

The data contains socio-economic characteristics obtained from administrative records of the Catholic church (such as gender, marital status, and age). Furthermore, the data contains detailed information about individual donations to the fundraiser via bank transfers for the years 2006–2015. Accordingly, we can measure the donations during the first and second year after we conducted the experiment as well as the donation history for the eight previous years.

Furthermore, the data contains the postal address of each individual. We use the address to retrieve geo-spatial information about economic and cultural activities from Google maps API. In particular, we collect the number of restaurants, supermarkets, medical facilities, cultural facilities, and churches within 300 meters proximity around the home address. We web-scrape the distance from the home address to the main train station, city hall, main church, and airport. Furthermore, we retrieve the elevation of the home address.

Tables 1 shows the descriptive statistics for the donation amount and probability in the warm and cold list, respectively. During the first year after the experiment, the average donation is 16.02 Euro (maximum 450 Euro) in the warm list, compared to 0.18 Euro (maximum 200 Euro) in the cold list. The donation amount is highly right skewed and has excess kurtosis, suggesting there are a few extraordinary large donations. In the warm list, 49% of the individuals make a donation in the first year after treatment. In the cold list, only 1% of individuals donate.

A similar picture emerges when we consider the donation amount and probability during the first two years after the experiment. The average donation amount is 30.48 Euro in the warm list and 0.43 Euro in the cold list. The share of donors is 57% in the

without gift) was 1.16 Euro. This comprises 47 Euro-Cents for the postcards and envelopes, and 69 Euro-Cents additional costs for preparing the mailings and higher postage.

[8]The strata were defined based on list (warm vs. cold), gender, household type indicators, quintiles of individuals' predicted baseline willingness to give, and quintiles of age. To construct a proxy for the baseline willingness to give in the treatment year, we first regressed an indicator variable for giving in the year prior to the experiment on indicator variables for further lags of the giving indicator. We then used the estimated model to predict the probability of giving in the treatment year (out-of-sample).

Table 1: Descriptive statistics of donation amount and donation probability

	Mean	Std. dev.	Skew.	Kurt.	Min.	Max.
	(1)	(2)	(3)	(4	(5)	(6)
Panel A: Warm list						
1st year after the experiment:						
Donation amount (in Euro)	16.02	30.38	4.70	39.05	0	450
Donation dummy	0.49				0	1
1st and 2nd year after the experiment:						
Donation amount (in Euro)	30.48	53.19	4.96	49.23	0	900
Donation dummy	0.57				0	1
Panel B: Cold list						
1st year after the experiment:						
Donation amount (in Euro)	0.18	3.00	38.39	2'049.4	0	200
Donation dummy	0.009				0	1
1st and 2nd year after the experiment:						
Donation amount (in Euro)	0.43	4.85	23.12	779.6	0	240
Donation dummy	0.017				0	1

Notes: This table shows descriptive statistics for our main outcomes: (a) the donation amount and (b) dummies indicating whether a person donated or not. We consider the warm list (Panel A) and cold list (Panel B) separately. Further, we track our outcomes over two periods: the first year after and the first two years after the experiment. For dummies the first moment is sufficient to infer the entire distribution.

warm list, and 2% in the cold list.

Table A.1 in the Online Appendix A shows the means and standard deviations of the observable characteristics of prospective donors. The descriptive statistics differ substantially between the warm and cold list. On average, cold list individuals are younger, have a higher likelihood of being singles, and tend to have a shorter residency duration in the urban area. Individuals in the warm list made on average four donations with a total donation amount of 126 Euro over the course of the last eight years before the experiment. By construction, cold list individuals have a donation history of zero donations. Individuals in the cold list live on average closer to the city centre (closer to city hall and main station) than individuals in the warm list. In the proximity of the home address (300 meters radius), cold list individuals have on average more access to restaurants, supermarkets, medical and cultural facilities, and churches than warm list individuals.

4 Empirical Strategy

The empirical strategy is based on three steps. First, we randomly allocate the fundraising instrument, which we described already in the last section. Second, we use machine learning algorithms to estimate the optimal targeting rule in a random subsample of the experimental data and retain the remaining data. Third, we extrapolate the optimal targeting rule to the retained data and apply off-policy-learning techniques to assess the out-of-sample performance of the optimal targeting rule.

4.1 Conditional Average Treatment Effects

In order to introduce targeting, we begin by discussing Conditional Average Treatment Effects (CATEs). CATEs identify heterogeneous treatment effects based on observable characteristics. Therefore, they are a natural starting point before the introduction of targeting rules in the next section.

We use the potential outcome framework (Rubin, 1974) to describe the parameters of interest. The potential outcomes framework is particularly useful to study targeting, because it allows us to describe an individual's reaction under different (counterfactual) treatment conditions. The treatment variable D_i indicates whether a fundraising gift was sent to individual i (for $i = 1, ..., N$) or not, with

$$D_i = \begin{cases} 1 & \text{when a gift was sent, and} \\ -1 & \text{otherwise.} \end{cases}$$

$Y_i(1)$ denotes the potential donations in response to the mailer with a fundraising gift. $Y_i(-1)$ denotes the potential donations in response to the mailer without a fundraising gift. Under the stable unit treatment value assumption (SUTVA), the observed donations are

$$Y_i = Y_i(-1) + \frac{1 + D_i}{2} (Y_i(1) - Y_i(-1)).$$

The individual causal effects are

$$\delta_i = Y_i(1) - Y_i(-1).$$

In an ideal world, a charity would know δ_i. It could then (exclusively) assign the gift to individuals for whom δ_i exceeds the costs of the gift. However, the fundamental problem of causal analysis is that δ_i is unobservable, because either $Y_i(1)$ or $Y_i(-1)$ is unobservable. Nevertheless, it is possible to identify and estimate group averages of δ_i. For example, the average treatment effect (ATE), $\delta = E[\delta_i] = E[Y_i(1) - Y_i(-1)]$, is the expected average donation effect of the fundraising gift. Possibly, the effect of the fundraising gift is heterogeneous with regard to observable characteristics X_i. For example, Andreoni and Vesterlund (2001) and Andreoni *et al.* (2003) show that men and women differ in their donation behavior. The conditional average treatment effect (CATE),

$$\delta(x) = E[\delta_i | X_i = x] = E[Y_i(1) - Y_i(-1) | X_i = x],$$

is the expected average donation effect of sending the fundraising gift to an individual with characteristics x.

The ATEs and CATEs are identified from observable data under the stratified experimental design and the SUTVA.[9] The strata characteristics Z_i are relevant to achieve identification. In contrast, the exogeneous heterogeneity characteristics X_i are potentially associated with heterogeneous effects of the gift, but we do not need them for identification. The strata and heterogeneity characteristics are not necessarily equivalent, but they may overlap. To achieve identification, we have to assume that the stratified randomization was appropriately conducted, such that $p(z,x) = Pr(D_i = 1|Z_i = z, X_i = x) = Pr(D_i = 1|Z_i = z) = p(z)$ and $(Y_i(1), Y_i(-1)) \perp\!\!\!\perp D_i|Z_i = z$. Furthermore, we have to assume that the the probability to receive the gift is between zero and one, $0 < p(z) < 1$. Tables A.2 and A.3 in Online Appendix A document that the random assignment into treatment and control groups was successful in achieving balanced groups with respect to observable characteristics.

4.2 Targeting Rules

A targeting rule $\pi(X_i) \in \{-1, 1\}$ is a deterministic function which defines the assignment of the gift to prospective donors based on their observable characteristics X_i. Individuals with $\pi(X_i) = 1$ receive the solicitation letter with the gift and individuals with $\pi(X_i) = -1$ receive the solicitation letter without the gift. The purpose of the targeting rule is to maximise the expected net donation $P(\cdot)$ of the fundraising campaign (= expected donation amount − variable costs of the gift),

$$P(\pi(X_i)) := E\left[Y_i(\pi(X_i)) - \frac{1+\pi(X_i)}{2}c\right], \tag{1}$$

where $Y_i(\pi(X_i))$ is the donation amount of individual i under the policy rule $\pi(X_i)$ and c are the variable costs of the gift. We ignore fixed costs, as they do not alter the targeting rule.

To evaluate the improvement that we achieve by implementing an optimal targeting rule, we compare expected net donations under the optimized rule $P(\pi(X_i))$ to expected net donations under three benchmark rules: A rule that assigns the gift to everybody, a

[9]The following proofs that the CATEs are identified from observable data

$$\begin{aligned}\delta(x) &\overset{LIE}{=} E_{Z|X=x}[E[Y_i(1) - Y_i(-1)|Z_i = z, X_i = x]],\\ &\overset{Experiment}{=} E_{Z|X=x}[E[Y_i(1)|D_i = 1, Z_i = z, X_i = x] - E[Y_i(-1)|D_i = -1, Z_i = z, X_i = x]],\\ &\overset{SUTVA}{=} E_{Z|X=x}[E[Y_i|D_i = 1, Z_i = z, X_i = x] - E[Y_i|D_i = -1, Z_i = z, X_i = x]],\end{aligned}$$

where the first equality is an application of the law of iterative expectations (LIE), the second equality holds under the experimental design, and the last equality under the SUTVA. The identification proof for the ATEs follows from the LIE, $\delta = E[\delta(X_i)]$.

rule that does not assign the gift to anybody, and a rule with random allocation. The expected net donation of the rule $\pi(X_i) = \pi_1 = 1$ that everybody receives the gift, is $P(\pi_1) = E[Y_i(1)] - c$. The expected net donation of the rule $\pi(X_i) = \pi_{-1} = -1$ that nobody receives the gift is $P(\pi_{-1}) = E[Y_i(-1)]$. Finally, under the random allocation rule π_R each individual has a 50% probability to receive the gift. The random rule can be viewed as a default option when no information about the effectiveness of the fundraising instrument is available and the fundraiser has no preferences about the fundraising instrument. The expected net donation under the random rule is $P(\pi_R) = 1/2 \cdot (E[Y_i(1) + Y_i(-1)] - c)$.

Now we compare the net donations of the optimal targeting rule with the three benchmark rules. First, we consider the benchmark that everybody receiving the gift. The excess net donation of the optimal targeting rule is

$$Q_1(\pi(X_i)) := P(\pi(X_i)) - P(\pi_1) = E\left[\frac{\pi(X_i) - 1}{2}(\delta_i - c)\right].$$

Second, we compare the expected net donation of the optimal targeting rule with the benchmark that nobody receives the gift,

$$Q_{-1}(\pi(X_i)) := P(\pi(X_i)) - P(\pi_{-1}) = E\left[\frac{1 + \pi(X_i)}{2}(\delta_i - c)\right].$$

Third, the excess net donation of the optimal targeting rule compared to the benchmark of random allocation is

$$\begin{aligned} Q_R(\pi(X_i)) := \; & P(\pi(X_i)) - P(\pi_R) = \frac{1}{2} E\left[\pi(X_i)(\delta_i - c)\right] \\ = \; & (Q_1(\pi(X_i)) + Q_{-1}(\pi(X_i)))/2. \end{aligned} \tag{2}$$

Because $P(\pi_1)$, $P(\pi_{-1})$, and $P(\pi_R)$ are constant, the targeting rule $\pi(X_i)$ which maximises the net donations $P(\pi(X_i))$ also maximizes $Q_1(\pi(X_i))$, $Q_{-1}(\pi(X_i))$, and $Q_R(\pi(X_i))$.

