I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18259

# Using the MVPF to Allocate Treatment Under Imperfect Compliance and Supply-Side Constraints

Antonio Dalla-Zuanna
Kai Liu

NOVEMBER 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18259

# Using the MVPF to Allocate Treatment Under Imperfect Compliance and Supply-Side Constraints

**Antonio Dalla-Zuanna**
*Bank of Italy and IZA*

**Kai Liu**
*University of Cambridge and IZA*

NOVEMBER 2025



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## Using the MVPF to Allocate Treatment Under Imperfect Compliance and Supply-Side Constraints*

This paper shows how the Marginal Value of Public Funds (MVPF) can guide treatment allocation to improve social welfare. Under budget constraints, the optimal treatment targets individuals with MVPFs above a threshold that minimizes the opportunity cost of treatment. Using experimental data, we show that prioritizing high-MVPF groups under tight budgets can double Head Start's social benefits compared to random assignment. Analyzing joint allocation across early (Head Start) and late (Job Corps) skill investment programs, we find that exclusive investment in early interventions is not optimal unless substantially higher welfare weights are placed on young children.

**JEL Classification:** H43, I38, D61, J68, I26

**Keywords:** Marginal Value of Public Funds (MVPF), treatment allocation, budget constraints, welfare maximization

**Corresponding author:**
Antonio Dalla-Zuanna
Bank of Italy
Via Nazionale, 91
00184 Roma
Italy
E-mail: antonio.dalla.zuanna@gmail.com

* We thank Henrik Kleven and Shania Mustika for helpful discussions and comments. The views expressed in the article are those of the authors only and do not involve the responsibility of the Bank of Italy or the Eurosystem.

# 1 Introduction

The Marginal Value of Public Funds (MVPF) is the ratio of the marginal benefit of the policy to the net marginal cost to the government of the policy. MVPF has been estimated in a large number of empirical applications that span different fields of economics (Hendren and Sprung-Keyser, 2020).[1] One attractive feature of the MVPF is that one can compare "bang for the buck" across policies that affect the same or different groups of individuals. For instance, an MVPF of, say, 1.5 means that every $1 of net government spending provides $1.50 of benefits to the beneficiaries of the policy $A$. All else being equal, this policy is preferred to another policy $B$ with an MVPF of 1 because we can increase social welfare by transferring resources from $B$ to $A$.

So far, MVPF has been used to quantify the trade-off embodied in the policy decision of whether resources should be reallocated from one policy to another, or from one subgroup of beneficiaries to another within the same policy (Finkelstein and Hendren, 2020). Although characterizing this trade-off is informative for policy (as in the example above), it remains unclear *how* a policymaker should use MVPFs to guide resource allocation decisions, particularly when multiple groups and policies are involved. For example, now suppose there is a third policy $C$ with an MVPF of 2. Should the government transfer all resources from $A$ and $B$ to $C$? Or could it be optimal to keep some resources to $A$? The answer will likely depend on the size of the government budget, among other things. In fact, if the government budget is unlimited, then all three policies should be implemented because they improve welfare. Ultimately, the goal of the policymaker is to determine the entire set of treatment groups/policies upon which (limited) government resources should be directed.

In this paper, we provide a methodological framework that uses the MVPF as a metric to determine optimal policy decision rules. Specifically, we consider a welfare-maximizing policymaker choosing to offer a program to different subgroups of individuals under imperfect compliance and a government budget constraint. We show that optimal treatment assignments depend on the distribution of MVPF

[1] For example, the MVPF has been recently used to evaluate training programs (the Civilian Conservation Corps in Aizer et al., 2024), early education programs (pre-K programs in Kline and Walters, 2016; Cascio, 2023), means tested transfer programs (SNAP in Cook and East, 2023), health insurance coverage (Medicaid in Finkelstein et al., 2019), algorithmic interventions (Ludwig et al., 2024), and a broad range of climate-related policies (Hahn et al., 2024).

over subgroups and/or policies. The optimal decision is to offer the program (treatment) to those with MVPF above a cost-effectiveness threshold. The threshold captures the marginal *opportunity cost* of treatment, which is endogenously determined by the available resources. A tight budget implies a high threshold, meaning that the program should be offered to only subgroups with a high MVPF. A lower threshold follows from a more generous budget, in which case the policymaker should expand the program to include individuals with a smaller MVPF.

A practical implication of our results is that the MVPF benchmark may not be 1 or any fixed number without explicitly specifying the government budget constraint. In fact, regardless of budget, the only groups to which treatment should never be assigned are those with a negative marginal benefit of receiving treatment. Therefore, the relevance of the MVPF for policy depends on the size of the budget constraint. Our framework can deliver the welfare-maximizing MVPF benchmark, for any policies involving the allocation of single or multiple treatments.

