Title
Targeting impact versus deprivation

Authors
Haushofer, Johannes
Niehaus, Paul
Paramo, Carlos
et al.

Publication Date
2022-04-21

DOI
10.5072/FK22F7TN60

Series Name: WPS
Paper No.: 199
Issue Date: 21 April 2022

Targeting deprivation versus impact

Johannes Haushofer, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael Walker

Targeting impact versus deprivation*

Johannes Haushofer[1], Paul Niehaus[2], Carlos Paramo[3], Edward Miguel[3], and Michael Walker[3]

[1]Stockholm University
[2]University of California, San Diego
[3]University of California, Berkeley

April 21, 2022

ⓡ Randomized author order.

Abstract

Targeting is a core element of anti-poverty program design, with benefits typically targeted to those most "deprived" in some sense (e.g., consumption, wealth). A large literature in economics examines how to best identify these households feasibly at scale, usually via proxy means tests (PMTs). We ask a different question, namely, whether targeting the most deprived has the greatest social welfare benefit: in particular, are the most deprived those with the largest treatment effects or do the "poorest of the poor" sometimes lack the circumstances and complementary inputs or skills to take full advantage of assistance? We explore this potential trade-off in the context of an NGO cash transfer program in Kenya, utilizing recent advances in machine learning (ML) methods (specifically, generalized random forests) to learn PMTs that target both a) deprivation and b) high conditional average treatment effects across several policy-relevant outcomes. We find that targeting solely on the basis of deprivation is generally not attractive in a social welfare sense, even when the social planner's preferences are highly redistributive. We show that a planner using simpler prediction models, based on OLS or less sophisticated ML approaches, could reach divergent conclusions. We discuss implications for the design of real-world anti-poverty programs at scale.

*Acknowledgements: We thank Siwan Anderson, Arun Chandresekhar, Rajeev Dehejia, Larry Katz, David McKenzie, Jonathan Morduch, Suresh Naidu, Vira Semenova, Monica Singhal, Jack Willis, Kaspar Wüthrich and conference and seminar audiences at the 2021 ASSA and LACEA meetings, CEGA, CEPR Political Economy, Columbia, Harvard, NYU, Oxford CSAE conference, University of Auckland, University of British Columbia, University of Minnesota, UC Merced, UC San Diego, and USC for helpful comments and suggestions. We thank Justin Abraham, Aakash Bhalothia, Christina Brown, Genevieve Deneoux, Tilman Graff, Grady Killeen, Max Lauletta, Michelle Layvant, Layna Lowe, Anya Marchenko, Gwyneth Miner, Max Mueller, Priscilla de Oliviera, Robert On, Rachel Pizatella-Haswell, Emaan Siddique, Zenan Wang, Francis Wong and Kejian Zhao for excellent research assistance. Niehaus is a co-founder, former president (2012-17) and current board member of GiveDirectly. The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

1 Introduction

Targeting is a core element of anti-poverty program design in both poor and rich countries, with program benefits typically targeted to those households or individuals who are "deprived" in some sense, for instance, in terms of wealth, income, or living standards. There is a growing literature in development economics focused on how best to identify such deprived households to target them with anti-poverty programming, via proxy means tests (PMT), community input, ordeal mechanisms, "big data", and other approaches (Hanna and Olken, 2018; Alatas et al., 2012; Brown et al., 2018; Blumenstock et al., 2015, among others).

Yet we know conceptually that targeting the most deprived is only half the problem facing a social planner or policymaker. Welfare-maximizing allocations of scarce resources should generally depend both on how poor people are to begin with and also on how much they would benefit from receiving additional assistance. An implication of this conceptual logic is that we could safely focus on solely targeting deprived households if treatment effect magnitudes were nearly the same for everyone – in which case the benefits would be largest by targeting the poor due to the concavity of the social welfare function – but this would not be the case if there were meaningful treatment effect heterogeneity. For a simple example, targeting small business skills training to people who are unable (for any reason) to themselves run a business would not yield economic gains, and so would simply be a waste of resources. This is not an idle concern: there is growing evidence from the recent microfinance literature in development economics that the "poorest of the poor" may sometimes lack the circumstances or complementary inputs and skills to successfully invest their loans (de Mel et al., 2008; Bhattacharya and Dupas, 2012; Haushofer and Shapiro, 2016; Banerjee et al., 2015; Hussam et al., 2020), and more generally that heterogeneous treatment effects are empirically important (Meager, 2020). [1] These findings raise the question to what extent there is an impact/deprivation trade-off in targeting anti-poverty programs—echoing longstanding debates regarding the possible trade-off between equity and efficiency in the process of economic growth and development more generally (Alesina and Rodrik, 1994; Persson and Tabellini, 1994; Banerjee et al., 2002).

This potential tension is likely to be particularly relevant for cash transfers, an increasingly popular form of anti-poverty programming (Haushofer and Shapiro, 2016; Bastagli et al., 2016, among many others). Because cash transfers can be used so flexibly, there are many reasons to expect heterogeneous impacts across households, including across various outcomes that policymakers and planners would consider important (e.g., consumption,

[1] Other important political economy considerations regarding program targeting, for instance, to maximize politician votes (see Lindbeck and Weibull, 1987; Manacorda et al., 2011) are not our focus in this paper.

income, nutrition, etc.). Most immediately, non-homothetic household consumption preferences could lead to differential patterns of impact across poor and rich households (along dimensions of interest to the planner), as could different marginal propensities to save versus consume, various behavioral biases that could be more influential for deprived households, as well as gaps in the extent to which individuals are affected by market failures such as credit constraints, which are thought to be pervasive in low- and middle-income countries. Taken together, these observations raise the possibility that there may be a trade-off between the competing social welfare goals of assisting the most deprived and maximizing a program's average treatment effect.

This paper characterizes and quantifies this trade-off empirically in the context of a large-scale unconditional cash transfer program in rural Kenya; this program was previously described in Egger et al. (2019) (henceforth, EHMNW) and is similar in design to the project analyzed by Haushofer and Shapiro (2016). The program targeted an unusually large share ($\sim 1/3$) of households in treated villages using a simple PMT, allowing us to consider the potential merits of more nuanced PMT targeting within this set. We use a common set of "PMT-like" baseline characteristics to predict *both* how deprived households will be on a per-capita basis at endline if not treated, and also how impacted they will be by treatment. The machine learning approach we use—generalized random forests (GRF), building on recent advances in Wager and Athey (2018), Chernozhukov et al. (2018), and especially Athey et al. (2019) — thus treats the two prediction problems symmetrically, using an approach that we pre-specified on the AEA RCT Registry.[2] Specifically, we partition the study population of eligible (and thus relatively poor) households into the 50% most (vs. least) deprived, and the 50% most (vs. least) impacted by the cash transfer program, and examine the overlap between these groups and the possible trade-offs between targeting their members.[34]

We apply this approach to a set of pre-specified financial outcomes—consumption expen-

[2]See https://www.socialscienceregistry.org/trials/505 for more information.

[3]The analysis is related to theoretical work on Empirical Welfare Maximization (Manski, 2004; Kitagawa and Tetenov, 2018; Athey and Wager, 2021), although this work has not focused on the specific and policy-relevant trade-off between deprivation and impact that is central to this paper. More closely related is Björkegren et al. (2021), who study the inverse of our problem: given an actual targeting rule, they ask what we can infer about the preferences of the government that chose the rule. Hussam et al. (2020) examine treatment effects forecasts obtained via machine learning as a benchmark for those elicited from community members, and McKenzie and Sansone (2019) finds limited additional benefits from using machine learning methods over and above the predictive power of a few key covariates in predicting entrepreneurial success in Nigeria. Bertrand et al. (2021) employ ML and other approaches to evaluate how to improve the targeting of workfare programs in Ivory Coast.

[4]A small emerging literature in development economics examines the potential trade-off between deprivation and impact across alternative targeting paradigms. **?** compare PMT targeting to alternatives in a cash transfer program in Niger and do not find evidence of a trade-off. Basurto et al. (2020) show that chiefs in Malawi tasked with assisting the needy tend to target productive farm inputs to households that have higher returns to their use, relative to the allocation achieved by a strict PMT approach.

ditures, assets, and income—that are important objectives for development policymakers, as well as to measures of food security. In a first main finding, we document a substantial trade-off between targeting for deprivation versus for impact in the realm of household consumption: those predicted to be in the most deprived half of the sample (the "D group"), if untreated, indeed have lower per capita endline consumption (by 43%) than those predicted to be in the most impacted half (the "I group"). This difference would provide an initial rationale for targeting the most deprived. However, we then demonstrate that a trade-off exists, showing that the average treatment effect for consumption is 67% larger in the most impacted half of the sample compared to the most deprived half. These magnitudes differ somewhat across outcomes, indicating that the trade-offs facing policymakers may also depend on the key outcome of interest. For instance, the most deprived households have 83% lower asset holdings per capita than the most impacted households, but also have a 18% smaller treatment effect on assets. Similarly, the most deprived households have 46% lower income per capita than the most impacted households; but they also have a 16% smaller treatment effect on income.

In a statistical sense there are two distinct forces that contribute to these apparent trade-offs between targeting for deprivation and for impact. One is that predicted ATEs covary with predicted deprivation; more deprived households tend to have smaller predicted treatment effects. The second is that even *conditional* on predicted deprivation there is substantial variation in predicted treatment effects. Some deprived households experience unusually large program impacts, for example, even in the case (as for consumption) where the ATE is lower among deprived households *on average*. The ML approach in this paper allows us to predict which specific households are likely to be in this set, who are particularly attractive to target from a policymaker's vantage point.

We also explore which economic differences across households drive the observed heterogeneity in program impacts. A priori there is a wide range of possibilities, including differences in preferences (for instance, for saving vs. consumption) and in individual ability, opportunities, or "capability" (Sen, 1999), all of which could deliver different returns on investment. The patterns in our data indicate that the same households that have higher consumption gains also tend to experience larger effects on assets and income. It thus seems likely that some households were able to save and invest the cash in more productive activities than others, and that this yields a higher stream of income and consumption (and greater asset accumulation over time). These differences emerge quickly, generating a trade-off between targeting deprivation and impact even for households surveyed shortly after they received transfers. The households that are able to generate these larger impacts differ along observed characteristics: they tend to be larger households with more prime-age adults and younger household heads (who are perhaps better able to match labor input and human capital to the financial capital they received), as well as households with more assets and greater employment at baseline, perhaps because these characteristics reflect existing business opportunities or underlying ability.

In one of this study's central analyses, we then examine which groups of households in the data a planner with a given social welfare function would optimally select, and how this selection overlaps with conventional deprivation targeting. We find that, for conventional values of α in a constant absolute risk aversion (CARA) utility function, namely α in the range of zero to 0.015 (which includes curvature equivalent to log utility), the social planner generally selects a group that overlaps with both the most impacted (I) and the most deprived (D) substantially, but with more overlap with the most impacted households than with the most deprived group. Even at the upper end of the range of α values the policymaker would still target a majority of the most impacted households. In other words, the conventional policy approach of targeting the most deprived households may not be consistent with social welfare optimal targeting in our data, and this holds across the main financial outcomes considered. Intuitively, there is a greater degree of overlap between the households that are optimally targeted by the social planner and the D group as planner's preferences for redistribution increase (captured in higher values of α).

For the pre-specified food security index we do find some evidence that more deprived households experience larger treatment effects, suggestive of a "hierarchy of needs" (as in Maslow, 1943). The interpretation is subtle, however, as the index – which is similar to those commonly used in development economics and based on survey responses regarding lack of food – appears to capture per capita rather than total household food consumption. This is problematic in our setting since all households received the same amount of money, regardless of size, so that per-capita effects will mechanically tend to be smaller in larger households. If we simply examine total consumption of food instead, the patterns again indicate a trade-off between targeting for deprivation versus impact, consistent with trends for the financial outcomes. This suggests that food security patterns may be driven more by opportunities or preferences and associated dynamics, as with financial outcomes, rather than as following a hierarchy of needs.

One potentially important caveat to these results is the role that spillover effects may play. Egger et al. (2019) document a sizable transfer multiplier of 2.4 due to the cash transfer program in the study area. The existence of spillovers does not necessarily affect the interpretation of the main results—our conclusions would be the same if all households cause and experience the same additive spillovers, for example (at least for CARA social welfare functions). But the interpretation would change to the extent the results capture *predictable* differences in which households *experience* larger spillover effects. In two auxiliary tests for this, using data on both eligible and ineligible households and both within- and between-village exposure to treatment, we do not find evidence that our approach is able to detect such heterogeneity. This provides a degree of increased confidence in the main targeting results.

Finally, we contrast results obtained using GRF to those obtained using a simple OLS regression as well as classic ML approaches, specifically, LASSO (Tibshirani, 1996). While OLS estimates have been widely used in practice to design PMTs, it is well-known that they are not sufficiently regularized (Athey and Imbens, 2019)—and addressing over-fitting is one of the main benefits of ML methods including GRF—thus leading to more extreme values for predicted impact using OLS. Indeed, in our data OLS selects most deprived and most impacted groups similar to those selected by GRF, but yields far too optimistic predictions about how deprived and how impacted they will actually be. Using these predictions for policy-making could therefore lead to targeting that is far less (or in some cases possibly more) redistributive than using the ML approach. Perhaps more surprisingly, using LASSO mitigates this problem only slightly. This illustrates the value of using ML methods such as GRF designed to learn conditional average treatment effects directly, as opposed to using generic methods to learn conditional means in the treatment and control groups separately and then differencing these.[5]

An overall punchline is that the results do not imply that the most deprived households should always be the sole focus of anti-poverty program targeting, although that is the norm in practice. The data indicate that there are important trade-offs for policymakers to consider. Depending on the outcome measure they favor and the degree of redistributive preferences captured in the social welfare function, the planner might prefer mostly *not* targeting the most deprived households but instead focusing assistance on those predicted to experience the largest impacts.

That said, the findings in this study that motivate this logic apply to one intervention in a single setting, and one program in isolation. Considering a *portfolio* of anti-poverty interventions, targeting one towards the most impacted may *strengthen* the case for targeting others towards the most deprived. For example, an optimal strategy might involve targeting cash transfers to those who benefit most from them (in terms of future income gains), while simultaneously working to remove for the most deprived the barriers that limit their ability to benefit from assistance. Doing so may be particularly important for socially marginal-

[5]The issue here appears to be analogous to that identified by Abadie et al. (2018), who document a bias in conventional approaches to studying impact heterogeneity towards *negative* estimates of the relationship between impact and untreated outcomes. In contrast, our approach yields positive estimates.

ized groups (e.g., female headed households, migrants and members of ethnic or religious minorities) who may lack the same market opportunities as other households.

2 Conceptual framework

We study the problem of choosing which households h to receive treatment (e.g., program assistance) in order to maximize a social welfare function

$$\sum_h W(Y_h(T_h)) \tag{1}$$

Here Y_h is a real-valued outcome of interest such as consumption, wealth, or food security, which potentially depends on the household's assignment to receive treatment, indicated by $T_h \in \{0, 1\}$. For simplicity we will think for now of each household as having a single member, abstracting from variation in household size (which we will introduce when we map the framework to the data in Section 4). The function $W : \mathbb{R} \to \mathbb{R}$ satisfies $W' > 0$ so that higher values of each household's outcome are preferred, and $W'' \leq 0$ so that gains matter (weakly) more for households that are more deprived to begin with.

Using potential outcomes notation allows us to rewrite this objective as

$$\sum_h W(Y_h^0 + T_h \cdot \Delta_h)) \tag{2}$$

where $Y_h^{T_h} \equiv Y_h(T_h)$ and $\Delta_h \equiv Y_h^1 - Y_h^0$ is h's treatment effect. This reformulation highlights the potential tension between two distinct objectives: targeting benefits to those *worst-off* absent the intervention (i.e. have the smallest Y_h^0's), and targeting benefits to those who will be *most positively impacted* by the intervention (largest Δ_h's). These objectives are captured in a disciplined way here, in the sense that both are tightly linked through the function W; W determines both the strength of preference for targeting deprived households, and also the extent to which large treatment effects are discounted due to diminishing marginal benefits.[6]

One can interpret the criterion function (2), and in particular the variation in treatment effects, in two distinct ways. One is that W correctly represents households' preferences over their own outcomes, but that households face different opportunities and constraints. Some may possess investment opportunities that others lack, for example, so that they are able to increase their standard of living more after receiving treatment (a household cash transfer in our empirical application). In this case households might agree – from a vantage point behind

[6] One could extend the framework by incorporating ad hoc weights to capture other forms of distributive preference (e.g. for historically disadvantaged minorities) without qualitatively altering the main ideas.

a "veil of ignorance" in which they do not yet know their specific draw of (Y_h, Δ_h) – that (2) is the appropriate objective of policy. Alternatively, W may represent the preferences of a paternalistic planner or policymaker, which differ from those of the households themselves. For example, households' time preferences may vary, and the policymaker may prefer that they make relatively "patient" choices.[7] In this case, maximization of (2) would implement policy-maker rather than household preferences.

We consider how to balance the objectives captured by (2) subject to information constraints facing a typical policymaker. Specifically, we suppose that she cannot observe Y_h^0 and Δ_h in the full population. This reflects the costs of gathering data on complex outcomes such as consumption, the fact that claims about these outcomes are hard to verify, and (in the case of Δ_h) the more fundamental issue that she can never directly observe a household's counterfactual outcomes. Instead we suppose that she observes a set of covariates $X_h \in \mathbf{X}$ in the full population, as well as the realized outcomes $Y_h(T_h)$ from a representative *experimental sub-sample*. We think of X_h as representing the kinds of variables typically seen in proxy means tests used to target programs in low- and middle-income countries (LMICs), e.g. major assets, household size, number of children, sector of employment, etc. The planner uses these data to select a rule $r : \mathbf{X} \to \{0, 1\}$ determining assignment to treatment in the rest of the population, subject to any budget or enrollment constraints, for instance, that there is sufficient funding to treat a share ϕ of households in the population.

Data from this experimental sample enable the planner to consider targeting based on *predictions*:

$$\hat{Y}^0(X_h) \text{ of } \mathbb{E}[Y_h^0 | X_h] \tag{3}$$

$$\hat{\Delta}(X_h) \text{ of } \mathbb{E}[Y_h^1 - Y_h^0 | X_h] \tag{4}$$

obtained from these data. For example, one approach would be to target based (solely) on predictions of the endline outcome $\hat{Y}^0(X_h)$, using treatment rules of the form

$$r^D(X_h) = 1(\hat{Y}^0(X_h) \leq q_\phi^{\hat{Y}}) \tag{5}$$

where q_ϕ^Z denotes the ϕ'th percentile of the empirical distribution of given variable Z, and the D superscript denotes "deprivation". This is precisely what the proxy means testing approach to targeting does, and is widely used by policymakers in practice. Notice that this approach will be appealing in a social welfare sense if there is wide variation in predicted deprivation $\hat{Y}^0(X_h)$ while treatment effects are relatively homogenous.

[7] Paternalism over others' time preferences seems to be common, as for example Ambuehl et al. (2021) document in the lab.

An alternative would be use $\hat{\Delta}(X_h)$ to target the group predicted to be most impacted (denoted I) by treatment:

$$r^I(X_h) = 1(\hat{\Delta}(X_h) \geq q^{\hat{\Delta}}_{1-\phi}) \tag{6}$$

This approach is uncommon in practice, to our knowledge, but research interest in it is growing as statistical tools for predicting heterogenous treatment effects are developed. It is intuitively appealing if Y^0_h does not vary (much) relative to Δ_h or if the social welfare W is (nearly) linear in its argument.

Finally, the planner might make use of both predictions, ranking households by the incremental contributions to social welfare that treating them would induce given their predicted outcomes:

$$d\hat{W} \equiv W(\hat{Y}^0(X_h) + \hat{\Delta}(X_h)) - W(\hat{Y}^0(X_h)) \tag{7}$$

$$r^*(X_h) \equiv 1(d\hat{W} \geq q^{d\hat{W}}_{1-\phi}) \tag{8}$$

This rule r^* strikes a balance between targeting deprivation and impact, with the terms of the tradeoff governed by the curvature of W. The empirical approach in this paper allows us to explore this tradeoff quantitatively by examining the joint distribution of $(\hat{Y}^0(X_h), \hat{\Delta}(X_h))$ and how the particular households h selected for treatment vary depending on W.[8]

3 Study design

We study targeting in the context of a large-scale experimental evaluation of unconditional cash transfers to low-income rural Kenyan households, previously studied by EHMNW. That paper provides details on the setting and design which we briefly summarize here.

3.1 Setting: rural western Kenya

The study took place in three contiguous subcounties of Siaya County, a largely rural area in western Kenya, which the NGO GiveDirectly (GD) had selected based on its high poverty levels (Figure A.1). Within this area, GD selected rural (i.e., not peri-urban) villages in which it had not previously worked. This yielded a final sample of 653 villages spread

[8] In contrast, the Empirical Welfare Maximization literature (Manski, 2004; Kitagawa and Tetenov, 2018; Athey and Wager, 2021) focuses on predicting $W(Y(T, X))$ using X directly, yielding predictions $\hat{W}_h(T_h)$, and then selecting for treatment observations with high values of $\hat{W}_h(1) - \hat{W}_h(0)$. This approach yields useful guarantees about the asymptotic performance of the targeting rule, but obscures the policy relevant tradeoff between impact and deprivation that we wish to draw out here.

across 84 sublocations (the administrative unit above a village). The mean village consists of 100 households, and at baseline, the average household had 4.3 members, of which 2.3 were children. The average survey respondent was 48 years old and had about 6 years of schooling. 97% of households were engaged in agriculture; at endline, 49% of households in control villages were also engaged in wage work and 48% in self-employment. Transfers and data collection took place from mid-2014 to early 2017, a period of steady economic growth, relative prosperity, and political stability in Kenya.

3.2 Intervention

The enrollment of households was relatively inclusive. GD defined as eligible all households that lived in homes with thatched (as opposed to metal) roofs. GD then enrolled all households that met this criterion in villages assigned to treatment. Based on our household census data (described below), 35%-40% of households were eligible. This is far more inclusive than existing public programs in the area, which reached 1.3% of individuals and 6.5% of households in Siaya at the time.[9] That said, the results (described below) may still understate the potential to boost social welfare by targeting even less-deprived households

Eligible households received transfers totaling KES 87,000, or USD 1,871 PPP (USD 1,000 nominal), which constitutes 75 percent of mean annual household expenditure. All transfers were delivered via the mobile money system M-Pesa, and households selected the member they wished to receive them. Transfers were delivered in a series of three tranches: a token transfer of KES 7,000 (USD 151 PPP) sent once a majority of eligible households within the village had completed the enrollment process, followed two months later by the first large installment of KES 40,000 (USD 860 PPP). Six months later (and eight months after the token transfer), the second and final large installment of KES 40,000 was sent. Beyond this point transfers were non-recurring, i.e., no additional financial assistance was provided to recipient households after their third and final installment, and they were informed of this up front. Households in control villages did not receive transfers.