4.3 Estimation

4.3.1 Augmented Inverse Probability Weighting (AIPW)

An important ingredient for the optimal targeting rule is δ_i. As we mentioned before, δ_i is unobervable and cannot be estimated directly. However, an approximation score of δ_i can be sufficient to estimate the optimal targeting rule. The Augmented Inverse Probability Weighting (AIPW) score Γ_i is an example of such an approximation score.[10]

The AIPW score can be estimated in the following way. First, we estimate the so-

[10] Alternatively, Kitagawa and Tetenov (2018) suggest inverse probability weighting scores, and Beygelzimer and Langford (2009) propose to use offset weighting scores.

called nuisance parameters. These are the estimated conditional expectations of the potential donations with and without the gift, $\hat{\mu}_1(z) = \hat{E}[Y_i(1)|Z_i = z]$ and $\hat{\mu}_{-1}(z) = \hat{E}[Y_i(-1)|Z_i = z]$, and the estimated conditional probability that the gift was sent, $\hat{p}(z) = \widehat{Pr}(D_i = 1|Z_i = z)$, which is often called the propensity score. Second, we plug the nuisance parameters into the estimator of the AIPW score, which is $\hat{\Gamma}_i = \hat{\Gamma}_i(1) - \hat{\Gamma}_i(-1)$, with

$$\hat{\Gamma}_i(1) = \hat{\mu}_1(Z_i) + \frac{1+D_i}{2} \cdot \frac{Y_i - \hat{\mu}_1(Z_i)}{\hat{p}(Z_i)} \text{ and}$$
$$\hat{\Gamma}_i(-1) = \hat{\mu}_{-1}(Z_i) - \frac{D_i - 1}{2} \cdot \frac{Y_i - \hat{\mu}_{-1}(Z_i)}{1 - \hat{p}(Z_i)}.$$

Under the SUTVA and the experimental design, the expected value of the AIPW score identifies the average treatment effect $\delta = E[\Gamma_i]$. The conditional expectations of the AIPW score identify the CATEs $\delta(x) = E[\Gamma_i|X_i = x]$.[11]

The ATE estimator

$$\hat{\delta} = \frac{1}{N}\sum_{i=1}^{N} \hat{\Gamma}_i \tag{3}$$

is consistent, asymptotically normal, and semi-parametrically efficient under the requirement that the nuisance parameter estimators are consistent and converge sufficiently fast (e.g., Chernozhukov *et al.*, 2017; Robins *et al.*, 1994). In our application, we have precise information about the stratification process. Therefore, we use parametric nuisance parameter estimators which satisfy the requirements. In particular, we use a Logit to estimate the propensity score and OLS to estimate the conditional expectations of the potential donations with and without gift (Table C.1 in Online Appendix C reports the estimated coefficients of the different models). Another advantage of the AIPW score is its double robustness property. In particular, $\hat{\delta}$ remains consistent if one of the nuisance parameters is misspecified. In contrast to a multivariate OLS model, the AIPW estimator allows for heterogeneous treatment effects. Semenova and Chernozhukov (2017), Fan *et al.* (2019), and Zimmert and Lechner (2019) show that the AIPW scores can be used to estimate the CATEs $\hat{\delta}(x)$ under additional restrictions on the parameter space of X_i (see, e.g., Knaus, 2020, for a comprehensive review).

[11]For completeness, we sketch the identification proofs for the AIPW score in Online Appendix B (see also Knaus *et al.* (2020b) for a detailed discussion).

4.3.2 Estimation of the Optimal Targeting Rule

For the estimation of the optimal targeting rule, Athey and Wager (2021) propose to maximise the sample analogue of (2)

$$\pi^* = \text{argmax}_\pi \left\{ \frac{1}{2N} \sum_{i=1}^{N} \pi(X_i)(\hat{\Gamma}_i - c) \right\}, \tag{4}$$

where the unobservable individual causal effect δ_i is replaced with the estimated AIPW score $\hat{\Gamma}_i$. Alternatively, the objective function (4) can be formulated as the weighted classification estimator

$$\pi^* = \text{argmax}_\pi \left\{ \frac{1}{2N} \sum_{i=1}^{N} \pi(X_i) \cdot \text{sign}(\hat{\Gamma}_i - c) \cdot |\hat{\Gamma}_i - c| \right\}, \tag{5}$$

where $(\hat{\Gamma}_i - c) = \text{sign}(\hat{\Gamma}_i - c) \cdot |\hat{\Gamma}_i - c|$ (e.g., Beygelzimer and Langford, 2009; Zadrozny, 2003; Zhao *et al.*, 2012). The idea is to classify the sign of the net donation effects and weigh each observation by $|\hat{\Gamma}_i - c|$. The objective function is maximized when the signs of $\pi(X_i)$ and $(\hat{\Gamma}_i - c)$ are equal. In case some signs differ, misclassifications of individuals who respond strongly to the gift (i.e., individuals with large weights) reduce the net donations more than misclassification of individuals who do not respond strongly to the gift (i.e., individuals with small weights). Accordingly, the optimal targeting algorithm should prioritize individuals with large weights.

Athey and Wager (2021) show that the estimated optimal targeting rule π^* achieves asymptotically minimax-optimal regret (Manski, 2004) when the complexity of the targeting rules is restricted. In principal, any weighted classification estimator could be used to solve (5). We follow Athey and Wager (2021) and use shallow decision trees to estimate to targeting rule. Trees partition the sample into mutually exclusive strata based on the heterogeneity characteristics X_i. The tree depth restricts the complexity of the targeting rule, which makes trees suitable estimators in our context. In our main specifications, we follow Zhou *et al.* (2018) and use exact policy learning trees, with a search depth of two, to estimate the optimal targeting rule.[12] Standard Classification and Regression Trees (CARTs) select the partition with a greedy algorithm by adding recursively sample splits to the tree without anticipating later splits (e.g., Breiman *et al.*, 1984). In contrast, exact policy learning trees search for a fixed tree depth over all possible targeting rules. In contrast to CARTs, exact policy learning trees estimate the global optimum of (5).

It is possible to use relatively standard estimators, such as a weighted Logit regression, to solve (5). Decision trees have several advantages compared to the Logit for

[12]For implementation, we use the R package `policytree` (Sverdrup *et al.*, 2020).

the estimation of targeting rules. They select the relevant heterogeneity characteristics in a data-driven way. This is particularly relevant when we have no *a priori* domain knowledge about the relevant characteristics. Even if we know the relevant characteristics, there might be several highly correlated measures of these characteristics and it is *a priori* not clear which are the most relevant of them. In our application, for example, several geo-spatial characteristics are highly correlated and it is not clear which are the most relevant. In the extreme case, including too many highly correlated characteristics in a Logit regressions could cause multi-collinearity problems. Furthermore, it is typically unclear how to control for non-linear and interaction terms. Trees can automatically incorporate non-linear and interaction terms of the different characteristics without pre-coding. This minimizes the risk of overlooking important heterogeneities.

Having said this, we study the sensitivity of our results with regard to different estimation methods for the targeting rule. In particular, we consider Logit, Logit-Lasso, CART, and classification forest estimators.

4.3.3 Out-of-Sample Performance of the Targeting Rule

Once we have estimated the optimal targeting rule π^*, we can apply the sample analogy principal to estimate

$$\hat{P}(\pi^*(X_i)) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\Gamma}_i(\pi^*(X_i)) - \frac{1+\pi^*(X_i)}{2}c\right),$$

$$\hat{Q}_1(\pi^*(X_i)) = \frac{1}{N}\sum_{i=1}^{N}\frac{\pi^*(X_i)-1}{2}\left(\hat{\Gamma}_i - c\right),$$

$$\hat{Q}_{-1}(\pi^*(X_i)) = \frac{1}{N}\sum_{i=1}^{N}\frac{1+\pi^*(X_i)}{2}\left(\hat{\Gamma}_i - c\right), \text{ and}$$

$$\hat{Q}_R(\pi^*(X_i)) = \frac{1}{2N}\sum_{i=1}^{N}\pi^*(X_i)\left(\hat{\Gamma}_i - c\right).$$

These estimators are consistent, asymptotically normal, and semi-parametrically efficient (see the general framework of Chernozhukov *et al.*, 2018a).

To assess the out-of-sample performance of the targeting rule, we use a cross-fitting procedure. We randomly partition our data in $K = 20$ equally sized samples. We use $K-1$ partitions to estimate the targeting rule π^* and calculate $\hat{P}(\pi^*(X_i))$, $\hat{Q}_1(\pi^*(X_i))$, $\hat{Q}_{-1}(\pi^*(X_i))$, and $\hat{Q}_R(\pi^*(X_i))$ in the retained partition. We then repeat this procedure, discarding each of the K partitions once. In this way, we use the entire data efficiently. Finally, we report the average values of $\hat{P}(\pi^*(X_i))$, $\hat{Q}_1(\pi^*(X_i))$, $\hat{Q}_{-1}(\pi^*(X_i))$, and $\hat{Q}_R(\pi^*(X_i))$ over all the 20 partitions.

5 Results

Before we present our results for the optimal targeting rule, we discuss the average effects of the gift on donations. Furthermore, we show the heterogeneous effects of the gift on donations. Then we evaluate the net donation gains that the charity can achieve by implementing the optimal targeting rule and show which characteristics are relevant for the targeting rule.

5.1 Average and Heterogenous Treatment Effects

5.1.1 Average Treatment Effects (ATE)

Table 2 reports the ATE of the gift on donations during the first year after the experiment for the warm and cold list, respectively. For comparison, we report different ATE estimates that we obtain from unconditional and conditional OLS regressions and the AIPW estimator (3). In contrast to OLS, the AIPW estimator makes fewer functional form assumptions, allows for heterogeneous treatment effects, and is more robust to misspecification. However, the differences between the ATEs obtained from the different estimators are minor.

Table 2: Average treatment effects of the gift on donations

	Warm list			Cold list		
	OLS	OLS	AIPW	OLS	OLS	AIPW
	(1)	(2)	(3)	(4)	(5)	(6)
Average treatment effects	1.24	1.21	1.22	0.19***	0.19***	0.19*
	(1.25)	(1.16)	(1.15)	(0.07)	(0.07)	(0.10)
Average treatment effects net of costs	0.08	0.05	0.06	-0.97***	-0.97***	-0.97***
	(1.25)	(1.16)	(1.15)	(0.07)	(0.07)	(0.10)
Strata controls	No	Yes	Yes	No	Yes	Yes

Notes: This table shows the estimated average treatment effects of the gift treatment on donations. The first set of estimates uses the amount donated in the first year after the gift as an outcome variable (in Euro). The second set of estimates additionally substracts the gifts' costs from the donation amount. We report results for the following specifications: unconditional OLS (Columns 1 and 4), OLS with strata control variables (Columns 2 and 5), and AIPW (Columns 3 and 6). Because the AIPW model allows for heterogeneous treatment effects, this model represents our preferred specification. Standard errors are in parenthesis. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

Using the AIPW results, we find that sending the gift to everybody in the warm list would increase the average donation by 1.22 Euro in the first year, but the effect is statistically insignificant. To assess whether it is more profitable to send the gift to everybody or nobody in the warm list, we have to take into account the cost of sending the gift, which amounts to 1.16 Euro. The difference between the ATE and the cost of the gift is equivalent to the excess net donation of the rule that everybody receives the gift, π_1, compared to the rule that nobody receives the gift, π_{-1}, such that $P(\pi_1) - P(\pi_{-1}) = E[\delta_i] - c$. The excess net donation of sending the gift to everybody is only 0.06 Euro ($= 1.22 - 1.16$) and statistically insignificant. Accordingly, the profitability of the campaign is not systematically higher when we send the gift either to everybody or nobody in the warm list. Hence, for the charity in search of a dominant fundraising strategy, these results are not instructive.

