GE Effects of Cash Transfers: Pre-analysis plan for targeting analysis*

Johannes Haushofer†, Edward Miguel‡, Paul Niehaus§, and Michael Walker¶

September 1, 2017

Abstract

This document outlines hypothesis tests and outcomes for testing whether targeting transfers on the basis of deprivation leads to the highest per-dollar impact, using data collected from households as part of the General Equilibrium Effects (GE) project. This project is a randomized evaluation of an unconditional cash transfer program by the NGO *GiveDirectly* (GD) in Kenya. This document is part of a series of five pre-analysis plans filed to the AEA trial registry as part of the GE project, and focuses on estimating impacts and testing for differences among candidate targeting rules. We specify outcomes, variables used for targeting, and the hypothesis tests we will conduct.

1 Motivation and summary

Targeting is a core element of anti-poverty program design, with benefits typically targeted to those who are "deprived in some sense. For example, the administrator of a social cash transfer scheme might seek to target transfers to households with low per capita consumption or high levels of food insecurity.

One first-order and open question is whether this approach to targeting maximizes the per-dollar *impact* of the targeting programs. It is possible that it does - for example, if the hungriest people are those with the highest marginal propensity to consume food out of a transfer, then targeting them will maximize impacts on food security. It is also entirely

*AEA Trial Registry: AEARCTR-0000505, `https://www.socialscienceregistry.org/trials/505` We thank Justin Abraham for excellent research assistance, *GiveDirectly* for collaboration, the Busara Center for Behavioral Economics for survey development and piloting, and IPA-Kenya for data collection. This work has been funded by the Private Enterprise Development in Low-Income Countries (PEDL) initiative, the International Growth Centre, the Weiss Family Foundation, and an anonymous donor. Niehaus is a co-founder and president of *GiveDirectly*.

†Princeton University, NBER, and Busara Center for Behavioral Economics, haushofer@princeton.edu

‡UC Berkeley, emiguel@berkeley.edu

§UC San Diego, pniehaus@ucsd.edu

¶UC Berkeley, mwwalker@berkeley.edu

possible, however, that it does not - for example, if poor households are poor because they are heavily indebted, or experience difficulty exercising self-control because they are poor, they might in fact be *less* capable of converting transfers into improved food security.

In this note we pre-specify three pieces of analysis to address this question in the context of the General Equilibrium Effects (GE) evaluation (see Haushofer et al. (2017) for full details on the intervention, experimental design and data collection).[1] First, we empirically estimate the subgroups for whom impacts are largest using machine learning techniques. This enables us to identify the targeting rule that maximizes impact in a purely data-driven way, and to compare this rule to eligibility based on deprivation. Second, we specify an omnibus test of the null hypothesis that the average treatment effect for some outcome is maximized by targeting benefits to households with low initial values of that outcome. The test, which leverages recent results from the moment inequality literature, does not require us to "bet" on a specific alternative targeting rule, but rather tests against the generic alternative hypothesis that there is *some* feasible rule which performs better. Finally, we consider optimal policy when the planners objective is a non-linear function of outcomes, and hence may not coincide with maximizing average treatment effects. For example, a planner who maximizes a utilitarian social welfare function defined over may value small consumption gains by the poorest more than larger gains for the less poor, due to diminishing marginal utility. We derive the degree of curvature needed to rationalize deprivation-targeting for given outcomes.

We limit this analysis to our sample of eligible households in treatment and control villages. As we need to estimate treatment effects and target based on pre-treatment values, this analysis is restricted to eligible households that were surveyed at baseline as either an "initially-sampled" or "replacement" household.

2 Methods

2.1 Setting and hypotheses

Our unit of analysis in the household, indexed by $i = 1, \ldots, n$. Define data $W_i = (Y_i^1, Y_i^0, T_i, \mathbf{X}_i)$, where

- T_i in an indicator taking the value 1 for being in a treatment village (and hence being treated, as we restrict attention to eligible households)
- Y_i^1 is household i's potential outcome under treatment ($T_i = 1$)

[1] This "targeting" pre-analysis plan is part of a series of five pre-analysis plans based data from the GE baseline and first endline; the other analysis plans are also outlined in Haushofer et al. (2017).

