Welfare analysis with nonlinear budget sets and evidence from a large charitable fundraising experiment

Marco Castillo* Ragan Petrie†‡

February 2025

Abstract

The compensating variation of a subsidy requires knowing an individual's preference over the choice set without a subsidy and the choice set with the subsidy and a reduced income. We show how to identify these preferences directly using nonlinear budgets and estimate the average excess burden of a subsidy. Our approach affords heterogeneity and is applicable when differential methods are difficult to implement, i.e. when subsidies create notches. We illustrate the approach in a large field experiment with charities using threshold matches that pay a fixed subsidy to the charity when a donation exceeds a threshold. Threshold matches cost-effectively increased out-of-pocket donations, but the excess burden of the subsidy increased in the threshold level. The excess burden of a $500 threshold match is about 45c per dollar of subsidy, whereas this is 9c for a $25 threshold match. We assess the desirability of threshold incentives against alternative subsidies.

Keywords: welfare, excess burden, compensating variation, notches, field experiment, charitable giving

JEL codes: D60, H41, D64, C93

*Castillo: Department of Economics, Texas A&M University, USA; IZA; CESifo Research Network, Germany, marco.castillo@tamu.edu

†Petrie: Department of Economics, Texas A&M University, USA; CESifo Research Network, Germany, rpetrie@tamu.edu

‡We gratefully acknowledge the support and guidance of the Coleman Foundation and the generous time given by the charities participating in the field experiment. The study is registered at the American Economic Association Social Science Registry (AEARCTR-0005918) and has IRB approval from Texas A&M University (IRB2016-0721M).

1 Introduction

A subsidy's compensating variation is the income reduction that makes the individual indifferent between the budget without the subsidy and the subsidy with a reduced income. Intuitively, we can learn preferences over these budgets if we let individuals choose between them or from their choices when the union of both sets is available and preferences are monotonic. Suppose an individual chooses from a budget that is the union of an original budget and one with a subsidy but a reduction in income. If the individual chooses a bundle that is only available on the portion of the subsidized budget, they reveal the reduction in income is not large enough to forgo the subsidy. If we observe a population choosing from sets constructed this way, as income is reduced, we could learn the distribution of income reductions that would make individuals indifferent between the subsidy with a reduced income and no subsidy at all. That is, we could learn the distribution of the compensating variation by observing individuals' choices over various nonlinear budget sets. We build on this intuition to non-parametrically identify the average compensating variation of subsidies. We then apply our result to the estimation of the compensating variation of threshold matches in a large fundraising field experiment with 26 charities and over 112,000 participants.

The textbook treatment of welfare evaluation of price changes (e.g. Mas-Collel et al., 1995, Section 3.I) discusses how to use a compensated demand to integrate over the changes in prices and recover the change in consumer's welfare (see Hausman, 1981, for an influential implementation of this approach). This is possible because the compensated demand of a good is the derivative of a money metric indirect utility function with respect to its price. Welfare measures, however, can be recovered from direct comparisons of choice sets as well (see Mas-Collel et al., 1995, p. 82). These direct comparisons are always feasible but are especially useful when differential methods are challenging to implement. This revealed preference approach underlies many elicitation tasks in experimental economics studies. Also, the approach is feasible in many applied problems as similar rankings are often recoverable from nonlinear budgets produced by policy changes. In this case, the structural object to recover is not a compensated demand, which might not be identified, but the proportion of people who are better off with the subsidies as income is reduced. A revealed preference argument is used in Bhattacharya (2015) to derive the welfare implications of

price changes in discrete choice problems. We exploit the same approach in nonlinear budget sets and obtain new identification results.

Our empirical application is threshold subsidies. There is a growing literature on the effects of notches (Kleven and Waseem, 2013; Kleven, 2016), and an open question is whether or not notches are desirable as a policy tool (e.g. Blinder and Rosen, 1985; Slemrod, 2013). Notches are common in many economic environments and are often accompanied by subsidies that are provided if a threshold is reached (see Kleven, 2016), i.e. home energy rebate programs, quantity discounts, gas guzzler tax, etc.[1] Notches in the choice set might be useful when individuals do not notice marginal prices (e.g. Ito, 2014), but as noted by Blinder and Rosen (1985), notches trade-off large excess burdens, i.e. deadweight losses, on a few individuals with small excess burdens on many individuals. We apply our approach to estimation of compensating variation to the case of charitable donations when there is a threshold match and use a field experiment to attain exogenous variation in these subsidies. This context is important since we know little about the efficiency costs of charitable giving fundraising schemes. Threshold matches create discontinuities in the budget set of donors by offering additional monies to charities if out-of-pocket donations meet or exceed a prespecified threshold.[2] We make progress in deriving welfare effects in this context because our approach does not require estimating a demand equation or elasticity.

The intuition of our approach is as follows. When the nature and distribution of preferences are unrestricted, how much donations exceeding a threshold vary with incentives is informative of the compensating variation of the threshold match. Intuitively, those whose donation does not meet a threshold match reveal that the match is not generous enough to increase their out-of-pocket donation. Those who respond to the match reveal their utility has improved. Consider now that the threshold is increased and the match decreased to keep the amount of money the charity receives constant if a donation meets the threshold exactly. As we discuss in the paper, this mimics a decrease in income, keeping the subsidy received by the charity constant.

[1] Notches are also common in social programs and other tax structures, i.e. Medicaid eligibility, social security, U.S. savers credit. In these cases, the subsidy is lost if income surpasses a threshold. This would result in bunching below the threshold. In settings such as home energy rebates or quantity discounts, bunching is above the threshold. We discuss income-based subsidies in Appendix A.

[2] An example of a threshold match offer to a donor is: if the donor donates $25 or more, the charity receives a fixed match of $25. Funds for matches can come from large donors and grants. This differs from popular dollar-for-dollar matches where every dollar donated is matched with another dollar.

The proportion of donors meeting the new threshold reveals that they would have met it with a less generous match. The less generous subsidy not only mimics a reduction in income but also reveals what the donor would have done if their income had been reduced. This implies that their compensating variation must be no less than the increase in their out-of-pocket donation.

We implement this approach as a field experiment. A field experiment guarantees independence between budget sets and preferences which is needed for identification. We implement our study in partnership with a private foundation and 26 charities in Chicago. We randomly assigned one of nine threshold incentive match offers, or a control of no match, to over 112,000 unique individuals and observed donation rates and amounts donated. We note that threshold match incentives can increase or decrease out-of-pocket donations. Donors may raise or lower their donations to meet a threshold since the additional dollars given above one threshold, but below another, are priced at one. With the additional dollars from the match, the total amount of money received by the charity could still increase, even when out-of-pocket donations have gone down. Importantly, donation behavior around the notches created by our field experiment allows for identification of the welfare impact of the subsidies.

The field experiment raised over $519,000 in donations, and an additional $250,000 was spent on matches paid to the charities by a private foundation. The average donation, conditional on making a gift, was $235 (20% higher than previous years), and the donation rate was 1.6%.[3] There is significant bunching at the treatment donation thresholds for each match offer. This indicates individuals shifted their donations to meet the threshold so the charity would receive a match.[4] The variation in bunching generated by our field experiment is essential to estimate the welfare effects of the subsidies.[5]

Our estimates are based on variations in the distribution of positive donations induced by the experiment treatments, combined with data on historical donations to the charities made without match incentives. We focus on positive donations for two reasons: there is not a significant extensive margin effect across treatments, and

[3]Donation rates for email charity fundraising campaigns such as ours are typically between 1-2%

[4]Reporting categories (Harbaugh, 1998) produce empirically similar donation behavior. Suggested amounts to donate or default donations also move donors to give the specified amount presented to them (Croson and Shang, 2008; Reiley and Samek, 2019; Altmann et al., 2019).

[5]Thresholds matches in this study were new to the participating charities. Subsequent campaigns using threshold matches confirm their effectiveness in fundraising (Castillo et al., 2023; Castillo and Petrie, 2020).

we do not have a way to assess the success rate of past donation appeals because we do not know the extent to which individuals were asked but did not donate. Our estimates using both data sources are very close to those obtained using only data from the experiment, but more precise. Using experimental data from all nine treatments and past donations, we obtain causal estimates of the probability of meeting or exceeding the threshold as a function of thresholds and matches by estimating a function with polynomials on both variables. We then use numerical integration to recover the expected value of the compensating variation of threshold matches. Confidence intervals are estimated with bootstrap.[6]

To illustrate our results, consider three distinct threshold matches: 1) $25 subsidy to the charity if the donation is equal or greater than $25, 2) $100 subsidy to the charity if the donation is equal or greater than $100, and 3) $500 subsidy to the charity if the donation is equal or greater than $500. The matches were chosen to lie along the budget set in which the price of donating is halved. These threshold matches can be considered a discretization of such a budget set. Our estimates (95% CI) of the excess burden on consumer welfare per dollar of subsidy are: $0.095 ($0.090, $0.099) for the $25 match, $0.329 ($0.323, $0.335) for the $100 match, and $0.458 ($0.434, $0.470) for the $500 match. This represents a sizeable variation in consumer welfare, especially for large subsidies. These results are robust to sample selection and strategic considerations by donors. An alternative method to measure excess burden is a linear approximation, assuming preferences are separable and incentives are offered separately. We show that this approach can under- and over-estimate welfare losses.

Our estimates show that the welfare effects of the subsidies are heterogeneous. The welfare cost is higher when donors are incentivized to donate large amounts, holding constant the ratio of the match to the threshold. The smaller excess burden of low threshold matches is partly because most donors would satisfy the threshold without the incentive, i.e., the subsidies do not alter the amount donated for these donors (see Boomhower and Davis, 2014). The key result from our analysis is that the excess burden increases in the threshold level, even accounting for infra-marginal donors. If high thresholds are more effective at increasing out-of-pocket donations, policymakers face a trade-off between increasing donations and decreasing donors'

[6]We sample donors with replacement and estimate each replication's bunching function and welfare impacts.

welfare. These considerations are important in the charitable sector where charities compete for donations (see Andreoni and Payne, 2003). While less effective or costlier for charities, other subsidy schemes, such as popular dollar-for-dollar matches, might be more attractive to donors. We discuss these trade-offs further in the paper.

It is known that welfare effects are not identified from behavior on linear budget sets when individual heterogeneity is allowed (Hausman and Newey, 2016, 2017). We overcome this identification problem by using nonlinear budget sets. With non-convex budget sets, we can recover preferences by observing behavior over the upper bound of the union of budget sets. As shown by Varian (1982) for the case of linear budget sets, the implicit preference relation over budget sets can be used to derive welfare effects. Our identification strategy is general and can be used to evaluate other subsidies. For instance, it is possible to identify the welfare effect of a linear match subsidy by varying the threshold at which the subsidy applies. Welfare is recovered by observing the change in subsidy claims as the threshold at which the subsidy applies varies. It is also possible to identify the welfare effect of an income-based lump-sum transfer from the observed response (i.e. bunching below the threshold income) when the transfer size is reduced, keeping the threshold income constant. The literature on revealed preference has long recognized the additional identifying power of nonlinear budget sets (e.g. Matzkin, 1991; Forges and Minelli, 2009), and our experiment adds to this literature.[7]

The main contribution of our paper is the development of a novel nonparametric estimate of the average compensating variation and excess burden of a subsidy for consumption, which we implement as a threshold match for charitable giving. Our results are based on a choice model that imposes minimal conditions on preferences.[8] Our approach exploits that nonlinear budget sets implicitly provide comparisons of alternative incentives, allowing us to identify preferences over the underlying budget sets. Because identifying welfare effects requires comparisons of budget sets, nonlinear budget sets offer a simple way to recover preferences over them. The added identifying power of nonlinear budget sets underlies our nonparametric identification results. The key structural parameter is how the probability of exceeding a threshold varies with

[7]Nonlinear budget sets can be used to detect non-convex preferences (Matzkin, 1991), test for rational behavior and play (Forges and Minelli, 2009; Cox et al., 2008), test for the separability of preferences (Castillo and Freer, 2023), and detect the presence of nuisance costs (Castillo et al., 2023).

[8]For estimation purposes, our application imposes ancillary conditions, i.e., convexity.

the location of the threshold and size of incentives. Variation in both dimensions is necessary for identification. The main insight of our approach is that to identify the compensating variation of the subsidy, we need to be able to identify policy changes that mimic changes in income. If those policy changes are available, we can define a structural choice probability that is used to estimate the distribution of the compensating variation of the subsidy.

Our paper also contributes to the recent literature on empirical welfare analysis (Kroft et al., 2020; Kamat and Morris, 2022; Hninn et al., 2017). This literature exploits Bhattacharya (2015)'s results on welfare analysis in discrete choice. Threshold matches resemble discrete choice in that they create discontinuities. However, our identification results differ because we exploit nonlinear budget sets' identifying power. Indeed, we can think of the nonlinear budget sets we create as affording to observe the kind of comparisons that would allow identification of welfare effects in ordered choice when only linear budgets are available (see Bhattacharya, 2015, Section 3). Our approach is also related to Deb et al. (2023), who introduce a preference relation over prices to evaluate the welfare implications of their change. Our results, however, stem from the nature of available choice sets, not preferences.

Our findings speak to the desirability of using notches as a policy tool. We take a first step in answering when notches are desirable by showing how to estimate welfare effects non-parametrically. We implement our approach in the context of charitable giving, the context in which Blinder and Rosen (1985) suggested a possible trade-off between fundraising and donors' welfare. We confirm their intuition.[9] We are unaware of other studies providing estimates comparing the welfare effect of notches.[10]

The literature on optimal taxation of charitable donations (Saez, 2004; Diamond, 2006) stresses the need to understand the welfare implications of subsidies for donations to determine optimal policies. Recent literature, however, shows practical and theoretical challenges to identifying the welfare impact of linear incentives for charitable donations (see Hungerman and Ottoni-Wilhelm, 2021). For instance, welfare effects of linear incentives cannot be identified from price elasticity in an impure im-

[9] Blinder and Rosen (1985) considered tax incentives that could be redirected to consumption. In our application, incentives can be used only for donations.

[10] We do not have access to participants' income, so we cannot construct a social welfare function to evaluate alternative policies (see Bhattacharya and Komarova, 2024; Bhattacharya, 2024, for a detailed discussion of this issue). This is a data limitation that might be irrelevant in other applied contexts.

pact giving model. Because our approach would identify the welfare effect of a linear match from the reaction to marginal subsidies as the qualifying threshold varies, and not from the price and income elasticities, we can overcome the identification problem identified by Hungerman and Ottoni-Wilhelm (2021). This shows that our approach can help deal with thorny identification issues in this literature. Linear consumption subsidies that are triggered when meeting a threshold are frequently observed in policy environments and threshold matches are an example of this in the nonprofit sector.[11]

Another contribution of our paper is to the literature on charitable fundraising (e.g., List and Lucking-Reiley, 2002; Eckel and Grossman, 2008; Landry et al., 2006; Karlan and List, 2007; Landry et al., 2010; Huck and Rasul, 2011; Huck et al., 2015; Adena and Huck, 2022; Hungerman and Ottoni-Wilhelm, 2021; Castillo et al., 2023). Much of this literature focuses on the motives to give, the effectiveness of alternative fundraising schemes, and donor retention. Research, however, has shown that charitable giving is not always welfare enhancing (Andreoni et al., 2017; DellaVigna et al., 2012; Butera et al., 2022). Our approach measures welfare losses among individuals who derive positive utility from contributing to charitable causes but suffer a loss in welfare because of the nature of the subsidies. The welfare measures we derive are useful in evaluating the optimal design of subsidies (Saez, 2004; Diamond, 2006) and are in addition to welfare losses from social pressure (DellaVigna et al., 2012).

A key component of our approach is estimating the structural choice equation, i.e., donors' behavior as a function of changes in nonlinear budgets. This is, in principle, a formidable task (see Blomquist et al., 2023). As discussed in more detail in the paper, we do not need to identify average demand; we only need to identify a threshold-crossing equation. We show how to exploit the maintained assumptions on individual preferences to identify threshold-crossing of a single notch from the reaction to multiple notches. This allows us to produce a more efficient experimental design.