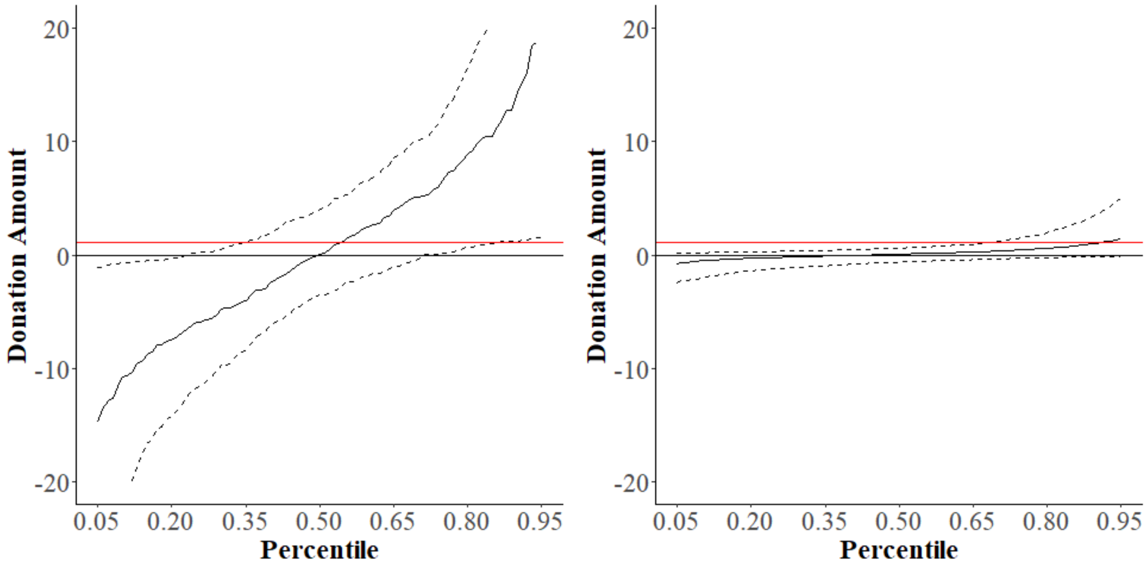
A different picture emerges for the cold list. The ATE of the gift on donations is only 0.19 Euro. However, after accounting for the costs of the gift, sending the gift would cause a statistically significant loss of 0.97 Euro per donor ($= 0.19 - 1.16$) compared to sending no gift. Accordingly, in the cold list not sending the gift to anybody is, on average, more profitable than sending the gift to everybody. Nevertheless, in the presence of heterogeneous effects, the optimal targeting rule might outperform the rule that nobody receives the gift.

5.1.2 Heterogeneous Treatment Effects

An important ingredient for the success of the optimal targeting rule is the presence of heterogeneous treatment effects. In particular, optimal targeting can only outperform the rules that either everybody or nobody receives the gift (π_1 or π_{-1}) if some individuals increase and others decrease their net donations in response to the gift. Accordingly, it makes sense to investigate heterogeneous treatment effects before moving to the results of optimal targeting. We use the sorted effects approach of Chernozhukov *et al.* (2018b) to visualize the distribution of the effects of the gift on donations during the first year after the experiment. To estimate the sorted effects, we specify a simple linear OLS regression including all observable characteristics and their interaction with the treatment dummy. When the sorted effects model shows heterogeneous effects that are below and above the costs of the gift, the optimal targeting approach has the potential to be successful. However, we cannot use the sorted effects model directly for targeting, because this approach is suitable to estimate the distribution of the effects, but not suitable to design targeting rules based on observable characteristics.

Figure 2 reports the sorted effects for the warm and cold list, respectively. On the ordinate are the donations. On the abscissa are the percentiles of the effect size of the gift on donations, moving from the smallest effects on the left side to the largest effects on the right side. The figure reveals a large amount of heterogeneity in the warm list. At the 95 percentile, donations increase by 20.62 Euro in response to the gift, which provides supporting evidence for the sequential reciprocity hypothesis (Dufwenberg and Kirchsteiger, 2004; Falk, 2007). At the 5 percentile, donations decrease by 14.76 Euro in response to the gift, pointing to the possibility that even a 'warm' gift (postcards) may

Figure 2: Sorted effects



(a) Warm list

(b) Cold list

Notes: This figure shows the heterogeneity in the effect of the gift on the donation amount. To that end, it sorts the estimated conditional average treatment effects by size. It then plots the size of the treatment effect in Euro (vertical axis) against the percentiles of the effect size (horizontal axis). The red horizontal line represents the cost of the gift (1.16 Euro). The solid line depicts the sorted effects. We report results between the 5 and 95 percentiles. The dashed lines report uniformly valid 95% confidence intervals, which build on a multiplier bootstrap and 500 replications (see Chernozhukov *et al.*, 2018b, for details).

change the donors' perception of the relationship with the fundraiser from a communal to an exchange norm (Yin *et al.*, 2020). The red horizontal line represents the cost of the gift. In the ideal case, the gift should be sent to all individuals for whom the effect on donations exceeds the costs, and not be sent to individuals with an effect size below the costs.[13] The sorted effects exceed the costs above the 62 percentile. This suggest that the gift should be targeted at the 38% of individuals with the largest effects.

The figure also shows that the effect heterogeneity in the cold list is much smaller than in the warm list. The effects do not differ significantly from zero at all percentiles. The donation amount decreases by 0.75 Euro at the 5 percentile and increases by 1.39 Euro at the 95 percentile. The sorted effects of the gift on the donation amount exceed the costs above the 92 percentile. Accordingly, the gift should be targeted at the 8% of individuals with the largest effects, but the vast majority of individuals in the cold list should not receive the gift. Below the 69 percentile, the effects on the donation amount are significantly lower than the costs.

In Tables D.1 and D.2 of Online Appendix D we report the means of all characteristics for the groups with the 10% largest and smallest sorted effects for the warm and cold list, respectively. We do not find any significant pattern for the warm list. However, there is a

[13]This is called the empirical success rule, which was proposed by Manski (2004).

tendency that individuals with the 10% largest effects donated less before the experiment and live on a somewhat lower altitude than individuals with the 10% smallest effects. In the cold list, individuals with the 10% largest effects live significantly closer to the city center than individuals with the 10% smallest effects.

5.2 Effectiveness of Targeting Rules

In order to evaluate the effectiveness of the targeting rules, we randomly split the sample into $K = 20$ folds. Then we use a cross-fitting procedure to optimize the targeting rule in $K - 1$ folds and evaluate the out-of-sample performance of the targeting rule in the retained fold. To use the data efficiently, we repeat the procedure discarding each fold once and report the average out-of-sample performance of the targeting rule across all folds. Because the out-of-sample evaluation is robust against over-fitting, this procedure prevents us from reporting spurious relationships and adds to the external validity of our results.

Results for the warm list. Table 3 documents the out-of-sample performance of the optimal targeting rule in the warm list. In Panel A, we report the results for the net donation amount during the first year after the gift was sent, which is the outcome variable that we maximize with the optimal targeting rule. The optimal targeting rule assigns the gift to 33% of the individuals in the warm list (column 1). This is fairly close to the share of individuals with a sorted effect that exceeds the costs of the gift (see Figure 2). The average net donations under the optimal targeting rule is 17.61 Euro (column 2). This is 2.14 Euro (13.8%) more than when everybody would receive the gift (column 3), 2.20 Euro (14.3%) more than when nobody would receive the gift (column 4), and 2.17 Euro (14.1%) more than when the gift would be randomly allocated to one half of the individuals in the warm list (column 5). Accordingly, the optimal targeting rule is significantly more profitable than all three benchmark policies.

Beyond the first-year's net donation amount, the optimal targeting rule may also affect other outcome variables, even though we do not train the algorithm on these variables. We label these influences on alternative outcome variables 'second order effects'. In particular, we investigate how the optimal targeting rule affects donations in the long term (first and second year pooled), as well as the donation probability during the first and second year. We report second order effects in Panel B of Table 3.

The positive effects of the optimal targeting rule on net donations are persistent during the first two years after the gift was send. This suggests our findings cannot be explained by pull-forward effects. We also find positive effects on the donation probability, but the effects are small and statistically insignificant when compared to the benchmark that everybody receives the gift. This suggests that the difference in the net donation

Table 3: Out-of-sample performance of targeting rule in the warm list

	Share treated	Expected outcome value under optimal targeting	Optimal targeting vs. everybody	nobody	random
	(1)	(2)	(3)	(4)	(5)
Panel A: Results for primary outcome variable					
Net donation amount (1st year)	0.33	17.61*** (0.97)	2.14*** (0.82)	2.20*** (0.81)	2.17*** (0.58)
Panel B: Results for secondary outcome variables					
Net donation amount (1st and 2nd year)		32.94*** (1.66)	2.33* (1.41)	3.75*** (1.41)	3.04*** (0.10)
Donation probability (1st year)		0.503*** (0.013)	0.007 (0.013)	0.025** (0.010)	0.016* (0.008)
Donation probability (1st and 2nd year)		0.582*** (0.013)	0.001 (0.013)	0.017* (0.009)	0.009 (0.008)

Notes: This table documents the out-of-sample performance of our optimal targeting rule, focusing on the warm list. As previously described, the goal of our optimal targeting strategy is to maximize donations net of costs. Panel A reports the expected consequences of our rule for our main outcome "net donations." Panel B, instead, focuses on secondary outcomes. The interpretations of the columns are as follows: Column 1 reports the share of individuals that, according to the rule, should receive the gift. Column 2 reports the expected value of the outcomes under optimal targeting. For example, we expect that, under optimal targeting, the donations net of costs would be 17.61 Euro. Column 3-5 show how optimal targeting changes the outcomes relative to three benchmark scenarios: everybody receives the gift (Column 3), nobody receives the gift (Column 4), and the gift is randomly assigned to half of the sample (Column 5). Methodologically, optimal targeting rules are estimated with exact policy learning trees with a search depth of two (Zhou *et al.*, 2018). Donations are measured in Euro. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

amount between the optimal targeting rule and the benchmark that everybody receives the gift is driven by intensive rather than extensive margin effects. In contrast, when we compare the optimal targeting rule to the benchmark that nobody receives the gift, the extensive margin effects seem to be relevant as well. For example, during the first year the donation probability increases by 3 percentage points (5%) relative to this benchmark.

Results for the cold list. Table 4 shows the out-of-sample performance of the optimal targeting rule in the cold list. The optimal targeting rule assigns the gift to only 1% of the individuals in the cold list (column 1). The average net donation amount under the optimal targeting rule is 0.15 Euro (column 2). This is 0.97 Euro more than when everybody would receive the gift (column 3) and 0.48 Euro more than when the gift would be randomly allocated to one half of the individuals (column 5) in the cold list (the net donation amount would be negative under both benchmarks). However, the optimal targeting rule does not outperform the benchmark that nobody receives the gift (column 4). The net donations are virtually identical under both rules, which is not surprising, because the targeting rule assigns the gift only to very few prospective donors. The effects are very precisely estimated in the cold list, which comprises more than 17,000 individuals. This suggests that optimal targeting is not systematically more profitable than sending the gift to nobody. While this insight is not too surprising, it is worth noting that it was not clear *a priori*. In fact, we demonstrate that the optimal targeting approach is suitable to establish this finding.