- Y_i^0 is household i's potential outcome without treatment ($T_i = 0$)
- $\mathbf{X}_i \in \mathcal{X}$ is an m-dimensional vector of finitely-valued observable covariates, with probability density function f

Note that we only observe Y_i^1 (Y_i^0) if $T_i = 1$ ($T_i = 0$).

Consider some partition of $\mathcal{X}$ into two member sets E and $\mathcal{X} \setminus E$, denoting values of $\mathbf{X}_i$ that make a given household eligible or ineligible, respectively, under some targeting rule. We will be specifically interested in rules that define households as eligible if they have low values of some measure of well-being (defined below). The null hypothesis is that this rule maximizes the average treatment effect among rules that condition on $\mathbf{X}_i$ and assign the same number of households $|E|$ to treatment. This can be stated as

$$H_0 : \mathbb{E}[Y_i^1 - Y_i^0|\mathbf{X}_i = \mathbf{x}] \geq \mathbb{E}[Y_i^1 - Y_i^0|\mathbf{X}_i = \mathbf{x}']\forall \mathbf{x} \in E, \mathbf{x}' \notin E \tag{1}$$

In other words, there is no eligible recipient for whom the expected effect of treatment is strictly less than that for any ineligible recipient. If this condition were to fail then we could lower the probability that a household with characteristics $\mathbf{x}$ receives treatment and raise the probability that one with characteristics $\mathbf{x}'$ receives treatment in a way that keeps the total number of households treated constant and increases total impact.

2.2 Estimating the ATE-maximizing eligibility rule

We first estimate the targeting rule that maximizes the average treatment effect (ATE) among the set of rules that make $|E|$ households eligible. We do so using the causal tree (CT) method proposed by Athey and Imbens (2016), which estimates the conditional average treatment effect by partitioning the covariate space with a modified regression tree algorithm. We will use either a single causal tree as in Athey and Imbens (2016), or a causal forest (a set of causal trees) following Wager and Athey (2017).

With either approach, the end result is a set of "leaves" that partition the data and have similar average treatment effects. We rank leaves by their ATE and calculate the number of households in each leaf to identify an candidate targeting rule which makes eligible the minimum set of leaves with the highest ATE such that at least $|E|$ households are assigned to treatment, using randomization to break ties. For example, if $|E| = 100$, the 4 leaves with the highest ATE contain 90 individuals, and the leaf with the next-highest ATE contains 20 individuals, the targeting rule would randomly assign half the members of this marginal leaf to be eligible. We then calculate the overall ATE for this targeting rule and test for equality with the ATE obtained by targeting the individuals in E.

The CT method confers several advantages for exploratory analysis. First, it avoids overfitting by optimizing a cross-validation criterion for comparing causal effects. Second, the method allows us to identify heterogeneity without specifying baseline characteristics ex ante. This can be done both over the set of variables that we use for targeting, as well as the full set of variables in our data. We will use an "honest" approach for this analysis, whereby the data is split into testing and estimation subsamples, and observations do not appear in both. The downside of this approach is that splitting the sample may lead to a reduction in statistical power. We therefore complement it with a more traditional hypothesis test which exploits the entire data set.

2.3 Hypothesis testing framework

We next conduct a formal test of the null hypothesis H_0. To do so, we define

$$m_{\mathbf{x},\mathbf{x}'}(W_i) = \left[Y_i^1 \cdot \frac{T_i}{\mathbb{P}(T_i = 1)} - Y_i^0 \cdot \frac{1 - T_i}{1 - \mathbb{P}(T_i = 1)}\right] \left[\frac{\mathbf{1}(\mathbf{X}_i = \mathbf{x}')}{f(\mathbf{x}')} - \frac{\mathbf{1}(\mathbf{X}_i = \mathbf{x})}{f(\mathbf{x})}\right] \tag{2}$$