Our paper does not discuss estimation issues extensively because our application is straightforward, however there are estimation issues worth mentioning. A researcher has limited access to policy changes in most applications, including ours. It would be important to derive nonparametric bounds on the structural choice probabilities

[11]The threshold matches we analyze differ from those analyzed in Kleven and Waseem (2013) in that our incentives do not create a set of dominated choices.

based on the maintained assumptions on preferences as discussed in Hausman and Newey (2017) and Blomquist et al. (2023). These implications will likely depend on the application but would add to the credibility of results based on finite variations in policies. While we do not implement this bound approach, we discuss how this could be implemented from available results in the literature. In particular, we discuss how recent results on nonparametric analysis of random utility models (Kitamura and Stoye, 2018; Deb et al., 2023) can be combined with testable implications of convex preferences (Cherchye et al., 2014) to derive bounds on the structural choice equation. These inequalities can be used to extrapolate behavior from a finite set of relevant policies.

The study was implemented as a natural field experiment (Harrison and List, 2004). Donors were not aware they were part of a study and were offered threshold matches as they normally would receive in a fundraising campaign. This maximizes the naturalness of our setting (List, 2020). Attrition is not an issue since we have administrative records of donations made to the charities during the experiment and in the past. The study is based on a large set of donors and a varied set of 26 charities, in terms of mission and size, providing evidence that is more likely to replicate in other settings. Importantly, our results are applicable to other environments where consumption subsidies create discontinuities in choice sets.

The paper is organized as follows. Section 2 outlines the identification and estimation of welfare effects of threshold matches. Section 3 describes our estimation strategy. Section 4 explains the experimental design and implementation, and Section 5 discusses results from the field experiment. Section 6 presents welfare estimates. Section 7 discusses robustness analyses. Section 8 concludes.

2 Theory

This section shows how nonlinear budget sets can be used to identify the average compensating variation of a subsidy non-parametrically and, consequently, the excess burden on welfare. Appendix A provides proofs of these results. Section 2.1 discusses the general approach. Section 2.2 applies the general approach to the specific case of a threshold match in charitable giving. Section 2.3 discusses the application to multiple thresholds.

2.1 Identification

2.1.1 Budgets and preferences

We specialize in the case of a target good g and composite numeraire good c. We assume that budgets have a non-empty interior, are bounded, downward closed,[12] and can be described by an increasing, continuous function $h : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ such that $B = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : h(x) \leq 0\}$ (see Forges and Minelli, 2009). In the case of a linear budget set, we have $h(x) = px - y$ with $(p, y) \gg 0$. We let $u(c, g; \eta)$ represent the utility of consuming (c, g) and assume it is continuous, monotone increasing[13] and strictly quasi-concave.[14] We allow for unobserved individual heterogeneity by letting utility depend on a vector η of unknown dimension. We assume there is an absolutely continuous function, $F(\cdot)$, representing the marginal distribution of η.

2.1.2 Revealed preferences

The pair (B_k^i, x_k^i) is an observation for individual i in choice problem k. We interpret $x_k^i \in B_k^i$ as the bundle chosen when budget B_k^i was available such that $h_k^i(x_k^i) = 0$. We denote a finite collection $D_i = (B_k^i, x_k^i)_{k=1,\ldots,K_i}$ a dataset for individual i. We suppress the superscript i when there is no risk of confusion. A data set D_i is rationalized by the utility function $u : \mathbb{R}_+^2 \to \mathbb{R}$ if, for all $k = 1, \ldots, K_i$, $x_k^i \in argmax_{x \in B_k^i} u(x)$. We say that x_k is revealed preferred to x_j, or $x_k R x_j$, if $h_k(x_j) \leq h_k(x_k) = 0$. Let H be the transitive closure of the relation R. We follow Forges and Minelli (2009) in defining the Generalized Axiom of Revealed Preference (GARP) in the setting with nonlinear budget sets.

Definition (GARP): The dataset $D = (x_k, B_k)_{k=1,\ldots,K}$ satisfies GARP if, for any k, j, $x_k H x_j$ implies $h_j(x_k) \geq 0$.

Under these conditions, if a dataset D_i satisfies GARP, an increasing and continuous function exists that rationalizes the dataset D_i (Forges and Minelli, 2009; Cherchye et al., 2014). The following results underlie our identification strategy.

Observation 1: Let (B_j, x_j) and (B_l, x_l) be two observations such that $B_j \cap B_l \subsetneq$

[12] That is, if $y \in B$ then $y' \in B$ for every $y' \leq y$.

[13] A function $f : A \subseteq \mathrm{R}^n \to \mathrm{R}$ is weakly monotone increasing, or non-decreasing if $f(x) \leq f(y)$ when $x \leq y$; and monotone increasing if it is weakly monotone increasing and $f(x) < f(y)$ when $x \ll y$.

[14] Strict quasi-concavity is unnecessary for the main identification results. It is used to aid identification in the case of multiple threshold matches.

B_j and let (B_m, x_m) for $B_m = B_j \cup B_l$ be the observation when the union of both sets is available. If $x_m \in B_l \setminus B_j$, then $x_l \in B_l \setminus B_j$.

Proof: See Appendix A.

Observation 1 considers the case where two budget sets intersect, but B_l is not contained in B_j. GARP implies that the choice from B_l should not belong to the intersection of the budget sets. The observation states that if the choice in the union of the sets is also in B_l, then this reveals the behavior in B_l as well. This choice does not reveal what an agent would do in budget B_j.

Definition (Preferences over budget sets): We say that budget set B_l is preferred to budget set B_j by individual i, denoted $B_l \succeq_i B_j$, if there is a $x \in B_l \setminus B_j$ such that for all $y \in B_j$, xHy.

Observation 2: Let (B_j, x_j) and (B_l, x_l) be two observations such that $B_j \cap B_l \subsetneq B_j$ and let (B_m, x_m) for $B = B_j \cup B_l$ be the observation when the union of both sets is available. If $x_m \in B_l \setminus B_j$, then $B_l \succeq_i B_j$.

Proof: See Appendix A.

Given our definition of the revealed preferred relation, Observation 2 directly implies that budget set B_l is preferred to B_j since there is an element in B_l which is preferred to all elements in B_j. The main insight is that we can determine individual preferences over budget sets from choices on their union and do not need to observe choices on each individual budget set. We can then determine the proportion of individuals who are better off from their choices. We use this result in the next section to estimate welfare effects.

Definition (Structural choice probability): Let $\mathcal{N} = \{1, \ldots, N\}$ be a random sample of individuals choosing from budget $B_m = B_l \cup B_j$. Define the structural choice probability as $q_N(B_l, B_j) = N^{-1} \sum_{i \in \mathcal{N}} 1[x^i \in B_l \setminus B_j]$, where $1[A]$ equals 1 if statement A is true. Let $Pr(B_l \succeq_i B_j)$ be the probability that budget B_l is preferred to budget B_j in the population from which sample $\mathcal{N}$ is drawn.

Observation 3: $lim_{N \to \infty} E[q_N(B_l, B_j)] = Pr(B_l \succeq_i B_j)$.

Proof: See Appendix A.

Observation 3 shows that the structural choice probability is a voting criterion over choice sets that can be consistently estimated from experimental data.

2.1.3 Welfare effects

We now show how to estimate the compensating variation of a subsidy. We define a budget set as $B(y, p, \zeta)$ where y denotes income, p prices, and $\zeta \in A \subset \mathbb{R}^K$, A compact, is a vector describing the subsidies. We assume that $B(y, p, \zeta) = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : h(x|y, p, \zeta) \leq 0\}$ with $g : \mathbb{R}_+ \times \mathbb{R}_+ \times A \to \mathbb{R}$ an increasing, continuous function in the first two arguments. Note that, for expositional reasons, we now make explicit that function h depends on parameters (y, p, ζ). For instance, in the case of threshold matches, that we discuss in detail in Section 2.2, we have that $\zeta = (\tau, m)$ when incentives are present and $\zeta = (0, 0)$ when they are absent. Here, τ is the minimum amount needed to be spent on target good g to get an additional amount of good g of size m.

Let the initial set of incentives be ζ_0 and consider a set of incentives ζ_1 where $B(y, p, \zeta_0) \subset B(y, p, \zeta_1)$. This assumption reflects that we are dealing with a subsidy, which we usually associate with larger choice set. Equation (1) defines the compensating variation (CV).

$$CV = \{Z : max_{(c,g)\in B(y,p,\zeta_0)} u(c, g; \eta) = max_{(c,g)\in B(y-Z,p,\zeta_1)} u(c, g; \eta)\} \tag{1}$$

CV is the reduction in income needed to make the agent indifferent between the original incentives ζ_0 and the new incentives ζ_1 and income $y - CV$. The previous discussion shows that those who choose a bundle not in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$ reveal that $B(y, p, \zeta_0) \succeq B(y, p, \zeta_1)$ and those who choose a bundle in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$ reveal that $B(y, p, \zeta_1) \succeq B(y, p, \zeta_0)$. This implies that those who choose a bundle not in $B(y, p, \zeta_0) \setminus B(y, p, \zeta_1)$ have a compensating variation equal to zero. We show next how to construct a family of budget sets to reveal the compensating variation of those who choose a bundle in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$.

We define a family of budget sets ordered by reductions in income Δ: $\{B_\Delta\}$ where $B_\Delta = B(y, p, \zeta_0) \cup B(y - \Delta, p, \zeta_1)$. Given our assumption, we have that $\bigcap_{\Delta\geq 0} B_\Delta = B(y, p, \zeta_0)$, i.e. the choice set reverts to the original budget set as income is reduced. Consider those choosing an option in $B(y - \Delta, p, \zeta_1) \setminus B(y, p, \zeta_0)$ when $\Delta = 0$, i.e. they changed their behavior due to the incentives. We can conclude that their CV is not negative. This is due to revealed preferences and preferences being monotone. Suppose $\Delta > 0$, i.e., available options are reduced, we can conclude that those who switch to a bundle in $B(y, p, \zeta_0)$ not in $B(y - \Delta, p, \zeta_1) \setminus B(y, p, \zeta_0)$ must have a CV no

larger than Δ. A choice in $B(y, p, \zeta_0)$ is preferred to all choices in $B(y, p-\Delta, \zeta_1)$. Note that the utility obtained by this new choice must be equal to that of the optimal choice in $B(y, p, \zeta_0)$ in isolation.[15] Due to the monotonicity and continuity of preferences, this switch will happen when Δ is large enough.[16] The sequence of budgets B_Δ allows us to find the reduction in income that makes a person indifferent between the original choice and a choice with the new incentives but less income.

We define the structural choice probability as:

$$q(\Delta, \zeta_1, \zeta_0) = \int_\eta \mathbf{1}\{\{argmax_{(c,g)\in B_\Delta} u(c, g; \eta)\} \cap \{B_\Delta \setminus B(y, p, \zeta_0)\} \neq \emptyset\} dF(\eta) \quad (2)$$

This measures the number of people who benefit from ζ_1 as Δ increases. By the definition of B_Δ there is a Δ' large enough such that $B_{\Delta'} = B(y, p, \zeta_0)$, and there is a Δ' such that everyone reverts to their original choices with the original subsidy. We then have that $1 - q(\Delta, \zeta_1, \zeta_0)$ reveals the probability that the CV is no higher than Δ. By construction, $q(\Delta, \zeta_1, \zeta_0)$ is decreasing in Δ, so $1 - q(\Delta, \zeta_1, \zeta_0)$ provides a distribution of the CV. This leads to the following proposition.

Proposition 1: Let ζ_1 be a subsidy such that no agent is worse off with respect to original subsidy ζ_0. Let $q(\Delta, \zeta_1, \zeta_0)$ be the structural choice probability as defined above. Let $E[CV|\zeta_1, \zeta_0]$ be the average compensating variation of subsidy ζ_1. Then, $E[CV|\zeta_1, \zeta_0] = \int_{\Delta\geq 0} q(\Delta, \zeta_1, \zeta_0) d\Delta$.

Proof: See Appendix A.

The discussion above has concentrated on estimating the compensating variation of subsidies. The approach introduced in this paper can also be adapted to calculate the equivalent variation of the subsidy. To calculate the equivalent variation, we need, instead, to define a family of budget sets ordered by increments in income Δ: $\{B_\Delta\}$ where $B_\Delta = B(y + \Delta, p, \zeta_0) \cup B(y, p, \zeta_1)$.

We conclude this section by remarking that the ability to identify the average compensating variation stems not from assumptions on preferences but from access to behavior in a richer family of choice sets. As we will show, these richer choice sets

[15] If not, the original choice was not optimal since the new choice was available but not chosen.

[16] The maximum is continuous and decreasing in Δ since the constraint correspondence varies continuously in Δ and preferences are continuous and monotone increasing (see Mas-Collel et al., 1995, Theorem M.K.6). There is at least one value of Δ at which utility equals the optimal utility at ζ_0 since $\bigcap_{\Delta\geq 0} B_\Delta = B(y, p, \zeta_0)$.

occur naturally.

2.2 Threshold match subsidies

We illustrate the usefulness of the approach developed above in analyzing the welfare effect of threshold match subsidies in the context of charitable giving. With a threshold match subsidy, when an individual makes a donation at or above a certain dollar threshold, the charity receives a fixed dollar subsidy.[17] We consider an individual consuming two goods: a composite good (c) and donations received by a charity (g). We let $u(c, g; \eta)$ represent the donor's utility function over bundle (c, g). η is a vector of an unknown dimension representing individual preference heterogeneity (see Hausman and Newey, 2016). We do not restrict the way η enters the utility function. We assume preferences are continuous, monotone increasing, and strictly quasi-concave.

Figure 1 represents a consumer choice when a threshold match is available. The y-axis is consumption, and the x-axis is total donations received by the charity, i.e. the out-of-pocket donation by the consumer and any match given to the charity. A threshold match offers a match $m(\tau)$ if an out-of-pocket donation equals or exceeds τ. Without a match, the consumer faces the linear budget constraint in black, attains utility level I_0 and donates g^* to the charity. When the match is available, this creates a notch in the budget constraint. At the threshold τ, the budget shifts to the right by the size of the match, $m(\tau)$. The black budget line shows this for donations below τ, and the new budget line is in blue for donations $\geq \tau$. The donor attains utility level I_1 and donates τ to the charity, i.e. "bunches" at the threshold. The charity receives $\tau + m(\tau)$.

The compensating variation (CV) for a price decrease is the amount of income that would need to be taken from the consumer at the new prices to return to their original utility. With threshold matches, the new budget to calculate the CV is defined by two variables: a threshold out-of-pocket donation (τ) and the match the charity receives if the threshold is met ($m(\tau)$). The CV of the match subsidy is the income the consumer would give up to return to the utility level without the match (I_0). Equation (3) provides a formal definition of the CV for a donor with income y, prices normalized to be equal to one and the charity receiving a threshold match of $m(\tau)$ if a donation meets or exceeds τ. $\mathbf{1}\{A\}$ is an indicator function that equals

[17] For example, a threshold match is: if an individual donates \$100 or more, the charity receives a fixed match of \$100. Matching funds can come from large donors and grants.