The second order effects on the net donation amount during the first two years after the gift was send are similar to the main results. We find again no evidence for pull-forward or delay effects. Furthermore, the targeting rule reduces the donation probability relative to the benchmark that everybody receives the gift. This suggests that the positive intensive margin effects overcompensate the negative extensive margin effects. We find no significant extensive margin effects of the targeting rule compared to the benchmark that nobody receives the gift.

Table 4: Out-of-sample performance of targeting rule in the cold list

	Share treated	Expected outcome value under optimal targeting	Optimal targeting vs. everybody	nobody	random
	(1)	(2)	(3)	(4)	(5)
Panel A: Results for primary outcome variable					
Net donation amount (1st year)	0.014	0.15***	0.97***	-0.005	0.48***
		(0.02)	(0.10)	(0.012)	(0.05)
Panel B: Results for secondary outcome variables					
Net donation amount (1st and 2nd year)		0.44***	0.96***	0.04	0.50***
		(0.07)	(0.13)	(0.06)	(0.07)
Donation probability (1st year)		0.009***	-0.007***	0.001	-0.003**
		(0.001)	(0.003)	(0.001)	(0.001)
Donation probability (1st and 2nd year)		0.017***	-0.006*	0.001	-0.003
		(0.001)	(0.003)	(0.001)	(0.002)

Notes: This table documents the out-of-sample performance of our optimal targeting rule, focusing on the cold list. As previously described, the goal of our optimal targeting strategy is to maximize donations net of costs. Panel A reports the expected consequences of our rule for our main outcome "net donations." Panel B, instead, focuses on secondary outcomes. The interpretations of the columns are as follows: Column 1 reports the share of individuals that, according to the rule, should receive the gift. Column 2 reports the expected value of the outcomes under optimal targeting. For example, we expect that, under optimal targeting, the donations net of costs would be 0.014 Euro. Column 3-5 show how optimal targeting changes the outcomes relative to three benchmark scenarios: everybody receives the gift (Column 3), nobody receives the gift (Column 4), and the gift is randomly assigned to half of the sample (Column 5). Methodologically, optimal targeting rules are estimated with exact policy learning trees with a search depth of two (Zhou *et al.*, 2018). Donations are measured in Euro. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

5.3 Which Characteristics Determine Net Donors?

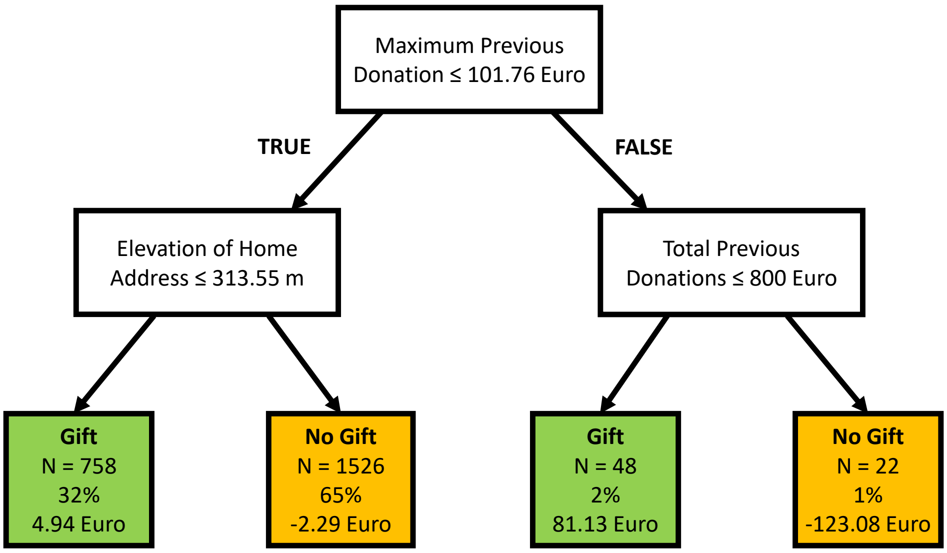
We use socio-economic characteristics, donation history, and geo-spatial information for optimal targeting, but the characteristics may not be equally important. Furthermore, compared to actual donors, charities might often have much less information about prospective donors, and data collection is costly. In the extreme (but quite realistic) case, a charity may only have address data before sending out written solicitations to expand its donor base. Accordingly, it is an important question which characteristics are relevant for determining the optimal targeting rule.

A natural way to describe the determinants of the targeting rule is to plot the decision tree. However, because we use a cross-fitting approach to evaluate the out-of-sample performance of optimal targeting, we estimate 20 different decision trees. Figure 3 plots exemplary the decision tree of the optimal targeting rule that we obtain when we use the entire sample (without cross-fitting) for the warm and cold list, respectively. In the warm list, the tree splits are based on the previous donation amount and the elevation at the home address. In the cold list, the tree splits are based on the number of restaurants in the proximity of the home address, the distance to the city hall and the distance to the airport. The trees do not use the socio-economic characteristics.

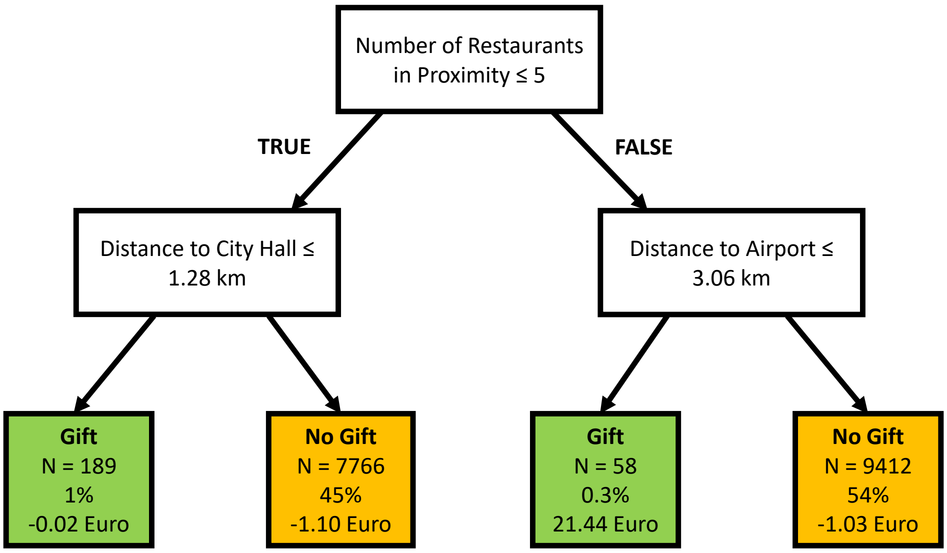
To interpret the role of the split variables in the decision tree, it is important to notice that the optimal targeting approach does not identify the causal effect of donor characteristics on net donations. Instead, it identifies the association between donor characteristics and the causal effect of the gift on net donations. Accordingly, we have to interpret the results of the decision trees with caution. Alternative decision trees could have a similar impact on net donations, especially in the presence of highly correlated characteristics (as in our application).

To obtain a better overview about the influence of the different characteristics, Table 5 and 6 report the means and standard deviations of all observed characteristics for the individuals target and not targeted by the algorithm in the warm and cold list, respectively. In the cold list are large differences in the socio-economic characteristics between individuals targeted and not targeted by the algorithm, but we find no large differences in socio-economic characteristics in the warm list. In the cold list, the algorithms tends to target more women, singles, and newly settled residents. In the warm list, the individuals targeted by the algorithm donated on average more and more frequently before the experiment than the individuals not targeted by the algorithm. In the warm and cold list, we find the largest differences in the geo-spatial information. Individuals targeted by the algorithm live on a lower altitude than individuals not targeted. In the concrete city we study, the topology is correlated with distance to the city center. Concretely, individuals living in lower altitudes live on average closer to the city hall and the main church. There are differences between the warm and cold list when we consider prox-

Figure 3: Illustration of decision tree



a) Warm list



b) Cold list

Notes: This figure illustrates one exemplary decision tree. For each terminal leaf, it reports the total and relative number of observation and the effect of the gift on net donations during the first year after the experiment. To derive the decision trees, we estimate exact policy learning trees with a search depth of two (Zhou *et al.*, 2018).

Table 5: Characteristics of individuals targeted by the algorithm in the warm list

	Individuals targeted by the algorithm				Std.
	Yes		No		Diff.
	Mean	Std. Dev.	Mean	Std. Dev.	
	(1)	(2)	(3)	(4)	(5)
Socio-economic characteristics					
Female dummy	0.507		0.539		6.459
Single dummy	0.503		0.496		1.464
Widowed dummy	0.050		0.052		0.974
Age (in years)	68.08	18.23	68.72	18.34	3.488
Duration residency (in years)	7.423	1.690	7.439	1.659	0.951
Donation history before the experiment					
Num. donations prev. 8 years	4.097	2.827	3.900	2.829	6.934
Max. don. prev. 8 years (in Euro)	39.94	44.61	34.05	41.89	13.63
Total don. prev. 8 years (in Euro)	130.98	150.69	123.39	187.39	4.460
Donations 1 year ago (in Euro)	22.69	36.49	19.53	34.60	8.891
Donations 2 years ago (in Euro)	17.91	30.30	16.89	28.77	3.466
Donations 3 years ago (in Euro)	16.61	27.49	15.62	27.52	3.593
Donations 4 years ago (in Euro)	16.71	28.33	15.37	27.21	4.813
Donations 5 years ago (in Euro)	15.69	24.48	15.03	29.96	2.410
Geo-spatial information about home address					
Elevation (in meters)	308.66	6.266	321.38	9.524	157.80
In 300 meters proximity:					
Number of restaurants	10.86	13.30	6.528	7.711	39.88
Number of supermarkets	1.062	1.371	1.086	1.362	1.748
Number of medical facilities	10.17	13.95	9.298	12.041	6.703
Number of cultural facilities	0.240	0.796	0.050	0.241	32.27
Number of churches	1.166	1.515	0.934	1.460	15.60
Distance to main station (in km)	3.247	2.521	3.245	1.867	0.053
Distance to city hall (in km)	2.927	2.237	3.196	1.856	13.11
Distance to main church (in km)	2.986	2.365	3.218	1.836	10.99
Distance to airport (in km)	5.427	1.236	5.483	1.960	3.408
Travel time to main station (in min.)	18.42	11.79	17.50	7.560	9.371
Observations	787		1'567		

Notes: The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays. For dummy variables the first moment is sufficient to infer the entire distribution. Rosenbaum and Rubin (1983) classify absolute standardized difference (std. diff.) of more than 20 as "large".

imity to economic and cultural facilities. In the warm list, the individuals targeted by the algorithm have more restaurant, medical and cultural facilities, and churches in the proximity of their homes than the not individuals targeted by the algorithm. In the cold list, we find the opposite relationship (with the exception of medical facilities).