Note that while we never observe both Y_i^1 and Y_i^0 in our data, we do observe $m_{\mathbf{x},\mathbf{x}'}(W_i)$ for all i. Taking expectations, and noting that treatment status T_i is independent of potential outcomes (Y_i^1, Y_i^0) and of covariates $\mathbf{X}_i$ due to random assignment, we have

$$\mathbb{E}[m_{\mathbf{x},\mathbf{x}'}(W_i)] = \mathbb{E}[Y_i^1 - Y_i^0 | \mathbf{X}_i = \mathbf{x}'] - \mathbb{E}[Y_i^1 - Y_i^0 | \mathbf{X}_i = \mathbf{x}] \tag{3}$$

In other words, m is a function whose expectation is the difference between average treatment effects for different values of X_i. The null hypothesis is then simply

$$H_0 : \mathbb{E}[m_{\mathbf{x},\mathbf{x}'}(W_i)] \leq 0 \text{ for all } \mathbf{x} \in E, \mathbf{x}' \notin E \tag{4}$$

Since $\mathcal{X}$ is finite, this represents a finite set of moment inequality restrictions, and thus fits within the moment inequality testing framework of Canay and Shaikh (2016). We follow approaches they describe to construct a statistical test. Specifically, we use the T^{max} test statistic from Canay and Shaikh (2016) and conduct inference using the two-step procedure in Romano et al. (2014). To construct the test statistic, define the empirical analogue to $\mathbb{E}[m_{\mathbf{x},\mathbf{x}'}(W_i)]$ as

$$\overline{m}_{n,\mathbf{x},\mathbf{x}'} \equiv \mathbb{E}_{\hat{P}_n}[m_{\mathbf{x},\mathbf{x}'}(W_i)] \tag{5}$$

where $\hat{P}_n$ is the empirical distribution of W_i., and define its empirical standard deviation as

$$\hat{\sigma}_{n,\mathbf{x},\mathbf{x}'} \equiv \sqrt{\mathrm{Var}_{\hat{P}_n}[m_{\mathbf{x},\mathbf{x}'}(W_i)]} \tag{6}$$

Then define the test statistic

$$T_n^{max} = \max\left\{\max_{\mathbf{x}\in E,\mathbf{x}'\notin E} \frac{\sqrt{n}\overline{m}_{n,\mathbf{x},\mathbf{x}'}}{\hat{\sigma}_{n,\mathbf{x},\mathbf{x}'}}, 0\right\} \tag{7}$$

This test statistic and procedure offer the advantage of behaving well and being computationally tractable for a large number of moment inequalities. Following Romano et al. (2014), we set $\beta = \alpha/10$ and bootstrap the first stage 499 times.

2.4 Social welfare function analysis

If we reject the null hypothesis that targeting the bottom $N\%$ leads to the highest ATE, it may still be that targeting the bottom $N\%$ maximizes some social welfare function which weights gains to those with low baseline value of the outcome more heavily. We therefore next examine what SWFs would rationalize targeting the lowest $N\%$. We do this in two ways.

First, we ask for all outcomes how much a social planner would need to over-weight the outcomes of deprived households to prefer targeting them, despite larger impacts for other less-deprived households. Specifically, suppose we have identified a subset $E' \in \mathcal{X}$ of the same size as E within which the average treatment effect is higher than that within E. Define their disjoint components as $\tilde{E} \equiv E \setminus E'$ and $\tilde{E}' \equiv E' \setminus E$, respectively. Necessarily, the average treatment effect in $\tilde{E}'$ is strictly greater than that in $\tilde{E}$. We calculate by how much more a social planner would need to weight outcomes gains for households in $\tilde{E}$ than those in $\tilde{E}'$ to nevertheless prefer to make the former eligible, calculating

$$\frac{\max_{\mathbf{x}'\in\tilde{E}'} \mathbb{E}_{\hat{P}_n}[Y_i^1 - Y_i^0|\mathbf{X}_i = \mathbf{x}']}{\min_{\mathbf{x}\in\tilde{E}} \mathbb{E}_{\hat{P}_n}[Y_i^1 - Y_i^0|\mathbf{X}_i = \mathbf{x}]} \tag{8}$$

For instance, if the highest ATE in $\tilde{E}'$ is 2 times larger than the smallest ATE in $\tilde{E}$, then a planner must weight the corresponding group in $\tilde{E}$ twice as highly as the corresponding group in $\tilde{E}'$ to rationalize deprivation-targeting. We will also report the equivalent of (8) with the max and min operators replaced by the mean, which yields the average amount by which the planner would need to over-weight group $\tilde{E}$ relative to $\tilde{E}'$.