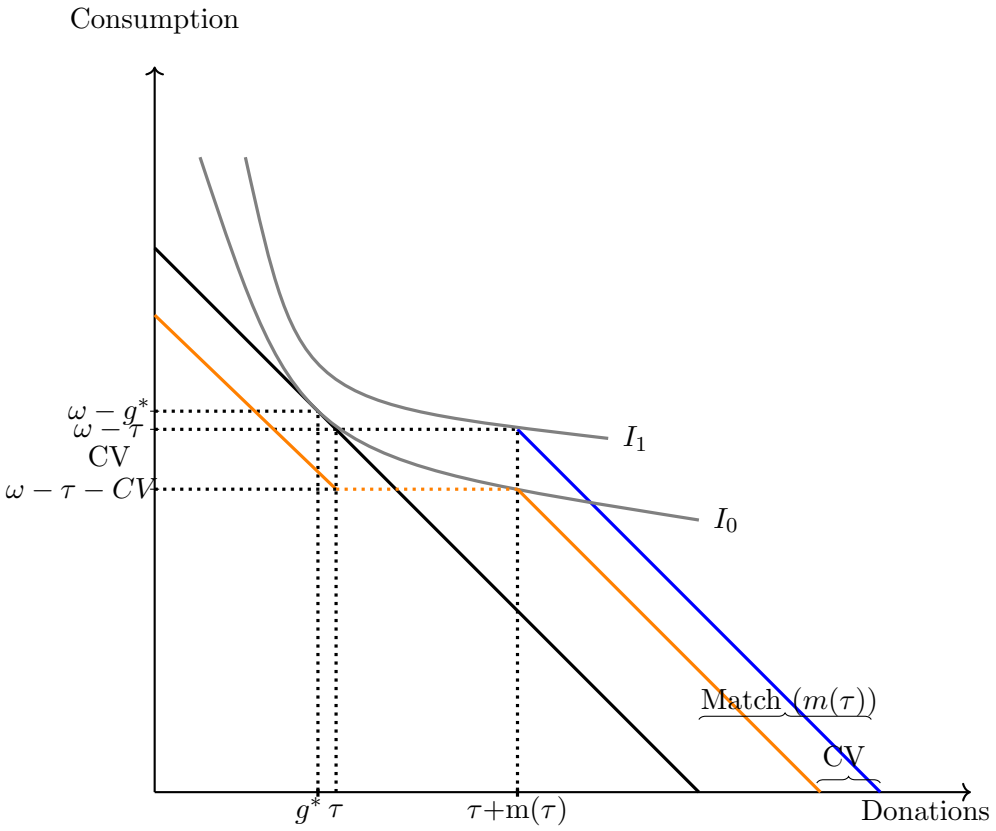


Figure 1: IDENTIFYING CV

1 if statement A is true and zero if not. This is graphically represented in Figure 1 by reducing the donor's income by CV without eliminating the threshold match, i.e., moving from the blue line to the orange line. The donor attains the same level of utility (I_0) at the new budget with threshold matches and reduced income as at the original budget without threshold matches.

$$CV = \{Z : max_{(c,g):c+g\leq y} u(c,g;\eta) = max_{(c,g):c+g\leq y-Z+\mathbf{1}\{g\geq\tau+m\}m(\tau)} u(c,g;\eta)\} \quad (3)$$

As explained in Section 2.1.3, the same information can be recovered from how a donor reacts to specific changes in threshold matches. Assume we could observe the behavior of a consumer to a new threshold match. The new threshold match offer is a match of $m(\tau) - CV$ that is given to the charity if the consumer's out-of-pocket donation equals or exceeds $\tau + CV$. This is the vertical shift in the budget constraint

from blue to the orange lines in Figure 1. This new threshold match keeps constant the amount of money that the charity receives ($\tau + m(\tau)$) but increases the out-of-pocket donation the consumer makes and, thus, reduces the match subsidy paid to the charity. This policy change, therefore, mimics a reduction in income.

In other words, CV is the maximum amount the consumer would pay to retain the original utility level and still receive the threshold match for the charity. The same arguments go through for donors who would have donated $\tau < g^* < \tau + m(\tau)$ without incentives and bunch at (or exceed) τ with incentives.[18] These donors prefer donation τ and the charity receiving $\tau + m(\tau)$ than their original donation plus $m(\tau)$. That is, they increase the amount the charity receives and reduce their out-of-pocket donation. Monotonicity of preferences implies that they revert to their original choice for a large enough reduction in the match. Figure B.1 illustrates these cases.

Suppose we observe the distribution of donations as a function of threshold matches holding constant the amount of money a charity receives. In particular, consider we modify threshold τ and match $m(\tau)$ to be instead $\tau + s$ and $m(\tau) - s$. Those who donate at least $\tau + s$ when the match is $m(\tau) - s$ reveal that their compensating variation is at least s. Those who switch from donating τ to their previous donation level reveal their compensating variation is less than s. One minus the proportion of donors reaching the threshold out-of-pocket donation as the match becomes less generous then reveals the cumulative distribution of the compensating variation.[19] Such a distribution can be used to calculate the expected compensating variation of a threshold match and the associated excess burden of the subsidy. This intuition is stated formally below.

Proposition 2: Let τ_i and $m(\tau_i)$ be a threshold match policy such that $\tau_i + m(\tau_i) = G$. Let $s_i \in [0, m(\tau)]$. Define the structural choice probability $q(G, s_i) = P_N(g_i + \mathbf{1}[g_i \geq \tau_i + s_i](m(\tau_i) - s_i) \geq G)$ as the probability that total donations exceed amount G when a match of $m(\tau_i) - s_i$ is awarded to any out-of-pocket donation exceeding $\tau_i + s_i$. Let $E[CV|G, m]$ be the expected value of the compensating variation of a threshold match with threshold $G - m$ and match m. Then, $E[CV|G, m] = \int_0^m q(G, x)dx$. Consequently, the expected excess burden of the subsidy is $E[CV|G, m] - m \times q(G, 0)$.

[18] The argument does not change for those who increased donation from below τ to above it. We do not require these donors to bunch at τ.

[19] To be precise, we are describing the donor's willingness to pay for the original match $m(\tau)$. Appendix A shows the equivalence.

Proof: See Appendix A.

The central intuition behind this result is that both thresholds and matches can be changed to mimic changes in income and, therefore, identify the CV. We need to observe changes in behavior as we alter thresholds and matches but keep their sum constant. Our experimental design randomly assigns threshold-subsidy combinations to individuals, thus providing the needed choice data to estimate the structural choice probability and CV.

This approach can be adapted to other subsidies, such as marginal subsidies to donate. A marginal subsidy is defined by a threshold $\tau \geq 0$ and subsidy $s > 0$, which reduces the price of each dollar above τ by s.[20] Figure B.2 shows how this could be done. A decrease in income, without changing the subsidy, mimics raising the minimum donation to receive the marginal subsidy. Suppose preferences for donations are a normal good. In that case, total donations (out-of-pocket donations plus the subsidy) will weakly increase for those below the threshold and strictly increase for those above the threshold.[21] Convexity of preferences implies that there is a threshold $\tau^* > \tau$ at which donors revert to their original donations. The observed probability of exceeding the threshold as the threshold is increased, therefore, reveals the (decumulative) distribution of the CV of the original subsidy.

2.3 Multiple thresholds

The analysis thus far has shown how to point identify the compensating variation for one threshold. In practice, individuals often face multiple threshold-subsidy pairs, i.e. charitable giving threshold matches, energy rebate programs. In the case of multiple thresholds, the compensating variation is obtained by simultaneously modifying all thresholds and incentives in the set of incentives. This is done by increasing all thresholds and reducing all matches by the same amount to mimic the analysis in Section 2.2. As before, subsidies cannot turn into taxes, i.e., they are bound by the original budget line (see Figure 1).

Multiple threshold matches create complications in estimating structural parameters because decisions depend on the shape of the entire budget (Blomquist et al.,

[20] When $\tau = 0$, this is the popular dollar-for-dollar match that many charities offer as an incentive to donate in fundraising campaigns.

[21] A violation of WARP can be constructed for these donors if total donations do not increase (see Cherchye et al., 2020).

2023). For instance, choosing one threshold might depend on the position and incentives of other thresholds. We show how to simplify this by exploiting the maintained assumptions on preferences.

Figure 2 depicts a budget with two threshold matches: τ_1 and τ_2. Any out-of-pocket donation exceeding τ_1, but less than τ_2, receives a match equal to τ_1. Any out-of-pocket donation exceeding τ_2 receives a match equal to τ_2. Consider the donor who would donate an amount $g_0 < \tau_1$. This person might find total donation (out-of-pocket donation plus match) g_1 to be preferred to g_0. In this example, if a donation in segment C of the budget dominates a donation in segment A of the budget, then from convexity, there is a donation in segment B that also dominates it (e.g., $\overline{g}$). This can be seen by observing that parts of segments B and C lie above the blue line that is a linear combination of donation g_0 and g_1. Anything above the blue line is strictly preferred. This implies that information about segment B is sufficient to determine if a donor has an incentive to reach a threshold. Information about segment C is redundant.[22] This says that the behavior of the donor, who would have donated $g_0 < \tau_1$, facing only threshold τ_1 is observationally equivalent to that of a donor facing thresholds τ_1 and τ_2. As discussed in Blomquist et al. (2023), the convexity of preferences implies that optimal behavior in a non-convex segment of the budget captures the behavior over the whole budget.

This representation suggests that we can model behavior based on the threshold to be attained, the marginal match to be attained, and the preceding threshold. In practical terms, we can use several threshold matches to identify the behavior of a threshold match in isolation if we place the threshold matches along the same budget line. We use this approach in the design of our experiment and for estimation to fully exploit the limited variation in incentives in our experiment. We show that our simplified model fits the data well.

A distinct advantage of our estimation approach is that it can be adapted to account for intertemporal substitution. For instance, suppose a donor faces two sets of threshold incentives over time. The relevant budget set for a donor trying to maximize the total amount donated over the two periods will be a budget that reflects the maximum match attainable for each pair of donations adding up the same out-of-pocket

[22] While it is redundant to examine choices in segment A that are dominated, it is not redundant if decisions depend on how salient incentives are. Since we do not know how preferences are distributed, this is not without loss of generality. We discuss this point further.

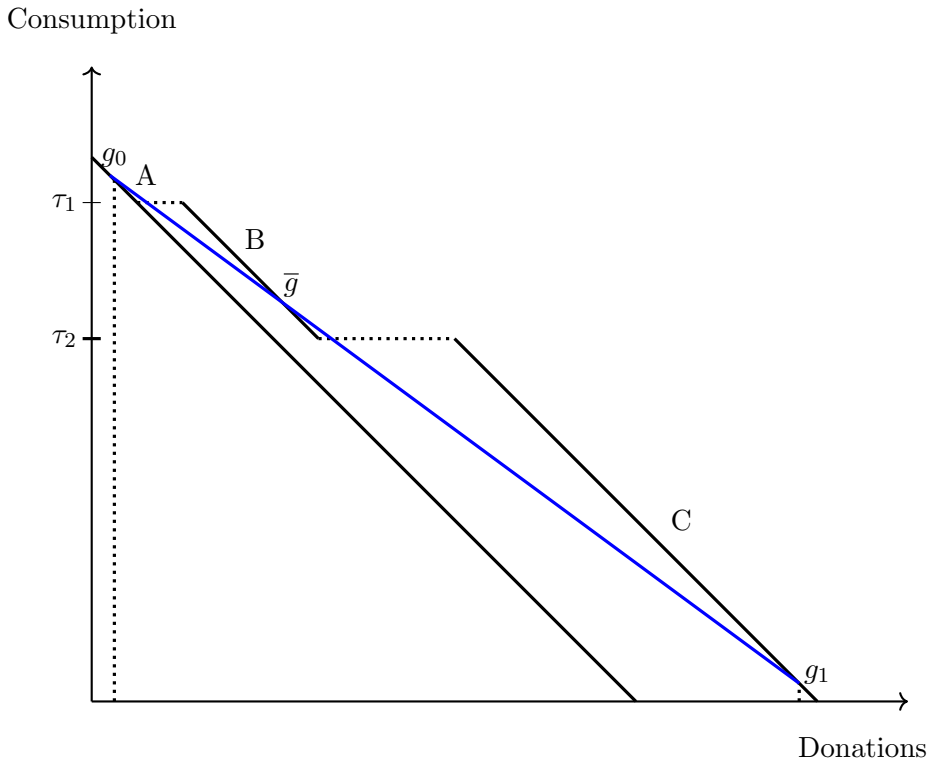


Figure 2: CHOICE DIMENSION REDUCTION UNDER CONVEXITY OF PREFERENCES

donations.[23] While the budget sets are more complex, they share the properties of the static problem and can be used to estimate responses to matches when donors intertemporally substitute. We find little support in our setting for the assumption of intertemporal substitution.

3 Estimation

The main object of interest is the probability of reaching or exceeding a threshold match as a function of the threshold and match: $Pr(g_i + \mathbf{1}[g_i > \tau_i]m_i \geq G)$ where $G = \tau_i + m_i$. Access to this structural choice probability allows us to estimate the compensating variation and excess burden. While, in principle, we can estimate this object non-parametrically, we use a flexible parametric model to increase precision and for extrapolation.[24] In particular, we posit that $Pr(g_i + \mathbf{1}[g_i > \tau_i]m_i \geq G)$ can be

[23] Figure B.3 exemplifies this construction.

[24] For comparison, in the Results section, we present non-parametric consumer surplus estimates using linear interpolation. These estimates can be interpreted as excess burden under the assumption

approximated by a polynomial on m_i and G. Following the discussion in the previous section, if there is an ordered set of matches and donation thresholds $\{(m_k, G_k = \tau_k + m_k)\}_k$, we assume that $Pr(g_i + \mathbf{1}[g_i > \tau_k]m_k \geq G_k) = f(m_k - m_{k-1}, G_k)$ with $m_0 = 0$. That is, donating above threshold G_k depends on the marginal incentive to exceed it, $m_k - m_{k-1}$, and G_k.[25]

To implement this approach, we define a series of dummy variables $Y_{k,i}$ that equal 1 if an out-of-pocket donation exceeds G_k without incentives or an out-of-pocket donation exceeds $G_k - m_k$ with incentives and equals 0 otherwise. We regress $Y_{k,i}$ on a polynomial of the log of G_k and the interaction of a polynomial of the log of G_k and $\frac{m_k - m_{k-1}}{G_k}$. The marginal match is divided by G_k to standardize the subsidy relative to the threshold. We use polynomials because they allow us to extrapolate behavior to thresholds and matches unavailable in the experimental data. Polynomials of order three are estimated to account for nonlinearities. To account for repeat measures per donor, we cluster standard errors by donor.

This estimation approach follows that in Pop-Eleches and Urquiola (2013) who note that individuals facing several thresholds, in the context of educational decisions, can be simultaneously below and above different thresholds. The approach exploits all the information on donation behavior available to us. We use $Pr(g_i + \mathbf{1}[g_i > \tau_i]m_i \geq G)$, the estimated structural choice probability, to estimate welfare parameters based on $q(G, m_i, s_i)$. To calculate standard errors, we use bootstrap to generate samples where we re-estimate the model and corresponding welfare parameters. The proposed approach is simple to implement and guarantees that the probability of exceeding a threshold is monotonic in the incentives to do so.

Identification: Calculating the welfare effects of subsidies to donate requires estimating counterfactual behavior if subsidies or thresholds are changed. The reaction to these two factors is not identified if participants face a unique threshold and match and these two are identical. Our data contains two sources of identification. First, some treatments vary subsidies without varying thresholds. Second, treatments vary the position of some thresholds while keeping others constant. If participants react to the marginal subsidy, as argued in the previous section, this provides an additional

of separability of preferences.

[25] In most treatments in our experiment, the preceding threshold, the marginal subsidy and the threshold under evaluation are not linearly independent since they are placed on the same budget line. We, therefore, estimate the model using only the marginal subsidy and threshold level as explanatory variables.

source of identification on the effect of subsidies on threshold crossing. We show that a simple behavior model predicts within-sample variation in our experiment (Figure 5).

Bounds: In the case of discrete choice, the structural choice probability is not non-parametrically identified with a finite set of points (see Bhattacharya and Komarova, 2024; Kamat and Norris, 2020). This is also true in our context. Hence, we use a flexible parametric model that is identified in our data. In the case of discrete choice, Bhattacharya (2021) provides shape restrictions on choice-probability functions consistent with economic rationality. These are shape restrictions on the relationship between choices, income and prices. Our dataset does not include information on donors' income, so the only theoretical restriction on choice probabilities to consider is that they are decreasing in price. As we discuss later, our estimates are consistent with the probability of crossing a threshold that increases in the subsidy amount.

More generally, the object we need to identify is the proportion of people who are better off with a subsidy but a reduced income. Therefore, we need to predict agents' behavior in counterfactual nonlinear budget sets. Cherchye et al. (2014) show that if preferences are convex and budget sets are the finite unions of polyhedral convex sets, consistency with rational behavior is equivalent to the solution of a linear program. This means we can extrapolate behavior for a large family of policies (see Hausman, 1985). Knowledge of how rational agents react to policy changes can then be used to estimate a random utility model as in Deb et al. (2023) and calculate bounds on the proportion of people better off in the new budget. This implies that our approach can be implemented fully nonparametrically by exploiting variations in related policies. We note, however, that our approach requires recovering the entire structural choice probability since we are interested in the average compensating variation. An approach similar to that used by Mogstad et al. (2018) to bound treatment effects can be added to construct bounds on this object when policy variation is limited.