Table 6: Characteristics of individuals targeted by the algorithm in the cold list

	Individuals targeted by the algorithm				Std.
	Yes		No		Diff.
	Mean	Std. Dev.	Mean	Std. Dev.	
	(1)	(2)	(3)	(4)	(5)
Socio-economic characteristics					
Female dummy	0.558		0.503		11.04
Single dummy	0.713		0.642		15.24
Widowed dummy	0.024		0.017		4.518
Age (in years)	47.58	21.00	48.41	19.30	4.132
Duration residency (in years)	5.677	2.964	5.973	2.818	10.22
Geo-spatial information about home address					
Elevation (in meters)	313.47	8.183	316.15	10.34	28.71
In 300 meters proximity:					
Number of restaurants	5.482	3.524	10.40	11.67	57.08
Number of supermarkets	0.888	1.122	1.296	1.502	30.77
Number of medical facilities	13.73	10.20	10.68	13.17	25.89
Number of cultural facilities	0.040	0.196	0.146	0.532	26.43
Number of churches	1.100	1.017	1.177	1.538	5.957
Distance to main station (in km)	1.995	0.890	2.874	2.028	56.14
Distance to city hall (in km)	1.364	0.645	2.816	1.885	103.02
Distance to main church (in km)	1.588	0.743	2.803	1.932	83.03
Distance to airport (in km)	4.143	1.038	5.567	1.642	103.68
Travel time to main station (in min.)	12.50	6.312	16.18	8.680	48.45
Observations	251		17'174		

Notes: The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays. For dummy variables the first moment is sufficient to infer the entire distribution. Rosenbaum and Rubin (1983) classify absolute standardized difference (std. diff.) of more than 20 as "large".

5.4 Which Data Sources are Relevant?

From a practical perspective, and taking into account the costs of data collection and handling, a charity will likely prefer a targeting rule which uses only data from a single source. It is therefore worth noting that the decision trees in the warm list suggest that the donation history and geo-spatial information are important for optimal targeting, while the socio-economic characteristics are less important. In addition, by definition there is no donation history to be considered in the cold list. A possible conjecture is that the optimal targeting does not outperform the benchmark that nobody receives the gift, because we have not enough variation in the data in the absence of a donation history. A natural way to test this conjecture is to estimate the optimal targeting rule in the warm list without making use of the donation history, and compare the resulting out-of-sample performance to the rule that uses all observable characteristics. Because many of the observed characteristics are highly correlated, it might be that the geo-spatial information is sufficient to design optimal targeting rules (because all observed characteristics are somehow correlated with economic circumstances).

In Table 7, we use this procedure to determine the relevant data sources in the warm list. We distinguish between characteristics that originate from the three different data sources: socio-economic characteristics, donation history, and geo-spatial information. We estimate optimal targeting rules which use each data source one-by-one. As before, we benchmark the policy rule to the scenario that everybody receives the gift (column 3), nobody receives the gift (column 4), and allocation (column 5). Additionally, in column 6 we benchmark the new targeting rules to the optimal targeting rule obtained when using all data sources (which corresponds to the optimal targeting rules reported in Table 3).

The results suggest that, as expected, the targeting rule is not outperforming the benchmark assignment rules when only the socio-economic characteristics are used. Furthermore, the performance of the targeting rule which uses only the socio-economic characteristics is significantly worse than the targeting rule which exploits all data sources. In contrast, using only the donation history or, alternatively, only the geo-spatial information does not affect the performance of the targeting rule compared to using all three data sources. This suggests that in the warm list, the donation history and the geo-spatial variables are substitutes for each other in terms of their ability to drive the targeting rule's performance. Accordingly, the results for the warm list suggest that the geo-spatial information is a good candidate to substitute the donation history in the cold list.

Adding the socio-economic variables to one of the two other covariate groups does not improve their single-variable group performance. Using jointly the characteristics about the donation history and the geo-spatial information is equivalent to the targeting rule which has access to all three data sources, because the socio-economic characteristics are not used anyway. This suggests the socio-economic characteristics do not improve the policy rule at all.

Table 8 reports the performance of the targeting rule in the cold list when we restrict the data sources either to the socio-economic characteristics or the geo-spatial information. The results do not change when we use only the geo-spatial information compared to using both covariate groups, because the targeting rule is not based on socio-economic characteristics anyway. This suggests the socio-economic characteristics are also redundant in the cold list. We conclude from this that, given the data available, not sending the gift to anybody is a dominant strategy for the cold list. Together, our results suggest that relying on widely available geo-spatial information alone, we can significantly improve the profitability of gifts as a fundraising tool by targeting net donors, particularly in the warm list.

Table 7: Relevant data sources in the warm list

Share treated	Expected donations under optimal targeting	Optimal targeting vs. everybody	nobody	random	all variables
(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Socio-economic characteristics					
0.55	15.71***	0.24	0.29	0.27	-1.91**
	(0.79)	(0.86)	(0.77)	(0.58)	(0.89)
Panel B: Donation history					
0.12	17.20***	1.73**	1.79**	1.76***	-0.41
	(0.97)	(0.82)	(0.81)	(0.58)	(0.55)
Panel C: Geo-spatial information					
0.49	17.40***	1.93**	1.98**	1.95***	-0.22
	(0.91)	(0.84)	(0.79)	(0.58)	(0.61)
Panel D: Socio-economic characteristics and donation history					
0.11	17.05***	1.58*	1.64**	1.61***	-0.56
	(0.96)	(0.84)	(0.79)	(0.58)	(0.56)
Panel E: Socio-Economic characteristics and geo-spatial information					
0.48	16.97***	1.50*	1.55*	1.53***	-0.64
	(0.89)	(0.84)	(0.80)	(0.58)	(0.62)
Panel F: Donation history and geo-spatial information					
0.33	17.61***	2.14***	2.20***	2.17***	0
	(0.97)	(0.82)	(0.81)	(0.58)	

Notes: This table documents the out-of-sample performance of targeting rules that only rely on selected subsets of our variables, focusing on the warm list. The rule in Panel A is based on socio-economic characteristics only, in Panel B on the donation history, in Panel C on geo-spatial information, in Panel D on socio-economic characteristics and the donation history, in Panel E on socio-economic characteristics and geo-spatial information, and in Panel F on the donation history and geo-spatial information. The interpretations of the columns are as follows: Column 1 reports the share of individuals that, according to the respective rule, should receive the gift. Column 2 reports the expected donations under optimal targeting. Column 3-5 show how optimal targeting changes the outcomes relative to three benchmark scenarios: everybody receives the gift (Column 3), nobody receives the gift (Column 4), and the gift is randomly assigned to half of the sample (Column 5). Column 6 compares the rule that only uses the subset of variables to the rule that relies on the full set of data. Methodologically, the rules are estimated with an exact policy learning trees with a search depth of two (Zhou *et al.*, 2018). Donations are measured in Euro. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

5.5 Sensitivity Analysis

We investigate the sensitivity of the optimal targeting rules with regard to different estimation approaches. We consider alternative depths of the exact policy learning tree (depth one and three).[14] We compare the results of the exact policy learning trees to

[14]For the cold list, exact policy learning trees with a depth of three are infeasible due to computational constraints.

Table 8: Relevant data sources in the cold list

Share treated	Expected donations under optimal targeting	Optimal targeting vs. everybody	nobody	random	all variables
(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Socio-economic characteristics					
0.015	0.14***	0.96***	-0.014**	0.47***	-0.01
	(0.02)	(0.10)	(0.007)	(0.05)	(0.014)
Panel B: Geo-spatial information					
0.014	0.15***	0.97***	-0.005	0.48***	0
	(0.02)	(0.10)	(0.012)	(0.05)	

Notes: This table documents the out-of-sample performance of targeting rules that only rely on selected subsets of our variables, focusing on the cold list. The rule in Panel A is based on socio-economic characteristics only and in Panel B geo-spatial information. The interpretations of the columns are as follows: Column 1 reports the share of individuals that, according to the respective rule, should receive the gift. Column 2 reports the expected donations under optimal targeting. Column 3-5 show how optimal targeting changes the outcomes relative to three benchmark scenarios: everybody receives the gift (Column 3), nobody receives the gift (Column 4), and the gift is randomly assigned to half of the sample (Column 5). Column 6 compares the rule that only uses the subset of variables to the rule that relies on the full set of data. Methodologically, the rules are estimated with an exact policy learning trees with a search depth of two (Zhou *et al.*, 2018). Donations are measured in Euro. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

the results from standard CARTs (Breiman *et al.*, 1984). For the CARTs, we also consider one version in which we use cross-validation to select the tree depth in a data-driven way.[15] Furthermore, we consider the weighted Logit as a standard estimator as well as Logit-Lasso (Hastie *et al.*, 2016) and classification forest (Breiman, 2001) as alternative machine learning estimators.[16] For the Logit estimator, we consider two different model specifications. In the baseline specification, we include all observed characteristics linearly (24 variables in the warm list and 16 variables in the cold list). In the flexible specification, we additionally include squared terms of continuous variables and first-order interactions between most characteristics (321 variables in the warm list and 149 variables in the cold list). The Logit-Lasso selects the relevant characteristics from the flexible model specification.[17]

Tables E.1 and E.2 in Online Appendix E report the results of the sensitivity analysis for the warm and cold list, respectively. For the warm list, the optimal targeting rules

[15]The Gini index is used for tree splitting. We use a 10-fold cross-validation procedure to select the optimal tree depth. In the warm list, the number of terminal leaves vary between four and nine across the 20 different cross-fitted trees, with an average of 4.6 terminal leaves. In the cold list, the number of terminal leaves vary between two and eight across the 20 different cross-fitted trees, with an average of 2.3 terminal leaves.

[16]We build 1,000 trees for the classification forest. For each tree, we draw a 50% random subsample with replacement and randomly select 50% of the baseline characteristics. The Gini index is used for tree splitting. We restrict the minimum size of the terminal leaf to 50 observations.

[17]We specify the penalty λ of the Logit-Lasso that minimizes the misclassification error using a 10-fold cross-validation approach.

of all alternative estimators yield lower net donations than our main specification. The exact policy learning tree with depth one and the Logit with the flexible model specification have the lowest out-of-sample performance. CART with cross-validated tree depth is the only alternative estimator that significantly outperforms the benchmark allocation rules that either everybody or nobody receives the gift. For the cold list, Logit-Lasso and CART with cross-validated tree depth yield 0.01 Euro higher net donations than our main specification. However, our main finding that the dominant strategy is to not send the gift to anybody is maintained.

6 Conclusion

We study the optimal targeting of fundraising instruments by exploiting data from a large-scale natural field experiment. The experiment allows a machine learning algorithm to learn the relationship between individual characteristics and the response of potential donors to small unconditional gifts from a charity. Because the gifts are randomly assigned in the experiment, we can trace out the causal effect on donations of targeting individuals on the basis of this learned relationship, and compare the profitability of this approach to alternative uniform strategies (sending out gifts to all potential donors, or not sending out any gifts).

We find that in the warm list, optimal targeting increases the charity's profits by about 14%. In the cold list, optimal targeting does not raise donations sufficiently to justify the additional costs of the fundraising instrument. We further show that we can achieve optimal targeting in the warm list even if we rely only on widely available geospatial information that can be linked to address data. We also document that the improvements in profitability achieved through optimal targeting exploit existing heterogeneities in donors' responses to fundraising activities. Notably, the existence of these heterogeneities has been suggested before by mixed evidence on the effectiveness of unconditional gifts on giving (Falk, 2007; Yin *et al.*, 2020; Alpizar *et al.*, 2008).