Second, for expenditure specifically (and household income, which can be thought of as an alternative proxy for consumption) we also estimate for what utility functions a utilitarian social planner would choose bottom-$N\%$ targeting. Specifically, we consider the following parametrized families of utility function:

$$\text{(CRRA)} \qquad u(y_i) = \frac{y_i^{1-\rho}}{1-\rho} \tag{9}$$

$$\text{(CARA)} \qquad u(y_i) = -e^{-\alpha_{CRRA} y_i} \tag{10}$$

$$\text{(Inequity Aversion)} \quad u(y_i) = y_i - \alpha_{IA} \frac{1}{n-1} \sum_{j \neq i} \max[y_j - y_i, 0] - \beta \frac{1}{n-1} \sum_{j \neq i} \max[y_i - y_j, 0]. \tag{11}$$

CRRA (Equation 9) and CARA (Equation 10) utility functions capture the idea of diminishing marginal returns in a standard way; this could reflect either individual risk-aversion or a social planners taste for equality. Inequity-aversion preferences (Equation 11) are those proposed by Fehr and Schmidt (1999) with α and β fixed; one could thus think of the exercise as identifying the parameters for which a median voter would prefer bottom-N% targeting. For the sake of exposition let θ represent the parameter(s) of an abstract utility function.

For each utility function, we then define a range of parameter values that spans those commonly found in the relevant literature, and define a set of grid points that evenly divide this space. Let θ be one of these values for an abstract utility function parameter. For each of these grid points θ, we then estimate a transformed outcome $\tilde{Y}_i = u(Y_i, \theta)$. We then calculate the average treatment effect on this transformed outcome assuming targeting rule E targeting the bottom N%, and alternatively the average effect assuming the alternative targeting rule E'. We record and report for this value of θ which targeting rule yields a higher average effect on the transformed outcome.

The parameter ranges consider will include

- $\rho : [0.5, 1]$. Estimates of ρ in Kenya have been found between 0.5 and 0.98 using the participant pool at the Busara Center for Behavioral Economics (Balakrishnan et al., 2015).
- $\alpha_{CRRA} : [0, 0.0005]$. Balakrishnan et al. (2015) estimates a value of α_{CRRA} of 0.0001.
- $\alpha_{IA} : [0, 4]$. This is the range considered in Fehr and Schmidt (1999).
- $\beta : [0, 0.6]$. This is the range considered in Fehr and Schmidt (1999).

If we find that either rule E or E' is uniformly preferred within one of these ranges we will expand the range of the search to identify indifference values of θ, and compare these to the values found in the literature.

3 Variable and null hypothesis selection

We select outcome variables and null hypotheses rules to be broadly consistent with the stated objectives and targeting principles of many real-world programs, keeping in mind that there is great variation in these parameters across programs and we are unlikely to exactly match any particular program.

3.1 Selecting outcome variables and null hypotheses

We focus on four outcome variables:

1. **Total consumption expenditure in last 12 months**:
2. **Total household income in the last 12 months**
3. **Total value of non-land assets**
4. **Food security index**

The construction of these variables is defined in the appendix.

For each outcome, we define the null hypothesis as targeting households with low baseline values of that outcome. We make one exception to this rule: because we did not measure household expenditure at baseline, we instead define the null for the expenditure outcome as targeting households with low baseline income levels. We define "low" as being in the bottom 40% of the empirical distribution in our data. We chose this cutoff to reflect the fact that in comparable data from the KLPS, 43% of households meeting GiveDirectly's eligibility criteria (i.e. having homes with thatched roofs) were receiving some form of government assistance, but will also assess sensitivity of our conclusions by re-running analysis at each decile of the empirical distribution.