4 Field Experiment

4.1 Design

The experiment is designed to collect data that allow causal identification of how an individual responds to a threshold match offer. The responses at various thresholds are used to identify the welfare effects of threshold matches. Our partner charities use nine possible threshold match offers, and we randomly offer one to a potential donor. The threshold match offers are chosen based on prior knowledge of the distribution of donations typically received by the charities partnering in the study and the best guess, based on consultation with fundraising experts, of where the thresholds should be placed to affect giving behavior.[26] By varying the threshold levels and matches across the nine possible offers, we alter incentives to donate across individuals. This gives the exogenous variation needed to identify how donations respond to incentives.

To test these threshold matches in the charitable giving market, we partnered with a private foundation and 26 charities in Chicago. Since we would like to know how individuals react to a threshold match offer from a charity in a natural setting, the charity delivered the match offers during a normal fundraising campaign. To this end, we worked with the charities to embed the threshold match offers in a personalized email fundraising campaign designed and sent by the charity to its respective supporters.[27] The offer was sent via personalized email on November 1, 2017. If the supporter makes a donation above the lowest threshold anytime between November 1-3, the charity receives the specified match. In addition, the supporter is informed of another match offer available on November 28 (Giving Tuesday) in the same email. Two match offers set apart in time were included to satisfy the constraints of our

[26] To determine the thresholds used in the experiment, we use data from a 2016 fundraising campaign in Chicago, with 42 charities and over 5,000 individual donations, to assess the underlying distribution of donations. There was no price variation in the campaign so we have no estimate of the underlying price elasticity of donations. A model of impact giving with constant price elasticity is used to assess possible behavioral responses to incentives, assuming that observed donations were the *true* distribution of donations, for different price elasticities. We searched over various thresholds, up to $500. The estimates showed that a match at $500 would have a small effect on out-of-pocket donations if demand was inelastic and a 1-to-1 match would have a negative effect. Thus, the highest threshold in all treatments in the field experiment is at $500. The estimates also showed that additional lower thresholds would decrease average out-of-pocket donations. Given model uncertainty and the interest by the private foundation funding the matches to attract new small donors, we included smaller thresholds as well.

[27] Supporters are individuals on the charity's email list. They could be previous donors or not.

partner's programming goals.[28] The November 28 match offer is identical for everyone. The supporter could give November 1-3 and Nov 28, and both donations would be matched. The two budget sets share the same properties outlined in the theory section as a single budget and can be used in the estimations. As discussed in the results section, we find little support for intertemporal substitution.

The match incentive treatments are shown in Table 1. All match incentives included three donation tiers, the two listed and a third tier that offered a $500 match for any donation of $500 or more. So, for example, the third row of the table shows the 50-100 treatment: the match offer is "a donation of $50-$99 receives a $50 match, a donation of $100-$499 receives a $100 match, and a donation of $500 or above receive a $500 match." The treatment name for each match incentive is listed in the first column, and the number of observations for each treatment is listed in the last column. In addition to the match incentive conditions, every charity included a control group that received a personalized email that did not include a match offer for donations made Nov 1-3 but informed the donor of the Nov 28 match.[29] The text in the body of the email describing the match incentive and the subject line of the email are identical for all 26 charities (see Appendix C).

4.2 Implementation

The fundraising campaign was conducted by sending personalized emails to each charity's supporters. To do this, each charity supplied us with a list of their supporters, including an email address and first and last name. We removed duplicate email addresses or names within a charity, and if there were duplicate email addresses or names across charities, we randomly assigned one charity to that email address or name and removed the email address or name from the remaining charities. These measures assure that each email address and individual in our experiment only received one

[28] A simpler design would offer the nine different matches on Nov 1 and no matches on Nov 28. However, this was not possible. The goal of our partner, the private foundation that provided the matches for the study, was to support charities' fundraising efforts on Giving Tuesday (Nov 28). Varying match offers on Giving Tuesday was too complicated from a programming perspective. As a compromise, the design offers a set of experimental treatment matches and a set of fixed matches spaced apart in time.

[29] Matching funds for the November campaigns were capped at $10,000 per charity, and a donor could only receive one match during the Nov 1-3 window and one match during the Nov 28 window. This information was included in the email text. See Appendix C.

Table 1: Experimental Treatments

Treatment		Tier 1		Tier 2		
Number	Name	Donation	Match	Donation	Match	Obs
1	25-100	\$25-\$99	\$25	\$100-\$499	\$100	12,625
2	35-100	\$35-\$99	\$35	\$100-\$499	\$100	12,615
3	50-100	\$50-\$99	\$50	\$100-\$499	\$100	12,619
4	25-75	\$25-\$74	\$25	\$75-\$499	\$75	12,612
5	25-150	\$25-\$149	\$25	\$150-\$499	\$150	12,604
6	50-150	\$50-\$149	\$50	\$150-\$499	\$150	12,623
7[a]	25*-100	\$25-\$99	\$35	\$100-\$499	\$100	8,024
8[a]	35*-100	\$35-\$99	\$45	\$100-\$499	\$100	8,020
9[a]	50*-100	\$50-\$99	\$60	\$100-\$499	\$100	8,022
Control		No Nov 1-3 match, informed of GT match				12,588
Total						112,352

Notes: [a] Treatments 7-9 were only used by four large charities.

match incentive offer from a charity on November 1.[30]

After we removed duplicates, for each charity, we randomly assigned their supporter list to either one of the treatments or the control condition, so that a supporter never received more than one of the match offers. The first six treatments in Table 1 were used with all 26 charities, and the final three treatments were also included for four large charities.[31] This means each charity's supporters experienced all of the first six treatments and the control condition.[32] This approach allows us to identify price

[30]The study is registered at the AEA RCT Registry (AEARCTR-0005918). Human subjects approval was completed at Texas A&M University (IRB2016-0721M). In addition, the researchers provided each charity with an memorandum of understanding detailing the steps that would be taken to preserve the confidentiality of data from the charities and their supporters.

[31]These last three treatments are identical to the first three but with a match for the lowest threshold that is equal to the threshold + \$10. (e.g. 25*-100 has a match of \$35 for a donation between \$25-\$99, a match of \$100 for a donation between \$100-\$499 and a match of \$500 for a donation of \$500 or above). We did not include these three treatments for all charities because the charities varied in size (i.e. number of supporters), and we did not have enough power for some charities to assign all 9 treatments and be able to detect effects across treatments. The number of supporters for a charity ranged from 500 to 57,000 across the 26 charities, with most charities having around 2,500 supporters. The large charities had $>$ 3,500 supporters.

[32]The only background data we have on supporters to check the balance across treatment assignments are previous donations. Roughly 55% of supporters donated at some point in the previous

effects in our analysis separately from any existing heterogeneity across charities. We control for small and large charities and the treatments in our estimations.

The charities developed a fundraising email that was consistent with their image and mission, and we inserted the wording of the incentive offer so that the match offer text was identical across charities. The wording and layout of the appeal sent in the email for the treatment groups and the control group, and an example email sent by one of the charities, are in Appendix C. We sent the emails to supporters, on behalf of the charity, using a mail management program (e.g. MailChimp or Constant Contact). All the email delivery and random assignment to treatment were managed by us, however, the email received by the supporter looked like it came from the charity.[33] The charities were blind to treatment assignment, so they did not know which match offer any particular supporter received.

Over 112,000 emails were sent to that many unique individuals on November 1, and over $250,000 was spent on matches during November. Each charity provided us the data for all donations made to the charity for November 2017, including the donor's name, email address (if available), donation date, and donation amount. They also provided us with donation data for the previous three years. We use this to classify individuals in the study as previous or new donors. Our final experimental data set includes all individuals treated in the field study (i.e., received an email on November 1), whether or not they donated during November, the date and amount of the donation, and whether or not they donated in the past and the amount.

5 Results from the field experiment

We examine giving behavior in the early window (Nov 1-3) when the different threshold matches were offered and on Giving Tuesday (Nov 28) when everyone faced the same price reduction for giving.

Table 2 combines data for both periods and shows the number of donations, donation rate, total out-of-pocket donations, and average out-of-pocket donation for the nine treatments, the control group, and the average across the first six treatments,

three years, and this is not statistically significantly different across the first six treatments and control.

[33]The "From:" field in the email had the charity's name and email address, and there was no mention that this was a field study. From the point of view of the supporter, this was a fundraising email sent by the charity.

as all charities received these. Over $519,000 were raised in these two periods in November and 1.6% of all supporters donated. The average unconditional donation was $4.61, and the average donation conditional on having donated was $265. The donation rate is similar for the control group and the first six treatment groups, and the average unconditional donation in the control group tends to be higher than in the treatment groups. This is due to some large donations in the control group. The last column lists the average trimmed conditional donation (i.e. donations <$10,000) and shows donations were more similar (e.g. $275 in the control and $233 in the first six treatments). There are no robust treatment effects on average donations compared to the control group.[34] The last three treatments that only large charities experienced yielded a lower donation rate and average donation.[35]

Most donations (74%) were made on Giving Tuesday (Nov 28). There were 1,761 unique donors across the two periods, and despite there being two opportunities to donate with a match, most donors only gave once. Of all donors, 124 (7%) gave in both periods, and 1,637 individuals gave either during Nov 1-3 or on Nov 28.[36] The distribution of donations for the two periods, for those who only gave once, are shown in Figure 3.[37] Panel (b) shows that, on Giving Tuesday (Nov 28), the distribution of donations is similar across all treatment groups and the control group. That is, those who chose to donate on Giving Tuesday behaved similarly and independently of the personalized email received on Nov 1. In contrast, during the Nov 1-3 period, donations are responsive to the treatment match incentives.

Focusing on donations made in the Nov 1-3 period (Figure 3, Panel (a)), when the nine threshold matches were offered, we do see evidence of bunching at the threshold donation level to receive a match. This bunching is the behavior needed to identify the excess burden of offering a threshold match for donations. There is differential bunching at $25, $35 and $50 across the treatments. For example, donors in the 25-100 group are at least 4.6 times more likely to donate $25 than those in the control,

[34]The p-values associated with pairwise tests of equality of means are: p-value(25-100 v control) = 0.9658, p-value(35-100 v control) = 0.1685, p-value(50-100 v control) = 0.2365, p-value(25-75 v control) = 0.5718, p-value(25-150 v control) = 0.0579, p-value(50-150 v control) = 0.6340.

[35]Donation behavior is similar in the first six treatments when comparing the large charities that received all nine treatments to those that received the first six. Donation rates are lower in large charities, but average trimmed donations are not statistically different by treatment between the two groups.

[36]Castillo et al. (2020) found a similar pattern of giving across time.

[37]We omit treatments 25*-100, 35*-100 and 50*-100 because of small numbers.

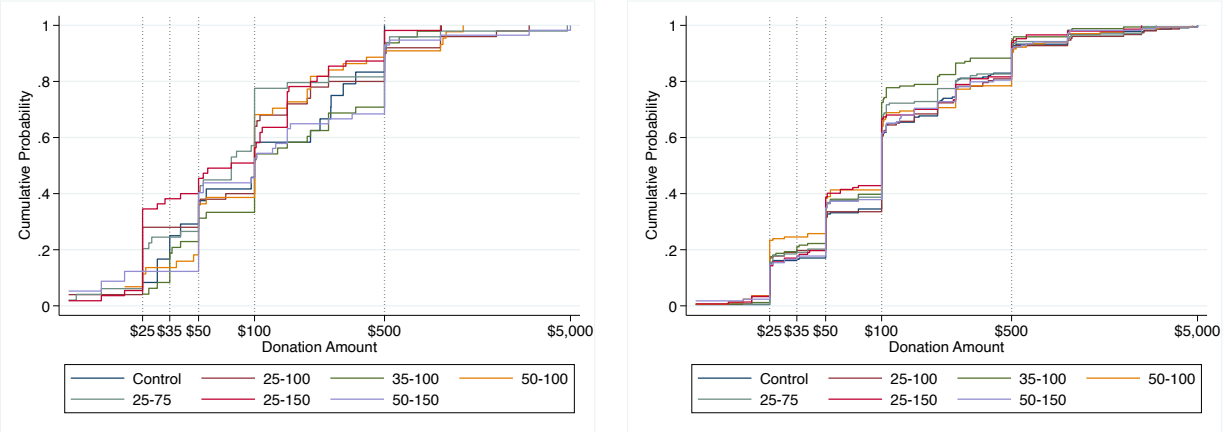
Table 2: Donations - Nov 1-3 and Nov 28

Treatment	Num Donations	Donation rate	Total Donations	Avg uncond	Avg cond	Avg cond (trim)
25-100	249	1.8	$87,752	$6.94	$352	$273
35-100	268	1.9	$56,892	$4.50	$212	$212
50-100	270	1.8	$69,981	$5.53	$259	$223
25-75	267	1.9	$75,861	$6.01	$284	$248
25-150	253	1.8	$49,337	$3.91	$195	$195
50-150	277	1.9	$70,017	$5.53	$253	$253
25*-100	32	0.4	$3,899	$0.49	$122	$122
35*-100	44	0.5	$7,053	$0.88	$160	$160
50*-100	38	0.4	$7,308	$0.91	$192	$192
Control	259	2.0	$91,061	$7.23	$352	$275
Treatments (1-6, avg)	1,584	1.8	$409,840	$5.40	$259	$233
Total	1,957	1.6	$519,161	$4.61	$265	$235

Notes: The last column reports average conditional donation trimmed at donations less than $10,000.

35-100, or 50-100 groups. A similar pattern emerges for donations at $35 for those in the 35-100 group and at $50 for those in the 50-100 group. At the second threshold of $100, there are at least three times more donations in the treatment groups relative to the control group.

In sum, the field experiment yields evidence of bunching at the treatment thresholds and no significant treatment effects on average out-of-pocket donations. These findings show that donors shift their donations to meet thresholds and impact the total amount received by the charity (i.e., their donation plus the match). In the next section, we show that the campaign led to a significant increase in out-of-pocket donations compared to the same donors' previous donations.



(a) Nov 1-3 (b) Nov 28

Figure 3: Distribution of donations

Note: The figure includes data for individuals who only gave once. Distributions show log of donation amount. Thresholds used in the experiment, i.e. $25, $35, $50, $100, $500, are marked as vertical lines in the figures.

6 Excess burden

We present the estimation of compensating variation and excess burden on consumer welfare of threshold matches. Estimation requires data on the distributions of donations with different threshold matches and without threshold matches. We construct the baseline distribution of donations from over 50,000 donors who donated to the participating charities in the two years before the experiment. While the model can be estimated using experimental data only, including previous donations increases the precision of the estimates. Our main results use these pooled data on donations. Results using experimental data only are discussed in Section 7 and presented in Appendix B.

6.1 Description of the estimation data

We start by discussing features of the experimental data compared to the baseline data of two years of past donations since these data determine the quality of our estimations.

First, we test the equality of distributions of baseline donations, donations between Nov 1-3 in the 9 treatments plus Control and donations on Giving Tuesday

(Kolgomorow-Smirnov test p-values are reported in Table B.1). The distributions of donations for the experiment treatments are different from baseline for most treatments but not for the control group. The distribution of donations on Giving Tuesday is different from baseline. These results are consistent with threshold matches altering the distribution of donations but also indicate a potential loss of power if we do not include information from past donations. Thus, we use these pooled data from the experiment and past donations to estimate the structural choice probability as outlined in Section 3.[38]

Next, we focus on baseline donations and compare these to donations on Giving Tuesday in our experiment to illustrate the shift in the distribution of donations due to threshold matches. Figure 4 shows the decumulative distribution function of baseline donations and the decumulative distribution function of Giving Tuesday donations. We use the decumulative function because it behaves like a demand equation and clearly illustrates the shift in demand with a price reduction. The graph excludes donations below \$10 and above \$5000.[39] Donors face three thresholds on Giving Tuesday: \$25, \$100, and \$500. Reaction to the experimental incentives is observed in the shift in the distribution of donations to the right and an increase in donations at or above the incentivized donation thresholds. We do not find distortions in the distribution of donations above the thresholds. While this is consistent with donors not reducing their out-of-pocket donations when they planned to donate in excess of the threshold, we do not have formal proof that this is the case. Distortions above the threshold might be due to donors lowering their out-of-pocket donations or increasing them to reach the next threshold.

The distribution of donations above the highest threshold (\$500) is virtually identical in the Baseline and on Giving Tuesday. This is consistent with out-of-pocket donations not decreasing for those planning to donate at least \$500.