We conclude that exploiting heterogenous responses of donors to fundraising instruments to optimize targeting can make charities significantly more effective. Furthermore, the machine-learning toolkit that we apply in this paper allows for an agnostic way to target fundraising efforts in a broad variety of contexts. In particular, to optimize targeting, charities do not need to develop a theoretical foundation, or make strong assumptions on the functional relationship of different potential influences on giving. We are, therefore, confident that this paper suggests an accessible way forward to improve the profitability of fundraising.

References

ADENA, M. and HUCK, S. (2019). Personalized Fundraising: A Field Experiment on Threshold Matching of Donation. *WZB Discussion Paper No. SP II 2019-306*.

ALDASHEV, G., MARINI, M. and VERDIER, T. (2014). Brothers in Alms? Coordination Between Nonprofits on Markets for Donations. *Journal of Public Economics*, **117**, 182–200.

— and VERDIER, T. (2010). Goodwill Bazaar: NGO Competition and Giving to Development. *Journal of Development Economics*, **91** (1), 48–63.

ALPIZAR, F., CARLSSON, F. and JOHANSSON-STENMAN, O. (2008). Anonymity, Reciprocity, and Conformity: Evidence From Voluntary Contributions to a National Park in Costa Rica. *Journal of Public Economics*, **92** (5-6), 1047–1060.

ANDINI, M., CIANI, E., DE BLASIO, G., D'IGNAZIO, A. and SALVESTRINI, V. (2018). Targeting With Machine Learning: An Application to a Tax Rebate Program in Italy. *Journal of Economic Behavior and Organization*, **156**, 86–102.

ANDREONI, J. (1988). Privately Provided Public Goods in a Large Economy: The Limits of Altruism. *Journal of Public Economics*, **35** (1), 57–73.

—, BROWN, E. and RISCHALL, I. (2003). Charitable Giving by Married Couples. Who Decides and Why Does It Matter? *Journal of Human Resources*, **38** (1), 111–133.

— and PAYNE, A. (2003). Do Government Grants to Private Charities Crowd Out Giving or Fund-Raising? *American Economic Review*, **93** (3), 792–812.

— and — (2011). Is Crowding Out Due Entirely to Fundraising? Evidence From a Panel of Charities. *Journal of Public Economics*, **95**, 334–343.

— and — (2013). Charitable Giving. In R. C. M. F. Auerbach, Alan and E. Saez (eds.), *Handbook of Public Economics*, vol. 5, *1*, Elsevier, pp. 1–50.

— and VESTERLUND, L. (2001). Which Is the Fair Sex? Gender Differences in Altruism. *Quarterly Journal of Economics*, **116** (1), 293–312.

ASCARZA, E. (2018). Retention Futility: Targeting High Risk Customers Might Be Ineffective. *Journal of Marketing Research*, **55** (1), 80–98.

ATHEY, S. and IMBENS, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, **11**, 685–725.

— and WAGER, S. (2021). Policy Learning with Observational Data. *Econometrica*, **forthcoming**.

BEKKERS, R. and WIEPKING, P. (2011). Who Gives? a Literature Review of Predictors of Charitable Giving. Part One: Religion, Education, Age, and Socialization. *Voluntary Sector Review*, **2** (3), 337–365.

— and — (2012). Who Gives? a Literature Review of Predictors of Charitable Giving. Part Two: Gender, Marital Status, Income, and Wealth. *Voluntary Sector Review*, **3** (2), 217–245.

BEYGELZIMER, A. and LANGFORD, J. (2009). The Offset Tree for Learning With Partial Labels. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 129–138.

BREIMAN, L. (2001). Random Forests. *Machine Learning*, **45** (1), 5–32.

—, FRIEDMAN, J., STONE, C. and OLSHEN, R. (1984). *Classification and Regression Trees*. CRC Press.

CASALE, D. and BAUMANN, A. (2015). Who Gives to International Causes? A Sociodemographic Analysis of US Donors. *Nonprofit and Voluntary Sector Quarterly*, **44** (1), 98–122.

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*, **107** (5), 261–265.

—, ESCANCIANO, J., ICHIMURA, H., NEWEY, W. and ROBINS, J. (2018a). Locally-Robust Semiparametric Estimation. *arXiv:1608.00033*.

—, FERNÁNDEZ-VAL, I. and LUO, Y. (2018b). The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages. *Econometrica*, **86** (6), 1911–1938.

DE VRIES, N., REIS, R. and MOSCATO, P. (2015). Clustering Consumers Based on Trust, Confidence and Giving Behaviour: Data-Driven Model Building for Charitable Involvement in the Australian Not-for-Profit Sector. *PloS One*, **10** (4), 1–28.

DONG, L., RATTI, C. and ZHENG, S. (2019). Predicting Neighborhoods' Socioeconomic Attributes Using Restaurant Data. *PNAS*, **116** (31), 15447–15452.

DROUVELIS, M. and MARX, B. (2021). Can Charitable Appeals Identify and Exploit Belief Heterogeneity? *CESifo Working Paper No. 8855*.

DUFWENBERG, M. and KIRCHSTEIGER, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, **47** (2), 268–298.

ECKEL, C., HERBERICH, D. and MEER, J. (2018). It's Not the Thought that Counts: A Field Experiment on Gift Exchange and Giving at a Public University. In *Econ. Philanthropy*, MIT Press.

FALK, A. (2007). Gift Exchange in the Field. *Econometrica*, **75**, 1501–1511.

—, BECKER, A., DOHMEN, T., ENKE, B., HUFFMAN, D. and SUNDE, U. (2018). Global Evidence on Economic Preferences. *Quarterly Journal of Economics*, **133** (4), 1645–1692.

FAN, Q., HSU, Y.-C., LIELI, R. P. and ZHANG, Y. (2019). Estimation of Conditional Average Treatment Effects With High-Dimensional Data. *arXiv:1908.02399*.

FARROKHVAR, L., ANSARI, A. and KAMALI, B. (2018). Predictive Models for Charitable Giving Using Machine Learning Techniques. *PloS One*, **13** (10), 1–14.

GLAESER, E., KIM, H. and LUCA, M. (2020). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity. In Abraham, Jarmin, Moyer and Shapiro (eds.), *Big Data for 21st Century Economic Statistics*.

—, KOMINERS, S., LUCA, M. and NAIK, N. (2018). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. *Economic Inquiry*, **56** (1), 114–137.

GNEEZY, U., KEENAN, E. A. and GNEEZY, A. (2014). Avoiding Overhead Aversion in Charity. *Science*, **346** (6209), 632–635.

HARBAUGH, W., MAYR, U. and BURGHART, D. (2007). Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations. *Science*, **316** (5831), 1622–1625.

HASSELL, H. and MONSON, J. (2014). Campaign Targets and Messages in Direct Mail Fundraising. *Political Behavior*, **36** (2), 359?376.

HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2016). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

HITSCH, G. and MISRA, J. (2018). Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation. *SSRN preprint 3111957*.

KANG, J., 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*, p. 1443?1448.

KITAGAWA, T. and TETENOV, A. (2018). Who Should be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, **86** (2), 591–616.

KLEINBERG, J., LUDWIG, J., MULLAINATHAN, S. and OBERMEYER, Z. (2015). Prediction Policy Problems. *American Economic Review*, **105** (5), 491–495.

KNAUS, M. (2020). Double Machine Learning Based Program Evaluation Under Unconfoundedness. *arXiv:2003.03191*.

—, LECHNER, M. and STRITTMATTER, A. (2020a). Heterogeneous Employment Effects of Job Search Programs: A Machine Learning Approach. *Journal of Human Resources*, **forthcoming**.

—, — and — (2020b). Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence. *Econometrics Journal*, **forthcoming**.

KNITTEL, C. and STOLPER, S. (2019). Using Machine Learning to Target Treatment: The Case of Household Energy Use. *NBER Working Papers 26531*.

KOTLER, P. and LEVY, S. (1969). Broadening the Concept of Marketing. *Journal of Marketing*, **33**.

KOTTASZ, R. (2004). Differences in the Donor Behavior Characteristics of Young Affluent Males and Females: Empirical Evidence from Britain. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, **15** (2), 181?203.

LANDRY, C., LANGE, A., LIST, J., PRICE, M. and RUPP, N. (2010). Is a Donor in Hand Better Than Two in the Bush? Evidence From a Natural Field Experiment. *American Economic Review*, **100** (3), 958–983.

LIST, J. A. and PRICE, M. K. (2012). Charitable Giving Around the World: Thoughts on How to Expand the Pie. *CESifo Economic Studies*, **58** (1), 1–30.

MANSKI, C. (2004). Statistical Treatment Rules for Heterogeneous Populations. *Econometrica*, **72** (4), 1221–1246.

MEER, J. (2017). Does fundraising create new giving? *Journal of Public Economics*, **145**, 82–93.

MINDAK, W. and BYBEE, M. (1971). Marketing's Application to Fund Raising. *Journal of Marketing*, **35**.

MORGAN, J. (2000). Financing Public Goods by Means of Lotteries. *The Review of Economic Studies*, **67** (4), 761–784.

NAME-CORREA, A. and YILDIRIM, H. (2013). A Theory of Charitable Fund-Raising With Costly Solicitations. *American Economic Review*, **103** (2), 1091–1107.

NEWMAN, G. E. and SHEN, J. Y. (2012). The Counterintuitive Effects of Thank-You Gifts on Charitable Giving. *Journal of Economic Psychology*, **33** (5), 973–983.

OKTEN, C. and WEISBROD, B. (2000). Determinants of Donations in Private Nonprofit Markets. *Journal of Public Economics*, **75** (2), 255–272.

RAJAN, S., PINK, G. and DOWN, W. (2009). Sociodemographic and Personality Characteristics of Canadian Donors Contributing to International Charity. *Nonprofit and Voluntary Sector Quarterly*, **38** (3), 413–440.

RANGANATHAN, S. and HENLEY, W. (2008). Determinants of Charitable Donation Intentions: A Structural Equation Model. *International Journal of Nonprofit and Voluntary Sector Marketing*, **13** (1), 1–11.

ROBINS, J., ROTNITZKY, A. and ZHAO, L. (1994). Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association*, **89** (427), 846–866.

ROCKOFF, J., JACOB, B., KANE, T. and STAIGER, D. (2011). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance Policy*, **6** (1), 43–74.

ROSE-ACKERMAN, S. (1982). Charitable Giving and "Excessive" Fundraising. *The Quarterly Journal of Economics*, **97** (2), 193–212.

ROSENBAUM, P. and RUBIN, D. (1983). The Central Role of Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70** (1), 41–55.

RUBIN, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66** (5), 688–701.

SARGEANT, A. and LEE, S. (2002). Trust and Relationship Commitment in the United Kingdom Voluntary Sector: Determinants of Donor Behavior. *Psychology and Marketing*, **21** (8), 613–635.

— and WOODLIFFE, L. (2007). Gift Giving: An Interdisciplinary Review. *International Journal of Nonprofit and Voluntary Sector Marketing*, **12** (4), 275–307.

SCHLEGELMILCH, B. (1988). Targeting of Fundraising Appeals – How to Identify Donors. *European Journal of Marketing*, **22** (1), 31–40.