3.2 Selecting candidate targeting variables

While our alternative hypothesis is very broad ("there is some partition of the type space $\mathcal{X}$ that yields a higher ATE than the partition (E, I)"), it still requires a design choice in that we have to choose the variables of which $\mathcal{X}$ is the product space. For example, if the null hypothesis is that transfers to those with low food security have the biggest impact, then food security must necessarily be one of the dimensions of $\mathcal{X}$, but there are many other dimensions whose inclusion is a choice (occupation of household head, ownership of specific assets, number of children, etc.)

In choosing variables to include as candidates on which to condition targeting, we follow two principles. First, we wish to include variables with a high a priori probability of predicting impacts. For example, based on evidence to date on the impacts of microcredit, we might expect ownership of a business to be predictive of impacts on earnings. Second, we wish to restrict attention to variables on which a government might plausible condition targeting. For example, we might rule out measures of time preference even though these are likely to predict impacts on the grounds that no government has used them to target benefits in the past or seems likely to do so in the future.

To achieve these goals, we begin with the full set of variables that (a) are in our baseline data and (b) have been used to condition eligibility in at least one public-sector program. We will use as partitions the binary values of indicator variables and quartiles of quantitative variables. From among these we select a (possibly weak) subset in two steps. First, we pre-select by hand the following variables which we consider to be of intuitive interest:

- An indicator for a female-headed household
- An indicator for whether the respondent owns land
- The number of children in the household

These variables are so commonly used in assessing eligibility that we would not want to exclude then from the analysis. Second, we select from among the remaining eligible variables all that that increase the adjusted R-squared of a regression of the form

$$Y_i = \alpha + \beta I_i + X_i \gamma + \pi Z_i + \epsilon_i \tag{12}$$

where Y_i is the value of the outcome of interest at baseline, I_i is baseline income, X_i are the pre-selected covariates, and Z_i is the candidate covariate. The intuition for this criterion is that it identifies variables that predict the outcome conditional on income, and thus seem plausibly more likely to predict the extent to which additional income will affect the outcome.

References

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.

Athey, Susan and Guido Imbens, "Recursive partitioning for heterogeneous causal effects," *PNAS*, 2016, *113* (27), 7353–7360.

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela, "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments," September 2015.

Canay, Ivan and Azeem Shaikh, "Practical and theoretical advances in inference for partially identified models," cemmap working paper, Centre for Microdata Methods and Practice CWP05/16, London 2016.

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

Haushofer, Johannes, Edward Miguel, Paul Niehaus, and Michael Walker, "GE Effects of Cash Transfers: Pre-analysis plan for household welfare analysis," July 2017.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Wager, S. and S. Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *ArXiv e-prints*, July 2017.

A Outcome Construction

We construct the following outcomes for the targeting analysis, with references to the endline survey questions used in the construction.

1. **Total consumption expenditure in last 12 months**: sum of total food consumption in last 7 days (12.Q1 for items 12.1-18), frequent purchases in last month (12.19-29) and infrequent purchases in last 12 months (4.4.4, 12.30-38,39b), converted to yearly values. This corresponds to summary measure in 5.2 in Haushofer et al. (2017).

2. **Total household income in the last 12 months**: Sum of total profits from agriculture and livestock in the last 12 months plus total profits from non-ag. business in the last 12 months plus the total after-tax value of wages, salaries and in-kind transfers earned in the last 12 months. These are created as follows:

   (a) *Total profits from agriculture and livestock in the last 12 mo.*: Due to data constraints, we construct this measure slightly differently at baseline and endline.

   At endline, we take the sum of:

- Crop production (7.16) valued at market prices if not reported in monetary units,[2]
- Pastoral output sold (7.6aa)
- Own production of poultry, eggs and beef in the last 12 months, calculated as number of months consuming own production (12.Q5) x monetary value of typical weekly consumption when consuming own production (12.Q6) x 4 for chicken/duck/poultry, beef and eggs)

Net of:

- salaries paid to workers outside the household (7.12)
- agricultural inputs (spending on tools, fertilizer, irrigation, animal medicine, improved/hybrid seeds, agricultural insurance) (7.13.a-f)
- the rental cost of land for agricultural purposes (acres rented for agriculture (6.8.b) times months rented (6.8.c) times monthly rent (6.8.d)).[3]