6.2 Prediction model

As described in Section 3, we estimate the probability of exceeding the threshold as a function of a polynomial of order three in the threshold size (in logs) and the percent

[38]The target donations we use in the estimation are: \$50, \$70, \$100, \$150, \$200, \$300, and \$1000. The targets reflect the experimental treatments, i.e., \$50 is the amount going to the charity when the threshold donation to get a match is \$25 and the match is \$25.

[39]The administrative data does not separate donations from fees, so we exclude low donations, i.e., those below \$10. A small number of donations, i.e. 0.25 of a percentage point, is above \$5,000.

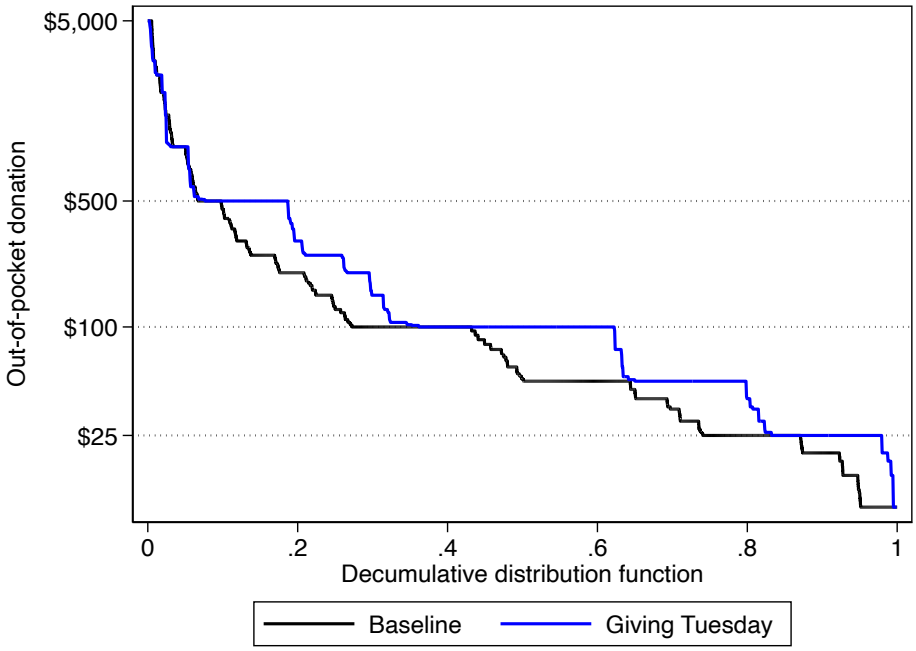


Figure 4: Threshold incentive effects on out-of-pocket donations

discount to reach the threshold (ΔMatch/Threshold). The percent discount is used to make it comparable across different threshold levels.

We first show in Figure 5 that the model fits the data quite well. Panel (a) compares donation behavior on previous Giving Tuesdays at baseline, for observed and predicted donations, and on Giving Tuesday during the experiment, for observed and predicted donations. We see the model tracks behavior during and prior to the experiment quite closely. Panel (b) compares observed and predicted donations for each incentivized threshold offered November 1-3. Predictions tend to underestimate the probability of exceeding the threshold but respect the patterns in the data of declining donations as the threshold rises.

Table 3 reports estimates of the structural choice equation. This is a Probit model of the probability of a donation meeting or exceeding a threshold controlling for the threshold level and the match amount as a proportion of the threshold. Columns (1) - (4) in the table report estimates of the model for different samples. Columns (5) - (6) report different models. The predictive power of the estimates is similar across different samples and methods. As expected, the probability of exceeding a threshold decreases in the threshold (ln(Threshold)) and increases in the match

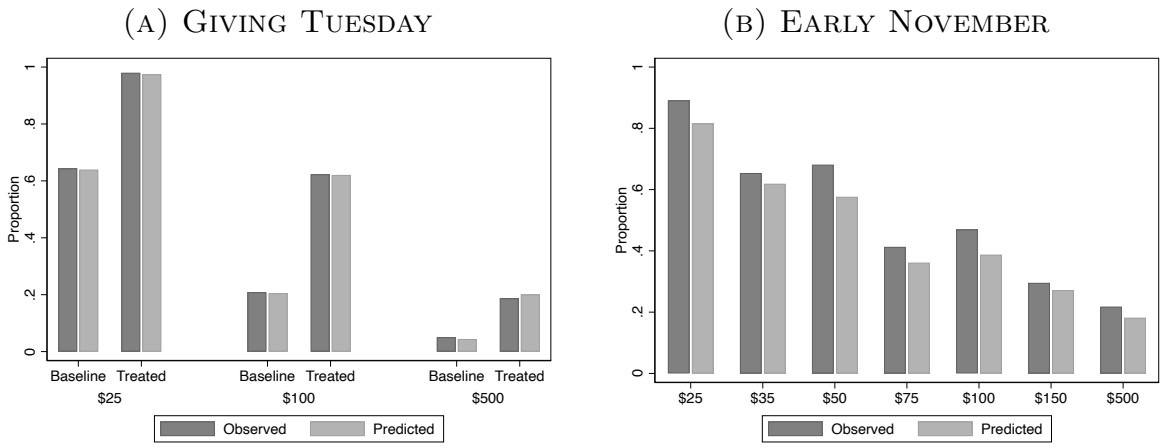


Figure 5: Fit of prediction model

Notes: The figure compares observed and predicted behavior. Panel (a) compares the observed number of donations exceeding the $25, $100, and $500 thresholds in the baseline data and during the experiment on Giving Tuesdays compared to model predictions. Panel (b) compares the observed number of donations exceeding the $25, $35, $50, $75, $100, $150, and $500 thresholds compared to model predictions..

(Match/Threshold). We obtain noisier estimates when we restrict the regressions to donations made only in November, Column (3), or only in November 2017, Column (4), when the experiment took place. We use the model in Column (1) for our main results on the excess burden and present alternative estimates in Appendix B.

6.3 Excess burden estimates

We present two sets of estimations of the excess burden on consumer welfare of threshold matches. First, as a benchmark, we present estimates using linear interpolation. These estimates are fully non-parametric. For this approach, we assume that the change in the probability between no matches and the offered match is linear. While straightforward, this approach is limited as it does not consider that donors react to marginal matches nor fully incorporate information from treatments with different thresholds and matches. Also, these are valid estimates of excess burden only if preferences are separable and incentives are offered separately, not jointly as in our experiment. Second, we present estimations based on the prediction model estimates from Table 3. This approach produces estimates of excess burden without needing

Table 3: Prediction model

	(1) All	(2) Donated once	(3) Nov. only	(4) Nov 2017 only	(5) Dynamic (OLS)	(6) OLS
ln(Threshold)	-3.057	-3.065	-2.419	-5.184	-4.950	-1.769
	(0.178)	(0.178)	(0.542)	(0.945)	(0.193)	(0.043)
ln(Threshold) squared (sq.)	0.335	0.337	0.261	0.708	0.685	0.225
	(0.032)	(0.032)	(0.098)	(0.167)	(0.035)	(0.007)
ln(Threshold) cubed	-0.014	-0.014	-0.012	-0.036	-0.035	-0.010
	(0.002)	(0.002)	(0.006)	(0.010)	(0.002)	(0.000)
Match/Threshold	1.632	1.691	4.869	0.706	6.505	-5.643
	(1.402)	(1.459)	(1.471)	(1.740)	(0.803)	(0.336)
Match/Threshold×ln(Threshold)	0.788	0.744	-0.411	0.795	-1.315	2.551
	(0.506)	(0.528)	(0.530)	(0.631)	(0.297)	(0.133)
Match/Threshold×ln(Threshold) sq.	-0.109	-0.107	-0.004	-0.096	0.071	-0.245
	(0.044)	(0.047)	(0.047)	(0.056)	(0.026)	(0.012)
Constant	8.261	8.276	6.776	12.552	11.563	4.746
	(0.322)	(0.323)	(0.981)	(1.745)	(0.345)	(0.081)
Obs	880,686	878,454	548,10	16,875	1,753,668	880,686
Donors	49,882	49,634	4,826	1,875	49,882	49,882
R2	0.229	0.229	0.279	0.304	0.229	0.243
F-test: Thresholds	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
F-test: Shares	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Note: Probit model estimates where the outcome variable is the probability of exceeding the threshold. Donor-level clustered standard errors in parentheses. Column (1) includes all the data, Column (2) includes the sample that donated only once, Column (3) includes only donations made in November, Column (4) includes only donations made in November 2017, Column (5) includes all the data and is estimated as a panel with donations in early November and Giving Tuesday, and Column (6) includes all the data and is estimated using OLS.

additional assumptions.

The first estimates are shown in Figure 6. The figure presents linear interpolation results using baseline donations and donations on Giving Tuesday when the threshold donations for matches were at \$25, \$100 and \$500.[40] The figure shows this calculation. For example, for the \$25 threshold, the figure shows the change in the propensity to donate the target amount without incentives and with incentives, i.e. moving from a propensity of 0.63 to 0.99. The line connecting the two points represents the linear interpolation of demand.[41] The main pattern is that the subsidy's excess burden increases in the threshold. The excess burden of a threshold match of \$25 is 17c per dollar of subsidy, while the excess burden of a threshold match of \$500 is 37c per dollar of subsidy. In other words, a donor would be willing to pay for 83% of a \$25 match that goes to the charity but only 64% of a \$500 match.

[40] We do not use data from November 1-3 because different donors faced different thresholds.

[41] The change in the propensity to donate the target amount can also be seen in Figure 4.

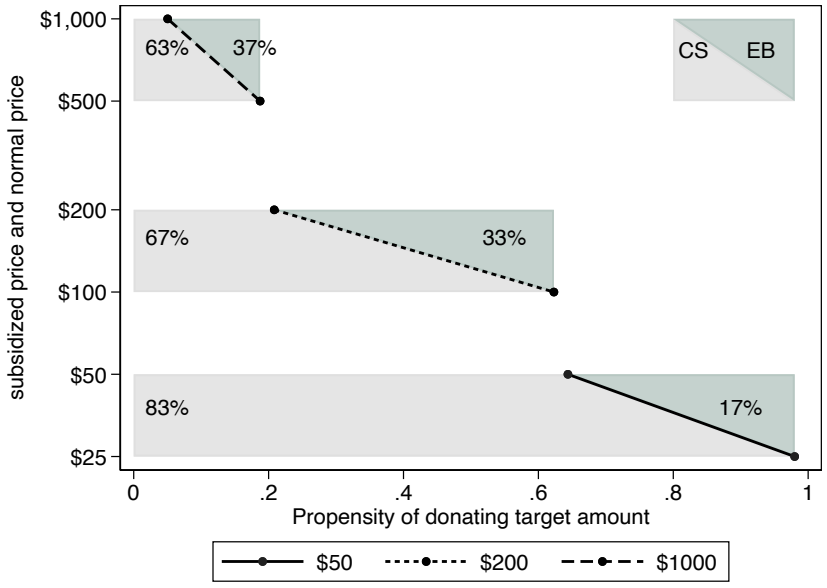


Figure 6: Excess Burden using linear interpolation

Notes: EB is excess burden. CS is consumer surplus.

The second estimates using the Probit model results from Table 3, column (1), are presented in Table 4. These estimates are calculated in the counterfactual case where each threshold match is offered in isolation. This is our preferred specification, and we discuss estimates assuming the three thresholds were offered simultaneously at the end of this section. The compensating variation is calculated using numerical integration. Confidence intervals are calculated using 1,000 bootstrap replications. In particular, we select a random sample of donors with replacement and re-estimate the Probit model.

Table 4 shows the same patterns as Figure 6: the excess burden per dollar of subsidy increases in the threshold level. The estimates, however, show a steeper increase in the excess burden as the level of the threshold increases. The excess burden of a threshold match of $25 is 9c per dollar of subsidy, while the excess burden of a threshold match of $500 is 45c per dollar of subsidy. The fact that the excess burden of threshold matches differs by the threshold level indicates that welfare evaluation of alternative subsidies must account for heterogeneous preferences. More importantly, ignoring the welfare of donors would severely distort policy choices. A policy that

Table 4: Excess Burden using prediction model

	$25 threshold $25 match			$100 threshold $100 match			$500 threshold $500 match		
	Est.	2.5%	97.5%	Est.	2.5%	97.5%	Est.	2.5%	97.5%
Subsidy per donor ($)	24.6	24.4	24.7	76.5	73.9	79.2	131.1	116.3	145.7
Compensating variation ($)	22.2	22.0	22.5	51.3	49.6	53.1	71.6	65.5	77.7
Excess burden (EB) ($)	2.3	2.2	2.4	25.2	24.1	26.2	59.5	50.6	68.3
EB per dollar of subsidy ($)	0.095	0.090	0.099	0.329	0.323	0.335	0.454	0.434	0.470
Percent exceeding threshold	0.982	0.976	0.988	0.765	0.739	0.792	0.262	0.233	0.291

increases donations by offering a high threshold donation/match also carries a larger proportional excess burden.[42]

So far, we have shown average excess burdens. What is the excess burden for those who actually change their behavior due to the threshold match? We show in Section 2 that any donor who donates above the total amount received by the charity (threshold plus match) has an excess burden equal to zero. The burden then is distributed among those who could have potentially changed their behavior. Using past donations at or above the threshold plus the match as a benchmark, we see that 51.3 percent, 19.6 percent, and 5.8 percent, respectively, donated above $50, $200, and $1,000. Adjusted estimates of the excess burdens on the marginal donors who would actually change their behavior in the presence of a match incentive are 19.5c, 40.9c, and 48.2c per dollar donated.[43] This demonstrates that the excess burden on those who change their behavior is larger and increasing. The result is not due to the lack of additionality in low thresholds (Boomhower and Davis, 2014).

We also estimate welfare effects using the Probit model results from Table 3, column (1), assuming the three thresholds were offered simultaneously (Table B.2). The estimates are lower than the row sum of those in Table 4. This is because the joint estimates consider that donors respond to marginal incentives, and lower thresholds reduce incentives to donate larger amounts. Also, excess burden estimates are larger because donors only react to marginal subsidies, given our behavioral assumption.

[42] Table 4 shows that 26 percent of donors would donate $500 or more if offered a match of $500. This is four percent without incentives (see Figure 5).

[43] We obtain these numbers by dividing the excess burden by the proportion of donors who donated below the threshold.

6.4 Policy evaluation

The estimates from Table 3 can be used to evaluate the impact of threshold matches on charitable donations and welfare. To do this, we report the total out-of-pocket donations received by charities, the standard in the nonprofit sector to assess subsidized donations, and the excess burden imposed on donors. Table 5 reports results using the data from baseline donations for all 26 participating charities and the donations received on Giving Tuesday during the experiment. The welfare effects are calculated assuming that three threshold matches were offered simultaneously (see Table B.2): \$25 match for the charity for a donation that meets or exceeds \$25, \$100 for \$100, and \$500 for \$500, respectively.[44]

Out-of-pocket donations when threshold matches are available are about twenty percent higher than average past out-of-pocket donations (\$244.9 versus \$205.7). If we count the increase in out-of-pocket donations (\$39.2) plus the match (\$191.8), this results in \$231.0 more dollars going to the charity per donation. Charities received 78c in matching funds per dollar donated (\$191.8/\$244.9).[45] This illustrates the accounting efficiency of threshold matches: it cost 78c to increase out-of-pocket donations by \$1. Donors gave an additional \$39.2 because of the match incentives, and this generated an increase in utility valued at \$104.2 because of the match (compensating variation). However, it also creates an excess burden of \$87.7, i.e. subsidy minus compensating variation.

These findings illustrate that policymakers need to balance the welfare of the donors (the excess burden of a subsidy) and the welfare of the donees (the amount of money received by charities). Also, charities competing for donors need to consider the donors' welfare as they plan their incentive schemes to curry new donors and retain existing ones. Some fundraising schemes might be effective in the intensive margin but costly in the extensive margin, especially if donors find donating to alternative charities, with more attractive schemes from their private point of view, more welfare-enhancing.

Finally, the estimates can be used to evaluate the desirability of using threshold

[44] Our measures of welfare are absent any utility costs to the donor for being asked to donate, i.e. social pressure. Our experiment was not designed to measure these costs. We note that such costs to ask, in our setting, would need to be larger than the sum of the extra donations generated and the compensating variation for welfare to be negative.