— and DIAMANTOPOULOS, A. (1997). Characteristics Affecting Charitable Donations: Empirical Evidence from Britain. *Journal of Marketing Practice: Applied Marketing Science*, **3** (1), 14–28.

SEMENOVA, V. and CHERNOZHUKOV, V. (2017). Simultaneous Inference for Best Linearpredictor of the Conditional Average Treatment Effect and Other Structural Functions. *arXiv:1702.06240*.

SHAPIRO, B. (1973). Marketing for Non-profit Organisations. *Harvard Business Review*, **3**.

SHELLEY, L. and POLONSKY, J. (1971). Do Charitable Causes Need to Segment Their Current Donor Base on Demographic Factors?: An Australian Examination. *International Journal of Nonprofit and Voluntary Sector Marketing*, **71** (1), 10–29.

SRNKA, K., GROHS, R. and ECKLER, I. (2003). Increasing Fundraising Efficiency by Segmenting Donors. *Australasian Marketing Journal*, **11** (1), 70–86.

SVERDRUP, E., KANODIA, A., ZHOU, Z., ATHEY, S. and WAGER, S. (2020). Policytree: Policy Learning via Doubly Robust Empirical Welfare Maximization over Trees. *Journal of Open Source Software*, **5** (50), 1–6.

TINKELMAN, D. and MANKANEY, K. (2007). When is Administrative Efficiency Associated With Charitable Donations? *Nonprofit and Voluntary Sector Quarterly*, **36**, 41–64.

VESTERLUND, L. (2003). The Informational Value of Sequential Fundraising. *Journal of public Economics*, **87** (3-4), 627–657.

WIEPKING, P. and JAMES, R. (2013). Why Are the Oldest Old Less Generous? Explanations for the Unexpected Age-Related Drop in Charitable Giving. *Ageing and Society*, **33** (3), 486–510.

YIN, B., LI, Y. and SINGH, S. (2020). Coins Are Cold and Cards Are Caring: The Effect of Pregiving Incentives on Charity Perceptions, Relationship Norms, and Donation Behavior. *Journal of Marketing*, **84** (6), 57–73.

ZADROZNY, B. (2003). *Policy Mining: Learning Decision Policies From Fixed Sets of Data*. Ph.D. Thesis,University of California, San Diego.

ZHAO, Y., ZENG, D., RUSH, A. and KOSOROK, M. (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, **117** (499), 1106–1118.

ZHOU, Z., ATHEY, S. and WAGER, S. (2018). Offline Multi-Action Policy Learning: Generalization and Optimization. *arXiv:1810.04778v2*.

ZIMMERT, M. and LECHNER, M. (2019). Nonparametric Estimation of Causal Heterogeneity Under High-Dimensional Confounding. *arXiv:1908.0877*.

Online Appendix to
”Optimal Targeting in Fundraising: A Machine Learning Approach”

Tobias Cagala, Ulrich Glogowsky,
Johannes Rincke, Anthony Strittmatter*

February 12, 2021

Sections:

A. Descriptives and Balance of Observables

B. Identification with AIPW Scores

C. Nuisance Parameters

D. Additional Results of the Sorted Effects Model

E. Sensitivity Analysis

*Cagala: Deutsche Bundesbank (tobias.cagala@bundesbank.de); Glogowsky: Johannes Kepler University Linz and CESifo (ulrich.glogowsky@jku.at); Rincke: University of Erlangen-Nuremberg (johannes.rincke@fau.de); Strittmatter: CREST-ENSAE, Institut Polytechnique Paris and CESifo (anthony.strittmatter@ensae.fr).

A Descriptives and Balance of Observables

Table A.1: Means and standard deviations of observable characteristics

	Warm-list		Cold-list	
	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Socio-economic characteristics				
Female dummy	0.53		0.50	
Single dummy	0.50		0.64	
Widowed dummy	0.05		0.02	
Age (in years)	68.51	18.30	48.40	19.32
Duration residency in urban area (in years)	7.43	1.67	5.97	2.82
Donation history before the experiment				
Number of donations previous 8 years	3.97	2.83	0	
Max. donations previous 8 years (in Euro)	36.02	42.90	0	
Total donations previous 8 years (in Euro)	125.9	176.0	0	
Donations 1 year ago (in Euro)	20.59	35.27	0	
Donations 2 years ago (in Euro)	17.23	29.29	0	
Donations 3 years ago (in Euro)	15.95	27.51	0	
Donations 4 years ago (in Euro)	15.82	27.59	0	
Donations 5 years ago (in Euro)	15.25	28.24	0	
Geo-spatial information about home address				
Elevation (in meters)	317.1	10.46	316.1	10.32
In 300 meters proximity:				
Number of restaurants	7.98	10.14	10.33	11.61
Number of supermarkets	1.08	1.36	1.29	1.50
Number of medical facilities	9.59	12.72	10.72	13.13
Number of cultural facilities	0.11	0.51	0.14	0.53
Number of churches	1.01	1.48	1.18	1.53
Distance to main station (in km)	3.25	2.11	2.86	2.02
Distance to city hall (in km)	3.11	2.00	2.79	1.88
Distance to main church (in km)	3.14	2.03	2.79	1.93
Distance to airport (in km)	5.46	1.75	5.55	1.64
Travel time to main station (in minutes)	17.81	9.20	16.13	8.66
Observations	2,354		17,425	

Notes: The donation history in the cold list is zero, because we only measure the donations to the specific fundraiser we cooperate with. The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays. For dummy variables the first moment is sufficient to infer the entire distribution.

Table A.2: Balance of observable characteristics in warm list

	Treatment group		Control group		Std.
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.
	(1)	(2)	(3)	(4)	(5)
Socio-economic characteristics					
Female dummy	0.53		0.53		1.22
Single dummy	0.50		0.49		2.04
Widowed dummy	0.05		0.06		3.98
Age (in years)	68.57	18.31	68.45	18.31	0.69
Duration residency (in years)	7.41	1.70	7.46	1.63	3.13
Donation history before the experiment					
Num. donations prev. 8 years	3.94	2.81	3.99	2.85	1.98
Max. don. prev. 8 years (in Euro)	36.49	46.63	35.54	38.81	2.22
Total don. prev. 8 years (in Euro)	126.64	181.52	125.21	170.31	0.82
Donations 1 year ago (in Euro)	21.19	39.73	19.98	30.13	3.41
Donations 2 years ago (in Euro)	17.28	29.55	17.18	29.03	0.32
Donations 3 years ago (in Euro)	16.11	28.13	15.80	26.88	1.13
Donations 4 years ago (in Euro)	16.30	28.79	15.34	26.34	3.48
Donations 5 years ago (in Euro)	14.83	28.53	15.67	27.95	2.96
Geo-spatial information about home address					
Elevation (in meters)	317.4	10.58	316.9	10.35	4.89
In 300 meters proximity:					
Number of restaurants	7.86	10.15	8.10	10.14	2.42
Number of supermarkets	1.11	1.37	1.05	1.36	4.83
Number of medical facilities	9.52	12.52	9.66	12.91	1.03
Number of cultural facilities	0.11	0.52	0.11	0.50	0.22
Number of churches	1.03	1.50	1.00	1.46	1.83
Distance to main station (in km)	3.24	2.04	3.25	2.18	0.59
Distance to city hall (in km)	3.09	1.93	3.12	2.06	1.15
Distance to main church (in km)	3.13	1.96	3.15	2.10	0.87
Distance to airport (in km)	5.44	1.77	5.49	1.74	2.47
Travel time to main station (in min.)	17.72	8.80	17.90	9.58	1.94
Observations	1'180		1'174		

Notes: The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays. For dummy variables the first moment is sufficient to infer the entire distribution. Rosenbaum and Rubin (1983) classify absolute standardized difference (std. diff.) of more than 20 as "large".

Table A.3: Balance of observable characteristics in cold list

	Treatment group		Control group		Std.
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.
	(1)	(2)	(3)	(4)	(5)
Socio-economic characteristics					
Female dummy	0.50		0.50		0.14
Single dummy	0.64		0.64		0.23
Widowed dummy	0.02		0.02		0.35
Age (in years)	48.34	19.25	48.41	19.33	0.34
Duration residency (in years)	5.96	2.82	5.97	2.82	0.36
Geo-spatial information about home address					
Elevation (in meters)	316.2	10.42	316.1	10.31	0.40
In 300 meters proximity:					
Number of restaurants	10.02	11.29	10.38	11.66	3.17
Number of supermarkets	1.31	1.50	1.29	1.50	1.14
Number of medical facilities	10.40	12.91	10.77	13.17	2.85
Number of cultural facilities	0.14	0.51	0.15	0.53	1.10
Number of churches	1.13	1.49	1.18	1.54	3.27
Distance to main station (in km)	2.89	2.02	2.86	2.02	1.69
Distance to city hall (in km)	2.81	1.88	2.79	1.88	0.90
Distance to main church (in km)	2.81	1.93	2.78	1.93	1.19
Distance to airport (in km)	5.55	1.65	5.55	1.64	0.07
Travel time to main station (in min.)	16.25	8.80	16.11	8.64	1.64
Observations	2'283		15'142		

Notes: The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays. For dummy variables the first moment is sufficient to infer the entire distribution. Rosenbaum and Rubin (1983) classify absolute standardized difference (std. diff.) of more than 20 as "large".

B Identification with AIPW Scores

To proof that $\delta = E[\Gamma_i]$ and $\delta(x) = E[\Gamma_i | X_i = x]$, it is sufficient to proof that $E[Y_i(1)|X_i = x] = E[\Gamma_i(1)|X_i = x]$ and $E[Y_i(-1)|X_i = x] = E[\Gamma_i(-1)|X_i = x]$. We focus on $E[\Gamma_i(1)|X_i = x]$ here:

$$
\begin{aligned}
E[\Gamma_i(1)|X_i = x] =& E\left[\mu_1(Z_i) + \frac{1+D_i}{2} \cdot \frac{Y_i - \mu_1(Z_i)}{p(Z_i)} \middle| X_i = x\right], \\
=& E\left[\frac{1+D_i}{2} \cdot \frac{Y_i}{p(Z_i)} \middle| X_i = x\right] + E\left[\left(p(Z_i) - \frac{1+D_i}{2}\right) \cdot \frac{\mu_1(Z_i)}{p(Z_i)} \middle| X_i = x\right], \\
=& E_{Z|X=x}\left[E\left[\frac{1+D_i}{2} \cdot \frac{Y_i}{p(Z_i)} \middle| X_i = x, Z_i = z\right]\right] \\
&+ E_{Z|X=x}\left[E\left[\left(p(Z_i) - \frac{1+D_i}{2}\right) \cdot \frac{\mu_1(Z_i)}{p(Z_i)} \middle| X_i = x, Z_i = z\right]\right], \\
=& E_{Z|X=x}\left[E\left[\frac{1+D_i}{2} \cdot \frac{Y_i}{p(z,x)} \middle| X_i = x, Z_i = z\right]\right] \\
&+ \underbrace{E_{Z|X=x}\left[E\left[\left(p(z,x) - \frac{1+D_i}{2}\right) \cdot \frac{\mu_1(Z_i)}{p(z,x)} \middle| X_i = x, Z_i = z\right]\right],}_{=0} \\
=& E_{Z|X=x}E\left[\left[\frac{1\{D_i = 1\}Y_i}{p(z,x)} \middle| X_i = x, Z_i = z\right]\right], \\
=& E_{Z|X=x}E[[Y_i | D_i = 1, X_i = x, Z_i = z]], \\
=& E_{Z|X=x}E[[Y_i(1) | D_i = 1, X_i = x, Z_i = z]], \\
=& E_{Z|X=x}E[[Y_i(1) | X_i = x, Z_i = z]], \\
=& E[Y_i(1) | X_i = x].
\end{aligned}
$$

In the first equality, we use the definition of $\Gamma_i(1)$. In the second equality, we just make a rearrangement. In the third equation, we apply the law of iterative expectations. In the fourth equality, we exploit that $p(z) = p(z,x)$ because of our experimental design (only the characteristics in Z_i have an impact on the probability to receive the gift). Note that the second right side term cancels, because $p(z,x) = E[(1+D_i)/2 | X_i = x, Z_i = z]$. In equality five, we replace $(1+D_i)/2$ with the indicator function $1\{D_i = 1\}$. In equality six, we make a backward application of the discrete law of iterative expectations. In equality seven and eight, we use $(Y_i(1), Y_i(-1)) \perp\!\!\!\perp D_i | Z_i = z$ the conditional independence assumption. In equality nine, we make a backward application of the law of iterative expectations. This finishes the proof that $E[Y_i(1)|X_i = x] = E[\Gamma_i(1)|X_i = x]$. The proof that $E[Y_i(-1)|X_i = x] = E[\Gamma_i(-1)|X_i = x]$ is analogous.