We apply our framework to assess the optimal allocation of two major skill investment programs for disadvantaged young people in the United States: the Head Start program, the largest preschool education program for age 3–5 children from disadvantaged families, and the Job Corps program, the largest vocationally focused education and training program for disadvantaged youth. We use data from randomized experimental evaluations of each program, the Head Start Impact Study (HSIS) and the National Job Corps Studies (NJCS). In both studies, the applicants were randomly assigned to either a treatment group where individuals were offered the opportunity to enroll in the focal program, or a control group where individuals had no access to the program during a period of embargo. Exploiting the experimental variation is crucial in that a market-level change in offer probabilities can be inferred from an individual-level randomized trial with a fixed offer probability (Kline and Walters, 2016). It also implies that we can derive MVPF from a program expansion that is invariant to the scale of policy change and hence the metric is relevant for a non-marginal change in the policy.[2]

We first consider the optimal treatment assignment problem for each program separately, exploiting

[2]We acknowledge that, even in experimental settings, this statement may be violated if a large reform entails changes to program features (such as accessibility) or program take-up mechanism (such as peer effects), which we have abstracted from in this paper.

heterogeneity in the MVPF across subgroups of individuals within each program. The subgroups are defined from predetermined characteristics that are easily observed by the policymaker. Our basic approach constructs subgroups by drawing upon existing empirical evidence on the heterogeneity of treatment effects in our applications. We also use a machine-learning approach, where the groups are constructed through recursive partitioning based on individual treatment effects estimated using the generalized random forest (GRF) estimator of Athey et al. (2019). Considering that both programs already target disadvantaged individuals, we do not consider subgroup-specific social welfare weights.[3] We find that for tight budgets only groups with an MVPF above 1 are treated, hence both policies deliver "bang for the buck". However, as the budget constraint relaxes, also groups with MVPF below 1 are included because the policies are welfare improving for them. For the few groups of a negative MVPF, the optimal strategy is not to target them regardless of the budget constraint.

Under optimal treatment assignment, we show that social benefit is a concave function of program costs – the slope of the benefit-cost curve is the MVPF of the marginal group to be treated. With a tight budget, the optimal strategy involves treating groups with higher MVPF. As the budget constraint is relaxed, treatment is extended to groups with progressively lower MVPF, resulting in a diminishing slope of the benefits-cost curve. By comparison, social benefit is linear in program cost with random assignment, as expanding the program draws a random set of individuals into the program. By prioritizing groups with markedly high MVPF, optimal targeting generates substantially greater social benefits, particularly under stringent budget constraints. For instance, the benefits derived from treating the five groups with the highest MVPF in the Job Corps are approximately 1.5 times greater than those obtained by allocating the same resources to randomly selected individuals. Put it in another way, the cost of treating these five groups is approximately 260 mln $; achieving equivalent benefits through random assignment would require extending treatment to nearly the entire eligible population, at a cost exceeding twice as much (681 mln $). Using the machine-learning approach to define subgroups can generate even larger welfare gains than the basic manual grouping approach, primarily when the bud-

[3]Job Corps targets young individuals (between 16 and 24 years of age) who are receiving welfare or food stamps or have an income less than 70% below the DOL's "lower living standard income level". Similarly, Head Start is targeted to poor families (with income below the federal poverty line), although agencies are allowed to admit up to 10% of children from wealthier families. There is no fee for attending Head Start or Job Corps.

get is tight. This is because the recursive partitioning based on the predicted individual-level treatment effect is capable of uncovering subpopulations with high program returns.

We highlight the role of imperfect compliance in driving policy decisions and welfare. On the one hand, non-compliance to the treatment (never-takers) frees up the budget constraint, meaning that policymakers can over-target relative to the constraint. On the other hand, crossovers from the control group to the treatment group (always-takers) reduce the effective budget available to policymakers. In further analysis, we also consider control group substitution due to the existence of close alternatives, where many control group individuals end up receiving similar services to the focal program (Heckman et al., 2000; Kline and Walters, 2016; Chan et al., 2024). Expanding the focal program not only generates a direct program provision cost, but also induces savings (positive fiscal externalities) from reduced participation in close alternatives. We find that these savings may enable the policymaker to expand the program offering, achieving the same level of social benefit with a reduced budget. Accounting for fiscal externalities can also alter the distribution of MVPFs and shift the composition of target subgroups - particularly when the degree of control group substitution varies substantially across groups.

Lastly, we combine the two datasets to study the optimal assignment of the two programs jointly, subject to an aggregate budget constraint that specifies the budget ceiling of the two programs combined. Here the planner faces an additional tradeoff of assigning different types of treatment: expanding one program may come with an opportunity cost of having to reduce the size of the other program. Although Job Corps has smaller MVPF than Head Start on average, we show that an optimal allocation involves offering both programs to different subgroups of the population, even when the budget is at low and intermediate levels. This results from MVPF heterogeneity; some groups of potential Job Corps participants exhibit especially high returns relative to certain groups of potential Head Start participants. Advocates of early childhood investment have argued that government resources should re-allocate to early childhood skill investment programs because they generate larger returns than skill investment programs at later ages. We find that such a reallocation would be justified only if the planner assigns a social welfare weight to Head Start participants that is at least 2.7 times that of Job Corps participants.