[45] In the experiment, matches were about 50c per dollar donated. This is because matches paid out per charity were capped at \$10,000. The estimates for matches (\$191.8) is unconstrained.

Table 5: Policy evaluation of $25:$100:$500 thresholds

	Treatment effect		
	# of donations	Mean donation	S.E.
Pre-treatment	95,979	205.7	1.7
Treatment	1,875	244.9	11.5
	Program costs per donor		
	Matches		Excess Burden
Total	191.8		87.7
	New donations (1)	Matches (2)	(1)+(2)
	39.2	191.8	231.0

matches compared to alternative incentives. In particular, consider donors are offered a linear match, e.g., for each dollar donated, the charity receives a fixed monetary subsidy as a match that would produce the same amount of donations received by the charity as with a threshold match. For the sake of argument and consistency with estimates on our sample of donors, we assume that the constant price elasticity of a linear match is -1. This implies that to obtain a similar fundraising result, donors should be offered a subsidy of 90c per dollar donated.

What is the welfare effect of such a linear subsidy? We follow Hungerman and Ottoni-Wilhelm (2021) who show that the welfare impact of a match cannot exceed the welfare effect of an equivalent rebate. Using their estimated elasticity for a rebate (-0.2), we can approximate the consumer surplus of the match to be about $114 while the average match would be $185. This alternative leads to higher welfare for consumers ($114 versus $98.4) but a higher cost in subsidies ($185 versus $145.8). Given the heterogeneity of welfare estimates, we expect these calculations to vary as the threshold set is changed. We should also point out that these comparisons depend on the assumptions made to identify the welfare effects of matches and rebates.

7 Robustness

We assess the robustness of our results by calculating the compensating variation and excess burden under alternative conditions. We investigate if the main results vary if

the estimation sample is restricted to the experimental sample alone and if we allow for dynamic choices.

7.1 Sample selection

We check the predictive accuracy of the main model using only donation data from November 2017. The model fit is noisier than that in Figure 5, but follows similar patterns.[46] We re-estimate Table 4, and estimates are similar (Table B.3). Most importantly, these checks reproduce the pattern of increasing excess burden as threshold levels increase. We remark that these estimates use variation in donations induced by the experiment only. Our main results, therefore, are not due to sample selection or require baseline data.

7.2 Modeling assumptions

Our main estimations in Table 3 assume that donations are timed to be efficient, i.e., donations are made to maximize the total amount of money received by the charity given the out-of-pocket donation.[47] For instance, given the two match offers are separated in time in the experiment, a strategic donor facing a match of $100 for a donation of $100 or more in both periods might split a donation of $200 in two across periods. The donor would obtain $200 in matches for the charity rather than $100 if they made a single donation of $200 in one period. Indeed, donors do not appear to always maximize the total matches attained (Figure B.6).

We estimate the welfare parameters assuming donors do not time donations to maximize matches (shown in Table 3, column (5)). Estimates of the excess burden of threshold subsidies are smaller (Table B.4). However, we reproduce the pattern that higher thresholds are associated with larger excess burdens.

8 Conclusions

We show how to use nonlinear budgets to identify individuals' preferences over other budget sets and the welfare impact of subsidies. Estimates can be recovered from the

[46] Figure B.4 compares observed and predicted behavior.

[47] Figure B.5 shows the predictive power of a model estimated under the assumption that donors maximize total donations to charities over time.

empirical reaction to exogenous variation in consumption thresholds and subsidies to reach such thresholds. We apply our approach to data from a large field experiment using threshold matches for charitable donations: if a donor's donation met or exceeded a pre-specified threshold, the charity received a fixed dollar match. We find that the excess burden of threshold matches increases in the threshold level, even accounting for non-additionality. For the marginal donor, the excess burden doubles (from 19.5c to 40.9c per dollar donated) when the threshold increases from $25 to $100 for an equivalent match for the charity. The problem charities face is formally equivalent to that of a policymaker providing threshold incentives to affect behavior in other choice environments, i.e., home energy efficiency rebate programs, U.S. Saver's Credit program, etc. The methods developed here can be used to conduct welfare analyses of those types of policies as well.

Threshold incentives can be effective at changing behavior, as we found in our study, i.e. it cost $0.78 to subsidize one additional dollar of charitable donation, especially if linear incentives are less salient. There is a trade-off, though. In our setting, threshold incentives increased donations at all threshold levels, which helped the recipients of donor generosity, and donors revealed they were willing to pay a portion of the match subsidy so that the charity would receive more donations. However, this is nonlinear. Independent of higher thresholds' effectiveness in raising money, they carry a larger excess burden. That is, donors are willing to pay a smaller portion of the match at a higher threshold for a gain in utility.

How should we interpret our approach and results from a behavioral economics perspective? Some of the assumptions on preferences might be in contradiction with reported behavioral biases (see List et al., 2023, on left-digit bias), individuals might be inattentive (Chetty et al., 2009), and price responsiveness might be heterogeneous and depend on the magnitude of the change (Taubinsky and Rees-Jones, 2018), to name a few. As our paper shows, donations in our experiment bunch at some thresholds more than others. Our estimates, however, exploit previous donations where these patterns are present and vary the location of the matches away from frequent numbers. We also exploit a dynamic feature of our experiment (attentive donors could increase total donations by splitting their out-of-pocket donations over time) to construct a measure of "money left on the table" which we expect to be correlated with behavioral biases. We find that donors leave money on the table, but this is not related to the magnitude of the incentives, suggesting that the pattern of increasing

excess burden is not due to these issues.

More importantly, a question for future research is whether our approach could be used to obtain welfare measures that account for behavioral biases (see Farhi and Gabaix, 2020; Rees-Jones, 2024). We think so. First, richer experimental settings can include budget sets with and without dominated choices (see Kleven, 2016) to assess the importance of violations of rationality.[48] Second, nonlinear budget sets can be used to estimate the distribution of nuisance costs and frictions nonparametrically (see Castillo et al., 2023). Finally, since our approach can be used in some settings where the elasticities might be difficult to estimate or not informative (see impact giving),[49] it might help behavioral economists detect deviations from standard assumptions.

[48] An example of a subsidy with dominated choices is threshold incentives in home-energy rebate programs.

[49] We thank Mark Ottoni-Wilhelm for pointing out this connection to impact giving.

References

Adena, M., Huck, S., 2022. Personalized fundraising: A field experiment on threshold matching of donations. Journal of Economic Behavior & Organization 200, 1–20.

Altmann, S., Falk, A., Heidhues, P., Jayaraman, R., Teirlinck, M., 2019. Defaults and donations: Evidence from a field experiment. Review of Economics and Statistics 101, 808–826.

Andreoni, J., Payne, A., 2003. Do government grants to private charities crowd out giving or fund-raising? American Economic Review 93, 792–812.

Andreoni, J., Rao, J.M., Trachtman, H., 2017. Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. Journal of Political Economy 125, 625–653.

Bergstrom, K., Dodds, W., Rios, J., forthcoming. Welfare analysis of changing notches. American Economic Journal: Economic Policy .

Bhattacharya, D., 2015. Nonparametric welfare analysis for discrete choice. Econometrica 83, 617–649.

Bhattacharya, D., 2021. The empirical content of binary choice models. Econometrica 89, 457–474.

Bhattacharya, D., 2024. Nonparametric approaches to empirical welfare analysis. Journal of Economic Literature 62, 554–593.

Bhattacharya, D., Komarova, T., 2024. Incorporating social welfare in program-evaluation and treatment choice. Review of Economics and Statistics , 1–45.

Blinder, A.S., Rosen, H.S., 1985. Notches. American Economic Review 75, 736–747.

Blomquist, S., Hausman, J.A., Newey, W.K., 2023. The econometrics of nonlinear budget sets. Annual Review of Economics 15, 287–306.

Boomhower, J., Davis, L.W., 2014. A credible approach for measuring inframarginal participation in energy efficiency programs. Journal of Public Economics 113, 67–79.

Butera, L., Metcalfe, R., Morrison, W., Taubinsky, D., 2022. Measuring the welfare effects of shame and prides. American Economic Review 112, 122–168.

Castillo, M., Freer, M., 2023. A general revealed preference test for quasilinear preferences: theory and experiments. Experimental Economics 26, 673–696.

Castillo, M., Petrie, R., 2020. Optimal incentives to give. Technical Report. IZA Discussion Paper.

Castillo, M., Petrie, R., Samek, A., 2020. Time to give: A field experiment on intertemporal giving.

Castillo, M., Petrie, R., Wardell, C., 2023. Barriers to charitable giving. Journal of Public Economics 224, 104945.

Cherchye, L., Demuynck, T., De Rock, B., 2014. Revealed preference analysis for convex rationalizations on nonlinear budget sets. Journal of economic theory 152, 224–236.

Cherchye, L., Demuynck, T., De Rock, B., Surana, K., 2020. Revealed preference analysis with normal goods: Application to cost-of-living indices. American Economic Journal: Microeconomics 12, 165–188.

Chetty, R., Looney, A., Kroft, K., 2009. Salience and taxation: Theory and evidence. American economic review 99, 1145–1177.

Cox, J.C., Friedman, D., Sadiraj, V., 2008. Revealed altruism 1. Econometrica 76, 31–69.

Croson, R., Shang, J., 2008. The impact of downward social information on contribution decisions. Experimental Economics 11, 221–233.

Deb, R., Kitamura, Y., Quah, J.K., Stoye, J., 2023. Revealed price preference: theory and empirical analysis. The Review of Economic Studies 90, 707–743.

DellaVigna, S., List, J.A., Malmendier, U., 2012. Testing for altruism and social pressure in charitable giving. Quarterly Journal of Economics 127, 1–56.

Diamond, P., 2006. Optimal tax treatment of private contributions for public goods with and without warm glow preferences. Journal of Public Economics 90, 897–919.

Eckel, C., Grossman, P., 2008. Subsidizing Charitable Contributions: A Natural Field Experiment Comparing Matching and Rebate Subsidies. Experimental Economics 11, 234–252.

Farhi, E., Gabaix, X., 2020. Optimal taxation with behavioral agents. American Economic Review 110, 298–336.

Forges, F., Minelli, E., 2009. Afriat's theorem for general budget sets. Journal of Economic Theory 144, 135–145.

Harbaugh, W., 1998. The prestige motive for making charitable transfers. American Economic Review P&P 88, 277–282.

Harrison, G., List, J., 2004. Field experiments. Journal of Economic Literature 42, 1009–1055.

Hausman, J.A., 1981. Exact consumer's surplus and deadweight loss. The American Economic Review 71, 662–676.

Hausman, J.A., 1985. The econometrics of nonlinear budget sets. Econometrica: Journal of the Econometric Society , 1255–1282.

Hausman, J.A., Newey, W.K., 2016. Individual heterogeneity and average welfare. Econometrica 84, 1225–1248.

Hausman, J.A., Newey, W.K., 2017. Nonparametric welfare analysis, in: Aghion, P., Rey, H. (Eds.), Annual Review of Economics. volume 9 of *Annual Review of Economics*, pp. 521–546.

Hninn, S.T., Shinji, K., Keisuke, K., Yuichiro, Y., 2017. A nonparametric welfare analysis via a randomized conjoint field experiment: an application to water quality improvement and the floating settlements on inlay lake, myanmar. IDEC DP2 Series 7, 1–33.

Huck, S., Rasul, I., 2011. Matched fundraising: Evidence from a natural field experiment. Journal of Public Economics 95, 351–362.

Huck, S., Rasul, I., Shephard, A., 2015. Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model. American Economic Journal: Economic Policy 7, 351–362.

Hungerman, D.M., Ottoni-Wilhelm, M., 2021. Impure impact giving: Theory and evidence. Journal of Political Economy 129, 1553–1614.

Ito, K., 2014. Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. American Economic Review 104, 537–563.

Kamat, V., Morris, S., 2022. Estimating Welfare Effects in a Nonparametric Choice Model: The Case of School Vouchers. Technical Report. arXiv:2002.00103v5.

Kamat, V., Norris, S., 2020. Estimating welfare effects in a nonparametric choice model: The case of school vouchers. Technical Report. arXiv preprint arXiv:2002.00103.

Karlan, D., List, J.A., 2007. Does price matter in charitable giving? evidence from a large-scale natural field experiment. American Economic Review 97, 1774–1793.

Kitamura, Y., Stoye, J., 2018. Nonparametric analysis of random utility models. Econometrica 86, 1883–1909.

Kleven, H.J., 2016. Bunching, in: Aghion, P and Rey, H (Ed.), Annual Review of Economics. volume 8, pp. 435–464.

Kleven, H.J., Waseem, M., 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. Quarterly Journal of Economics 128, 669–723.

Kroft, K., Luo, Y., Mogstad, M., Setzler, B., 2020. Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice. Technical Report. National Bureau of Economic Research.

Landry, C.E., Lange, A., List, J.A., Price, M.K., Rupp, N.G., 2006. Toward an understanding of the economics of charity: Evidence from a field experiment. Quarterly Journal of Economics 121, 747–782.

Landry, C.E., Lange, A., List, J.A., Price, M.K., Rupp, N.G., 2010. Is a donor in hand better than two in the bush? evidence from a natural field experiment. American Economic Review 100, 958–983.

List, J., 2020. Non Est Disputandum De Generalizability? A Glimmpse Into the External Validity Trial. Technical Report. NBER Working Paper 27535.

List, J., Lucking-Reiley, D., 2002. The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. Journal of Political Economy 110, 215–234.

List, J.A., Muir, I., Pope, D., Sun, G., 2023. Left-digit bias at lyft. Review of Economic Studies 90, 3186–3237.

Mas-Collel, A., Whinston, M., Green, J., 1995. Microeconomic theory. 198 madison avenue, new york, new york, 10016: Oxford university press.

Matzkin, R.L., 1991. Axioms of revealed preference for nonlinear choice sets. Econometrica: Journal of the Econometric Society , 1779–1786.

Mogstad, M., Santos, A., Torgovitsky, A., 2018. Using instrumental variables for inference about policy relevant treatment parameters. Econometrica 86, 1589–1619.

Pop-Eleches, C., Urquiola, M., 2013. Going to a better school: Effects and behavioral responses. American Economic Review 103, 1289–1324.

Rees-Jones, A., 2024. Behavioral incentive compatibility and empirically informed welfare analysis: An introductory guide. Journal of Economic Perspectives 38, 155–174.

Reiley, D., Samek, A., 2019. Round giving: A field experiment on suggested charitable donation amounts in public television. Economic Inquiry 57, 876–889.

Saez, E., 2004. The optimal treatment of tax expenditures. Journal of Public Economics 88, 2657–2684.

Slemrod, J., 2013. Buenas notches: lines and notches in tax system design. eJournal of Tax Research 11, 259–283.

Taubinsky, D., Rees-Jones, A., 2018. Attention variation and welfare: theory and evidence from a tax salience experiment. The Review of Economic Studies 85, 2462–2496.

Varian, H.R., 1982. The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society , 945–973.

APPENDICES INTENDED FOR ONLINE PUBLICATION

A Proofs

This section collects proofs for Section 2. We repeat arguments for completeness.

A.1 Budgets and preferences

We specialize in the case of a target good g and composite numeraire good c. We assume that budgets have a non-empty interior, are bounded, downward closed,[1] and can be described by an increasing, continuous function $h : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ such that $B = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : h(x) \leq 0\}$ (see Forges and Minelli, 2009). In the case of a linear budget set, we have $h(x) = px - y$ with $(p, y) \gg 0$. We let $u(c, g; \eta)$ represent the utility of consuming (c, g) and assume it is continuous, monotone increasing[2] and strictly quasi-concave.[3] We allow for unobserved individual heterogeneity by letting utility depend on a vector η of unknown dimension. We assume there is an absolutely continuous function, $F(\cdot)$, representing the marginal distribution of η.

A.2 Revealed preferences

This section provides the results upon which our identification strategy is built. The pair (B_k^i, x_k^i) is an observation for individual i in choice problem k. We interpret $x_k^i \in B_k^i$ as the bundle chosen when budget B_k^i was available such that $h_k^i(x_k^i) = 0$. We denote a finite collection $D_i = (B_k^i, x_k^i)_{k=1,\ldots,K_i}$, a dataset for individual i. A data set D_i is rationalized by the utility function $u : \mathbb{R}_+^2 \to \mathbb{R}$ if, for all $k = 1, \ldots, K_i$, $x_k^i \in argmax_{x \in B_k^i} u(x)$. We say that x_k is revealed preferred to x_j, or $x_k R x_j$, if $h_k(x_j) \leq h_k(x_k) = 0$. Let H be the transitive closure of the relation R. We follow Forges and Minelli (2009) in defining the Generalized Axiom of Revealed Preference (GARP) in the setting with nonlinear budget sets.