3.3 Experimental design and data

The study employed a two-level randomization design. First, we randomly assigned sublocations (or in some cases, groups of sublocations) to high or low saturation status, resulting in 33 high- and 35 low-saturation groups. Within high (low) saturation groups, we then randomly assigned two-thirds (one-third) of villages to treatment. Randomization was well-

[9] Data provided by GiveDirectly, originally from the Government of Kenya's Single Registry for Social Protection.

balanced with respect to an array of household demographic and economic characteristics (see Table A.1 and EHMNW).

We first conducted a baseline household census in all villages, which serves as a sampling frame and classifies household eligibility status. The census was designed to mimic GD's censusing procedure but was conducted by independent (non-GD) enumerators across both treatment and control villages for consistency. The census identified 65,385 households with a total baseline population of 280,000 people in study villages.

Within one to two months after the census, and before the distribution of any transfers to each village, we conducted baseline household surveys. These targeted a representative sample of eight households eligible to receive a transfer and four ineligible households per village. When households contained a married or cohabiting couple, we randomly selected one of the partners as the target survey respondent. We conducted a total of 7,848 baseline household surveys between September 2014 and August 2015, of which 5,123 (66%) were of eligible and 2,722 (34%) were ineligible households, in line with the sampling targets.

We later conducted endline household surveys, targeting all households that had been surveyed at baseline, as well as those that were sampled but missed at baseline, and we attempted to survey the individual who was the baseline respondent. We conducted a total of 8,239 endline household surveys between May 2016 and June 2017, of which 5,423 (66%) were of eligible and 2,816 (34%) were ineligible households. We achieved high respondent tracking rates at endline, reaching over 90% of households in both treatment and control villages, and these rates do not systematically vary by treatment status (Table A.2).[10]

Endline surveys were timed between 9 and 31 months after each household's "experimental start date," meaning the month in which GD transfers were scheduled to start in its village if that village were assigned to treatment.[11] Figure A.2 illustrates the resulting distribution of time elapsed between the date when a given shilling was transferred to a household and the date that household's endline survey was conducted. The mode is roughly 13 months, but with substantial mass at both higher values and at zero lag (i.e., the household was surveyed in the same month as the final transfer). This implies that the data are informative about predicted deprivation and impact over a relatively wide range of time horizons post-transfer—certainly as compared to a typical PMT exercise that uses covariates to predict contemporaneous deprivation, i.e. with no lag. Below we examine how results vary for

[10] In addition to household surveys, the study also collected surveys of enterprises, market prices, and local government. EHMNW and Walker (2018) discuss these data and present additional results.

[11] All study villages, including control villages, were randomly ordered for data collection and (in the event they were assigned to treatment) for treatment. We use these orderings to assign experimental start dates. The median survey was conducted 19 months after the experimental start month, or about 11 months after the distribution of the last lump sum transfer; the 5th/95th percentiles of the gap ranged from 12 to 27 months since the experimental start date, or 4 to 19 months since the final transfer.

households surveyed at different time horizons after transfer receipt.

For the purposes of this paper, we focus primarily on eligible households that were surveyed at both baseline and endline, as we observe them under either treated or control conditions (at endline) and can use baseline values of household characteristics to predict both deprivation and impact. We also require households to have non-missing endline outcome data and baseline covariates.[12] These inclusion conditions yield an analysis sample of 4,749 *eligible* households. Relative to ineligible households, we note (as expected) that eligible households tend to have lower income and net assets on average, but also that there is substantial overlap between the distributions of economic outcomes in the two groups (Figure A.3). Data on the eligible households thus allow us to examine the relationship between deprivation and impact over a relatively wide range of economic conditions, including both among the very deprived as well as relatively well-off households.

We use baseline data on a set of 16 covariates (the vector X_h in the framework above) to predict endline outcomes. We selected variables that we found in other real-world proxy means tests used to target social protection problems and that exhibit meaningful variation in our data. The resulting list includes demographic measures (e.g., household size, indicators for children of various ages) and economic measures (e.g., ownership of major assets, employment status); Appendix B.1 provides the full list.[13]

We focus on four pre-specified outcomes at endline, including core household financial outcomes (namely, consumption expenditure, assets, and income) as well as an index of food security. Details of the construction of these aggregates are provided in Appendix B, and the project's pre-analysis plan (PAP) is posted on the AEA RCT Registry (at `https://www.socialscienceregistry.org/trials/505`). In the main analysis, we predict versions of these outcomes demeaned by the month in which the survey was conducted, in order to remove any effects due purely to correlation between predictors and survey timing, and then add back in the overall mean to all observations for interpretability; results are nearly unchanged without this precaution (see Tables D.1 and D.2).

The three financial outcomes—consumption expenditure, income, and assets—are defined at the household level, the same level at which treatment was assigned, so that they correctly capture the total effects of treatment as opposed to their per-capita analogous (which would under-weight impacts on individuals living in large households). Recall that cash transfers of

[12]Specifically, we exclude households for which more than 7 baseline covariates were missing (which only drops 3 observations). The Generalized Random Forest (GRF) statistical package (discussed below) handles missing covariate values by considering the missing status itself as a potential split on that variable, allowing missing values to be informative.

[13]As discussed in detail in the Appendix, we select predictors by hand rather than using the specific data-driven approach we had original pre-specified, as the latter was not well-defined and creates issues for inference. That said, the main results are all qualitatively robust to using a data-driven approach instead.

the same magnitude were provided to all treatment households regardless of the number of members. Taken together, these outcomes form a natural constellation given their connection via the household's budget constraint, and studying them in tandem allows us to relate the results to canonical dynamic models of consumption and investment. For example, if households vary in their marginal propensity to consume (MPC) as opposed to investing out of a transfer, then we would expect to see negative covariation between *initial* treatment impacts on consumption and accumulated assets. Over time, however, the households that invested more should realize higher levels of income, consumption, and assets. This effect would be especially strong if their higher levels of initial investment were in part the result of higher-return investment *opportunities*, as in this case differences in behavior and differences in returns would be mutually reinforcing.

Food security is an important public policy objective for many transfer programs (though these are usually structured as streams of small payments, as opposed the lump sum transfers studied here). It is also theoretically interesting as a case in which we might expect *a priori* to observe a relatively weak tradeoff between targeting on deprivation versus impact, given that the households most likely to spend on better nutrition are often those not eating enough (see for example Subramanian and Deaton, 1996). Unlike the total household financial outcomes noted above, the index of food security we use is arguably best interpreted as a *per capita* measure: typical constituent questions ask how many days (out of the past 7) family members experienced a negative outcome such as skipping meals, a quantity we would not expect to scale mechanically with household size (as for example total household food consumption would). Indeed, we will show below that results for the food security index parallel those for per capita food consumption, and that these both differ from results using total household food consumption.

3.4 Existing results

EHMNW report the overall average impacts of the GD program on recipient households, estimating positive ITT effects on each of the four outcomes we consider here, among others. They also find large spillovers onto untreated households, for example, substantial expenditure increases for non-recipient households and higher enterprise revenue in areas that received more cash transfers. Using these and related estimates, they derive the implied multiplier effect on overall economic activity, estimating a transfer multiplier of 2.4.

Given these spillover results, the analysis that follows should be interpreted as examining variation in *who* is selected for treatment, holding fixed the total *number* of local households treated. Spillover effects do not alter this analysis to the extent that they are approximately additive and invariant to the identity of the original transfer recipient. We cannot readily estimate the extent to which different kinds of households *generate* different spillovers; this would require an experiment even larger than our (already very large) one. We can, however, use several complementary strategies to assess the extent to which different kinds of households are *affected* differently by spillovers; we return to this issue below.

With respect to heterogeneity of treatment effects, EHMNW take the conventional approach of testing across a pre-specified, researcher-selected set of covariates (including, for example, respondent gender, age, marital status, and educational attainment, among others). They generally fail to reject homogeneity of treatment effects along these dimensions but are only moderately powered to detect effects (Figure A.4, reproduced from EHMNW). We therefore turn next to examining data-driven ML approaches to identifying features of the baseline data that (potentially) predict deprivation and impact.

4 Empirical methods

This section describes the empirical methods used to operationalize the ideas outlined in the conceptual framework. Broadly speaking, the approach is to (i) predict (per capita) outcomes absent treatment, and treatment effects, for each household as a function of its baseline covariates; (ii) classify households into groups based on whether they are or are not among the most deprived or most impacted households according to these predictions; and then (iii) measure deprivation and impact within the extremal groups selected by this procedure using simple OLS estimators. We discuss among other things the approach to regularization and to inference. The analysis follows a pre-analysis plan submitted to the AEA registry on 1 September 2017 prior to the estimation of treatment effects for these outcomes.[14]

Because the outcomes in the data are measured at the household and not the individual level, the analysis needs to account for variation in household size. Generalizing Equation 2 by interpreting Y_h as a household aggregate and denoting by n_h the size of household h, the planner's objective function is

$$\sum_h n_h W(Y_h(T_h)/n_h) = \sum_h n_h W(Y_h^0/n_h + T_h \cdot \Delta_h/n_h) \tag{9}$$

Note that in this empirical setting the size of the transfers (and thus the cost of treatment) are the same irrespective of household size. We would therefore expect per capita treatment effects to be mechanically smaller in larger households, but this does not mean that they

[14]See https://www.socialscienceregistry.org/trials/505.

Algorithm 1: Select most-deprived and most-impacted groups

Split data into K folds;
foreach $k \in K$ **do**
    Training data $\leftarrow (K-1)$ *other* folds ;
    $\{\hat{y}^{0,k} : \mathbf{X} \to \mathbb{R}\} \leftarrow$ predictor of y_h^0 learned from training data;
    Classify observations in bottom 50% of $\hat{y}^{0,k}(X_h)$ for h in fold k as most deprived (D);
    $\{\hat{\Delta}^k : \mathbf{X} \to \mathbb{R}\} \leftarrow$ predictor of Δ_h learned from training data;
    Classify observations in top 50% of $\hat{\Delta}^k(X_h)$ for h in fold k as most impacted (I);
end

are less attractive to target. Indeed the precise details of optimal targeting here depend on the interplay of the distribution of (n_h, Y_h^0, Δ_h) with the curvature of W, something that is captured in the welfare analysis. That said, the planner generally prefers to target households with large *absolute* treatment effects Δ_h and with low *per capita* outcomes absent treatment (denoted henceforth by $y_h^0 = Y_h^0/n_h$). To see this, note that for small treatment effects welfare is well-approximated by the first-order expansion

$$\sum_h W'(y_h^0) \cdot [Y_h^0 + \Delta_h \cdot T_h] \tag{10}$$

so that the incremental benefit of treating h is approximately $W'(y_h^0) \cdot \Delta_h$. We therefore begin the analysis by identifying the households predicted to be most deprived on a per capita basis, and those most impacted on an absolute basis.[15]

At the core of this approach is the classification procedure summarized in Algorithm 1. The procedure classifies every household in the dataset as either in or out of the set of households that would be most deprived absent treatment, and in or out of the set that would be most impacted by treatment.[16] This procedure aims to reduce the risk of over-fitting by classifying each observation h into groups without making any use of its own outcome Y_h; h is instead classified using a function learned only from folds of the data that do not include it. We set $K = 5$, and (to ensure results are not sensitive to the specific split into K folds) then repeat the entire procedure 150 times and report mean outcomes across these iterations.[17]

[15] We abstract from issues of intra-household inequality, which the data do not let us examine.

[16] In practice we learn models for endline per capita values using the full dataset (i.e., including both treated and control individuals) while including an indicator for treatment status among the predictors. Results are similar if the model is trained on control group data only (Tables D.4, D.5, and D.6).

[17] Most deprived and most impacted thresholds are defined for each fold using only their predictions to avoid overfitting concerns since these are not trained using that fold's data. Therefore, a higher number of folds leads to fewer data points being used to define these thresholds. On the other hand, a lower number of folds leads to fewer data points being used to train each random forest. Given our sample size, 5 folds leads

Predictions are formed by learning the regression function $\mathbb{E}[y_h^0|X_h]$ through random forests and the conditional average treatment effect (CATE) function $\mathbb{E}[Y_h^1 - Y_h^0|X_h]$ through causal forests, using the generalized random forests (GRF) package of Athey et al. (2019). We pre-specified an approach based on random forests as these are an attractive tool for uncovering heterogeneity in this setting.[18] Specifically, the dimensionality of our predictors is low relative to the number of observations and we do not see strong evidence of heterogeneity along dimensions that we (originally) thought might matter. Random forests are particularly well-suited for dealing with such non-sparse settings, and can account for complex non-linearities and interactions between the predictors.

At the same time, using a regularized method is important in an optimal targeting context to mitigate the risk of over-fitting. Naive methods—based for example on OLS—might claim to identify very deprived households or those with large treatment effects, leading to over-stated estimates of the overall anti-poverty impact of a program or to mis-estimation of the tradeoff between deprivation and impact. Regularized methods such as random forests help to address this risk.[19] We report forest-based results as our preferred estimates, and also benchmark these against results using OLS and alternative ML estimators in Section 5.6.

Given a classification of the sample into groups $S = D, I$, we define the following measures of performance. The **predicted averages** are the within-group means of GRF predicted values:

$$\bar{\hat{y}}^0(S) = \frac{1}{|S|}\sum_{i \in S} \hat{y}^0(X_i) \qquad \bar{\hat{\Delta}}(S) = \frac{1}{|S|}\sum_{i \in S} \hat{\Delta}(X_i) \tag{11}$$

These may or may not be consistent for the results a policymaker would actually obtain by targeting group S. While our procedure guards against over-fitting in forming predictions $\hat{Y}_h^0/n_h$ and $\hat{\Delta}_h$ for *individual* households, targeting requires us to take the additional step of *selecting groups* of households based on these predictions. This introduces the additional risk of a "winner's curse." To the extent there is even non-systematic error in the predictions, we

to reasonable subsample sizes for each of these steps. Note also that while we use common splits to learn $\hat{y}^0$ and $\hat{\Delta}$, we obtain essentially identical results if we use separate splits.

[18] The pre-analysis plan specified that we would implement the causal forests approach of Wager and Athey (2018) or methods that improved on it, if any were available by the time data were collected. We therefore implement Athey et al. (2019) which generalizes and extends Wager and Athey (2018). In parallel Chernozhukov et al. (2018) developed attractive methods for learning average treatment effects and characterizing units within *quantiles* of the treatment effect distribution; for our purpose here, however, we require the unit-level predictions that GRF provides.

[19] The GRF package in particular uses cross-fitting and an "honest" approach to growing trees to control over-fitting, and we add to this by classifying each observation without using data from its own fold. Random forests do require some tuning and, unlike for other ML procedures such as LASSO, optimal regularization procedures are not available. We selected tuning parameters from among two options: the GRF package defaults, and an alternative set suggested by one of the authors of the package as a way to provide stronger regularization (see `https://github.com/grf-labs/grf/issues/120#issuecomment-327276697`, accessed 31 August 2021). We use the latter as it provides a closer match between predicted and actual statistics.

will tend to select observations with extreme values of this error. For example, we will tend to classify households with high values of $Y_h^0 - \hat{Y}_h^0$ as deprived, and thus to over-estimate how deprived the most deprived group is.

To address this issue, we also calculate a separate set of **actual averages** which are simply group means (for y^0) or group average treatment effects (for Δ) estimated via OLS:

$$\overline{y}^0(S) = \frac{1}{|S|}\sum_{h \in S} y_h^0 \qquad \overline{\Delta}(S) = \frac{2}{|S|}\sum_{h \in S}\left(Y_h^1 T_h - Y_h^0(1 - T_h)\right) \tag{12}$$

This approach uses predictions of deprivation $\overline{\hat{y}}^0(S)$ and impact $\overline{\hat{\Delta}}(S)$ only to select groups, not to estimate outcomes within those groups. We interpret the comparison between predicated and actual averages as a measure of how successfully our approach predicts results in these groups, where smaller gaps are indicative of better performance.

We employ three distinct approaches to inference. First, for key statistics we report bootstrapped confidence intervals. These have the advantage that they can be asymmetric, reflecting the potential asymmetry involved in selecting maximal elements from a set of statistics.[20] Second, we follow Chernozhukov et al. (2018) by reporting the confidence intervals implied by the median standard error for actual averages as defined above and estimated via linear regression. Conditional on the group definitions for D, I, the control means and CATEs for these groups are asymptotically normal. Moreover, by reporting the median standard error across the 150 iterations we are accounting for the variation that results from the k-fold crossfitting procedure. Nevertheless, because the asymptotic properties of this approach follow from conditioning on the group definitions for D, I these standard errors do not account for the fact that these groups are selected endogenously using the data. Therefore the bootstrapped CIs are our preferred inference approach for the actual statistics. Third, to test the sharp null of *no* heterogeneity in treatment effects we use randomization inference. Specifically, we calculate via re-randomization (clustered at the village level, as in the original design) the probability of observing statistics as extreme as those we see under the null of a constant treatment effect. Following Ding et al. (2016), we consider a range of values for this constant effect, centered at the empirical estimate of the average treatment effect, and report the maximal p-value we observe in that range. We interpret this test not as a guide to optimal policy-making (which should exploit all information in the data) but as a diagnostic to help assess whether the observed values of the statistics of the groups defined

[20]GRF provides asymptotic inference for individual predictions $\hat{y}_h^0$ and $\hat{\Delta}_h$ but not for their joint distribution, so approaches like that proposed by Andrews et al. (2021) are not available. Due to computational limitations we only compute the bootstrap CIs for our main results. For some of our alternative model results, the point estimate of the statistic of interest does not lie within the bootstrap CI. This can occur when the estimator is biased as documented by Karlsson (2009).

through machine learning (D, I) are consistent with a null of no heterogeneity. This helps ameliorate concerns about whether differences in means across these groups are a result of over-fitting (in a setting of no heterogeneity but high outcome variance).

Diagnostics suggest that our procedure, and in particular the repeated 5-fold splitting, produces fairly stable results. Figure A.5 shows, for example, that the mean differences between treatment effects in the most deprived and most impacted groups remain more or less constant if we increase the number of splits from 150 to 300.[21] Figure A.6 shows that the classification of households into most deprived and most impacted groups are also quite stable, with most households assigned fairly consistently to either one or the other group.

5 Results

We next present results, beginning in Section 5.1 with estimates of average deprivation and impact for financial outcomes in the groups we identify as most deprived and most impacted. These estimates quantity the tradeoff (if any) between policies that target benefits *solely* on deprivation and on impact. We consider economic implications of these results in Section 5.2, examining what the joint distribution of treatment effects for different outcomes and the importance of different predictors suggest about the underlying forms of heterogeneity that drive our results. We then move in Section 5.3 to examining quantitatively how a policy-maker would trade off deprivation and impact in our sample given a social welfare function with a particular curvature. We then turn to the case of food security in Section 5.4 examine potential spillover effects in Section 5.5. Finally, Section 5.6 examines the performance of alternative statistical methods for learning deprivation and impact.

5.1 Deprivation versus impact

We begin with levels of deprivation, summarized in Table 1. We first note that actual outcomes (on which we focus) line up closely with those predicted by our model. Examining results for the most deprived group (Column 2), we see that actual averages are similar to and in fact consistently slightly *lower* than predicted averages. This suggests that our regularization and cross-fitting procedures are effective at mitigating over-fitting and "winner's curse" effects, which would tend to lead to over-optimistic predictions about the levels of deprivation we can identify.

Next, and consistent with the long tradition of work on targeting social programs to the

[21] We still report results for 150 splits, however, because we also need to do randomization inference and/or bootstrapping on these which is computationally costly at 150 splits and would be yet more so at 300.

most deprived using proxy means tests, the model identifies groups that are substantially poorer than average. For all three outcomes the average outcome among the most deprived (Column 2) is substantially lower than the overall average (Column 1) — by 30%, 74%, and 43% for per capita consumption, assets, and income respectively. Evidently the predictors contain enough information to identify a sub-population substantially more deprived than average, even among a population that has *already* been selected to be poorer than average using GD's coarser targeting criterion.

Targeting the most impacted, on the other hand, comes at a substantial cost in terms of targeting deprivation. Column 3 reports endline values in the absence of treatment for the group identified by the model as most impacted by treatment. In contrast to the most deprived group, the most impacted group is actually *better-off* than average for each outcome. Relative to the overall sample mean, their levels of per capita consumption, assets, and income are higher by 24%, 54%, and 6%, respectively. As a result, the differences in deprivation between the most deprived and most impacted groups are also large (Column 4). Targeting the most impacted would thus mean targeted substantially less deprived households. Yet how much this matters for welfare would depend on the social preferences of the planner (to which we will return shortly in Section 5.3).

The key question is then whether there are compensating gains in impact. We examine this in Table 2. We first examine treatment effects on the most deprived. For financial outcomes, impacts for this group (Column 2) are consistently below the overall average treatment effect (Column 1). In contrast, outcomes for the most impacted are (as expected) consistently *above* average (Column 3). The net result is that targeting the most impacted as opposed to the most deprived yields substantial gains in treatment effect—equal to 52%, 18%, and 16% of the overall average treatment effect for consumption, assets, and income, respectively (Column 4). Considering these results alongside those in Table 1, we observe a meaningful trade-off between targeting deprivation and targeting impact.

Visualizing the joint distribution of predicted deprivation (absent treatment) and predicted treatment effects can help reveal the patterns driving these results. Figure 1 presents these distributions along with locally smoothed regression fits (Figure A.7 presents the corresponding joint CDFs). We color-code each observation to indicate into which of four groups it falls, based on whether or not it is classified among the most-deprived (low values of $\hat{y}^0$) and among the most-impacted (high values of $\hat{\Delta}$). Observations in the upper-left quadrant are those that are both most impacted and most deprived, and thus targeted under either criterion. Those in the lower-right quadrant are targeted under neither criterion, while those in the upper-right and lower-left quadrants are those on which the two criteria disagree. Financial outcomes are plotted on a common vertical axis scale for comparability.