Due to time constraints in our baseline survey, we did not collect information on crop-by-crop production or own consumption of poultry, eggs and beef. To construct baseline measures, we generate predicted a) crop profits, b) poultry profits and c) livestock profit as a flexible function of quartic polynomials of:

i. amount of crop/poultry/livestock sold
ii. input costs (sum of spending on tools, fertilizer, irrigation, animal medicine, improved/hybrid seeds, agricultural insurance)
iii. (crops only) land used in agriculture
iv. (crops only) amount paid on land rental for agriculture
v. number of workers working in each agricultural activity in the last 12 months

The listed variables were all collected at baseline. We use out-of-sample data from households surveyed as part of the Kenya Life Panel Survey, Round 3 (KLPS-3) located in Busia and Siaya counties to estimate the relationship between these variables, as the KLPS-3 survey collected all of these variables. We then generate predicted agricultural and livestock profits for households at baseline in our sample.

(b) *Total profits from non-ag. business in the last 12 mo.*: Sum of self-reported profits

[2] Commodity prices obtained from market surveys. We use the price of the crop at the nearest market over the course of the endline survey. For any crops that are not included in our market survey, we use the median unit price for households within the same sublocation.

[3] We do not subtract off unpaid labor. It is not clear that respondents are doing this when we ask about self-employment profits, so we do not subtract from ag. profits for consistency.

(8.11b) for all businesses owned.[4]

(c) *Total after-tax value of wages, salaries and in-kind transfers earned last 12 mo.*: Sum of cash salary last month (9.10) plus total value of benefits last month (9.12) net of income tax paid last month (9.11), annualized to a yearly measure by multiplying by the number of months worked in last 12 months (calculated as month of survey minus employment start date (9.3) if employment working patterns (9.7) are full-time or part-time, and based on 9.7a if employment working patterns are seasonal.) for all workers in the household. We set this to zero if no household members are working for wages.[5]

This corresponds to summary measure in 5.3 in Haushofer et al. (2017).

3. **Total value of non-land assets**: sum of value of asset variables 6.13.a-z, 6.13.aa-hh, and value of loans given (10.8.b) net of total amount of loans taken (sum of variables 10.3.d, 10.4.a, 10.5.a, 10.6.a, and 10.7.b). This corresponds to summary measure 5.1 in Haushofer et al. (2017).

4. **Food security index**: weighted and standardized index (following Anderson (2008)) of the following food security outcomes, appropriately signed so that higher values represent greater food security:

   (a) *No. of days adults in household skipped meals or cut the amount of meals in the last 7 days*: Variable 11.11a.

   (b) *No. of days children in household skipped meals or cut the amount of meals in the last 7 days*: Variable 11.11b.

   (c) *No. of days adults in household gone entire days without food in the last 7 days*: Variable 11.12a.

   (d) *No. of days children in household gone entire days without food in the last 7 days*: Variable 11.12b.

   (e) *No. of days children in household gone to bed hungry in the last 7 days*: Variable 11.10b.

[4] We may also analyze this variable as a monthly measure using 8.11a. Our primary measure will use self-reported profits. We will also construct a measure where we replace missing values of self-reported profits with self-employment revenue (8.7b) net of total costs (sum of 8.15, 8.16.a-h, and 8.17.a-g times 12 plus annualized business license costs (8.8a times 12 divided by 8.8b, the number of months license valid) for all businesses owned. If we find high levels of missing values for 8.11b we may replace our primary measure with this constructed measure.

[5] We may also analyze this variable as a monthly measure.

(f) *No. of days adults in household gone to bed hungry in the last 7 days*: Variable 11.10a.

(g) *No. of meals eaten yesterday that included meat, fish, or eggs*: Sum of variable 11.2 and 11.3.

This corresponds to summary measure 5.10 in Haushofer et al. (2017), which can be referenced for full details on the index construction.

We winsorize the top 1% of monetary outcomes bounded below by zero and the top and bottom 1% of unbounded monetary outcomes. As a robustness check, we may conduct the analysis with raw and trimmed values as well.