[1] That is, if $y \in B$ then $y' \in B$ for every $y' \leq y$.

[2] A function $f : A \subseteq \mathrm{R}^n \to \mathrm{R}$ is weakly monotone increasing, or non-decreasing if $f(x) \leq f(y)$ when $x \leq y$; and monotone increasing if it is weakly monotone increasing and $f(x) < f(y)$ when $x \ll y$.

[3] Strict quasi-concavity is unnecessary for the main identification results. It is used to aid identification in the case of multiple threshold matches.

Definition (GARP): The dataset $D_i = (B_k^i, x_k^i)_{k=1,\ldots,K_i}$ satisfies GARP if, for any k, j, $x_k H x_j$ implies $h_j(x_k) \geq 0$.

Under these conditions, if a dataset D_i satisfies GARP, an increasing and continuous function exists that rationalizes the dataset D_i (Forges and Minelli, 2009; Cherchye et al., 2014). The following results underlie our identification strategy.

Observation 1: Let (B_j, x_j) and (B_l, x_l) be two observations such that $B_j \cap B_l \subsetneq B_j$ and let (B_m, x_m) for $B_m = B_j \cup B_l$ be the observation when the union of both sets is available. If $x_m \in B_l \setminus B_j$, then $x_l \in B_l \setminus B_j$.

Proof: By the way of contradiction, suppose $x_l \notin B_l \setminus B_j$. By revealed preference $x_l H x_m$, and, by GARP, $h_m(x_l) \geq h_m(x_m) = 0$.[4] Since $x_l \in B_j \cap B_l \subsetneq B_j$, we have that there is a $y \in B_j \setminus B_l$ such that $y > x_l$. We then have $h_m(x_l) < h_m(y) \leq h_m(x_m)$ which is a contradiction.

Definition (Preferences over budget sets): We say that budget set B_l is preferred to budget set B_j by individual i, denoted $B_l \succeq_i B_j$, if there is a $x \in B_l \setminus B_j$ such that for all $y \in B_j$, xHy.

Observation 2: Let (B_j, x_j) and (B_l, x_l) be two observations such that $B_j \cap B_l \subsetneq B_j$ and let (B_m, x_m) for $B_m = B_j \cup B_l$ be the observation when the union of both sets is available. If $x_m \in B_l \setminus B_j$, then $B_l \succeq_i B_j$.

Proof: By definition, $x_l H y$ for all $y \in B_j$. This already shows that $B_l \succeq_i B_j$. However, observation 1 says that this conclusion would not be contradicted had we only observed (B_l, x_l) instead. Note also that the preference relation over sets B_j and B_l cannot be determined from the observation of (B_j, x_j) and (B_l, x_l) alone.

The main insight from this section is that we can determine individual preferences over budget sets from choices on their union. In other words, we can determine the proportion of better-off individuals based on their choices. We use this result next to estimate welfare effects.

Definition (Structural choice probability): Let $\mathcal{N} = \{1, \ldots, N\}$ be a random sample of individuals choosing from budget $B_m = B_l \cup B_j$. Define the structural choice probability as $q_N(B_l, B_j) = N^{-1} \sum_{i \in \mathcal{N}} 1[x^i \in B_l \setminus B_j]$, where $1[A]$ equals 1 if statement A is true. Let $Pr(B_l \succeq_i B_j)$ be the probability that budget B_l is preferred to budget B_j in the population from which sample $\mathcal{N}$ is drawn.

Observation 3: $lim_{N \to \infty} E[q_N(B_l, B_j)] = Pr(B_l \succeq_i B_j)$.

Proof: Observation 2 establishes that $B_l \succeq_i B_j$ if $x_l \in B_l \setminus B_j$. The statement

[4] We see that B_m is a valid budget set by defining $h_m(x) = min\{h_j(x), h_l(x)\}$.

follows from assuming random sampling.

A.3 Welfare effects

We now show how to estimate the compensating variation of a subsidy. We define a budget set as $B(y, p, \zeta)$ where y denotes income, p prices, and $\zeta \in A \subset \mathbb{R}^K$, A compact, is a vector describing the subsidies. We assume that $B(y, p, \zeta) = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : h(x|y, p, \zeta) \leq 0\}$ with $h : \mathbb{R}_+ \times \mathbb{R}_+ \times A \to \mathbb{R}$ an increasing, continuous function in the first two arguments. Note that, for expositional reasons, we now make explicit that function h depends on parameters (y, p, ζ). For instance, in the case of threshold matches, we have that $\zeta = (\tau, m)$ when incentives are present and $\zeta = (0, 0)$ when they are absent. Here, τ is the minimum amount spent on target good g needed to get an additional amount of good g of size m.

Let the initial set of incentives be ζ_0 and consider a set of incentives ζ_1 where $B(y, p, \zeta_0) \subset B(y, p, \zeta_1)$. This assumption reflects that we are dealing with a subsidy, which we usually associate with larger choice set. Equation (1) defines the compensating variation.

$$CV = \{Z : max_{(c,g)\in B(y,p,\zeta_0)} u(c, g; \eta) = max_{(c,g)\in B(y-Z,p,\zeta_1)} u(c, g; \eta)\} \tag{1}$$

It is the reduction in income CV needed to make the agent indifferent between the original incentives ζ_0 and the new incentives ζ_1 and income $y - CV$. The previous discussion shows that those who choose a bundle not in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$ reveal that $B(y, p, \zeta_0) \succeq B(y, p, \zeta_1)$ and those who choose a bundle in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$ reveal that $B(y, p, \zeta_1) \succeq B(y, p, \zeta_0)$. This implies that those who choose a bundle not in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$ have a compensating variation equal to zero. We show next how to construct a family of budget sets to reveal the compensating variation of those who choose a bundle in $B(y, p, \zeta_1) \setminus B(y, p, \zeta_0)$.

We define a family of budget sets ordered by reductions in income Δ: $\{B_\Delta\}$ where $B_\Delta = B(y, p, \zeta_0) \cup B(y - \Delta, p, \zeta_1)$. Given our assumption, we have that $\bigcap_{\Delta\geq 0} B_\Delta = B(y, p, \zeta_0)$, i.e. the choice set reverts to the original budget set as income is reduced. Consider those choosing an option in $B(y - \Delta, p, \zeta_1) \setminus B(y, p, \zeta_0)$ when $\Delta = 0$, i.e. they changed their behavior due to the incentives. We can conclude that their CV is not negative. This is due to revealed preferences and preferences being monotone. Suppose $\Delta > 0$, i.e., available options are reduced; we can conclude that those who

switch to a bundle in $B(y,p,\zeta_0)$ not in $B(y-\Delta,p,\zeta_1)\setminus B(y,p,\zeta_0)$ must have a CV no larger than Δ. A choice in $B(y,p,\zeta_0)$ is preferred to all choices in $B(y-\Delta,p,\zeta_1)$. Note that the utility obtained by this new choice must be equal to that of the optimal choice in $B(y,p,\zeta_0)$ in isolation.[5] Due to the monotonicity and continuity of preferences, this switch will happen when Δ is large enough.[6] The sequence of budgets B_Δ allows us to find the reduction in income that makes a person indifferent between the original choice and a choice with the new incentives but less income.

We define the structural choice probability as:

$$q(\Delta,\zeta_1,\zeta_0)=\int_\eta \mathbf{1}\{\{argmax_{(c,g)\in B_\Delta}u(c,g;\eta)\}\cap\{B_\Delta\setminus B(y,p,\zeta_0)\}\neq\emptyset\}dF(\eta) \quad (2)$$

This measures the number of people who benefit from ζ_1 as Δ increases. By the definition of B_Δ there is a Δ' large enough such that $B_{\Delta'}=B(y,p,\zeta_0)$. There is a Δ' such that everyone reverts to their original choices with the original subsidy. We then have that $1-q(\Delta,\zeta_1,\zeta_0)$ reveals the probability that the CV is no higher than Δ. By construction, $q(\Delta,\zeta_1,\zeta_0)$ is decreasing in Δ, so $1-q(\Delta,\zeta_1,\zeta_0)$ provides a distribution of the CV. This leads to the following proposition.

Proposition 1: Let ζ_1 a subsidy such that no agent is worse off with respect to original subsidy ζ_0. Let $q(\Delta,\zeta_1,\zeta_0)$ the structural choice probability as defined above. Let $E[CV|\zeta_1,\zeta_0]$ be the average compensating variation of subsidy ζ_1. Then, $E[CV|\zeta_1,\zeta_0]=\int_{\Delta\geq 0}q(\Delta,\zeta_1,\zeta_0)d\Delta$.

Proof: Since $F_\Delta(x)=Pr(\Delta\leq x)=1-q(\Delta,\zeta_1,\zeta_0)$ and $\Delta\geq 0$, we have that $E[CV]=\int_0^\infty(1-F_\Delta(x))dx$. Then, $E[CV|\zeta_1,\zeta_0]=\int_{\Delta\geq 0}q(\Delta,\zeta_1,\zeta_0)d\Delta$.

In the paper, we construct $B_\Delta=B(y,p,\zeta_0)\cup B(y-\Delta,p,\zeta_1)$ by increasing the threshold and reducing the match by the same amount and define the structural choice probability as the proportion of those donating above their sum. In Figure B.2, we provide two examples of how to calculate the CV of two alternative policies. The left panel illustrates the construction of B_Δ for the case of a marginal subsidy for charitable donations. In this case, mimicking reductions in income is equivalent to

[5] If not, the original choice was not optimal since the new choice was available but not chosen.

[6] The maximum is continuous and decreasing in Δ since the constrained correspondence varies continuously in Δ and preferences are continuous and monotone increasing (see Mas-Collel et al., 1995, Theorem M.K.6). There is at least one value of Δ at which utility equals the optimal utility at ζ_0.

increasing the minimum donation to qualify for the subsidy without changing its magnitude. The structural choice probability is the proportion of donors donating above these thresholds. Note that the CV of a linear match subsidy could be calculated this way if we had access to the appropriate choice probability. The right panel of Figure B.2 illustrates the construction of B_Δ for the case of an income-based transfer as the one discussed in Bergstrom et al. (forthcoming).[7] In this case, mimicking reductions in income is equivalent to decreasing the transfer amount without changing the threshold (τ) to qualify for the transfer. The structural choice probability is the proportion of donors reporting an income below the threshold. The distribution of the CV is identified in both of these cases as long as the structural choice probability is identified. As in our application of subsidies in charitable giving, researchers might use policy changes to approximate these functions.

The discussion above has concentrated on estimating the compensating variation of threshold subsidies. The approach introduced in this paper can be adapted to calculate the equivalent variation of the subsidy as well. To calculate the equivalent variation, we need, instead, to define a family of budget sets ordered by increments in income Δ: $\{B_\Delta\}$ where $B_\Delta = B(y+\Delta, p, \zeta_0) \cup B(y, p, \zeta_1)$. In our experiment, this would have required offering matches of different magnitudes regardless of the donation size while keeping higher threshold matches constant. Because in our experiment, the position of lower matches resembles an unconditional match for those who donate larger amounts, the equivalent variation of large thresholds could be approximated. We do not pursue this in the paper.

A.4 Threshold match subsidies

Equation (3) provides a formal definition of the compensating variation (CV) for a donor with income y facing initial prices and receiving a threshold match of $m(\tau)$ if total donation minus the match exceeds τ. $\mathbf{1}\{A\}$ is an indicator function that equals 1 if statement A is true and zero if not.

$$CV = \{Z : max_{\tilde{g}} u(y-\tilde{g}, \tilde{g}; \eta) = max_{\tilde{g}} u(y-\tilde{g}+\mathbf{1}\{\tilde{g} \geq \tau + m(\tau)\}m(\tau) - Z, \tilde{g}; \eta)\} \quad (3)$$

Threshold match subsidies are a special case of the subsidies described in the

[7] To fit this case in the framework described above, we need to redefine variable y as $-y$.

previous section. The only possible complication is the budget not being defined by a continuous function. That means we need to redefine the budget set to allow for free disposal of the matches. In practice, this is not a problem since monotone-increasing preferences imply that no agent would throw away the matches if allowed to do so.

For g^* the optimal donation without subsidies, define the structural choice probability $q(\tilde{\tau}, m(\tilde{\tau}))$ for threshold $\tilde{\tau}$ and match $\tilde{m}$ as:

$$q(\tilde{\tau}, \tilde{m}) = \int_{\eta} \mathbf{1}\{g^* \leq argmax_g u(y - g + \mathbf{1}\{g \geq \tilde{\tau} + \tilde{m}\}\tilde{m}, g; \eta)\} dF(\eta) \tag{4}$$

This is the proportion of people who increase total donations (out-of-pocket plus match) above $\tilde{\tau} + \tilde{m}$. We want to calculate $Pr(CV \leq \Delta)$ for $\Delta \leq m(\tau)$. The results in the previous section imply that $Pr(CV \leq 0) = 1 - q(\tau, m(\tau))$. Those who increase total donations reveal an increase in utility and a positive CV, and those who do not reveal their CV is not positive. Suppose we could observe the reaction to a higher threshold $\tau + \Delta$ and a lower incentive $m(\tau) - \Delta$. This is equivalent to a decrease in income of magnitude Δ. Those whose total donations exceed $\tilde{\tau}+\tilde{m}$ at $(\tau+\Delta, m(\tau)-\Delta)$ reveal their gains in utility are larger than a loss of income of Δ. Their CV must be larger than Δ. We have that $Pr(CV \leq \Delta) = 1 - q(\tau + \Delta, m(\tau) - \Delta)$. We can continue this argument until $\Delta = m(\tau)$. To show that $Pr(CV \leq \Delta)$ is a proper cumulative probability function, suppose $\Delta < \Delta'$. We have that $q(\tau+\Delta, m(\tau)-\Delta) \geq q(\tau + \Delta', m(\tau) - \Delta')$ and $1 - q(\tau + \Delta, m(\tau) - \Delta) \leq 1 - q(\tau + \Delta', m(\tau) - \Delta')$ or $Pr(CV \leq \Delta) \leq Pr(CV \leq \Delta')$.

We are ready to state our Proposition. We redefine the structural choice probability $q(\cdot)$ to $\tilde{q}(\cdot)$ to make the policy being evaluated more salient.

Proposition 2: Define $\tilde{q}(G, s_i, m_i) = P_N(g_i + \mathbf{1}[G - m_i + s_i](m_i - s_i) \geq G)$ as the probability that total donations exceed amount G when a match of $m_i - s_i$ is awarded to any out-of-pocket donation exceeding $G - m_i + s_i$. Let $E[CV|G, m]$ be the expected value of the compensating variation of a threshold match with threshold $G - m$ and match m. Then, $E[CV|G, m] = \int_0^m \tilde{q}(G, x, m)dx$. Consequently, the expected excess burden of the subsidy is $E[CV|G, m] - m \times \tilde{q}(G, 0)$.

In sum, we have the CV of the policy or the willingness to pay for it. Since $F_{CV}(x) = Pr(CV \leq x)$ and $CV \geq 0$, we have that $E[CV] = \int_0^\infty (1 - F_{CV}(x))dx$. Then, $E[CV] = \int_0^m q(\tau + \Delta, m(\tau) - \Delta)d\Delta$. Section 2.1 shows how to estimate an upper bound for the CV and a lower bound for the excess burden.