One noticeable feature of the distributions for all three outcomes is that there is substantial variation in predicted impact *conditional* on predicted deprivation, and vice versa. Even absent any *systematic* relationship between impact and deprivation, this variation creates a trade-off between the two: some households happen to be high-impact and low-deprivation, while others happen to be low-impact and high-deprivation, and the planner must prioritize between these.

In addition to this variation, there is also some evidence of systematic covariation between deprivation and impact, particularly for consumption and assets. Here the slope of the non-parametric fit is positive, indicating that less-deprived households also tend to see larger gains when treated. This helps to explain the trade-off observed between group averages in Tables 1 and 2. The other financial variable, income, displays a slight positive relationship over most of its range, albeit more muted.

5.2 Economic interpretation

What economic forces give rise to the observed trade-off between deprivation and impact?

To explore this issue, we begin with the purely descriptive question of what household characteristics are *statistically* important predictors of deprivation and impact, as these may contain clues as to *economically* important mechanisms. Table 3 summarizes the predictive importance of each of the 16 predictors for explaining variation in both deprivation (Columns 2-4) and impact (Columns 6-8). We measure importance here (as does the GRF package) as a depth-weighted average of the share of splits created in the process of growing trees that split on this variable.[22] A value of 0.07 for "female head," for example, means that 7% of all the splits created (when growing trees) split on whether or not the household had a female head. Numbers in parenthesis indicate the rank of each predictor's importance within that column, and the signs indicate whether it predicts the outcome positively or negatively. The three most important predictors in each column are indicated in bold. For ease of interpretation, predictors are also grouped into two broad categories, demographic characteristics and financial characteristics.[23]

One striking pattern that emerges is the role of household size: it is the most important

[22]The formula is

$$\text{Importance}(x_j) = \Sigma_{k=1}^{4} \left[\frac{\Sigma_{\text{all trees}} \text{ number depth } k \text{ splits on } x_j}{\Sigma_{\text{all trees}} \text{ total number depth } k \text{ splits}} \right] \Big/ \Sigma_{k=1}^{4} k^{-2} \tag{13}$$

Note that this metric sums to 1 across all covariates in the model.

[23]Note that a treatment indicator appears as a predictor of y_h^0 since for efficiency we use all the data to learn these models, and then predict deprivation setting $T_h = 0$. We obtain similar results, however, if we use only control group data to predict y_h^0.

predictor of both deprivation and treatment effects for all three outcomes. This pattern is not mechanical: transfers are fixed irrespective of household size so there is no a priori reason to expect treatment effects to increase in household size. As for deprivation, household size is in the *denominator* of y_h^0 by construction, so that any measurement error will tend to induce a negative relationship, yet larger households still have noticeably higher per-capita values. These patterns call to mind the classic idea of scale economies in household production (Nelson, 1988; Deaton and Paxson, 1998), or of risk diversification, as households with more members may be better able to spare one to undertake risky, higher-return ventures. Consistent with this idea, the most impacted households have substantially more working-age adult members than do the most deprived across all primary outcomes (Figure A.8, Panel (a)).

For deprivation the second-most important predictor overall is also demographic: having an elderly member. Besides its immediate relevance to policy debates over the provision of old-age pensions and other forms of support, this also suggests that life cycle patterns of earning, spending and saving may be one of the economic drivers of differences between the deprived and the impacted. Indeed, sorting households by the age of the household head, we observe that the most deprived are disproportionately likely to be *either* young or old, while the most impacted are more likely to be either young adults or middle-aged (Figure A.8, Panel (b)). Note that we see this pattern even though age of household head is *not* itself a predictor in our model; the model appears to be "inferring" age from other covariates.[24]

To explore in more detail what differentiates the most-deprived and -impacted groups, we also examine baseline values of non-PMT characteristics and treatment effects on additional outcomes. Tables A.3, A.4, and A.5 present these statistics for consumption, assets, and income, respectively.[25] Generally speaking we see that (as expected) most-impacted household appear better-off on a range of baseline characteristics, and also that their larger household size reflects a greater number of working-age adults, consistent with the idea that having more laborers enables them to make greater use of capital. We do not see significant differences in effects on labor supply or occupational choice (employment v.s. self-employment), however, suggesting that it is the returns on these activities rather than the tendency towards them that drives differences in effects on primary outcomes. Note also that treatment

[24]Interestingly, land ownership is not a strong predictor of deprivation (or impact). This is partly because it simply does not vary greatly (with 85% of households owning land), but likely also because—unlike in some other agrarian settings—non-land holders in our context are likely to be profitably engaged in commerce or non-agricultural employment as opposed to working on other people's farms.

[25]Our algorithm assigns each household a distinct classification for each split of the data. For the sake of this exercise we give households an overall classification based on whether they are classified as most deprived (or most impacted) in 50% or more of these splits; in practice, however, most households' classifications are insensitive to splits (Figure A.6).

effects on inter-household transfers are not significantly different; the heterogeneity in effects does not appear to be simply heterogeneity in the tendency to share transfers with others.

The fact that treatment effects on different outcomes have strong common predictors—household size, in particular—suggests that they are likely to be positively related. Figure 2 shows exactly this. The lower triangular section presents scatterplots of predicted treatment effects on different outcomes, pairwise; the upper triangular section reports the corresponding pairwise correlations; and the diagonal shows the unconditional distribution of each treatment effect. We observe that impacts on financial outcomes are all strongly positively associated with each other: households that experience larger consumption gains also tend experience larger asset and income gains, and so on.

The positive relationship observed between treatment effects on the financial outcomes matches what we would expect to see emerging over time from canonical dynamic models of consumption and investment. Suppose that households differ either in their marginal propensity to invest (as opposed to consuming) out of their initial transfer, or in their returns on such investment. In the former scenario we might see a negative relationship between impacts on consumption and investment around the moment of transfer receipt (households in this setting typically spend transfers very quickly). Once investments begin yielding a return, however, we would expect in either scenario to see the households that invested more would have higher assets and incomes, and as a result higher consumption, than those that invested less. The pattern of results we see is consistent with this second, "post-investment" situation.[26]

That said, the observed trade-off between deprivation and impact materializes quickly. In Figure A.9 we split the sample into those surveyed recently and those surveyed late relative to their experimental start date (see above). Timing of surveys was randomly assigned, so that this comparison is unconfounded by other differences between households. We see the same, positive relationship between impact and counterfactual outcomes in both halves of the data, with the relationship if anything slightly stronger among those who had *recently* began receiving transfers. Differences in returns may thus play a larger role than differences in preferences in generating the observed heterogeneous effects.

5.3 Optimal policy under concave social welfare functions

The predicted levels of deprivation and treatment effects examined in Section 5.1 define the possibilities facing a social planner deciding whom to target. To see exactly what they imply

[26]As a point of contrast, Chowdhury et al. (2021) estimate that households that experienced larger treatment effects on assets from a graduation intervention in Bangladesh experienced smaller treatment effects on consumption.

for optimal policy, however, we need to translate variation in *levels* of deprivation in Table 1 into variation in the *marginal* social value of a unit increase in the outcome, i.e. of a given treatment effect.

To do this, we now make concrete the notion of social welfare discussed in Section 2, characterizing the households the planner would choose to treat given a specific social welfare function W. We work in particular with the constant absolute risk aversion (CARA) function

$$W(\hat{y}) = \begin{cases} (1 - e^{-\alpha\hat{y}})/\alpha & \alpha \neq 0 \\ \hat{y} & \alpha = 0 \end{cases} \tag{14}$$

which is is commonly used in applied work.[27] [28]

Interpreting W as a private utility function which the planner sums over agents, α is a private preference parameter that represents those agents' risk preferences, and we can draw on existing estimates of it. Estimates are available from a setting close to ours, the Busara Center lab in Nairobi, where Balakrishnan et al. (2020) estimate values of about 0.001. Of course, a social planner may have stronger redistributive preferences than this implies. We therefore consider a set of values ranging from 0 (risk neutral) to 0.015 (stronger concavity), nesting the Busara estimates but also allowing for substantially more curvature. This range includes most of the estimates in the literature review by Barseghyan et al. (2018), and (for intuition) corresponds to a range of certainty equivalents for a 50-50 gamble between \$0 and \$100 of between \$50 (i.e., for no risk aversion) and \$33 (at $\alpha = 0.015$).

We examine the ways in which social preferences W interact with the joint distribution of predicted deprivation and treatment effects in three ways. We first consider a binary choice between targeting the most deprived (as is currently the norm in practice) and the most impacted, and ask which of these the planner would prefer for given values of α. Note that these are both feasible policies in the sense that (by construction) these groups can be

[27]CARA preferences are one of three representations we pre-specified, along with Constant Relative Risk Aversion (CRRA) preferences and the inequality-averse preferences of Fehr and Schmidt (1999). We prefer results for CARA preferences because in a small minority of cases the predicted per-capita outcomes from our model are negative, and CARA allows us to include these observations (which would be undefined for CRRA). We obtain qualitatively similar results, however, if we truncate the predictions and use CRRA preferences. We have not attempted to compute inequality-averse welfare functions as these depend on pairwise comparisons that are computationally prohibitive in our setting (and are not widely used for social welfare analysis).

[28]An alternative to computing $W(\hat{y})$ is to first calculate $W(y)$ and then learn models to form predictions $\hat{W(y)}$ directly, as in the Empirical Welfare Maximization literature. Empirically we find that learning models perform relatively poorly on the transformed $W(y)$ data due to the wide range of numeric values they take on, however. Our application differs in this regard from the empirical application in Kitagawa and Tetenov (2018), for example, who consider maximization of the average treatment effect on (untransformed) earnings and in a setting where baseline household income is much higher than in ours.

selected using a targeting rule that maps solely from our list of PMT-like covariates. We next estimate via numerical search the critical value α_c at which the planner would be just indifferent between targeting the most deprived and the most impacted. Intuitively, for high enough values (and thus a sufficiently strong preference for redistribution) her priority will be deprivation, while for low enough values (and thus a strong emphasis on overall gains) her priority will be impact. Finally, we examine the set of individuals the planner would choose to treat if allowed to choose *any* targeting rule based on those covariates, and the extent to which this selection overlaps with both the most deprived and the most impacted groups.

The results indicate that *exclusively* targeting the most deprived does not generally yield the greatest welfare gains. In fact, for many plausible parameter values the planner would prefer targeting exclusively the most impacted to targeting exclusively the most deprived (Table 4, Column 3). The critical values at which the planner switches to targeting the most deprived are quite high (corresponding to relatively low certainty equivalents of $32, $41 and $37 for consumption, assets and income, respectively), implying that strong preferences for redistribution would be needed to justify targeting only the most deprived in this setting.

The same broad theme emerges when examining the groups the planner would choose if unconstrained. Columns 1 and 2 in each panel report the overlap of these groups with most deprived and most impacted, respectively. Overlap with the most impacted group is (tautologically) 100% when the planner maximizes the average outcome, i.e. $\alpha = 0$. Stronger preferences for redistribution (higher α) are associated with more overlap with the most deprived, and less overlap with the most impacted, as we would expect. But even with strongly redistributive preferences the planner chooses to target a substantial proportion of her transfers to individuals outside the most deprived group. For example, at $\alpha = 0.015$, which corresponds to a strong preference for redistribution, the share targeted outside the most deprived is 58%, 43%, and 37% for consumption, assets, and income, respectively. These conclusions are also robust to sampling variation; in Figure A.10 we plot estimates corresponding to those in Table 4 with bootstrapped 95% confidence intervals, and see that even at the endpoints of these intervals the planner still includes large shares of non-deprived households in the optimal targeted group. Overall, then, the data suggest that optimal targeting in this context should reflect heterogeneity in *both* deprivation and impact.

5.4 Food security

Food security is a narrower measure of well-being than overall consumption but also of obvious humanitarian and policy interest. Recall that we pre-specified as a measure of food security an index aggregating responses to questions about the number of days out of the past seven that family members experienced negative outcomes, such as skipping meals. As it is unclear whether to interpret this as a per capita or an aggregate measure, we examine results for this index alongside results for both per capita and total household food consumption. We define food consumption as the sum of expenditure on food items (including meals outside of the home) and the estimated market value of own-farm output consumed by the household.

Regardless of which measure is used, the procedure identifies a most deprived group that is at least somewhat more deprived than the average, and than the most impacted group (Table A.6). In terms of per capita food consumption, for example—arguably the conceptually most appropriate measure—the most deprived group's mean consumption is 47% lower than average and 33% lower than that in the most impacted group.

The trade-off with impact is somewhat less pronounced than for financial outcomes. For the food security index itself, estimated impacts are *the same* for the most deprived as the for the most impacted group (Table A.7). This is consistent with the intuitive, Maslovian idea that the poorest households are both most likely to be eating too little and also most likely to spend marginal income on food. For total food consumption, however—arguably the conceptually appropriate quantity here, since households of all sizes received transfers of the same magnitude—we again see a substantial trade-off, with impacts on the most impacted roughly twice as large as those on the most deprived.

Figure 3 makes the same point visually. For the food security index (and to a lesser extent for per capita food consumption) we observe a negative relationship, suggesting there might be little or no trade-off between deprivation and impact. But when we plot effects on total food consumption against deprivation measured in per capita terms, we again see a positive relationship similar to that we observed for our financial outcomes. One might worry that this is driven by consumption of "luxury" food items such as snacks or meals out, but we obtain similar flat to upward-sloping relationships even if we restrict attention to consumption of basic foodstuffs (e.g., staple grains).

Overall, when the appropriate analysis is carried out, the picture that emerges thus seems to be that—as for financial outcomes—there is a non-trivial trade-off between targeting the most impacted and the most deprived. Because absolute impacts tend to be larger for larger households, however, this point is obscured if we only examine impacts on *per capita* measures of food security (including the food security index, which behaves similarly to per capita food consumption).

5.5 Spillover effects

An important open question of interpretation concerns the role of spillover effects. Because treatment in the experiment we study was assigned at the village level, the (differential) effects of treatment that we document on a given household h could in principle reflect differences in both the *direct* effect of transfers to household h itself and also *indirect* effects of transfers to other households in the same village.

The key issue for our purposes is the extent to which indirect effects are predictably heterogeneous. As a concrete example, suppose that households that own businesses tend to benefit disproportionately when their villages are treated with cash transfers. To the extent this is because they invest their own transfers and grow their businesses, the correct inference is that reallocating transfers to them would increase average treatment effects. To the extent this is because they benefit from the shock to demand from their neighbors, however, reallocating transfers to them would have no effect.[29]

One way to assess the importance of this issue is to examine *ineligible* households. We have exactly the same data (predictors and outcomes) for these households as for eligible households, and can thus conduct exactly the same analysis. But in this case the interpretation of the results is unambiguous: because ineligible households did not receive transfers themselves, any predictable heterogeneity we find in the effects of assigning their *village* to treatment must reflect heterogeneous indirect effects. A caveat is that we surveyed roughly half as many ineligibles as eligibles, and thus cannot estimate effects as precisely for this group.

A second, complementary diagnostic is to examine eligible households whose villages were not treated, and focus on variation in their exposure to indirect effects from *outside* those villages. To construct a binary measure of this exposure, we calculate whether their neighborhood treatment intensity, as defined in Egger et al. (2019), is above or below median. We then re-run our analysis replacing the own-village treatment indicator with this high-exposure to transfers indicator and examine whether were able to predict patterns here similar to those in the main results.

Generally speaking, neither approach yields results similar to the main ones presented above. The models do not reliably predict heterogeneity, producing predicted effects on the most impacted that are quite different from estimated actual effects (Tables A.8 and A.9, Column 3). For consumption—the outcome where we found the strongest evidence

[29] Note that any common spillover component that affects all households in a village equally would not alter our welfare analysis, since under a CARA social welfare function a common additive term does not affect the planner's ranking of treatment assignments. Under alternative social welfare functions an additional adjustment would be needed.

for heterogenous effects and a deprivation-impact trade-off—both spillover exercises actually identify a most impacted group that is somewhat *less* impacted than the most deprived group. The strongest evidence for predictable differences in spillovers is for within-village spillovers on the income of ineligibles, where we estimate meaningful differences in average impacts between the groups, though the large observed differences here between the predicted and actual effects for the impacted group is a cause for doubt (Table A.8, Panel C).

Taken together, there is no strong evidence that the data and approach are able to detect heterogeneous spillover effects; this gives us more confidence that the main results are primarily picking up heterogeneity in direct effects. That said, both tests are indirect ways of getting at the root question of heterogeneous within-village spillovers onto the treated. It would be valuable to explore this issue directly in future work by applying methods like those we use here to data from a multi-level experimental design, in which treatment probabilities vary at both the individual and the community level. Such a study would need to be large enough (in terms of sample size) to to generate sufficient variation in the characteristics of households targeted for transfers across areas in order to conduct meaningful inference regarding the existence of heterogeneous spillover effects.

5.6 Alternative statistical learning methods

We close by comparing the performance of the GRF learning model to alternatives. We focus on two benchmarks in particular: Ordinary Least Squares (OLS) regression, and LASSO regression. OLS has been widely used in practice to learn scoring rules for PMT targeting, but is not designed for prediction and thus does not incorporate regularization to guard against over-fitting. LASSO does provide regularization but (like OLS) cannot directly learn treatment effects, as GRF does. Instead both of these approaches generate predictions of $\hat{Y}_h(1)$ and $\hat{Y}_h(0)$ separately, which can be used to construct an indirect estimate of the treatment effect as $\hat{\Delta}_h = \hat{Y}_h(1) - \hat{Y}_h(0)$. This is potentially problematic for our application since any "noise" in the calculation of $\hat{Y}_h(0)$ and $\hat{y}_h(0)$ (due to sampling variation, measurement error, and so on) that does not also appear in $\hat{Y}_h(1)$ will tend to mechanically generate negative correlation between $\hat{y}_h(0)$ and $\hat{\Delta}_h$, biasing us towards concluding that the most deprived are also most impacted.

Broadly speaking, both OLS and LASSO perform similarly to GRF at *identifying* deprived and impacted groups, but both are also over-optimistic in predicting how well they do so. For OLS (see Tables C.1 and C.2), the actual deprivation of the most deprived and actual impact on the most impacted are similar to those we obtain via GRF, but the most deprived are predicted to be more deprived (Table C.1, Column 2) and the most impacted are predicted to be more impacted (Table C.2, Column 3) than they actually are. For consumption, for example, OLS predicts that the average effect for the most impacted group will 72% larger than the overall average while in fact it is only 28% higher. Results using LASSO (Tables C.4 and C.5) are somewhat less over-optimistic, but the gap is still substantial: OLS over-estimates consumption impacts on the most impacted by 34% while LASSO overestimates them by 26%; for assets the corresponding overestimates are 44% and 32%; and for income, 129% and 93%.[30]

OLS and LASSO also generally find less of a trade-off between deprivation and impact. Comparing results for GRF (Figure 1) with those for OLS and LASSO (Figure C.1), we see a noticeably more negative relationship in the latter figure for all outcomes. The correlation between $\hat{y}_h(0)$ and $\hat{\Delta}_h$ for consumption, for example, is slightly *negative* using either OLS or LASSO ($\rho = -0.07$) while positive when using GRF ($\rho = 0.36$). This likely at least in part reflects the bias introduced by noise in predictions of $\hat{Y}_h(0)$. The overconfidence of OLS and LASSO predictions is also visually evident here, as predicted treatment effects (on the vertical axis) span a much wider range than do those produced by GRF. For effects on consumption, for example, the 10th-90th percentile range produced by GRF is \$213, while for LASSO it is \$715 and for OLS an even larger \$781.[31]

Overall, this comparison highlights the potential value of using methods that are both regularized (unlike OLS) and explicitly designed to learn heterogeneous treatment effects (unlike both OLS and LASSO). The results obtained when doing otherwise raise two specific concerns. First, selecting beneficiaries based on over-optimistic predictions may simply lead policymakers to get the trade-off between deprivation and impact wrong. We see some evidence of this in Table C.3, for example, where we re-characterize optimal policies using OLS (as opposed to GRF) predictions. For consumption, for example, OLS selects a larger share of the most deprived and a smaller share of the most impacted. And second, conditional on the groups targeted, over-optimism about targeting performance implies over-optimism about the overall welfare gains from implementing a given targeted program. Mistakes like this will tend to distort resource allocation towards PMT-targeted programming at the expense of other approaches to targeting (or other uses of public funds entirely).

As robustness checks we also consider several perturbations to data preparation methods, holding fixed the GRF algorithm for learning. These address sensitivity to the discretionary

[30] We obtain similar results when using elastic net rather than LASSO (not reported). This further suggests that the issue is not the specific predictive model but rather the approach of learning counterfactual outcomes separately as opposed to learning treatment effects directly.

[31] The same pattern holds for the other financial outcomes: the corresponding figures for assets are \$72 for GRF, \$289 for LASSO, and \$316 for OLS, while those for income are \$117 for GRF, \$454 for LASSO, and \$551 for OLS.

choices that are needed even when using (largely) machine learning methods. We see that results are qualitatively similar if we use un-demeaned outcomes as our target rather than time-demeaned outcomes (Tables D.1 and D.2) and if we learn deprivation using data on control eligible households only (Tables D.4, D.5, and D.6).

6 Conclusion

We ask whether targeting an anti-poverty program to the most "deprived" households, as is typically the case in real-world programs, has the greatest social welfare benefit, in the setting of an NGO cash transfer program in rural Kenya. A noteworthy innovation of our approach is the application of recently developed machine learning (ML) methods—specifically, generalized random forests—to learn the household characteristics that target either deprivation levels or high conditional average treatment effects across several outcomes that are prominent in development policy debates. A central finding is that exclusively targeting the most deprived households is only attractive in a social welfare sense under very strongly redistributive preferences.

A corollary is that, for more plausible redistributive preferences, a meaningful share of the households that are social welfare maximizing to target are not those predicted to be most deprived. The results imply that policymakers should carefully consider whether automatically targeting anti-poverty assistance, like cash grants, to the poorest of the poor is necessarily appropriate in their own setting. This issue, and the results of this study, are more relevant than ever given the large rise in social assistance programming (often in the form of cash assistance) during the COVID-19 health crisis (Gentilini et al., 2020), and that in many cases appear likely to outlive the pandemic.