C Nuisance Parameters

Table C.1: Coefficients of nuisance parameters

	Logit		OLS			
	Donation dummy		Donations without gift		Donations with gift	
	(1)		(2)		(3)	
	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
Warm list						
Female dummy	-0.009	0.023	-3.67**	1.69	-5.00**	1.91
Single dummy	0.009	0.023	0.31	1.74	1.22	1.98
Widowed dummy	-0.041	0.050	-1.29	3.54	2.40	4.41
Age quintiles:						
2nd quintile	0.005	0.029	-3.96*	2.13	1.22	2.42
3rd quintile	0.004	0.030	-5.68**	2.20	-3.20	2.51
4th quintile	0.006	0.031	-2.61	2.31	1.12	2.64
Baseline willingness to donate quintiles:						
2nd quintile	-0.002	0.031	-29.64***	2.33	-25.76***	2.65
3rd quintile	0.001	0.030	-24.81***	2.23	-19.02***	2.55
4th quintile	0.003	0.030	-12.73***	2.23	-7.94***	2.54
Previous experiment dummy	0.006	0.028	-1.66	2.09	-3.25	2.36
Intercept	0.498***	0.031	37.13***	2.27	32.44***	2.59
Cold list						
Female dummy	-0.003	0.046	-0.002	0.04	0.18	0.20
Single dummy	-0.003	0.052	-0.10**	0.05	-0.28	0.23
Widowed dummy	0.027	0.176	-0.01	0.17	0.93	0.77
Age quintiles:						
2nd quintile	0.009	0.065	0.09	0.06	0.22	0.28
3rd quintile	0.002	0.067	0.10	0.06	0.03	0.29
4th quintile	0.001	0.068	0.20***	0.06	0.45	0.30
Previous experiment dummy	-0.004	0.046	-0.02	0.04	-0.17	0.20
Intercept	-1.891***	0.068	0.14**	0.07	0.32	0.30

Notes: Donations (in Euro) are measured during the first year after the gift was send. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

D Additional Results of the Sorted Effects Model

Table D.1: Mean characteristics of the groups with the 10% largest and smallest treatment effects in the warm list

	10% Largest		10% Smallest		Difference		
	Mean	Std.Err.	Mean	Std.Err.	Mean	Std.Err.	JP-val
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Socio-economic characteristics							
Female dummy	0.49	0.05	0.41	0.05	0.08	0.08	1.00
Single dummy	0.55	0.05	0.49	0.05	0.06	0.08	1.00
Widowed dummy	0.11	0.03	0.04	0.03	0.06	0.04	0.94
Age (in years)	70.84	1.72	67.45	1.90	3.39	3.10	1.00
Duration residency (in years)	7.28	0.21	6.93	0.24	0.36	0.32	1.00
Donation history before the experiment							
Num. donations prev. 8 years	3.64	0.27	4.42	0.29	-0.78	0.43	0.79
Max. don. prev. 8 years (in Euro)	55.05	6.05	70.13	8.36	-15.08	12.37	0.99
Total don. prev. 8 years (in Euro)	161.7	24.79	269.5	30.57	-107.9	43.85	0.37
Donations 1 year ago (in Euro)	25.53	4.94	50.67	6.96	-25.13	9.29	0.24
Donations 2 years ago (in Euro)	19.36	4.11	41.82	5.46	-22.46	7.81	0.17
Donations 3 years ago (in Euro)	26.21	3.74	36.08	4.88	-9.88	7.08	0.97
Donations 4 years ago (in Euro)	23.70	3.94	29.41	4.74	-5.71	7.22	1.00
Donations 5 years ago (in Euro)	21.33	4.22	31.81	4.52	-10.49	7.04	0.95
Geo-spatial information about home address							
Elevation (in meters)	315.4	1.10	321.3	1.51	-5.98	2.11	0.18
In 300 meters proximity:							
Number of restaurants	8.96	1.38	7.50	1.19	1.46	2.26	1.00
Number of supermarkets	0.79	0.15	1.12	0.14	-0.32	0.23	0.98
Number of medical facilities	9.21	1.54	9.42	1.46	-0.21	2.50	1.00
Number of cultural facilities	0.35	0.08	0.24	0.09	0.11	0.14	1.00
Number of churches	1.15	0.19	1.31	0.18	-0.16	0.29	1.00
Distance to main station (in km)	3.70	0.24	3.49	0.26	0.21	0.40	1.00
Distance. to city hall (in km)	3.57	0.24	3.30	0.24	0.28	0.37	1.00
Distance to main church (in km)	3.60	0.23	3.37	0.24	0.23	0.39	1.00
Distance to airport (in km)	5.87	0.17	5.38	0.16	0.49	0.23	0.58
Travel time to main station (in min.)	19.75	1.11	17.99	1.05	1.75	1.84	1.00

Notes: Mean values of the characteristics in the groups with the 10% largest and smallest effects. We report joint p-values (JP-val), which account for simultaneous inference on several characteristic. We employ the so-called “single-step” methods for controlling the family-wise error rate (see, e.g., Chernozhukov *et al.*, 2018, for details). Standard errors are calculated with a multiplier bootstrap using 500 replications. The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays.

Table D.2: Mean characteristics of the groups with the 10% largest and smallest treatment effects in the cold list

	10% Largest		10% Smallest		Difference		
	Mean	Std.Err.	Mean	Std.Err.	Mean	Std.Err.	JP-val
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Socio-economic characteristics							
Female dummy	0.71	0.09	0.49	0.05	0.22	0.13	0.77
Single dummy	0.52	0.11	0.61	0.05	-0.09	0.12	0.99
Widowed dummy	0.11	0.02	0.06	0.01	0.05	0.03	0.66
Age (in years)	55.75	3.75	51.84	1.71	3.92	4.15	0.99
Duration residency (in years)	5.63	0.62	6.55	0.22	-0.92	0.70	0.94
Geo-spatial information about home address							
Elevation (in meters)	320.3	1.26	318.8	1.01	1.48	1.53	0.99
In 300 meters proximity:							
Number of restaurants	10.09	1.51	10.09	1.22	0.00	1.85	1.00
Number of supermarkets	1.34	0.18	1.22	0.16	0.12	0.25	1.00
Number of medical facilities	18.73	2.97	11.54	1.83	7.19	2.92	0.28
Number of cultural facilities	0.09	0.05	0.39	0.06	-0.30	0.08	0.02
Number of churches	1.12	0.30	1.42	0.14	-0.30	0.36	0.99
Distance to main station (in km)	3.33	0.28	3.87	0.25	-0.54	0.31	0.76
Distance to city hall (in km)	2.63	0.32	3.64	0.26	-1.02	0.33	0.09
Distance to main church (in km)	2.91	0.30	3.71	0.27	-0.79	0.32	0.28
Distance to airport (in km)	4.24	0.27	5.42	0.18	-1.18	0.39	0.10
Travel time to main station (in min.)	16.51	0.96	20.19	1.25	-3.68	1.22	0.10

Notes: Mean values of the characteristics in the groups with the 10% largest and smallest effects. We report joint p-values (JP-val), which account for simultaneous inference on several characteristic. We employ the so-called "single-step" methods for controlling the family-wise error rate (see, e.g., Chernozhukov *et al.*, 2018, for details). Standard errors are calculated with a multiplier bootstrap using 500 replications. The duration residency in the urban area is censored after 8 years. We measure travel time to the main station using public transportation at 9:00am on weekdays.

E Sensitivity Analysis

Table E.1: Results for alternative estimators in the warm list

	Share Treated	Net Donations	Optimal Targeting vs. Everybody	Nobody
	(1)	(2)	(3)	(4)
Logit				
Baseline model	0.47	16.19***	0.72	0.78
		(0.90)	(0.81)	(0.82)
Flexible model	0.43	15.45***	-0.02	0.04
		(0.80)	(0.91)	(0.71)
Logit-Lasso	0.83	15.86***	0.39	0.45
		(0.91)	(0.62)	(0.98)
Exact policy learning tree				
depth = 1	0.39	15.05***	-0.42	-0.36
		(0.77)	(0.98)	(0.61)
depth = 3	0.34	15.49***	0.02	0.08
		(0.88)	(0.87)	(0.76)
CART				
depth =2	0.11	16.10***	0.63	0.68
		(0.94)	(0.93)	(0.69)
Cross-validated depth	0.33	17.40***	1.93**	1.98**
		(0.96)	(0.83)	(0.80)
Classification forest	0.42	15.88***	0.41	0.47
		(0.83)	(0.89)	(0.73)

Notes: The outcome variable is donation amount (in Euro) during the first year after the gift was send. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

Table E.2: Results for alternative estimators in the cold list

	Share Treated	Net Donations	Optimal Targeting vs. Everybody	Nobody
	(1)	(2)	(3)	(4)
Logit				
Baseline model	0.047	0.15***	0.96***	-0.01
		(0.04)	(0.10)	(0.03)
Flexible model	0.058	0.11***	0.93***	-0.05*
		(0.03)	(0.10)	(0.03)
Logit-Lasso	0.0003	0.16***	0.97***	-0.002
		(0.02)	(0.10)	(0.001)
Exact policy learning tree				
depth = 1	0.07	0.06***	0.87***	-0.10***
		(0.02)	(0.10)	(0.01)
CART				
depth =2	0.042	0.10***	0.92***	-0.06***
		(0.02)	(0.10)	(0.02)
Cross-validated depth	0.0006	0.16***	0.97***	-0.001**
		(0.02)	(0.10)	(0.0003)
Classification forest	0.001	0.15***	0.97***	-0.005
		(0.02)	(0.10)	(0.003)

Notes: The outcome variable is donation amount (in Euro) during the first year after the gift was send. In the cold list, implementing exact policy learning tree with a depth of three is computationally infeasible. Standard errors are in parentheses. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

References

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and LUO, Y. (2018). The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages. *Econometrica*, **86** (6), 1911–1938.

ROSENBAUM, P. and RUBIN, D. (1983). The Central Role of Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70** (1), 41–55.