B Additional tables and figures

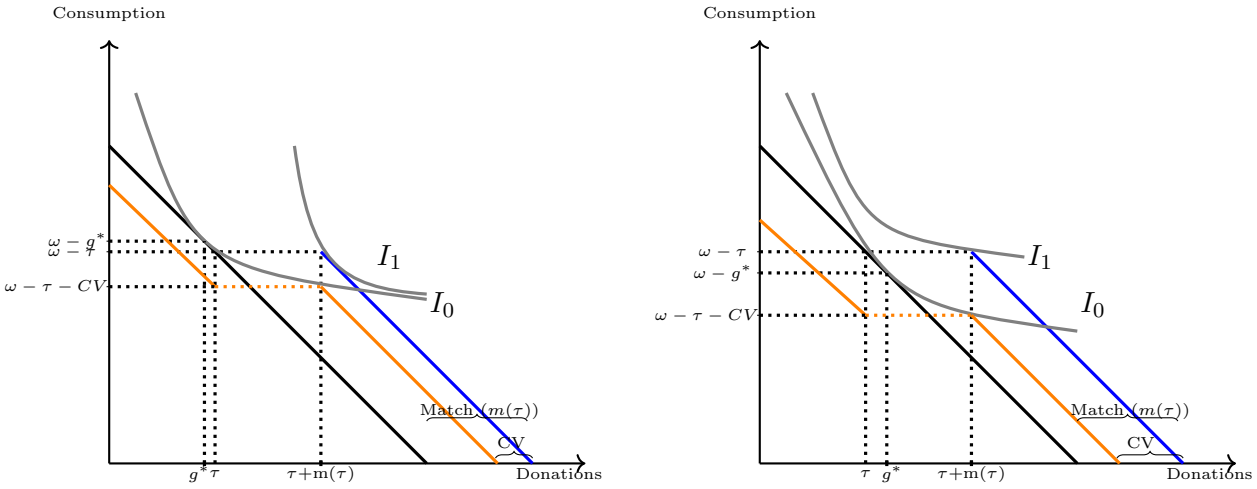
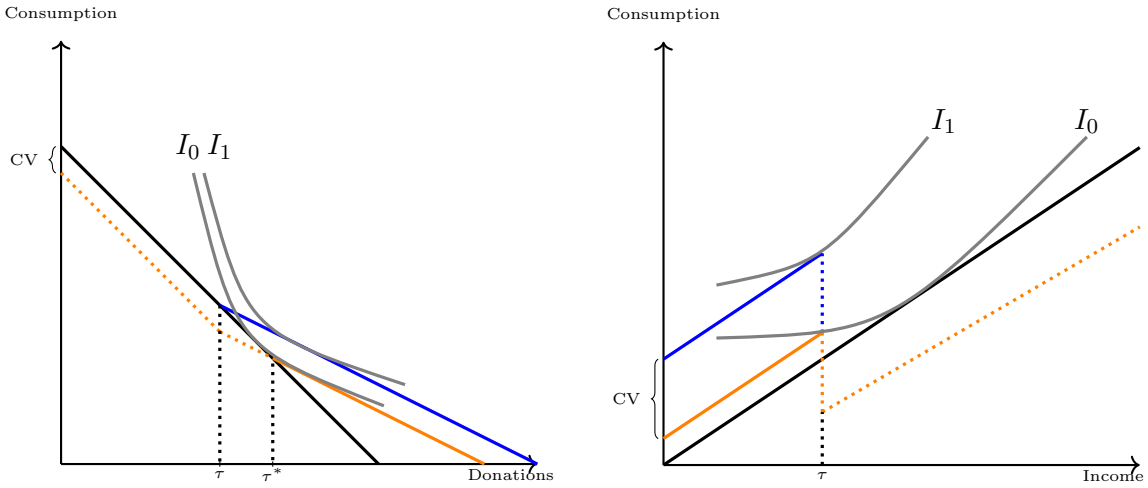


Figure B.1: Identifying CV

Notes: The figures illustrate how changes in threshold matches can be used to identify the compensating variation of the original subsidy. The left panel shows the case of a donor who increase out-of-pocket donations above the threshold (τ) and reverts to the original out-of-pocket donation (g^*) when the threshold match is sufficiently less generous. The right panel shows the case of a donor who decreases the out-of-pocket donations to meet the threshold (τ) and reverts to the original out-of-pocket donation ($g^* > \tau$) when the threshold match is sufficiently less generous.



Marginal subsidy | Income-based transfer

Figure B.2: Identifying CV - Additional examples

Notes: The figures illustrate how to construct policy changes that mimic changes in income. The ability to mimic changes in income is needed to identify the CV of subsidies. The left panel shows the case of a marginal subsidy to donate. In this case, a change in income can be mimicked by an increase in the minimum out-of-pocket donation at which the subsidy takes effect ($\tau \to \tau^*$), keeping the marginal subsidy constant. Those who donate less than τ with the original marginal subsidy have a CV equal to zero. The right panel shows the case of an income-based cash transfer. In this case, a change in income can be mimicked by a decrease in the cash transfer, keeping the minimum income to qualify for the policy (τ) constant. Those who report an income above τ with the original cash transfer have a CV equal to zero. As discussed in the paper, the new budgets are the union of the original budget and the budget with incentives but a reduced income. The new budgets so constructed allow for detection when individuals switch back to the original choice, revealing their CV.

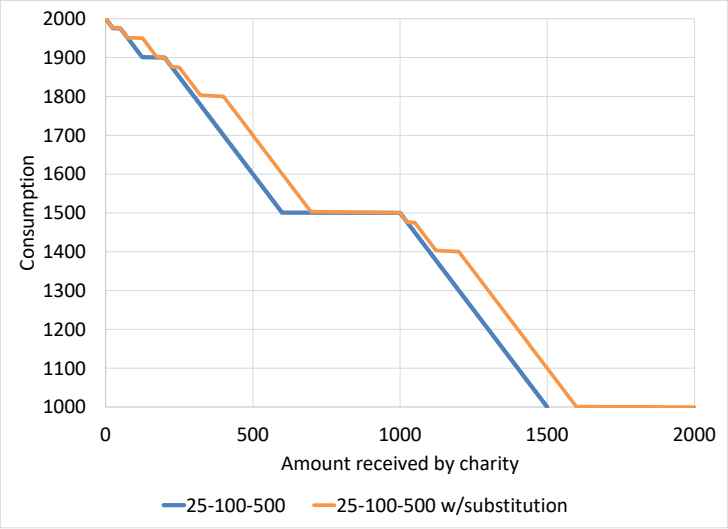


Figure B.3: Budget frontier for two consecutive threshold matches

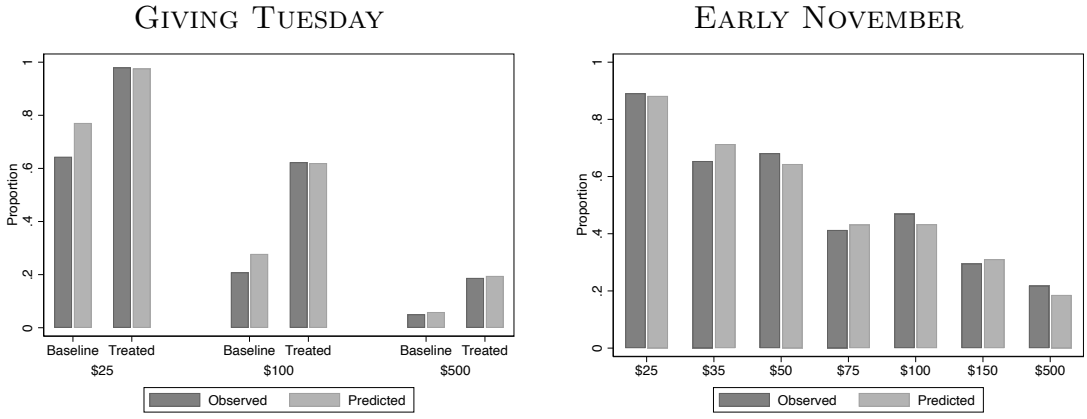


Figure B.4: Fit of Prediction Model - using only experiment data

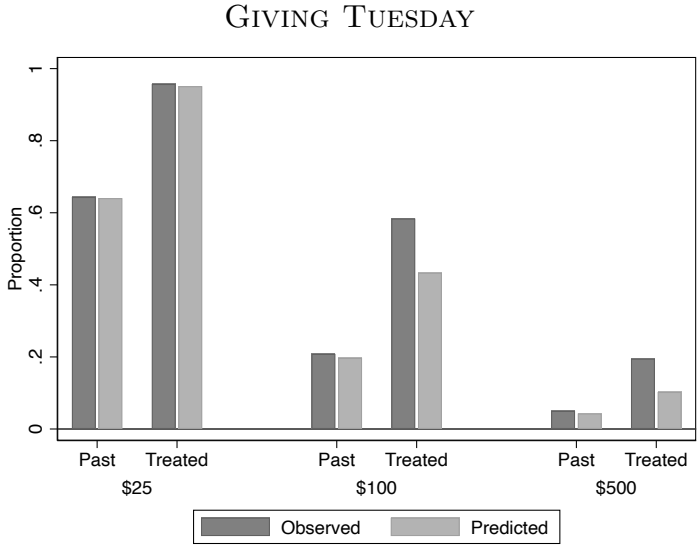


Figure B.5: Fit of Prediction Model - Dynamic model

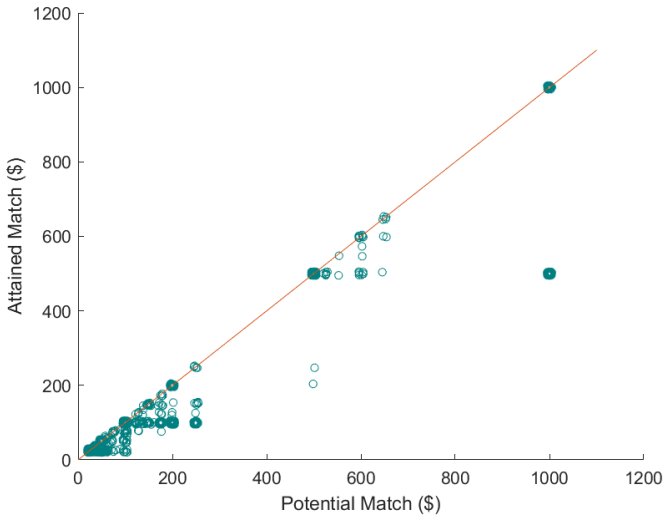


Figure B.6: Timing of donations and efficiency

Table B.1: Kolmogorov-Smirnov test of equality of distributions

	Past donations	Control	T1	T2	T3	T4	T5	T6	T7	T8	T9
Control	0.235										
T1	0.000	0.896									
T2	0.000	0.775	0.156								
T3	0.002	0.568	0.541	0.252							
T4	0.278	0.672	0.068	0.036	0.262						
T5	0.004	0.374	0.528	0.006	0.036	0.173					
T6	0.001	0.652	0.409	0.182	0.253	0.235	0.045				
T7	0.747	0.999	0.993	0.941	0.611	0.940	0.966	0.679			
T8	0.906	0.975	0.958	0.711	0.992	0.983	0.689	0.751	1.000		
T9	0.662	0.947	0.998	0.956	1.000	0.904	0.759	0.967	0.956	0.921	
Giving Tuesday	0.000	0.812	0.926	0.123	0.633	0.071	0.056	0.196	0.933	0.995	0.998

Table B.2: Welfare effects of multiple threshold subsidies offered simulataneously

	All ($25, $100, $500)		
	Est.	2.5%	97.5%
Subsidy ($)	191.8	179.6	203.4
CV ($)	104.2	99.6	108.5
EB ($)	87.7	79.9	94.9
EB per dollar	0.457	0.444	0.469
Additional above threshold	0.396	0.360	0.432

Table B.3: Welfare effects of alternative thresholds using only experimental data

	$25 threshold $25 match			$100 threshold $100 match			$500 threshold $500 match		
	Est.	2.5%	97.5%	Est.	2.5%	97.5%	Est.	2.5%	97.5%
Subsidy per donor ($)	24.6	24.4	24.7	74.9	70.2	79.8	130.4	106.7	153.6
Compensating variation ($)	23.1	22.6	23.6	54.3	49.1	59.8	77.2	60.0	94.1
Excess Burden (EB) ($)	1.4	1.1	1.8	20.5	19.3	22.0	53.2	45.0	61.6
EB per dollar of subsidy ($)	0.058	0.044	0.073	0.274	0.248	0.304	0.408	0.374	0.451
Percent exceeding threshold	0.983	0.974	0.990	0.749	0.702	0.798	0.261	0.213	0.307

Table B.4: Welfare effects of alternative thresholds allowing dynamic decisions

	$25 threshold $25 match			$100 threshold $100 match			$500 threshold $500 match		
	Est.	2.5%	97.5%	Est.	2.5%	97.5%	Est.	2.5%	97.5%
Subsidy per donor ($)	24.5	24.4	24.6	50.6	48.0	53.3	66.0	55.4	76.8
Compensating variation ($)	20.3	20.2	20.3	35.3	34.0	36.7	44.0	38.6	49.3
Excess burden (EB) ($)	4.2	4.1	4.3	15.3	14.0	16.7	22.0	16.8	27.5
EB per dollar of subsidy ($)	0.101	0.097	0.105	0.279	0.264	0.293	0.286	0.239	0.326
Percent exceeding threshold	0.963	0.957	0.969	0.528	0.495	0.560	0.116	0.098	0.134

Table B.5: Prediction model

	All none
ln(Threshold)	2.152***
	(0.387)
ln(Threshold) sq.	-0.532***
	(0.069)
ln(Threshold) cu.	0.035***
	(0.004)
Match/Threshold	-20.153***
	(1.499)
Match/Threshold×ln(Threshold)	7.654***
	(0.580)
Match/Threshold×ln(Threshold) sq	-0.617***
	(0.054)
Constant	-3.555***
	(0.703)
Obs	880,686
Donors	49,882
R2	0.0955
F-test: Thresholds	< 0.001
F-test: Shares	< 0.001

C Field experiment materials

This Appendix shows the wording for one of the treatments and control emails. Amounts in brackets change depending on the treatment.

C.1 Wording of email appeal - $50-$150-$500 treatment:

Subject line: Your donation Nov 1-3 will be matched

Message: We have two great matching offers in November − one that starts today and one on Giving Tuesday (November 28).

Thanks to a generous supporter, any donation of at least [$50] between now and November 3 will be matched.* [charity name] will receive a

- [$50] match if your donation is between [$50] - [$149]
- [$150] match if your donation is between [$150] - [$499]
- [$500] match if your donation is [$500] or above

On Giving Tuesday, a donation between $25-$99 will receive a $25 match, a donation between $100-$499 will receive a $100 match, and a donation of $500 or above will receive a $500 match.

If you are able to give [$50] or more today, your gift will be matched, and any donation of $25 or more on Giving Tuesday will still be matched.

[DONATE NOW]

* Today's match offer requirements: (1) Donations must be made using the same email address to which this message was sent, (2) This match offer can only be used by the recipient of this message. It cannot be shared with others, (3) Matches on Giving Tuesday are not subject to these two requirements, (4) Only one donation will be matched per period (Nov 1-3 and Giving Tuesday), (5) November matches subject to a $10,000 cap.

C.2 Wording of Control email:

Subject line: Your donation in November will be matched

Message: We have a great matching offer available on Giving Tuesday (November 28).

Thanks to a generous supporter, any donation of at least $25 will be matched.* [charity name] will receive a

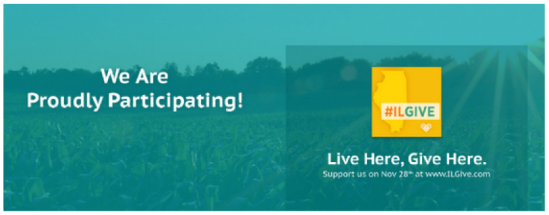
- $25 match if your donation is between $25 - $99
- $100 match if your donation is between $100 - $499
- $500 match if your donation is $500 or above

If you are able to make a donation on Giving Tuesday, remember that your gift of $25 or more will be matched.

[DONATE NOW]

* Match offer requirements: (1) Only one donation will be matched, (2) November matches subject to a $10,000 cap.

C.3 Example email sent by a charity ($50-$150-$500 treatment)



Your donation will be matched!

We have two great matching offers in November – one that starts today and one on Giving Tuesday (November 28).

Thanks to a generous supporter, any donation of at least $50 between now and November 3 will be matched.* will receive a

- $50 match if your donation is between $50 - $149
- $150 match if your donation is between $150 - $499
- $500 match if your donation is $500 or above

On Giving Tuesday, a donation between $25-$99 will receive a $25 match, a donation between $100-$499 will receive a $100 match, and a donation of $500 or above will receive a $500 match.

If you are able to give $50 or more today, your gift will be matched, and any donation of $25 or more on Giving Tuesday will still be matched.

Donate now

Figure C.1: Example email sent for Treatment 50-150