There are several important caveats. First, the results we present apply to large-scale cash grants, but patterns of impact, and the nature of the deprivation-impact trade-off, may plausibly differ for other types of assistance (e.g., subsidized credit or public health insurance). The rural Kenyan setting we study is also ethnically and religiously homogeneous and characterized by relatively limited inequality across households (within a village). In other settings with greater gaps in household living standards or salient social divisions, the benefits to targeting the poorest may be more pronounced. At the same time, in such settings, the gains from targeting those with the largest treatment effects may also be greater, and it is unclear which of these two effects outweighs the other.

Second, we measure endline outcomes (and thus treatment effects) starting immediately after transfer receipt and continuing up to 2 years after the start of transfer distributions. We see this as a strength relative to past work on targeting deprived households, which has often had to limit itself to using household characteristics to predict *contemporaneous* deprivation even while acknowledging that poverty is dynamic. But both targeting performance and the persistence of cash impacts might of course change over yet longer time horizons (Kondylis and Loeser, 2021). The longer-term effects of this particular cash transfer program are the subject of ongoing work (Egger et al., 2021).

Third, we caution that targeting assistance to those with the largest treatment effects may deepen existing inequalities. It appears that several marginalized subgroups in the population we study, e.g., widow-headed households or those with few or no prime-age adults, translate the cash grants into less substantial gains in future consumption, assets and income. It is possible that this finding might hold more generally: groups that are frequently marginalized or discriminated against (e.g., women, and ethnic or religious minorities, etc.) may not be able to leverage an assistance program as effectively as more favored groups that have other social advantages. The analytical approach we propose might, in this case, conclude that it is social welfare optimal to target assistance to precisely these favored groups, even though this decision to target assistance to those who would use it "effectively" will tend to reinforce existing social inequalities. Sustained assistance over a longer period of time might be needed to allow deprived and marginalized groups to take full advantage of the opportunities provided by an assistance program. This is beyond the scope of our study, given the one-time transfer and the static social welfare function that we employ, but could be a rationale for more aggressively targeting assistance to deprived groups, providing complementary forms of assistance, or extending cash assistance over longer time periods (as in an ongoing universal basic income study in the same region, Banerjee et al., 2020). The correct inference, other words, might be akin to the idea that "poorer households should be served by other interventions that credit" if they benefit less from credit (Morduch, 1999).

Despite these limitations, our hope is that the approach we propose can be used to reinvigorate real-world policy discussions around optimal targeting of social assistance. The use of richer data and sophisticated machine learning methods to target the households that are most likely to contribute to social welfare could potentially even help to build greater popular support for anti-poverty programs by convincing citizens that social benefits are being maximized (rather than targeting being driven by politicians' electoral considerations, say), although it may be a challenge to transparently and succinctly explain ML methods to citizens. Doing so might even make such programs more politically sustainable. In our view, it will be valuable to extend the approach in this study to other forms of assistance (beyond cash transfers), to other contexts, and to the use of alternative machine learning methods, and to ensure an active feedback loop with international development policymakers.

References

Abadie, Alberto, Matthew M. Chingos, and Martin R. West, "Endogenous Stratification in Randomized Experiments," *The Review of Economics and Statistics*, 10 2018, *100* (4), 567–580.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias, "Targeting the Poor: Evidence from a Field Experiment in Indonesia," *American Economic Review*, June 2012, *102* (4), 1206–40.

Alesina, Alberto and Dani Rodrik, "Distributive Politics and Economic Growth," *Quarterly Journal of Economics*, May 1994, *109* (2), 465–490.

Ambuehl, Sandro, B. Douglas Bernheim, and Axel Ockenfels, "What Motivates Paternalism? An Experimental Study," *American Economic Review*, March 2021, *111* (3), 787–830.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Andrews, Isaiah, Toru Kitagawa, and Adam McCloskey, "Inference on Winners," Technical Report, Harvard University December 2021.

Athey, Susan and Guido W. Imbens, "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 2019, *11* (1), 685–725.

_ **and Stefan Wager**, "Policy Learning With Observational Data," *Econometrica*, 2021, *89* (1), 133–161.

_ **, Julie Tibshirani, and Stefan Wager**, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148 – 1178.

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela, "How soon is now? Evidence of present bias from convex time budget experiments," *Experimental Economics*, 2020, *23* (2), 294–321.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman, "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, January 2015, *7* (1), 1–21.

— , **Michael Faye, Alan Krueger, Paul Niehaus, and Tavneet Suri**, "Effects of a Universal Basic Income during the pandemic," *Working Paper*, 2020.

Banerjee, Abhijit V., Paul J. Gertler, and Maitreesh Ghatak, "Empowerment and Efficiency: Tenancy Reform in West Bengal," *Journal of Political Economy*, April 2002, *110* (2), 239–280.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C Teitelbaum, "Estimating risk preferences in the field," *Journal of Economic Literature*, 2018, *56* (2), 501–64.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano, "Cash transfers: what does the evidence say? A rigorous review of programme impact and of the role of design and implementation features," 2016.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson, "Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi," *Journal of Public Economics*, 2020, *185* (104047).

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive Linear Step-up Procedures That Control the False Discovery Rate," *Biometrika*, 2006, pp. 491–507.

Bertrand, Marianne, Bruno Crépon, Alicia Marguerie, and Patrick Premand, "Do Workfare Programs Live Up to Their Promises? Experimental Evidence from Cote D'Ivoire," Working Paper 28664, National Bureau of Economic Research April 2021.

Bhattacharya, Debopam and Pascaline Dupas, "Inferring welfare maximizing treatment assignment under budget constraints," *Journal of Econometrics*, 2012, *167* (1), 168–196.

Björkegren, Daniel, Joshua Blumenstock, and Samsun Knight, "(Machine) Learning what Governments Value," *Working Paper*, 2021.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On, "Predicting poverty and wealth from mobile phone metadata," *Science*, 2015, *350* (6264), 1073–1076.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle, "A poor means test? Econometric targeting in Africa," *Journal of Development Economics*, 2018, *134*, 109–124.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val, "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India," Technical Report, National Bureau of Economic Research 2018.

Chowdhury, Reajul, Federico Ceballos-Sierra, and Munshi Sulaiman, "Grow the pie or have it? Using machine learning for impact heterogeneity in the Ultra-poor Graduation Model," Technical Report 170, Center for Effective Global Action 2021.

de Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 11 2008, *123* (4), 1329–1372.

Deaton, Angus and Christina Paxson, "Economies of Scale, Household Size, and the Demand for Food," *Journal of Political Economy*, 1998, *106* (5), 897–930.

Ding, Peng, Avi Feller, and Luke Miratrix, "Randomization inference for treatment effect variation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016, *78* (3), 655–671.

Egger, Dennis, Johannes Haushofer, Edward Miguel, and Michael Walker, "GE Effects of Cash Transfers: Pre-analysis plan for Endline 2 Household Welfare Analyses," 7 2021.

_ , _ , _ , Paul Niehaus, and Michael W Walker, "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya," Working Paper 26600, National Bureau of Economic Research December 2019.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

Gentilini, Ugo, Mohamed Almenfi, Ian Orton, and Pamela Dale, "Social protection and jobs responses to COVID-19," 2020.

Hanna, Rema and Benjamin A. Olken, "Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries," *Journal of Economic Perspectives*, November 2018, *32* (4), 201–26.

Haushofer, Johannes and Jeremy Shapiro, "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya*," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1973–2042.

Hussam, Reshmaan, Natalia Rigol, and Benjamin N Roth, "Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field," *Working Paper*, 2020.

Karlsson, Andreas, "Bootstrap methods for bias correction and confidence interval estimation for nonlinear quantile regression of longitudinal data," *Journal of Statistical Computation and Simulation*, 2009, *79* (10), 1205–1218.

Kidd, Stephen and Emily Wylde, "Targeting the Poorest: An assessment of the proxy means test methodology," Technical Report, AusAID 10 2011.

Kitagawa, Toru and Aleksey Tetenov, "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 2018, *86* (2), 591–616.

Kondylis, Florence and John Loeser, "Intervention Size and Persistence," 2021.

Lindbeck, Assar and Jörgen W. Weibull, "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, Jan 1987, *52* (3), 273–297.

Manacorda, Marco, Edward Miguel, and Andrea Vigorito, "Government Transfers and Political Support," *American Economic Journal: Applied Economics*, July 2011, *3* (3), 1–28.

Manski, Charles F., "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 2004, *72* (4), 1221–1246.

Maslow, A H, "A theory of human motivation," *Psychological Review*, 1943, *50* (4), 370–396.

McKenzie, David and Dario Sansone, "Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria," *Journal of Development Economics*, November 2019, *141*, 102369.

Meager, Rachael, "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature," 2020.

Morduch, Jonathan, "The Microfinance Promise," *Journal of Economic Literature*, December 1999, *37* (4), 1569–1614.

Nelson, Julie A., "Household Economies of Scale in Consumption: Theory and Evidence," *Econometrica*, 1988, *56* (6), 1301–1314.

Niehaus, Paul, Antonia Atanassova, Marianne Bertrand, and Sendhil Mullainathan, "Targeting with Agents," *American Economic Journal: Economic Policy*, February 2013, *5* (1), 206–38.

Persson, Torsten and Guido Tabellini, "Is Inequality Harmful for Growth?," *American Economic Review*, June 1994, *84* (3), 600–621.

Sen, Amartya., *Development as freedom*, Oxford: OUP, 1999.

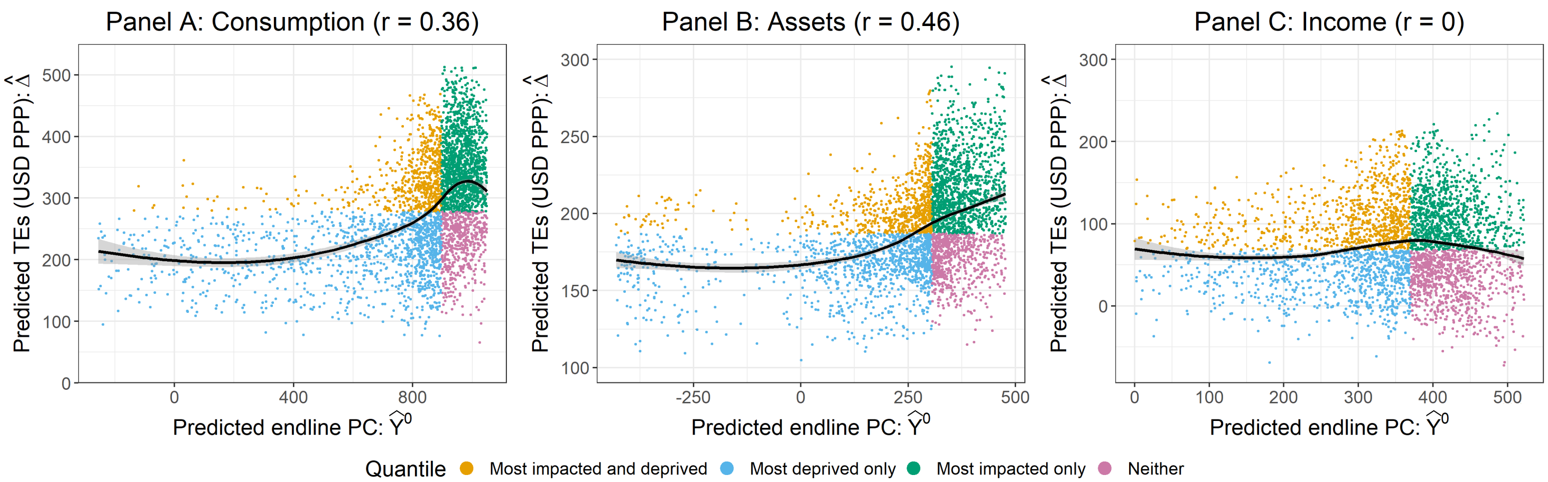
Subramanian, Shankar and Angus Deaton, "The Demand for Food and Calories," *Journal of Political Economy*, 1996, *104* (1), 133–162.

Tibshirani, Robert, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, 1996, *58* (1), 267–288.

Wager, Stefan and Susan Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

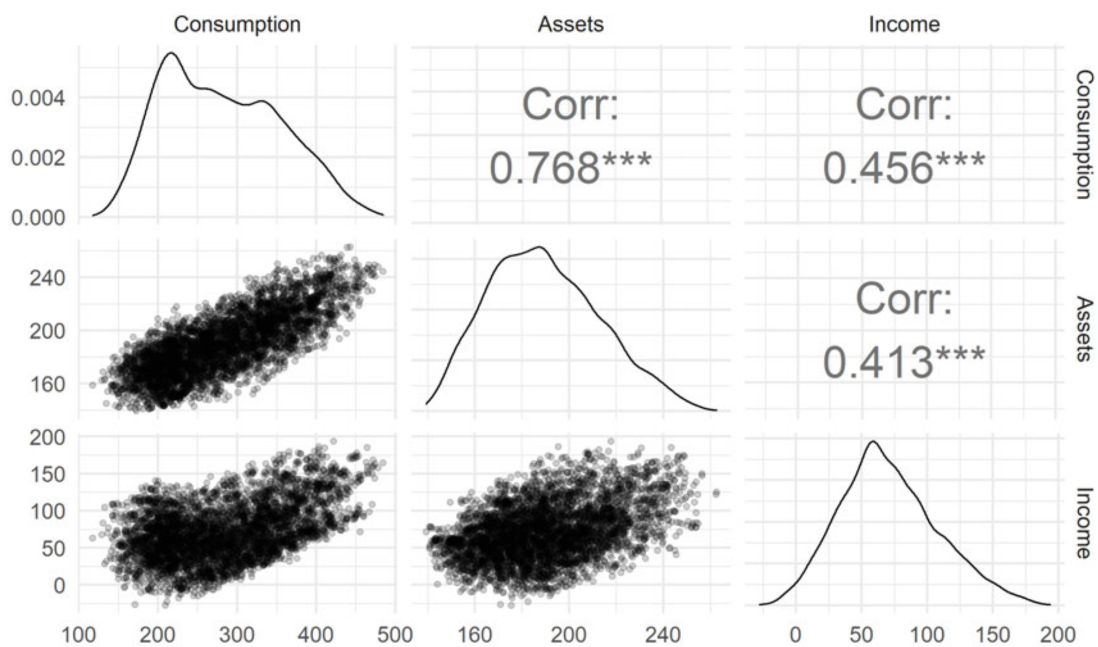
Walker, Michael, "Informal taxation and cash transfers: Experimental evidence from Kenya," 2018.

Figure 1: Predicted treatment effects ($\hat{\Delta}_h$) plotted against the predicted untreated per capita values ($\hat{y}_h^0$)



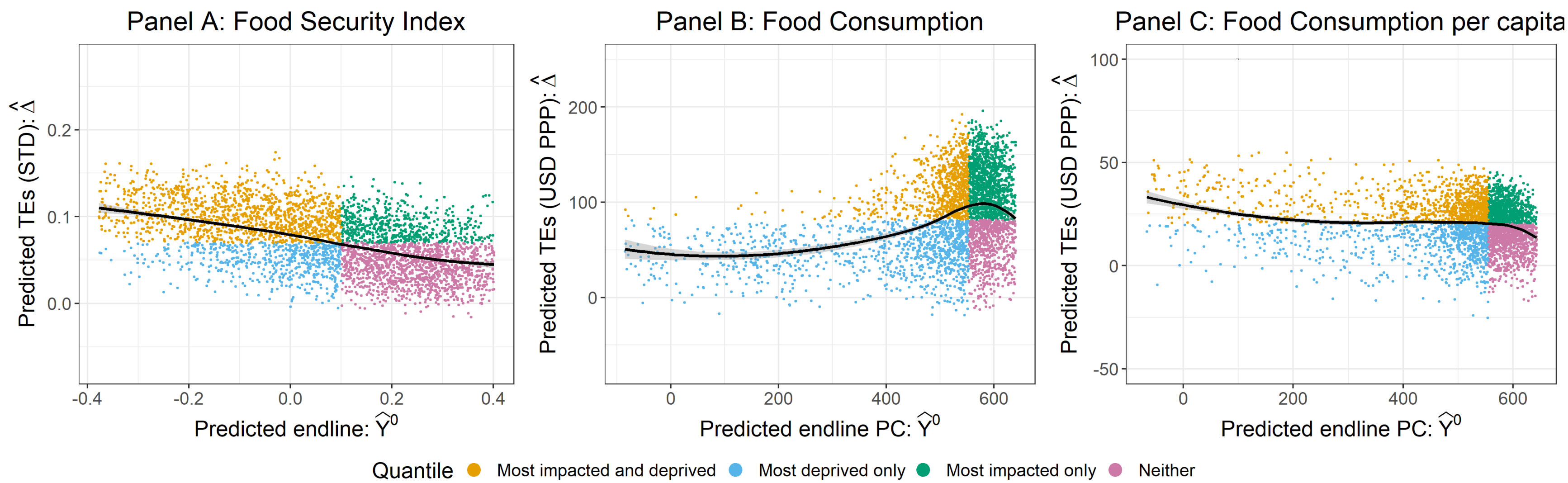
Notes: Each sub-figure plots predicted treatment effects for an outcome (y-axis) against the predicted endline values (x-axis) for that same outcome with a local regression line. As we generate 150 models per outcome, the figures presented are from the median model in terms of the difference in average treatment effects between the most deprived and most impacted groups for each outcome. Both predicted endline and predicted treatment effects are estimated from generalized random forest models with the same set of covariates. Predicted endline values and treatment effects are from models trained on time-demeaned data; a constant was added to the predicted endline outcomes so that the overall predicted mean matches the observed sample mean. The correlation (r) between predicted endline values and treatment effects for the median model is reported in the subfigure title.

Figure 2: Cross-outcome relationships in predicted treatment effects



Notes: This figure looks at correlations in predicted treatment effects across different outcomes for the main models presented in Figure 1 and Table 2. For each household we use the average prediction across the 150 models trained. The lower triangular section displays scatterplots of predicted treatment effects across outcomes. The upper triangular section displays the Pearson correlations. The diagonal displays the distribution of treatment effects for each outcome.

Figure 3: Predicted treatment effects ($\hat{\Delta}_h$) and untreated per capita values ($\hat{y}_h^0$) for food security



Notes: This figure demonstrates that the pre-specified food security index appears to behave more similarly to a per-capita measure than a household measure. Each sub-figure plots predicted treatment effects for an outcome (y-axis) against the predicted endline values (x-axis) for that same outcome with a local regression line. Both predicted endline and predicted treatment effects are estimated from generalized random forest models with the same set of covariates. Predicted endline values are from models trained on time-demeaned data; a constant was added to the reported statistics so that the overall predicted mean matches the observed sample mean.

Table 1: Predicted per capita untreated outcomes (y_h^0) by group

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	750	532	918	-386
Actual	729	512	906	-394
				(-533,-295)
				[-457,-332]
Panel B: Assets				
Predicted	228	76	331	-255
Actual	213	56	328	-272
				(-402,-222)
				[-301,-242]
Panel C: Income				
Predicted	308	180	319	-139
Actual	297	171	316	-145
				(-208,-40)
				[-179,-111]

Note: This table presents the group averages of actual and predicted per capita endline values among transfer-eligible households in treatment and control villages. Predicted values are based on generalized random forest models that i) were trained on time-demeaned data (a constant was added to the reported statistics so that the All statistic matches the observed sample mean) and ii) were trained to produce *predicted* endline values in the absence of treatment. While models are trained using data from both treatment and control households, for comparability statistics reported here restrict attention to control households for both actual and predicted values. Estimates reported in the table are the mean value of that statistic across the 150 models trained (see Appendix E for details). Column (1) reports overall averages of predicted and actual values. Columns (2) and (3) classify households on the basis of predicted deprivation and impact using our model; the most deprived group is the 50% of households with the lowest predicted per capita endline values, and the most impacted are the 50% of households with the highest predicted treatment effects. Column (4) reports the difference between the most deprived and the most impacted. We report the 95% CI for the actual difference statistic computed through empirical bootstrap in parentheses, and using the median standard error for the actual statistic, clustered at the village level, in brackets (Chernozhukov et al., 2018). All analyses are weighted by inverse sampling probabilities to be representative of the population of eligible households. $N = 2,367$.

Table 2: Predicted Average Treatment Effects (Δ_i) by group

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)	(5) RI p-value $I - I^C > 0$
Panel A: Consumption					
Predicted	281	247	346	-99	
Actual	310	241	402	-161	0.03
				(-300,4)	
				[-245,-78]	
Panel B: Assets					
Predicted	190	178	211	-33	
Actual	182	149	182	-33	0.58
				(-68,92)	
				[-80,14]	
Panel C: Income					
Predicted	71	68	104	-36	
Actual	85	78	92	-14	0.39
				(-51,186)	
				[-76,47]	

Note: This table reports treatment effects for transfer-eligible households. The *Actual* row denotes the average treatment effect of the group while *predicted* denotes the average of household-level predicted treatment effects from the generalized random forests (GRF) model. *Actual* averages are estimated using OLS and a group (deprived, impacted) indicator. Estimates reported in the table are the mean value of that statistic across the 150 models trained (see Appendix E for details). Column (1) reports overall treatment effects in this sample of eligible households. Columns (2) and (3) classify households on the basis of predicted deprivation and impact using our model; the most deprived group is the 50% of households with the lowest predicted per capita endline values, and the most impacted are the 50% of households with the highest predicted treatment effects. Column (4) reports the difference between the most deprived and the most impacted, with negative values representing a cost of targeting the most deprived relative to the most impacted. We report the 95% CI for the actual difference statistic computed through empirical bootstrap in parentheses, and using the median standard error for the actual statistic, clustered at the village level, in brackets (Chernozhukov et al., 2018). Column (5) reports randomization inference p-values for a test of heterogeneity under the null of homogeneous treatment effects, where each treated household has an individual treatment effect equal to $x \in [ATE - 3\sigma^2, ATE + 3\sigma^2]$, where ATE is the observed average treatment effect of the sample. The reported p-value is the maximum from searching over this grid of possible values of x. Note that each value of x defines a null. I^C denotes the complement of the most impacted. All analyses are weighted by inverse sampling probabilities to be representative of the population of eligible households. $N = 4,749$.

Table 3: Variable importance for predicting untreated outcomes and treatment effects

Variable	Mean	Predicted untreated outcomes (y_h^0)			Predicted treatment effects (Δ_h)		
		Consumption	Assets	Income	Consumption	Assets	Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Household demographics							
HH size	4.38	**0.68 (1,+)**	**0.71 (1,+)**	**0.42 (1,+)**	**0.24 (1,+)**	**0.23 (1,+)**	**0.22 (1,+)**
Female head	0.69	0.01 (11,-)	0.00 (14,-)	0.03 (5,-)	0.07 (5,+)	0.06 (7,+)	0.07 (6,-)
Has children	0.81	0.06 (4,+)	**0.05 (3,+)**	0.05 (4,+)	0.02 (15,-)	0.01 (16,+)	0.02 (15,-)
Has children in school	0.66	0.02 (5,+)	0.01 (7,+)	0.01 (9,+)	0.04 (10,+)	0.04 (11,+)	0.04 (10,+)
Has child under 3	0.50	0.00 (17,+)	0.00 (17,-)	0.00 (16,+)	0.06 (7,+)	0.06 (6,+)	0.07 (7,+)
Has child under 6	0.64	0.01 (13,+)	0.01 (10,+)	0.00 (13,+)	0.04 (12,+)	0.05 (10,+)	0.04 (13,-)
Widow	0.21	**0.06 (3,-)**	0.03 (4,-)	**0.19 (3,-)**	0.03 (14,+)	0.03 (12,-)	0.02 (14,+)
Has elder member	0.11	**0.09 (2,-)**	0.03 (5,-)	**0.22 (2,-)**	0.01 (16,+)	0.01 (15,+)	0.00 (16,-)
Treatment	0.50	0.01 (8,+)	0.01 (8,+)	0.00 (12,+)			
Panel B: Financial characteristics							
Employed	0.34	0.00 (14,-)	0.01 (12,-)	0.01 (11,+)	0.05 (9,+)	0.05 (8,+)	**0.09 (3,+)**
Self-employed	0.27	0.01 (10,+)	0.01 (11,+)	0.03 (6,+)	0.06 (6,-)	0.07 (5,+)	0.08 (4,-)
Has any livestock	0.26	0.00 (15,+)	**0.10 (2,+)**	0.00 (14,+)	**0.09 (3,+)**	**0.10 (2,+)**	0.06 (8,+)
Owns land	0.84	0.01 (12,-)	0.00 (15,-)	0.00 (15,-)	0.04 (13,+)	0.03 (14,+)	0.04 (11,+)
Owns 1/4 acre	0.82	0.00 (16,-)	0.00 (16,-)	0.00 (17,+)	0.04 (11,+)	0.03 (13,+)	0.04 (12,+)
Owns TV or radio	0.62	0.02 (6,+)	0.02 (6,+)	0.01 (10,+)	0.06 (8,-)	0.05 (9,+)	0.06 (9,-)
Meals yesterday	2.29	0.01 (7,+)	0.01 (9,+)	0.01 (8,+)	**0.10 (2,-)**	**0.10 (3,-)**	**0.09 (2,-)**
Meals with protein yesterday	0.43	0.01 (9,+)	0.01 (13,+)	0.02 (7,+)	0.07 (4,-)	0.07 (4,+)	0.07 (5,+)

Notes: Column (1) reports the unconditional mean of each variable at the baseline. Columns (2)-(5) report variable importance for endline predictions, and columns (6)-(9) report importance for predicted treatment effects. Variable importance is measured as the a depth-weighted average of the share of splits created in the process of growing trees that split on a particular variable (see Equation (13)). The first argument in parentheses is the variable importance ranking; the second argument is whether the predicted outcome increases (+) or decreases (−) when the variable is 1 versus 0 for indicators or a one standard deviation increase from the mean for continuous variables, fixing all other covariates to their mean. For each outcome, the top three variables by importance are in bold. $N = 4,749$.

Table 4: Overlap of socially optimal households to target with most deprived and most impacted

	(1)	(2)	(3)	(4)	(5)
CARA: α	CE	Most deprived	Most impacted	Choice	α_c
Panel A: Consumption					
0.0000	\$50.00	0.30	1.00	I	
0.0005	\$49.38	0.31	0.96	I	
0.0010	\$48.75	0.33	0.92	I	
0.0075	\$40.84	0.40	0.81	I	
0.0150	\$32.78	0.42	0.79	I	0.016
Panel B: Assets					
0.0000	\$50.00	0.31	1.00	I	
0.0005	\$49.38	0.35	0.91	I	
0.0010	\$48.75	0.39	0.83	I	0.007
0.0075	\$40.84	0.55	0.59	D	
0.0150	\$32.78	0.57	0.55	D	
Panel C: Income					
0.0000	\$50.00	0.47	1.00	I	
0.0005	\$49.38	0.48	0.97	I	
0.0010	\$48.75	0.50	0.93	I	
0.0075	\$40.84	0.59	0.73	I	0.011
0.0150	\$32.78	0.63	0.66	D	

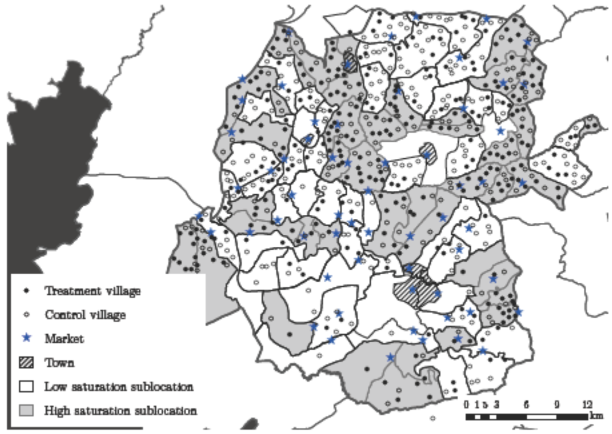
Notes: Column 1 denotes the certainty equivalent (CE) of a 50-50 lottery over \$0 or \$100 under the specified CARA α parameter value. Column 2 (3) reports the share of households belonging to I (D) that are also "socially optimal" for a planner to treat. Socially optimal households are those in the top 50% of households ranked by potential gains from treatment using a CARA utility function for the risk aversion parameter (α) given in the row label. Reported shares are the mean of 150 5-fold GRF iterations; median ratios are similar (not shown). Column 4 reports the welfare maximizing choice between targeting the most impacted (I) and the most deprived (D) for a given α value. Column (5) reports the critical value α_c, the mean minimum value of α required to rationalize a policy targeting the most deprived instead of targeting the most impacted across the 150 estimated models. Formally, $\alpha_c = \min(\{\alpha : SW(D;\alpha) \geq SW(I;\alpha)\})$.

A Additional exhibits

Figure A.1: Map of study area

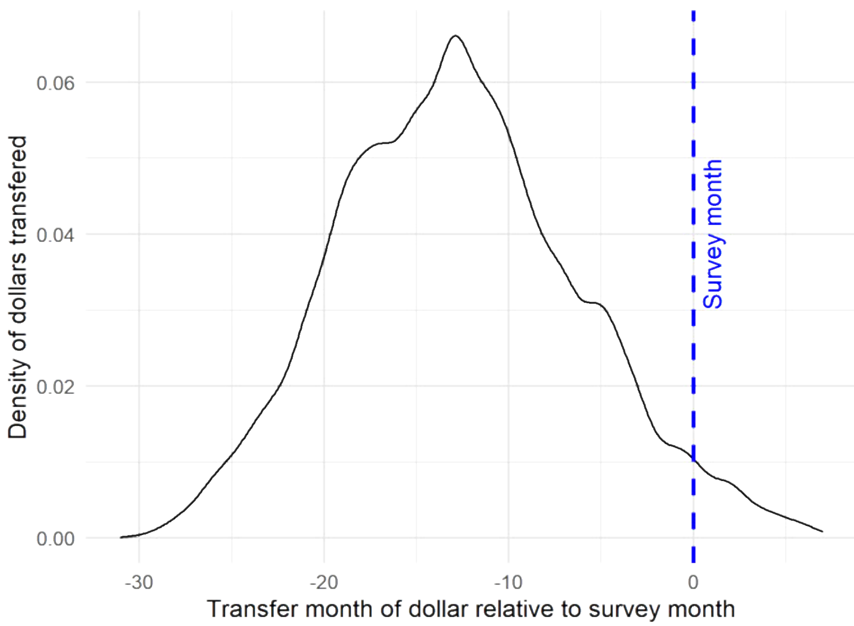
Panel A: Location of the study area in Kenya

Panel B: Mapping study villages



Notes: In Panel A, Siaya County, the county in which the study takes place, is represented by the shaded area. Panel B plots the location of study villages and sublocation treatment saturation status (from Egger et al. (2019).

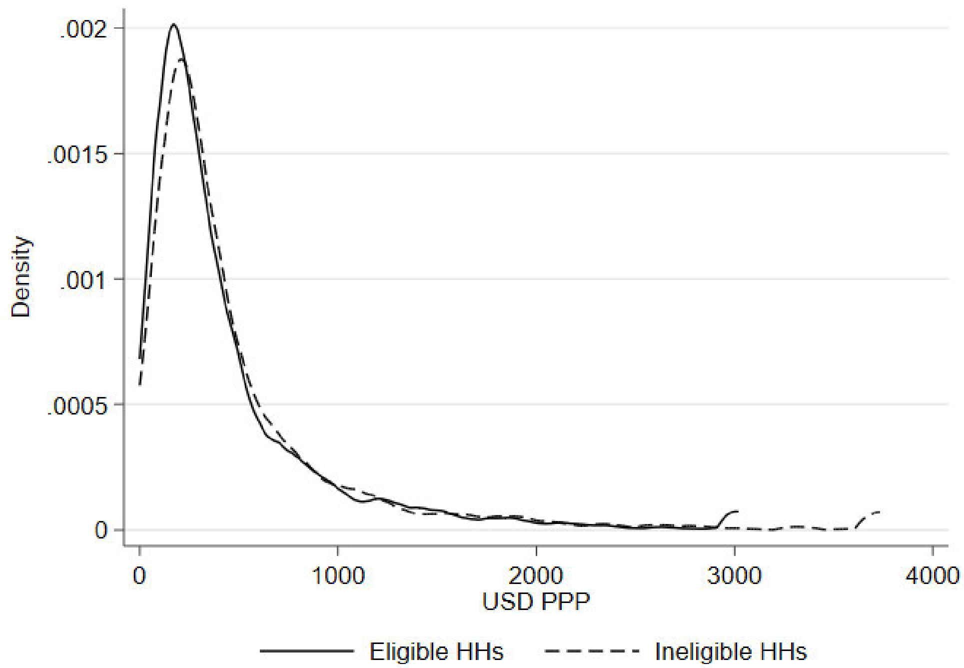
Figure A.2: Transfer date relative to endline survey time



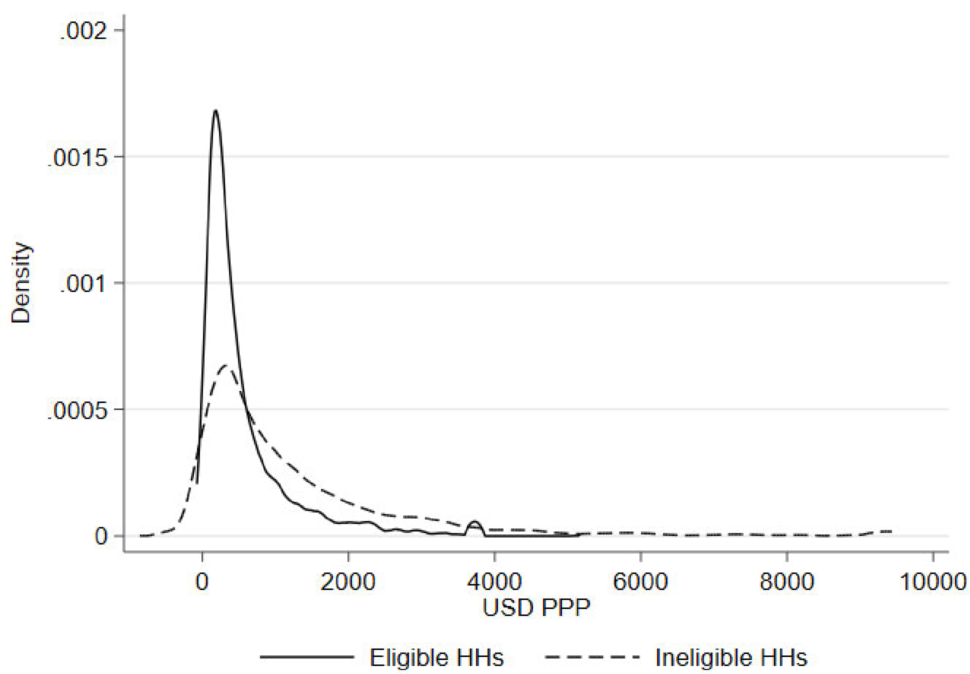
Notes: This figure plots the density of dollars transferred relative to the month when a household completed the endline survey (conducted by the research team) for treated households matched to treatment timing data from the NGO *GiveDirectly*.

Figure A.3: Kernel densities of baseline distributions

Panel A: Income

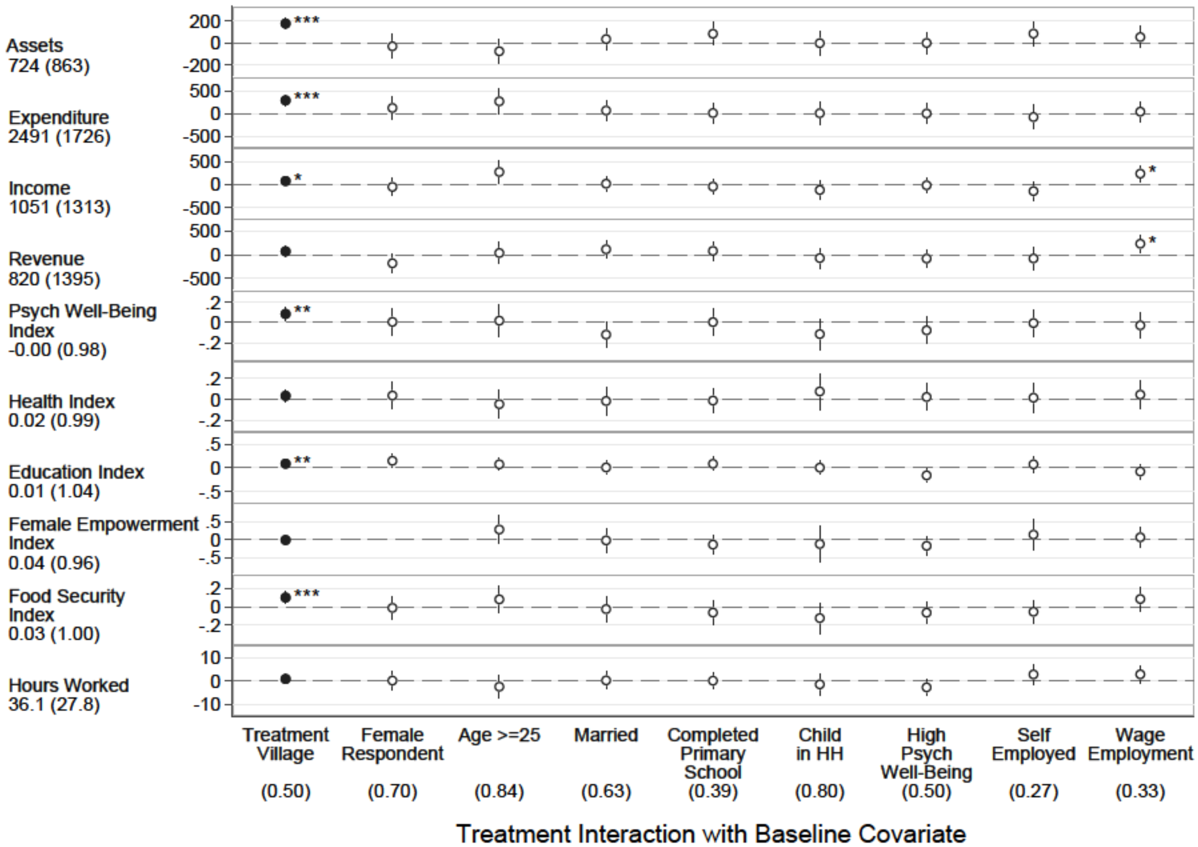


Panel B: Net assets



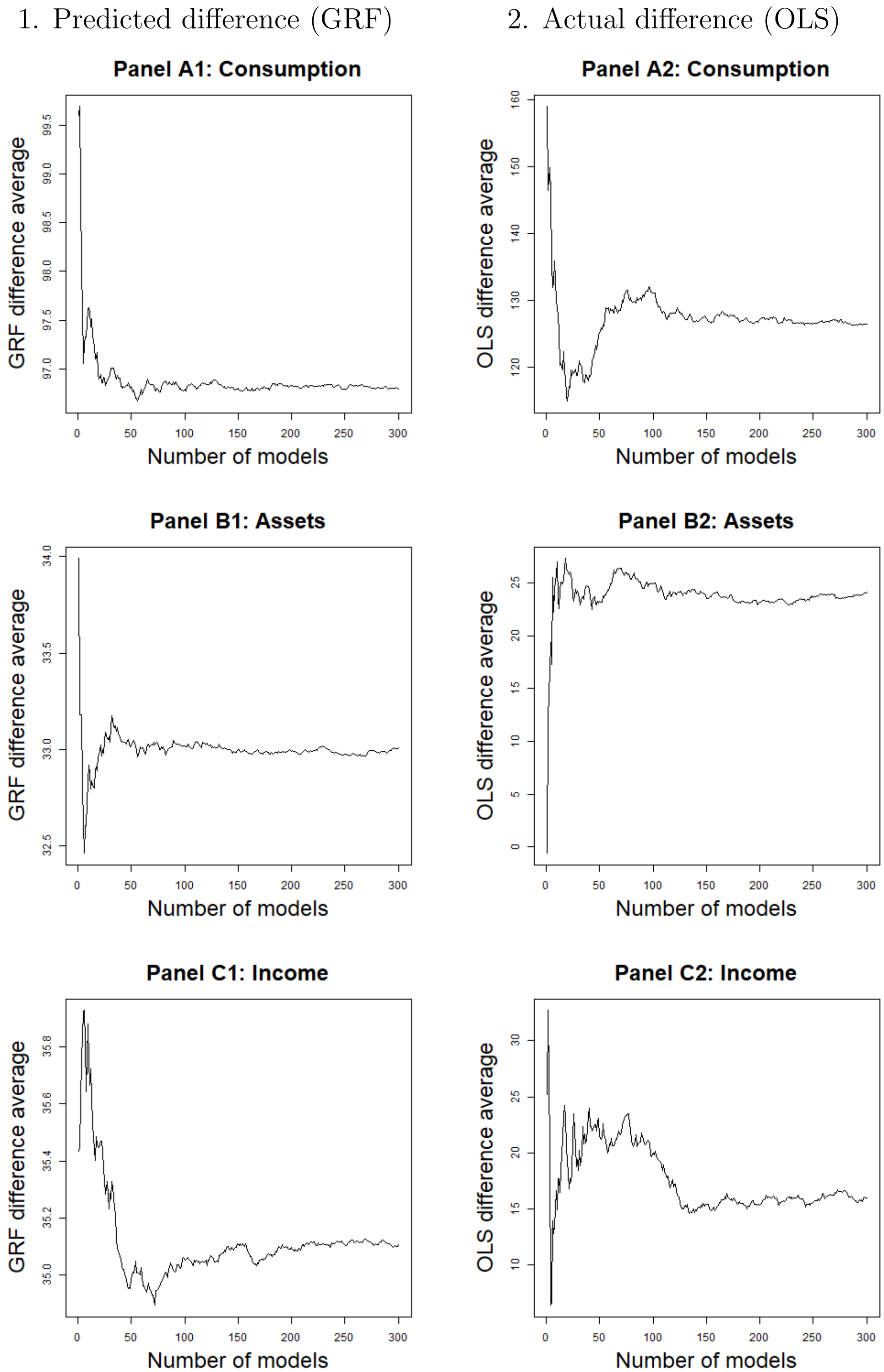
Notes: These figures plot the kernel densities (using an Epanechnikov kernel) of baseline income (Panel A) and net asset ownership (excluding land and housing) for transfer-eligible and ineligible households. There is noticeable overlap in the distributions for eligible and ineligible households for both of these outcomes.

Figure A.4: Previous heterogeneity results (EHMNW)



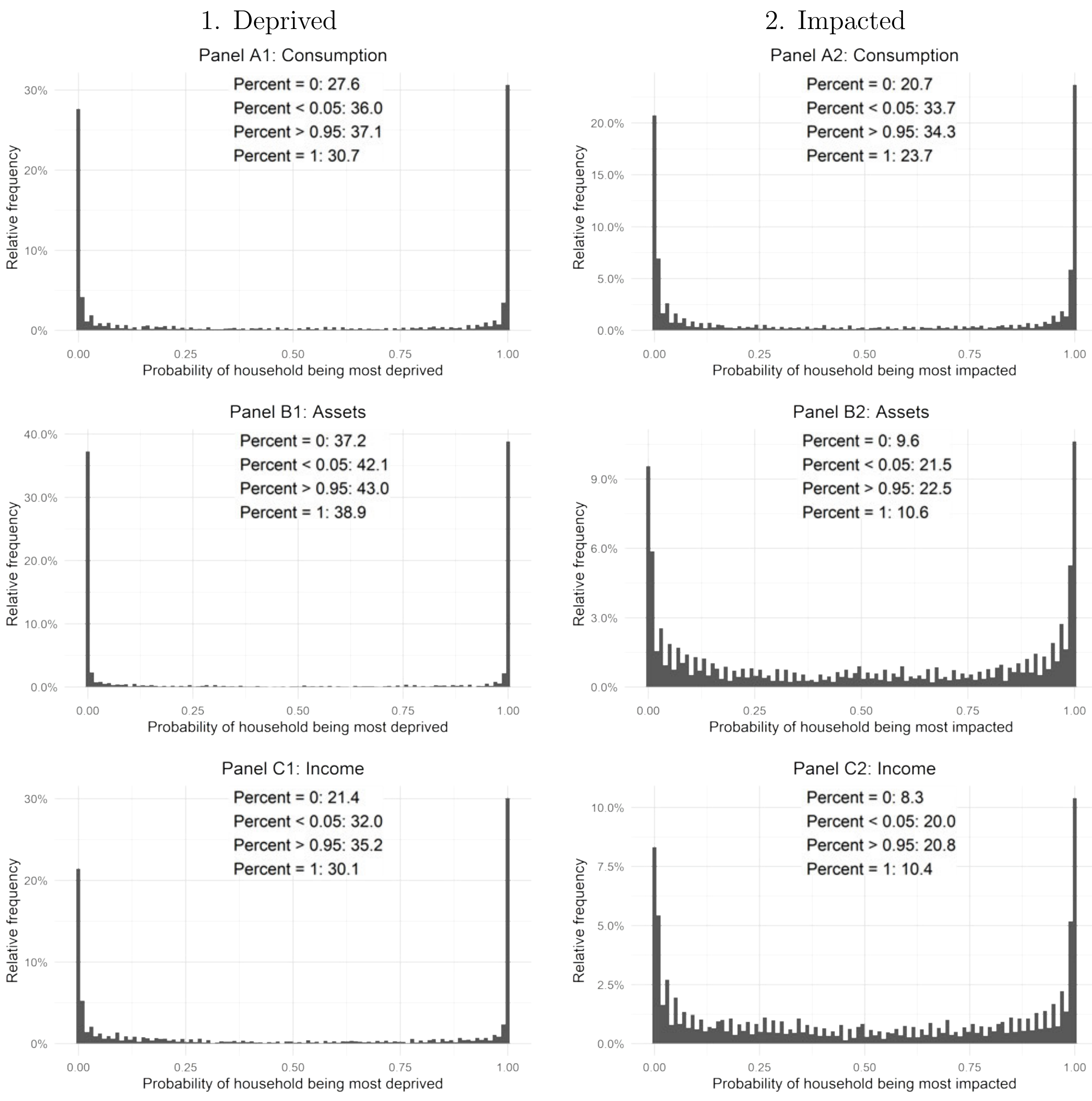
Notes: This figure, reproduced from Egger et al. (2019), presents estimates for treatment effect heterogeneity for eligible households in pre-specified primary outcomes along 8 pre-specified dimensions of heterogeneity. Each plotted coefficient is from a separate regression. Each row represents a separate primary outcome; the mean (SD) for eligible households in control, low saturation villages is reported below the outcome label. The first column (Treatment Village) plots estimated effects for the coefficient on an indicator for being in a treatment village, where the sample is restricted to eligible households and controls include an indicator for sublocation saturation status and the baseline value of the outcome variable (if available). Columns 2 through 8 plot the coefficient on the interaction term of the listed baseline covariate with the treatment village indicator; this interaction term and baseline covariate are added to the regression equation. Values in parentheses on the x-axis denote the mean of the baseline covariate. Standard errors are clustered at the village level. Reported significance levels correspond to FDR q-values, calculated following Benjamini et al. (2006). * denotes significance at 10 pct., ** denotes significance at 5 pct., and *** denotes significance at 1 pct. level.

Figure A.5: Model convergence



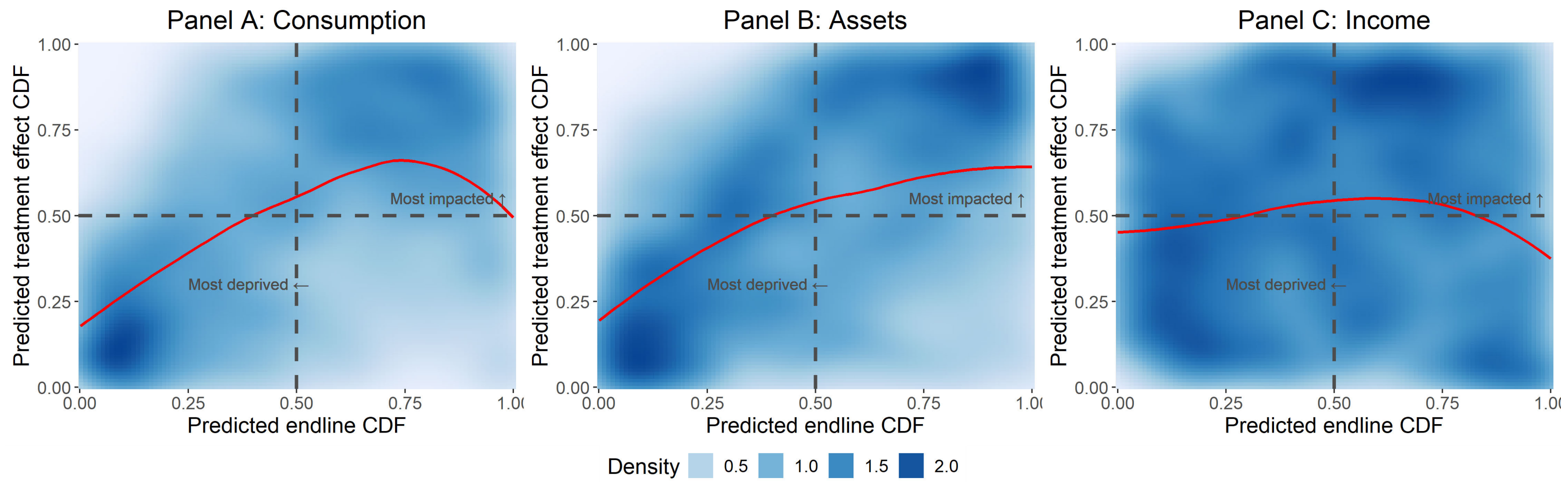
Notes: This figure presents the convergence of the difference of the predicted and actual average treatment effect between groups I, D as a function of the number of models being trained. Note that this statistic remains roughly constant between 150 and 300 models.

Figure A.6: Stability of most deprived and impacted classification



Notes: This figure presents the relative frequency of the probability that a household is classified as either most deprived (column 1) or most impacted (column 2) across 150 models. The large mass at/around 0 and 1 indicate that most households are being consistently classified to a group across models.

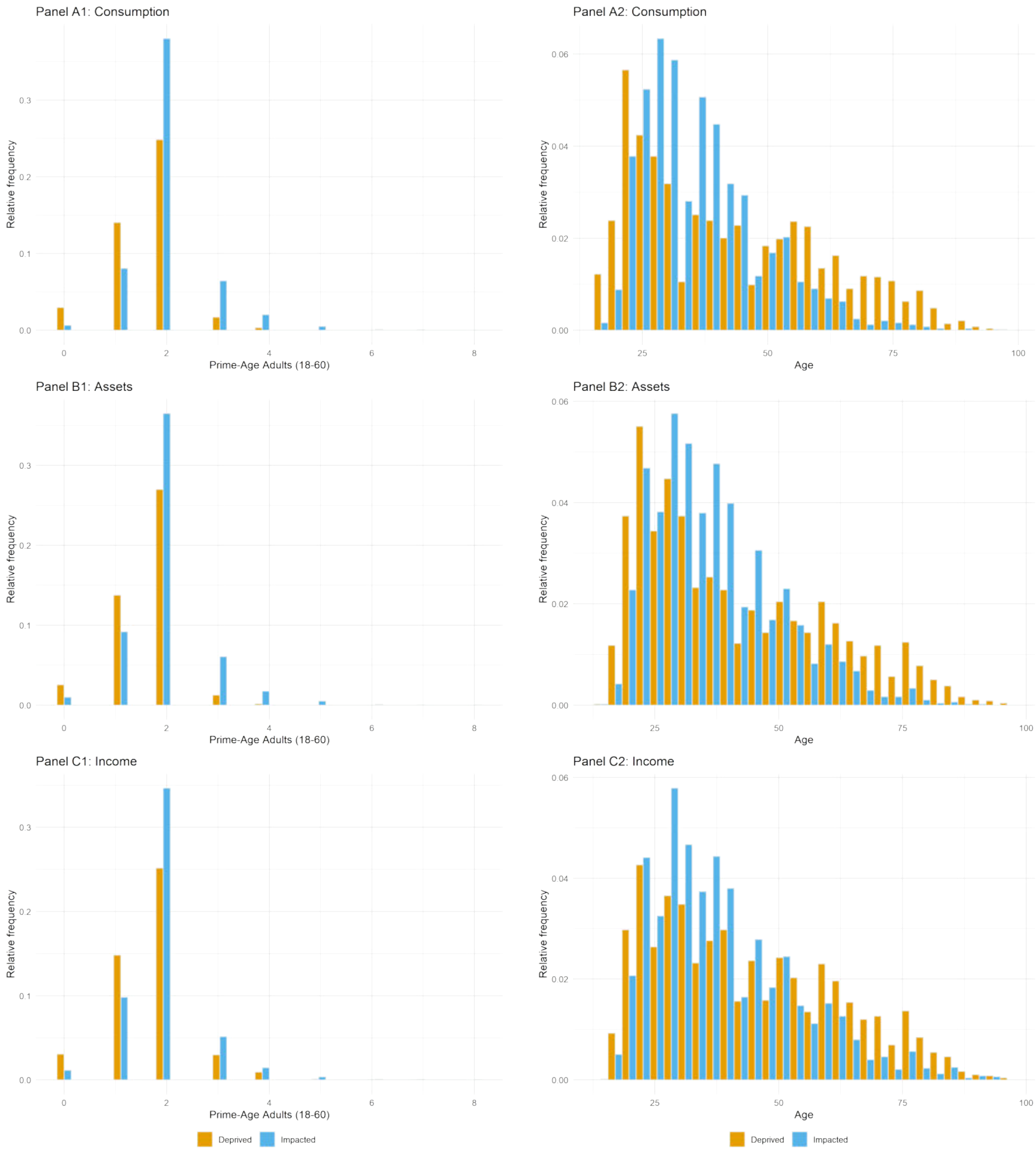
Figure A.7: CDF of predicted treatment effects against the CDF of predicted endline values for each outcome



Notes: This figure complements Figure 1 and Figure 3 (food security), but showing CDFs rather than levels. Each sub-figure shows the relationship between the CDF of predicted treatment effects against the CDF of the predicted endline values for each outcome with a local regression line and the two-dimensional density of observations.

Figure A.8: Demographic characteristics in the most impacted and deprived households

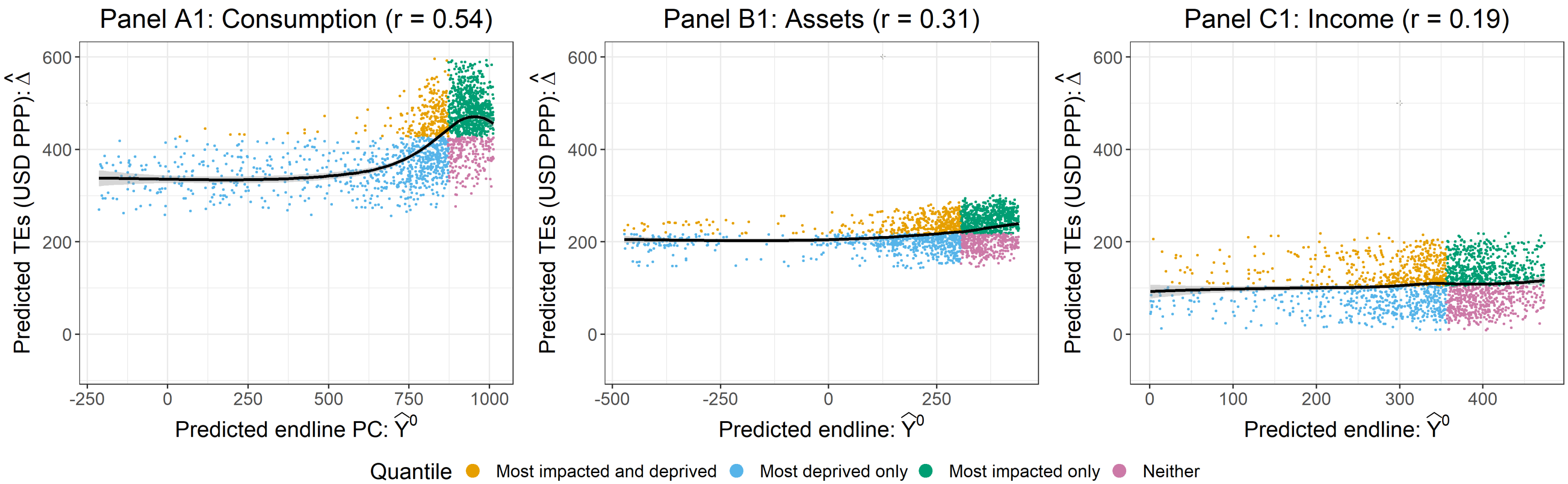
1. Number of prime-age adults (18-60) 2. Age of household head



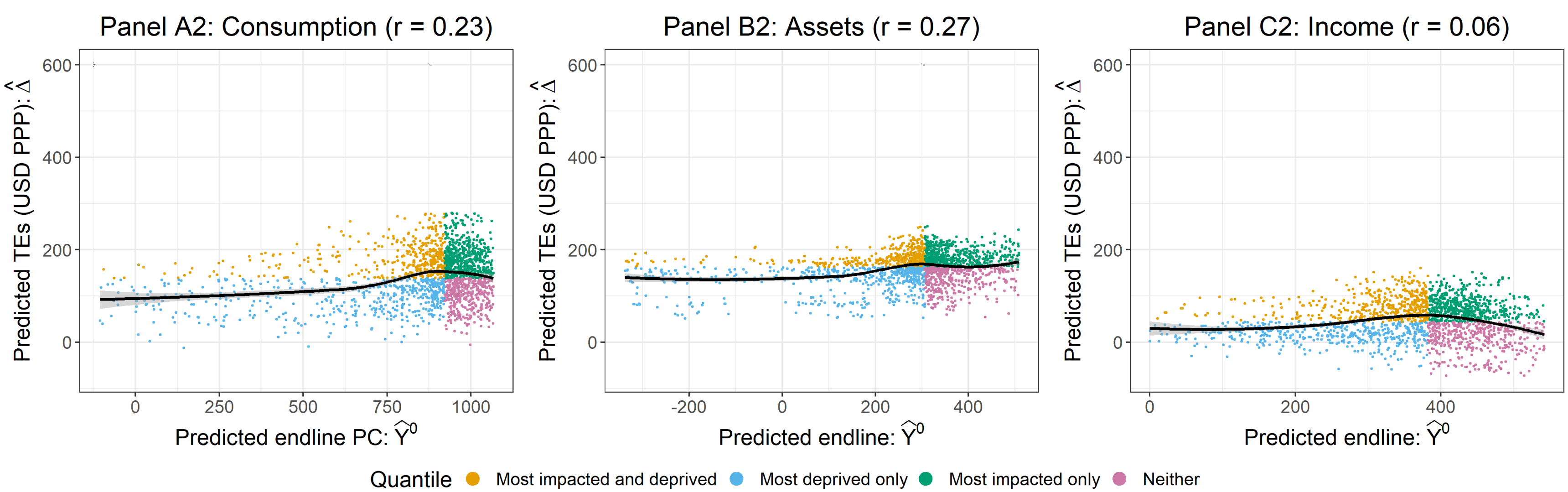
Notes: This figure plots the baseline distribution of two key demographic characteristics – the number of prime-age adults in a household (Column 1), and the age of the household head (Column 2), by classification into the most deprived and most impacted groups from the median model in terms of the difference in average treatment effects between the most deprived and most impacted groups for each outcome (as used in Figure 1).

Figure A.9: Predicted treatment effects ($\hat{\Delta}_h$) plotted against the predicted untreated per capita values ($\hat{y}_h^0$) by treatment timing

1. Recently treated households

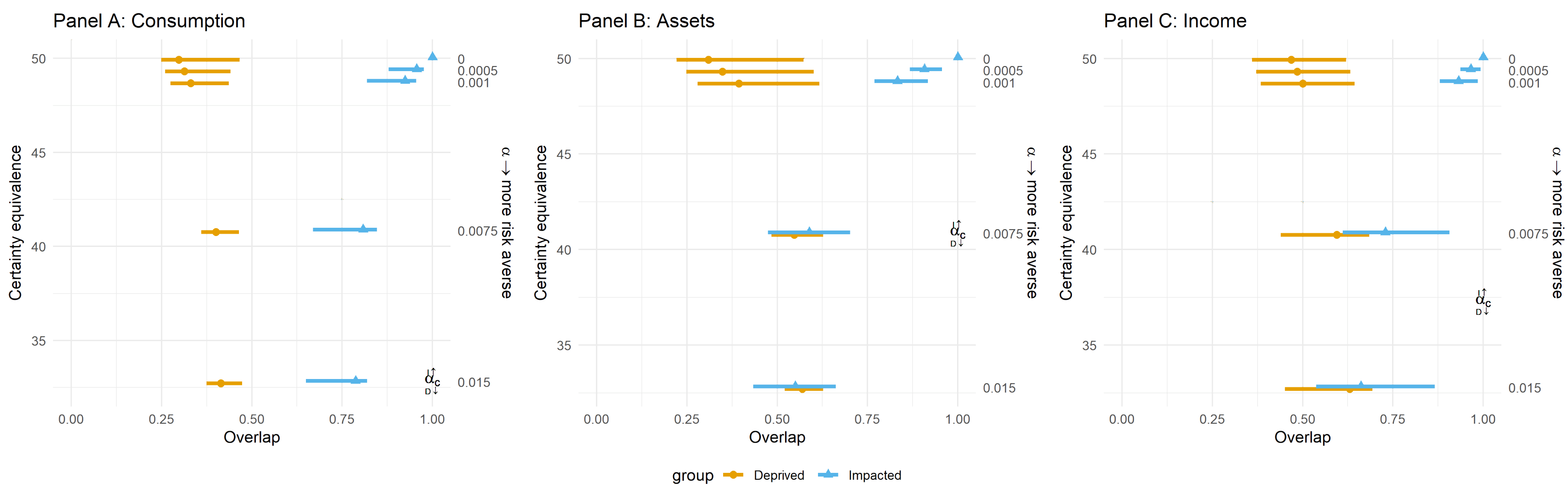


2. Non-recently treated households



Notes: This figure reproduces Figure 1 using models estimated separately for recently treated households (Row 1) and non-recently treated households (Row 2). For each household we computed the number of months between the survey and the experimental start date of their village. Households under the median of this timing variable are defined as recently treated, otherwise they are defined as non-recently treated.

Figure A.10: Overlap of socially optimal households to target with most deprived and most impacted



Notes: The left y-axis denotes the certainty equivalent (CE) of a 50-50 lottery over \$0 or \$100 under the specified CARA α parameter value given on the right y-axis. The figure plots point estimates and bootstrapped 95% confidence intervals of the share of I (D) that are also "socially optimal" for a planner to treat. Socially optimal households are those in the top 50% of households ranked by potential gains from treatment using a CARA utility function. α_c displays the minimum value of α required to rationalize a policy targeting the most deprived instead of targeting the most impacted. Formally, $\alpha_c = \min(\{\alpha : SW(D;\alpha) \geq SW(I;\alpha)\})$.

Table A.1: Household balance

	Eligibles			Ineligibles		
	(1) Control, Low Sat Mean (SD)	(2) Treatment Effect	(3) N	(4) Control, Low Sat Mean (SD)	(5) Treatment Effect	(6) N
Female	0.67	0.01	4,768	0.79	-0.02	2,458
	(0.47)	(0.02)		(0.41)	(0.02)	
Respondent aged 25 or older	0.83	0.01	4,755	0.97	-0.01	2,448
	(0.38)	(0.01)		(0.18)	(0.01)	
Is married	0.64	0.02	4,768	0.42	0.03	2,458
	(0.48)	(0.01)		(0.49)	(0.02)	
Completed primary school	0.41	0.01	4,768	0.29	0.02	2,458
	(0.49)	(0.02)		(0.45)	(0.02)	
Has child	0.80	0.01	4,768	0.68	0.04*	2,458
	(0.40)	(0.01)		(0.47)	(0.02)	
Self-employed	0.27	-0.00	4,768	0.28	0.01	2,458
	(0.45)	(0.01)		(0.45)	(0.02)	
Employed in wage work	0.36	-0.03	4,768	0.21	-0.03*	2,458
	(0.48)	(0.02)		(0.41)	(0.02)	
Total non-land, non-home assets, net loans (z-scored) (USD)	-0.18	0.01	4,768	0.41	0.08	2,458
	(0.65)	(0.02)		(1.36)	(0.06)	
Total household income in the last 12 months (z-scored) (USD)	0.01	0.01	4,768	0.07	0.02	2,458
	(0.89)	(0.03)		(1.02)	(0.05)	
Total business revenue in the last 12 months (z-scored) (USD)	-0.08	0.03	4,768	0.13	0.05	2,458
	(0.75)	(0.03)		(1.20)	(0.06)	
Psychological wellbeing index	-0.00	0.06*	4,765	-0.00	0.08*	2,458
	(0.99)	(0.04)		(0.99)	(0.05)	
Food security index	-0.01	0.00	4,768	-0.01	0.09**	2,458
	(1.00)	(0.03)		(1.01)	(0.04)	

Notes: Differences in baseline outcomes between households in treatment and control villages. Column (2) shows the baseline difference between eligible households in treated and untreated villages, and Column (5) shows the same difference for ineligible households. Columns (1) and (4) show the respective control means and standard deviations. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.2: Tracking and attrition for eligible HHs

	(1) Surveyed at endline	(2) Surveyed at baseline	(3) Surveyed at baseline & endline	(4) Initially sampled household	(5) Replacement household
Panel A: All households targeted at endline					
Treatment Village	0.006	0.010	0.006	0.005	-0.008
	(0.009)	(0.011)	(0.013)	(0.010)	(0.013)
High Saturation Sublocation	-0.003	-0.020	-0.019	0.006	-0.003
	(0.009)	(0.014)	(0.015)	(0.011)	(0.013)
Control, Low Sat Mean (SD)	0.903	0.879	0.814	0.853	0.168
	(0.297)	(0.326)	(0.389)	(0.355)	(0.374)
Observations	6,039	6,039	6,039	6,039	5,196
Panel B: Among households surveyed at endline					
Treatment Village		0.001	0.001	0.013	-0.015
		(0.011)	(0.011)	(0.010)	(0.013)
High Saturation Sublocation		-0.018	-0.018	0.005	-0.002
		(0.013)	(0.013)	(0.011)	(0.013)
Control, Low Sat Mean (SD)		0.902	0.902	0.845	0.172
		(0.298)	(0.298)	(0.362)	(0.377)
Observations		5,425	5,425	5,425	4,768
Panel C: Among households surveyed at baseline					
Treatment Village	-0.004		-0.004	0.008	-0.008
	(0.009)		(0.009)	(0.013)	(0.013)
High Saturation Sublocation	-0.000		-0.000	0.003	-0.003
	(0.009)		(0.009)	(0.013)	(0.013)
Control, Low Sat Mean (SD)	0.926		0.926	0.832	0.168
	(0.262)		(0.262)	(0.374)	(0.374)
Observations	5,197		5,197	5,197	5,196

Notes: This table reports differences in tracking and attrition by treatment status for baseline and endline household surveys. The main analysis sample used in this paper comprises households surveyed at both baseline and endline (Column 3). Initially sampled households are the initial eight eligible households targeted for surveys in a village. If these households were not available on the date of baseline household surveys, a replacement household was sought. Endline surveys targeted all initially-sampled households (regardless of baseline survey status) and replacement households.

Table A.3: Differences in characteristics and treatment effects by deprivation and impact group classification using consumption

	(1) All	(2) D	(3) $!D$	(4) $D-!D$	(5) I	(6) $!I$	(7) $I-!I$	(8) $D-I$
Panel A: Baseline demographics, Mean (SD)								
Household size	4.40	3.04	5.75	-2.71***	6.02	2.77	3.24***	-4.07***
	(2.21)	(1.60)	(1.86)	(0.06)	(1.66)	(1.30)	(0.05)	(0.05)
Number of prime-age adults in household	1.88	1.59	2.11	-0.52***	2.07	1.65	0.42***	-0.68***
	(0.77)	(0.70)	(0.74)	(0.02)	(0.79)	(0.67)	(0.03)	(0.03)
Number of children in household	2.46	1.38	3.53	-2.15***	3.82	1.10	2.72***	-3.33***
	(1.88)	(1.39)	(1.69)	(0.05)	(1.50)	(1.07)	(0.04)	(0.05)
Education level of household head	6.48	5.53	7.42	-1.89***	6.95	6.00	0.94***	-1.94***
	(3.72)	(3.88)	(3.30)	(0.12)	(3.28)	(4.07)	(0.12)	(0.14)
Age of household head	42.37	45.29	39.45	5.84***	40.59	44.17	-3.57***	6.46***
	(15.47)	(18.10)	(11.59)	(0.50)	(12.14)	(18.07)	(0.52)	(0.62)
Female household head	0.26	0.41	0.10	0.31***	0.21	0.31	-0.10***	0.28***
	(0.44)	(0.49)	(0.30)	(0.01)	(0.41)	(0.46)	(0.01)	(0.02)
Widow	0.18	0.34	0.03	0.32***	0.13	0.24	-0.12***	0.30***
	(0.39)	(0.48)	(0.16)	(0.01)	(0.33)	(0.43)	(0.01)	(0.01)
Household owns any livestock	0.27	0.17	0.36	-0.19***	0.38	0.15	0.23***	-0.29***
	(0.44)	(0.38)	(0.48)	(0.01)	(0.49)	(0.36)	(0.01)	(0.02)
Household owns land	0.84	0.84	0.84	0.01	0.89	0.80	0.09***	-0.06***
	(0.36)	(0.36)	(0.37)	(0.01)	(0.32)	(0.40)	(0.01)	(0.01)
Respondent self-employed	0.27	0.20	0.34	-0.14***	0.29	0.25	0.04**	-0.12***
	(0.44)	(0.40)	(0.47)	(0.01)	(0.45)	(0.43)	(0.01)	(0.02)
Respondent employed	0.33	0.31	0.35	-0.03**	0.37	0.29	0.08***	-0.08***
	(0.47)	(0.46)	(0.48)	(0.02)	(0.48)	(0.45)	(0.02)	(0.02)
Number of meals eaten yesterday	2.29	2.21	2.38	-0.17***	2.22	2.37	-0.16***	-0.01
	(0.68)	(0.69)	(0.66)	(0.02)	(0.66)	(0.69)	(0.02)	(0.03)
Number of meals with protein yesterday	0.43	0.34	0.52	-0.18***	0.40	0.46	-0.07***	-0.08***
	(0.60)	(0.55)	(0.64)	(0.02)	(0.59)	(0.62)	(0.02)	(0.02)
Panel B: Endline Treatment Effects (SEs)								
Respondent hours worked last week	1.05	1.97	-0.23	2.19	-0.17	2.12	-2.29	1.43
	(0.98)	(1.33)	(1.34)	(1.84)	(1.28)	(1.38)	(1.79)	(1.94)
Indicator for household self-employed	0.04**	0.04*	0.02	0.02	0.06***	0.01	0.05	-0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Indicator for household employed	-0.03	-0.03	-0.03	-0.00	-0.05**	-0.01	-0.04	0.02
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Interhousehold transfers received	9.85	-9.52	28.91***	-38.43***	27.68***	-8.34	36.02***	-41.18***
	(7.13)	(9.77)	(9.59)	(13.18)	(10.44)	(8.92)	(13.26)	(14.33)
Interhousehold transfers sent	7.96***	6.92**	8.13*	-1.21	9.41**	6.34	3.07	1.34
	(2.75)	(3.32)	(4.45)	(5.65)	(4.11)	(3.87)	(5.80)	(5.30)
Total value of loans taken in last 12 months	3.48	-5.41	10.73	-16.14*	8.18	-1.79	9.96	-10.87
	(4.03)	(4.40)	(6.89)	(8.22)	(6.06)	(5.61)	(8.38)	(8.90)
Total value of loans given in last 12 months	3.43***	3.86***	2.78**	1.08	2.77**	4.08***	-1.31	2.56
	(0.79)	(1.00)	(1.32)	(1.71)	(1.11)	(1.20)	(1.69)	(1.69)

Notes: This table presents differences in baseline characteristics (Panel A) and treatment effects estimated via OLS (Panel B) based on households' classification into most deprived (D) or most impacted (I) groups across our 150 models. Specifically, households are assigned to the most deprived (impacted) group if they are classified as deprived (impacted) in more than 50% of the 150 models.

Table A.4: Differences in characteristics and treatment effects by deprivation and impact group classification using assets

	(1) All	(2) D	(3) $!D$	(4) $D-!D$	(5) I	(6) $!I$	(7) $I-!I$	(8) $D-I$
Panel A: Baseline demographics, Mean (SD)								
Household size	4.40	2.94	5.82	-2.88***	5.65	3.16	2.49***	-3.71***
	(2.21)	(1.52)	(1.80)	(0.05)	(1.96)	(1.68)	(0.06)	(0.06)
Number of prime-age adults in household	1.88	1.61	2.09	-0.49***	2.05	1.68	0.36***	-0.61***
	(0.77)	(0.67)	(0.77)	(0.02)	(0.78)	(0.70)	(0.03)	(0.03)
Number of children in household	2.46	1.31	3.58	-2.27***	3.50	1.44	2.06***	-2.99***
	(1.88)	(1.32)	(1.66)	(0.05)	(1.74)	(1.39)	(0.05)	(0.05)
Education level of household head	6.48	5.71	7.22	-1.51***	6.89	6.08	0.81***	-1.59***
	(3.72)	(4.00)	(3.27)	(0.12)	(3.34)	(4.02)	(0.11)	(0.14)
Age of household head	42.37	44.20	40.60	3.60***	40.18	44.53	-4.35***	5.47***
	(15.47)	(17.86)	(12.51)	(0.51)	(12.13)	(17.92)	(0.49)	(0.60)
Female household head	0.26	0.38	0.14	0.23***	0.20	0.31	-0.11***	0.23***
	(0.44)	(0.48)	(0.35)	(0.01)	(0.40)	(0.46)	(0.01)	(0.02)
Widow	0.18	0.31	0.06	0.25***	0.12	0.25	-0.14***	0.26***
	(0.39)	(0.46)	(0.24)	(0.01)	(0.32)	(0.43)	(0.01)	(0.01)
Household owns any livestock	0.27	0.07	0.45	-0.38***	0.39	0.14	0.25***	-0.43***
	(0.44)	(0.26)	(0.50)	(0.01)	(0.49)	(0.35)	(0.01)	(0.02)
Household owns land	0.84	0.82	0.86	-0.04***	0.86	0.83	0.03***	-0.05***
	(0.36)	(0.38)	(0.35)	(0.01)	(0.35)	(0.38)	(0.01)	(0.01)
Respondent self-employed	0.27	0.22	0.32	-0.10***	0.38	0.16	0.22***	-0.22***
	(0.44)	(0.41)	(0.47)	(0.01)	(0.49)	(0.37)	(0.01)	(0.02)
Respondent employed	0.33	0.33	0.33	-0.01	0.44	0.22	0.21***	-0.15***
	(0.47)	(0.47)	(0.47)	(0.02)	(0.50)	(0.42)	(0.01)	(0.02)
Number of meals eaten yesterday	2.29	2.21	2.37	-0.16***	2.12	2.47	-0.35***	0.13***
	(0.68)	(0.69)	(0.66)	(0.02)	(0.63)	(0.68)	(0.02)	(0.03)
Number of meals with protein yesterday	0.43	0.39	0.47	-0.07***	0.46	0.41	0.05***	-0.09***
	(0.60)	(0.59)	(0.62)	(0.02)	(0.61)	(0.60)	(0.02)	(0.02)
Panel B: Endline Treatment Effects (SEs)								
Respondent hours worked last week	1.05	2.01	-0.07	2.09	1.33	0.70	0.64	0.07
	(0.98)	(1.33)	(1.36)	(1.86)	(1.26)	(1.36)	(1.73)	(1.89)
Indicator for household self-employed	0.04**	0.04	0.03	0.00	0.05**	0.02	0.03	-0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Indicator for household employed	-0.03	-0.02	-0.04*	0.02	-0.04*	-0.02	-0.03	0.02
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Interhousehold transfers received	9.85	-11.00	29.81***	-40.81***	23.81**	-3.97	27.78**	-45.18***
	(7.13)	(9.26)	(10.26)	(13.42)	(9.67)	(10.09)	(13.72)	(13.90)
Interhousehold transfers sent	7.96***	8.10**	7.55*	0.55	6.75*	9.08**	-2.32	4.73
	(2.75)	(3.81)	(3.99)	(5.57)	(4.03)	(3.73)	(5.47)	(5.11)
Total value of loans taken in last 12 months	3.48	1.43	4.85	-3.42	4.86	1.89	2.97	1.66
	(4.03)	(4.81)	(6.63)	(8.31)	(6.10)	(5.55)	(8.37)	(9.12)
Total value of loans given in last 12 months	3.43***	3.70***	3.08**	0.62	2.89**	3.93***	-1.05	2.12
	(0.79)	(1.04)	(1.28)	(1.70)	(1.23)	(1.11)	(1.72)	(1.82)

Notes: This table presents differences in baseline characteristics (Panel A) and treatment effects estimated via OLS (Panel B) based on households' classification into most deprived (D) or most impacted (I) groups across our 150 models. Specifically, households are assigned to the most deprived (impacted) group if they are classified as deprived (impacted) in more than 50% of the 150 models.

Table A.5: Differences in characteristics and treatment effects by deprivation and impact group classification using income

	(1) All	(2) D	(3) $!D$	(4) $D-!D$	(5) I	(6) $!I$	(7) $I-!I$	(8) $D-I$
Panel A: Baseline demographics, Mean (SD)								
Household size	4.40	3.56	5.21	-1.65***	4.97	3.86	1.11***	-2.50***
	(2.21)	(2.05)	(2.04)	(0.07)	(2.46)	(1.77)	(0.07)	(0.09)
Number of prime-age adults in household	1.88	1.66	2.06	-0.40***	1.98	1.79	0.19***	-0.53***
	(0.77)	(0.80)	(0.68)	(0.03)	(0.79)	(0.73)	(0.03)	(0.03)
Number of children in household	2.46	1.82	3.08	-1.26***	2.98	1.97	1.00***	-2.05***
	(1.88)	(1.73)	(1.81)	(0.06)	(2.09)	(1.51)	(0.06)	(0.07)
Education level of household head	6.48	5.33	7.59	-2.26***	6.55	6.41	0.13	-2.15***
	(3.72)	(3.87)	(3.21)	(0.12)	(3.63)	(3.81)	(0.13)	(0.16)
Age of household head	42.37	47.17	37.72	9.45***	41.82	42.90	-1.08**	9.51***
	(15.47)	(17.60)	(11.29)	(0.47)	(14.09)	(16.69)	(0.51)	(0.65)
Female household head	0.26	0.45	0.07	0.38***	0.22	0.29	-0.07***	0.41***
	(0.44)	(0.50)	(0.26)	(0.01)	(0.42)	(0.46)	(0.01)	(0.02)
Widow	0.18	0.37	0.00	0.37***	0.15	0.22	-0.06***	0.40***
	(0.39)	(0.48)	(0.00)	(0.01)	(0.36)	(0.41)	(0.01)	(0.02)
Household owns any livestock	0.27	0.21	0.32	-0.10***	0.34	0.19	0.15***	-0.23***
	(0.44)	(0.41)	(0.47)	(0.01)	(0.48)	(0.39)	(0.01)	(0.02)
Household owns land	0.84	0.85	0.84	0.01	0.93	0.76	0.17***	-0.15***
	(0.36)	(0.36)	(0.37)	(0.01)	(0.25)	(0.43)	(0.01)	(0.02)
Respondent self-employed	0.27	0.13	0.41	-0.28***	0.19	0.34	-0.15***	-0.11***
	(0.44)	(0.33)	(0.49)	(0.01)	(0.39)	(0.47)	(0.01)	(0.02)
Respondent employed	0.33	0.31	0.35	-0.04**	0.60	0.07	0.53***	-0.51***
	(0.47)	(0.46)	(0.48)	(0.02)	(0.49)	(0.26)	(0.01)	(0.02)
Number of meals eaten yesterday	2.29	2.15	2.43	-0.28***	2.29	2.30	-0.01	-0.24***
	(0.68)	(0.71)	(0.62)	(0.02)	(0.66)	(0.69)	(0.02)	(0.03)
Number of meals with protein yesterday	0.43	0.30	0.56	-0.26***	0.61	0.27	0.34***	-0.54***
	(0.60)	(0.53)	(0.65)	(0.02)	(0.64)	(0.51)	(0.02)	(0.02)
Panel B: Endline Treatment Effects (SEs)								
Respondent hours worked last week	1.05	1.31	0.39	0.92	0.26	1.69	-1.43	1.21
	(0.98)	(1.27)	(1.34)	(1.77)	(1.26)	(1.39)	(1.78)	(2.16)
Indicator for household self-employed	0.04**	0.02	0.05**	-0.03	0.06***	0.01	0.05	-0.06
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)
Indicator for household employed	-0.03	-0.02	-0.04*	0.02	-0.03	-0.02	-0.01	0.03
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)
Interhousehold transfers received	9.85	8.68	10.91	-2.23	13.92	6.33	7.58	-4.47
	(7.13)	(9.59)	(9.81)	(13.15)	(8.69)	(10.39)	(12.93)	(16.34)
Interhousehold transfers sent	7.96***	4.61	10.35**	-5.74	4.91	10.83**	-5.91	2.90
	(2.75)	(3.55)	(4.39)	(5.85)	(3.79)	(4.20)	(5.83)	(5.74)
Total value of loans taken in last 12 months	3.48	1.38	4.07	-2.69	9.45*	-1.99	11.44	-8.78
	(4.03)	(4.31)	(6.90)	(8.21)	(5.36)	(6.01)	(8.15)	(9.56)
Total value of loans given in last 12 months	3.43***	3.11***	3.46***	-0.34	3.59***	3.29***	0.29	1.32
	(0.79)	(0.91)	(1.29)	(1.59)	(1.13)	(1.11)	(1.58)	(1.93)

Notes: This table presents differences in baseline characteristics (Panel A) and treatment effects estimated via OLS (Panel B) based on households' classification into most deprived (D) or most impacted (I) groups across our 150 models. Specifically, households are assigned to the most deprived (impacted) group if they are classified as deprived (impacted) in more than 50% of the 150 models.

Table A.6: Predicted per capita untreated food outcomes (y_h^0) by group

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Food Security Index				
Predicted	0.06	-0.12	-0.05	-0.07
Actual	0.04	-0.16	-0.10	-0.06
				(0,0.16)
				[-0.11,-0.02]
Panel B: Total household food consumption				
Predicted	471	352	552	-200
Actual	467	351	549	-198
				(-278,-132)
				[-233,-163]
Panel C: Food consumption per capita				
Predicted	221	101	181	-80
Actual	220	126	187	-61
				(-117,91)
				[-83,-39]

Notes: This table reproduces Table 1 for food security-related outcomes. The food security index in an index of questions about the food consumption of adults and children over the past 7 days (see Appendix B for details). The rest of the details follow Table 1.

Table A.7: Predicted Average Treatment Effects for food outcomes (Δ_i) by group

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Food Security Index				
Predicted	42.65	44.19	59.69	-15.50
Actual	48.58	60.64	71.07	-10.43
				(-0.07,0.06)
				[-32.08,11.21]
Panel B: Total household food consumption				
Predicted	84	73	116	-43
Actual	100	64	126	-62
				(-138,48)
				[-111,-13]
Panel C: Food consumption per capita				
Predicted	20	23	29	-7
Actual	28	23	32	-9
				(-31,32)
				[-27,8]

Notes: This table reproduces Table 2 for food security-related outcomes. The food security index in an index of questions about the food consumption of adults and children over the past 7 days (see Appendix B for details). The rest of the details follow Table 2.

Table A.8: Within-village spillovers for inelegible households

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	91	86	118	-31
Actual	147	107	53	54
				(53,469)
				[-73,182]
Panel B: Assets				
Predicted	81	66	112	-46
Actual	105	31	83	-52
				(-179,294)
				[-176,71]
Panel C: Income				
Predicted	-8	-22	26	-48
Actual	25	-22	78	-100
				(-255,192)
				[-225,24]

Notes: This table reproduces Table 2 for ineligible households located within treatment villages, in order to look at within-village spillover effects onto ineligible households. While ineligible households were not treated themselves, we make use of the treatment status of their village to estimate models in the same method as Table 2. $N = 2,434$.

Table A.9: Cross-village spillovers for control village transfer-eligible households

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	9	5	32	-27
Actual	32	-22	-119	97
				(19,393)
				[-6,199]
Panel B: Assets				
Predicted	27	21	40	-19
Actual	27	-12	5	-18
				(-89,154)
				[-80,45]
Panel C: Income				
Predicted	101	102	116	-14
Actual	110	97	-20	117
				(149,443)
				[39,195]

Notes: This table reproduces Table 2 for control village transfer-eligible households, in order to look at cross-village spillovers. Here, treament is defined as whether the neighborhood treatment intensity (the amount transferred within 2 km of their village, from Egger et al. (2019)) for a household is above the median. The rest of the model estimation follows Table 2; we estimate causal forests using the indicator for above-median intensity as the definition of treatment. $N = 2,367$.

B Data & variable construction

B.1 Baseline predictors

Our baseline survey collected data on a number of household characteristics that might predict endline outcomes and treatment effects. From among these, we first selected a subset that have been documented as appearing in other proxy means tests used to target social protection programs in comparable low-income countries. Specifically, we retained all variables in the intersection of household-level variables in the PMTs studied by Kidd and Wylde (2011), Alatas et al. (2012), and Niehaus et al. (2013), for a total of 31 potential predictors. Among these we retained those that exhibit non-trivial amounts of variation in our data, yielding 24 potential predictors. Table B.1 summarizes this selection procedure.

In our pre-analysis plan we planned to further narrow our feature selection by keeping predictors that increased the adjusted R-squared of a regression predicting baseline outcomes. This procedure is not well-defined, however (since whether or not a variable increases the adjusted R-squared depends on what other variables are included). It also creates further complications for inference, since it uses the data once to select predictors before then using them again to form predictors. Our preferred approach is therefore to use a list of 16 covariates selected by hand, prior to any analysis, based on our knowledge of the local context. These covariates are

1. Household size
2. Respondent a widow
3. Respondent female
4. Household has children
5. Household has school-aged children
6. Household has children 3 or under
7. Household has children 6 or under
8. Household has an elderly (65+) member
9. Household owns any livestock
10. Household owns any land
11. Household owns more than 0.25 acres of land
12. Household owns TV or radio
13. Number of meals eaten yesterday
14. Number of meals with protein yesterday
15. Respondent self-employed
16. Respondent employed

As a robustness check we also examine results using a data-selected subset of 24 candidate predictors. Specifically, we run LASSO regressions using these 24 variables to predict our four main outcomes, and retain all variables selected by LASSO in any of these four models. This yields a list of 15 variables that is in fact quite similar to our hand-selected list (see Table B.1). Re-estimating our core specifications using this alternative list of predictors, we obtain results that are generally quite similar to our main results (Tables D.7, D.8, and D.9).

B.2 Outcomes

The endline survey contained detailed modules on economic activities such as household expenditures and crop production, asset ownership, psychological well-being, health and nutrition, and female respondents surveyed by a female enumerator were also administered a module on female empowerment and gender-based violence. We construct four aggregate outcomes using this data: consumption expenditure, income, assets, and a food security index.

Consumption expenditure is defined as the total annualized household expenditure. Income is total annualized household income, given by the sum of agricultural profits, profits from self-employment, and wage earnings. Assets is equal to the total value of household's assets, excluding land and houses. Each of these variables is winsorized at the 99th percentile and expressed in USD PPP terms.

The food security index is a weighted average of standardized food security covariates constructed according to Anderson (2008). The index is calculated using the following variables:

1. Number of days adults skipped or cut the amount of meals in the past 7 days
2. Number of days children skipped or cut the amount of meals in the past 7 days
3. Number of days adults went the entire day without meals in the past 7 days
4. Number of days children went the entire day without meals in the past 7 days
5. Number of days adults went to bed hungry in the last 7 days
6. Number of days children went to bed hungry in the past 7 days
7. Number of meals eaten yesterday that included meat, fish, or eggs

Table B.1: Selecting Proxy Means Test (PMT) Variables

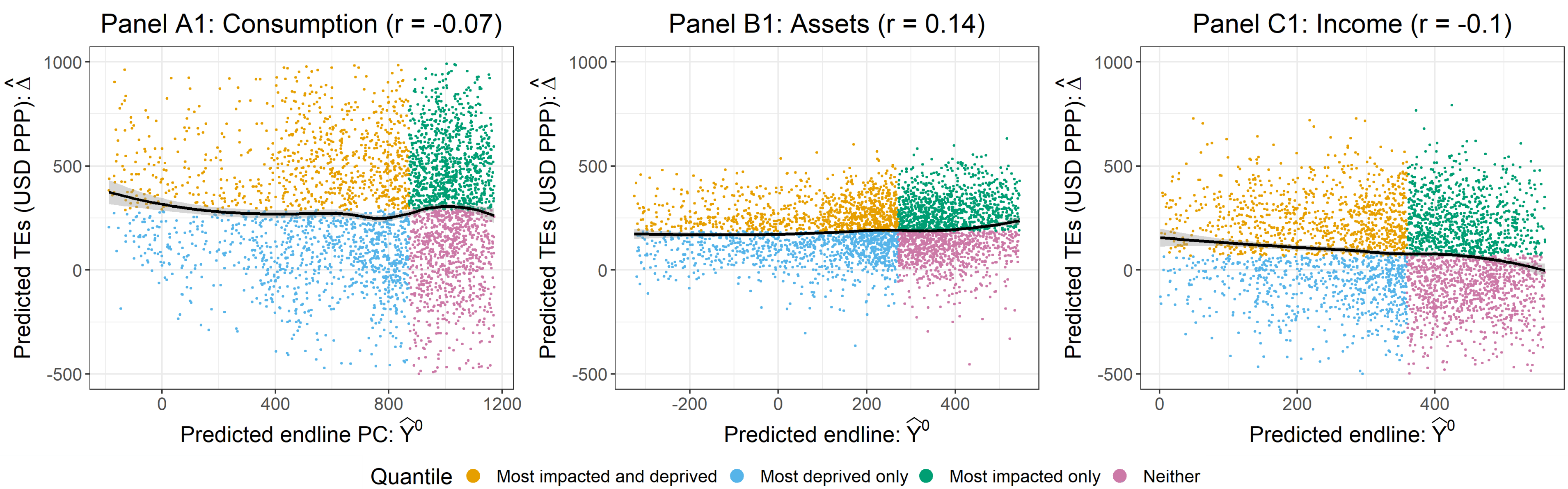
(1) Measure	(2) Source	(3) Collected in GE Baseline (1/0)	(4) Variation in GE baseline data (1/0)	(5) Selected (preferred) PMT list	(6) Selected via LASSO (robustness check)
Panel A: Human Capital					
Education of Household Head	KW, A	1	1		Indicator for above median edu for HH head
Highest level of education in household	KW, A	1	1		Indicator for above median edu for highest in HH
Female literacy	KW	0	–		
Number of children in school	KW, A	1	1		Number of children in school
Panel B: Demographic Characteristics					
Household Size	KW, A	1	1	Household size	Household size
Number of Children	KW, A	1	1	Indicator for children; Indicator for child under 3; Indicator for child under 6; Indicator for school-age children	Indicator for has children
Gender/marital status of head (e.g. widow)	KW, A	1	1	Respondent is female; Respondent is widow	Female household head; Respondent is widow
Age of household head	KW, A	1	1	Household has elderly member	Household head age; Household has elderly member
Dependency ratio	KW, A	1	1		
Panel C: Household assets					
Own home	KW, A	1	0		
Wall material	KW, A	1	0		
Roofing material	KW, A	1	0		
Floor material	A	1	0		
Number of rooms / floor space per-capita	KW, A	1	1		
Type of latrine / toilet	KW, A	1	0		
Water source	A, N	1	1		
Access to electricity	KW, A	1	0		
Gas connection	N	0	–		
Type of cooking fuel	KW, A	0	–		
Radio, television	KW, N	1	1	Indicator for owning TV or radio	Indicator for owning TV or radio
Telephone / Mobile phone	KW, N	1	1		Indicator for mobile phone
Cooker, heater, fan, air conditioning	KW	1	1		Above median appliance value
Furniture	KW	1	1		Above median furniture value
Bicycle, car, motorcycle	KW, N	1	1		Indicator for owning bicycle
Access to microcredit	A	1	1		
Panel D: Productive assets					
Landholding size	KW, N	1	1	Household owns land; Household owns more than 0.25 acres of land	
Livestock	KW	1	1	Indicator for owning livestock	Indicator for owning livestock
Use of fertilizer	KW	1	1		
Panel E: Livelihood options					
Agricutural or non-farm wage labor	KW	1	1	Respondent employed	Respondent employed
Non-farm independent business	KW	1	1	Respondent self-employed	Respondent self-employed
Agricultural production of cash or staple crops	KW	1	0		
Receipt of foreign remittances	KW	0	–		
Sector of work (informal, industry, or agriculture)	KW, A	1	1		
Annual income threshold	N	1	1		Above median total income
Government employee	N	0	–		
Food security (adults, children)		1	1	Number of meals eaten yesterday; Number of meals with protein yesterday	
Total variables		31	24	16	18

Notes: This table outlines variables that have been included in proxy means tests (PMTs) and their overlap with variables in this study's baseline survey. Column (1) reports the measures from the sources in Column (2), namely Kidd and Wylde (2011) (KW), Alatas et al. (2012) (A), and Niehaus et al. (2013) (N). Columns (3) and (4) denote whether or not similar variables were collected as part of the GE baseline survey, and if so, whether there is meaningful variation in the variable. Column (5) includes our preferred list of PMT-like variables, and column (6) reports the set of variables selected via LASSO among those in column (4).

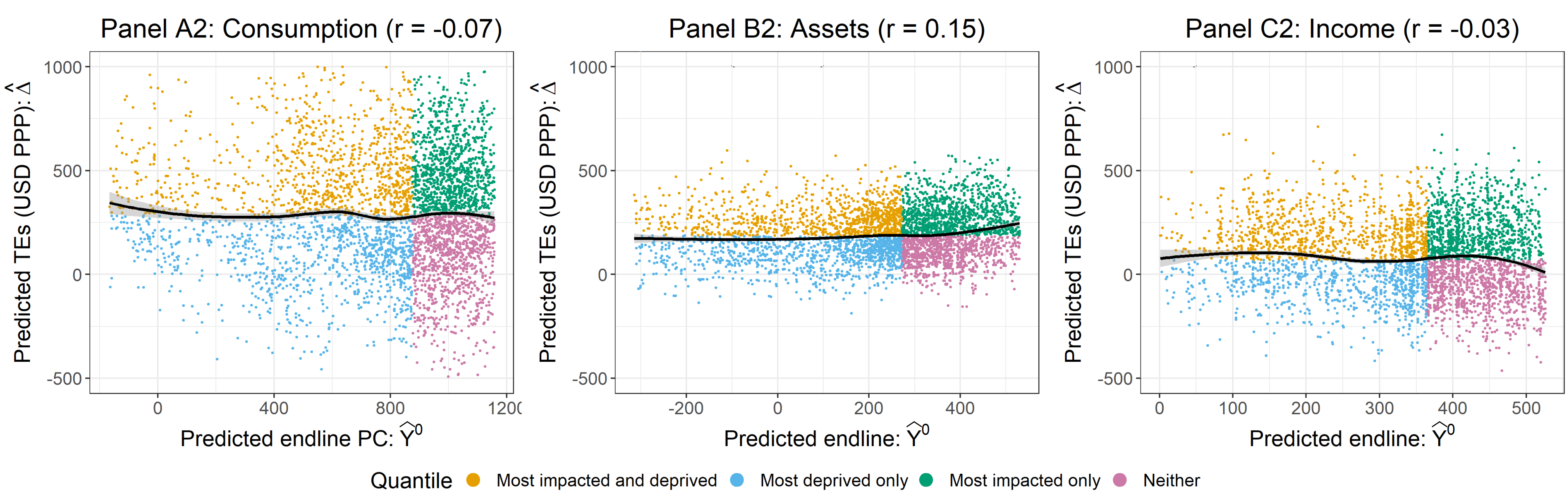
C Alternative learning models

Figure C.1: Predicted treatment effects ($\hat{\Delta}_h$) plotted against the predicted untreated per capita values ($\hat{y}_h^0$) for OLS and LASSO predictions

1. OLS



2. LASSO



Notes: These figures re-create Figure 1 using OLS (Row 1) and LASSO (Row 2) as prediction methods in contrast to generalized random forests.

Table C.1: Predicted per capita untreated outcomes (y_h^0) by group using OLS for prediction

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	729	433	720	-288
Actual	729	487	736	-249
				(-348,-145)
				[-296,-202]
Panel B: Assets				
Predicted	213	21	239	-219
Actual	213	46	253	-207
				(-289,-159)
				[-233,-181]
Panel C: Income				
Predicted	297	134	281	-147
Actual	297	173	304	-131
				(-187,-86)
				[-162,-100]

Notes: This table reproduces Table 1 but replaces GRF predictions for each fold by ordinary least squares (OLS) predictions. See Table 1 for more details.

Table C.2: Predicted Average Treatment Effects (Δ_i) by group using OLS for prediction

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	285	289	531	-242
Actual	310	341	398	-56
				(-164,166)
				[-133,20]
Panel B: Assets				
Predicted	190	179	287	-108
Actual	182	186	199	-14
				(-48,114)
				[-54,27]
Panel C: Income				
Predicted	75	98	244	-145
Actual	85	75	106	-32
				(-86,140)
				[-89,26]

Notes: This table reproduces Table 2 but replaces GRF predictions for each fold by ordinary least squares (OLS) predictions. See Table 2 for more details.

Table C.3: Overlap of socially optimal households with most impacted and deprived using OLS for prediction

	(1)	(2)	(3)	(4)	(5)
CARA: α	CE	Most deprived	Most impacted	Choice	α_c
Panel A: Consumption					
0.0000	\$50	0.48	1.00	I	
0.0005	\$49	0.48	0.96	I	
0.0010	\$49	0.48	0.93	I	
0.0075	\$41	0.53	0.79	I	
0.0150	\$33	0.55	0.75	I	← - - 0.806
Panel B: Assets					
0.0000	\$50	0.46	1.00	I	
0.0005	\$49	0.48	0.97	I	
0.0010	\$49	0.50	0.94	I	
0.0075	\$41	0.63	0.69	I	
0.0150	\$33	0.66	0.62	I	← - - 0.150
Panel C: Income					
0.0000	\$50	0.54	1.00	I	
0.0005	\$49	0.55	0.99	I	
0.0010	\$49	0.55	0.98	I	
0.0075	\$41	0.57	0.90	I	
0.0150	\$33	0.58	0.86	I	← - - 0.693

Notes: This table reproduces the social welfare analysis of Table 4 but using OLS predictions (Tables C.1 and C.2) rather than GRF predictions.

Table C.4: Predicted per capita untreated outcomes (y_h^0) by group using LASSO

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	729	440	711	-270
Actual	729	486	718	-232
				(-328,-120)
				[-278,-185]
Panel B: Assets				
Predicted	213	29	240	-212
Actual	213	47	245	-198
				(-281,-141)
				[-224,-172]
Panel C: Income				
Predicted	297	152	295	-143
Actual	297	172	306	-133
				(-193,-82)
				[-164,-102]

Notes: This table reproduces Table 1 but with LASSO predictions replacing GRF predictions for each fold.

Table C.5: Predicted Average Treatment Effects (Δ_i) by group using LASSO

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	286	294	511	-218
Actual	310	346	404	-58
				(-149,169)
				[-132,17]
Panel B: Assets				
Predicted	189	177	280	-103
Actual	182	179	212	-32
				(-78,77)
				[-73,9]
Panel C: Income				
Predicted	74	78	208	-130
Actual	85	79	108	-29
				(-79,155)
				[-87,29]

Notes: This table reproduces Table 2 but with LASSO predictions replacing GRF predictions for each fold.

Table C.6: Overlap of socially optimal households with most impacted and deprived using LASSO

	(1)	(2)	(3)	(4)	(5)
CARA: α	CE	Most deprived	Most impacted	Choice	α_c
Panel A: Consumption					
0.0000	\$50	0.50	1.00	I	
0.0005	\$49	0.50	0.96	I	
0.0010	\$49	0.50	0.93	I	
0.0075	\$41	0.53	0.79	I	
0.0150	\$33	0.55	0.74	I	0.148
Panel B: Assets					
0.0000	\$50	0.46	1.00	I	
0.0005	\$49	0.47	0.97	I	
0.0010	\$49	0.48	0.93	I	
0.0075	\$41	0.61	0.68	I	0.008
0.0150	\$33	0.64	0.61	D	
Panel C: Income					
0.0000	\$50	0.51	1.00	I	
0.0005	\$49	0.51	0.99	I	
0.0010	\$49	0.52	0.98	I	
0.0075	\$41	0.53	0.89	I	
0.0150	\$33	0.54	0.84	I	0.938

Notes: This table reproduces the social welfare analysis of Table 4 but with LASSO predictions (from Tables C.4 and C.5). See Table 4 for more details.

D Robustness checks

Table D.1: Predicting endline values (Y_i^0) by group, non-time-demeaned data

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	762	566	599	-33
Actual	752	542	572	-30
				(-58,-2)
Panel B: Assets				
Predicted	233	158	210	-53
Actual	224	163	200	-37
				(-57,-16)
Panel C: Income				
Predicted	313	212	311	-99
Actual	309	209	306	-97
				(-129,-64)

Notes: This table reproduces Table 1 using data that is not time-demeaned. Due to computational limitations we only compute the bootstrap CIs for our main results. Following Chernozhukov et al. (2018), we report the confidence intervals using the median standard error for the actual statistic, clustered at the village level. See Table 1 for more details.

Table D.2: Comparing ATEs (Δ_i) by Group, non-time-demeaned data

Statistic	(1) All	(2) Most deprived	(3) Most impacted	(4) Difference
Panel A: Consumption				
Predicted	291	264	355	-91
Actual	321	304	407	-102
				(-184,-20)
Panel B: Assets				
Predicted	195	200	217	-17
Actual	187	202	193	9
				(-28,47)
Panel C: Income				
Predicted	73	80	106	-26
Actual	88	101	103	-3
				(-56,51)

Note: This table reproduces Table 2 using training data that is not time-demeaned, and shows results remain similar. Following Chernozhukov et al. (2018), we report the confidence intervals using the median standard error for the actual statistic, clustered at the village level. See Table 2 for more details.

Table D.3: Overlap of socially optimal households with most deprived and most impacted using CRRA utility

	(1)	(2)	(3)	(4)
CRRA: ρ	Most deprived	Most impacted	Choice	ρ_c
Panel A: Consumption				
0.0000	0.30	1.00	I	
0.5000	0.33	0.93	I	
1.0000	0.35	0.89	I	
2.0000	0.38	0.85	I	2.99
4.0000	0.40	0.81	D	
Panel B: Assets				
0.0000	0.31	1.00	I	
0.5000	0.48	0.69	I	0.94
1.0000	0.53	0.61	D	
2.0000	0.56	0.56	D	
4.0000	0.57	0.54	D	
Panel C: Income				
0.0000	0.46	0.99	I	
0.5000	0.52	0.88	I	
1.0000	0.56	0.80	I	1.27
2.0000	0.60	0.72	D	
4.0000	0.63	0.65	D	

Notes: This table reproduces the social welfare analysis of Table 4, but using constant relative risk aversion (CRRA) utility. Column 1 (2) reports the share of households belonging to I (D) that are also "socially optimal" for a planner to treat. Socially optimal households are those in the top 50% of households ranked by potential gains from treatment using a CRRA utility function for the risk aversion parameter (ρ) given in the row label. Reported shares are the mean of 150 5-fold GRF iterations; median ratios are similar (not shown). Column 3 reports the welfare maximizing choice between targeting the most impacted (I) and the most deprived (D) for a given ρ value. Column (4) reports ρ_c, the mean minimum value of ρ required to rationalize a policy targeting the most deprived instead of targeting the most impacted across the 150 estimated models. Formally, $\rho_c = \min(\{\rho : SW(D;\rho) \geq SW(I;\rho)\})$.

Table D.4: Predicted per capita untreated outcomes (y_h^0) by group using untreated data for endline predictions

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	729	501	896	-395
Actual	729	511	903	-393
				(-456,-329)
Panel B: Assets				
Predicted	213	57	315	-258
Actual	213	56	326	-269
				(-299,-240)
Panel C: Income				
Predicted	298	168	301	-133
Actual	297	171	313	-143
				(-176,-110)

Notes: This table reproduces Table 1 by generating prediction models for endline outcomes using only data from transfer-eligible households in control villages. Due to computational limitations we only compute the bootstrap CIs for our main results. Following Chernozhukov et al. (2018), we report the confidence intervals using the median standard error for the actual statistic, clustered at the village level. See Table 1 for more details.

Table D.5: Predicted Average Treatment Effects (Δ_i) by group using untreated data for endline predictions

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	281	249	346	-96
Actual	310	287	402	-115
				(-199,-31)
Panel B: Assets				
Predicted	190	178	211	-33
Actual	182	144	182	-39
				(-85,8)
Panel C: Income				
Predicted	71	70	104	-35
Actual	85	96	92	4
				(-57,65)

Note: This table reproduces Table 2 where the classification of most deprived comes from models trained using only untreated data (i.e. transfer-eligible households in control villages), and shows results remain similar. The process for classifying most impacted households is the same as in Table 2. Following Chernozhukov et al. (2018), we report the confidence intervals using the median standard error for the actual statistic, clustered at the village level. See Table 2 for more details.

Table D.6: Overlap of socially optimal households with most impacted and deprived using untreated data for endline predictions

CARA: α	(1) Most deprived	(2) Most impacted	(3) Choice	(4) α_c
Panel A: Consumption				
0.0000	0.31	1.00	I	
0.0005	0.33	0.95	I	
0.0010	0.35	0.92	I	
0.0075	0.42	0.81	I	← 0.013
0.0150	0.44	0.79	D	
Panel B: Assets				
0.0000	0.32	1.00	I	
0.0005	0.35	0.91	I	
0.0010	0.40	0.84	I	← 0.006
0.0075	0.55	0.59	D	
0.0150	0.57	0.55	D	
Panel C: Income				
0.0000	0.49	1.00	I	
0.0005	0.50	0.97	I	
0.0010	0.51	0.94	I	
0.0075	0.59	0.74	I	← 0.011
0.0150	0.62	0.67	D	

Notes: This table reproduces the social welfare analysis of Table 4, but using the results based on estimating endline outcomes with untreated data from Tables D.4 and D.5. See Table 4 for more details.

Table D.7: Predicted per capita untreated outcomes (y_h^0) by group with LASSO selected covariates

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	752	541	922	-381
Actual	729	500	916	-416
				(-476,-356)
Panel B: Assets				
Predicted	231	72	326	-254
Actual	213	36	322	-287
				(-317,-256)
Panel C: Income				
Predicted	305	182	323	-141
Actual	297	158	328	-170
				(-213,-126)

Notes: This table reproduces Table 1 but with covariates (features) selected via LASSO (see Table B.1).

Table D.8: Predicted Average Treatment Effects (Δ_i) by group with LASSO selected covariates

Statistic	(1) All	(2) Most deprived (D)	(3) Most impacted (I)	(4) Difference (D)-(I)
Panel A: Consumption				
Predicted	260	237	308	-71
Actual	310	300	396	-96
				(-180,-12)
Panel B: Assets				
Predicted	182	171	200	-29
Actual	182	180	183	-3
				(-49,43)
Panel C: Income				
Predicted	62	61	89	-27
Actual	85	86	101	-15
				(-78,48)

Notes: This table reproduces Table 2 but with covariates (features) selected via LASSO (see Table B.1).

Table D.9: Overlap of socially optimal households with most deprived and most impacted with LASSO selected covariates

	(1)	(2)	(3)	(4)	(5)
CARA: α	CE	Most deprived	Most impacted	Choice	α_c
Panel A: Consumption					
0.0000	\$50	0.32	1.00	I	
0.0005	\$49	0.34	0.93	I	
0.0010	\$49	0.36	0.89	I	
0.0075	\$41	0.43	0.75	I	← - - 0.012
0.0150	\$33	0.44	0.73	D	
Panel B: Assets					
0.0000	\$50	0.32	1.00	I	
0.0005	\$49	0.37	0.89	I	
0.0010	\$49	0.42	0.80	I	← - - 0.005
0.0075	\$41	0.61	0.53	D	
0.0150	\$33	0.63	0.50	D	
Panel C: Income					
0.0000	\$50	0.49	1.00	I	
0.0005	\$49	0.51	0.96	I	
0.0010	\$49	0.53	0.93	I	
0.0075	\$41	0.64	0.71	I	← - - 0.008
0.0150	\$33	0.67	0.64	D	

Notes: This table reproduces the social welfare analysis of Table 4 but with features selected via LASSO (Table B.1), as in Tables D.7 and D.8). See Table 4 for more details.

E Algorithms

Algorithm 2:

```
Result: Mean statistic of the most impacted and most deprived
for i in 1...150 do
    Split data randomly into 5 folds*;
    for k in 1...5 do
        test data ← fold(k);
        train data ← fold(-k);
        rf ← randomforest(train data);
        cf ← causalforest(train data);
        Ŷ⁰(X_i) ← predict(rf, test data);
        Δ̂(X_i) ← predict(cf, test data);
        HH_k ← Households ∩ fold(k);
        for h in HH_k do
            if Ŷ⁰(X_i)[h] ≤ median(Ŷ⁰(X_i)) then
                deprived[h] = 1
            end
            if Δ̂(X_i)[h] > median(Δ̂(X_i)) then
                impacted[h] = 1
            end
        end
    end
    results[i] ← statistic(deprived, impacted, data);
end
result ← mean(results);
```

* The random splits into 5 folds can be done separately for random forests and causal forests, or they can share the same splits. Our preferred results use the same splits but we include results from the alternative option as a robustness check.

Algorithm 3: Randomization inference

Result: RI test (p)

STATISTIC ← result[Algorithm 2 on actual data]

Generate a GRID of values in the 99.9% CI of the ATE

for *ATE in GRID* **do**

  Construct potential outcomes under the null of a homogeneous treatment effect **if** *household i was treated* **then**

    $Y_0 = Y_{obs} - ATE$ $Y_1 = Y_{obs}$

  else

    $Y_0 = Y_{obs}$ $Y_1 = Y_{obs} + ATE$

  end

  for *ITERATION in 1...NUMBER_RI_ITERATIONS* **do**

    Randomly assign $\frac{1}{2}$ of villages a treatment status of 1

    if $treatment = 1$ **then**

      $Y = Y_1$ for all households in the village

    else

      $Y = Y_0$ for all households in the village

    end

    Run Algorithm 2 using the simulated outcome data Y

    RIresults[i] ← result[Algorithm 2 on simulated data]

  end

  $pvals[i] \leftarrow \frac{1}{len(RIresults)} \cdot |\{R \in RIresults : R > STATISTIC\}|$

end

$p \leftarrow \max\{x : x \in pvals\}$