This paper contributes to the empirical MVPF literature (Mayshar, 1990; Hendren and Sprung-Keyser, 2020; Finkelstein and Hendren, 2020), which interprets MVPF of different programs/subgroups in terms of implicit but intuitive trade-offs. As already mentioned above, we make these tradeoffs explicit and derive an MVPF threshold that arises endogenously from a program allocation problem of the policymaker under budget constraints. More broadly, our model can inform the full set of subgroups and/or programs that maximize social welfare under *any* pre-specified budget. A recent critique emphasizes that the MVPF evaluates programs for a fixed government budget, whereas traditional cost-benefit analysis also considers the benefits of expanding the size of the budget to finance new programs (García and Heckman, 2022). We assume that the policymaker is constrained by binding budget ceilings, but the size of the budget can vary. Indeed, our optimal policy decision rule is a (nonlinear) function of the size of the budget, from which we can characterize the social welfare function as the scale of the budget expands. That being said, we are agnostic about the revenue sources that finance a budget expansion and hence ignore any distortion from the expansion or contraction of the government budget (such as distortions from taxation if the budget expansion is financed by a tax rise; see Kleven and Kreiner, 2006).

The statistical problem analyzed in this paper relates to the literature on treatment assignment, in particular the plug-in approach that uses conditional treatment effects estimated from regression models.[4] Most related to us is Bhattacharya and Dupas (2012), who consider the problem of allocating a binary treatment based on observed covariates, subject to a capacity constraint that limits the fraction of the population that can be treated. An important way the current paper differentiates itself from Bhattacharya and Dupas (2012) is by deriving and interpreting the policy rule through the MVPF lens. Although both papers consider similar constrained optimization problems of treatment assignment, our framework allows treatment costs to vary across individuals due to imperfect compliance and heteroge-

[4]Manski (2004) assesses the welfare properties of statistical treatment rules by their maximum regret. He proposes Conditional Empirical Success rules, which assign persons to treatments that yield the best experimental outcomes conditional on alternative subsets of the observed covariates. Kitagawa and Tetenov (2018) propose a Empirical Welfare Maximization (EWM) rule, that can directly select an eligibility policy from a set of available policies based on the experimental data.

neous program take-up rates—features that are central to our empirical applications.[5] By comparison, Bhattacharya and Dupas (2012) assume perfect compliance for all individuals (100% take-up in the treatment group and 0% in the control group), which implies homogeneous program take-up and uniform treatment costs. Under constant treatment costs, ranking individuals by the expected benefit per dollar spent is equivalent to ranking them by their expected benefit from the program alone—the criterion used in Bhattacharya and Dupas (2012). However, when treatment costs are heterogeneous, the planner optimally allocates limited resources by prioritizing individuals for whom the *per-dollar*-spent benefit is highest. This distinction is essential for establishing the connection between the optimal treatment assignment rule and the MVPF defined from group-level changes in offer probabilities. We further extend the analysis to account for fiscal externalities arising from program substitution and to study the assignment problem across multiple programs, where the policymaker faces a binding budget constraint rather than a simple capacity constraint.

Empirical estimates of treatment effect heterogeneity are of particular interest to policymakers seeking to target policies on those most likely to benefit. A separate literature has investigated treatment targeting (profiling) as a means of allocating government services to individuals (Berger et al., 2001; Dehejia, 2005; Lechner and Smith, 2007; Frölich, 2008).

The paper is organized as follows. Section 2 formulates the treatment assignment problem and derives the optimal policy rules. Section 3 describes the HSIS and the NJCS samples, and presents descriptive evidence of compliance rates, and experimental impacts. Section 4 reports the results. Section 5 concludes.

# 2 Model: Optimal Treatment Assignment

Section 2.1 describes a basic treatment assignment problem, where the policymaker targets a single program to different groups of individuals under imperfect compliance and a government budget constraint. We then extend this basic problem in two directions. Section 2.2 accounts for fiscal exter-

[5] Imperfect compliance also changes the aggregate budget constraint (e.g., the control group also incurs positive program costs) and the opportunity cost of offering treatment to an additional individual.

nalities generated by program substitution, and Section 2.3 analyzes the problem of assigning multiple treatments that target different groups in the population.

## 2.1 Basic Set-up

Consider a social planner who chooses the probability of offering a treatment to an individual $i$ with characteristic $X_i = x$. The treatment, denoted by $Z_i$, is equal to 1 if an individual $i$ receives an offer to *access* the focal program (and hence in the treatment group) and 0 otherwise (in the control group). The value of $Z_i$ determines the assignment of the treatment, which can be directly affected by the planner. Program enrollment, on the other hand, is an individual's decision, and hence there is imperfect compliance to the treatment. Let $\mathsf{d}_i$ denote a binary program choice, where $\mathsf{d}_i = 1$ if the individual $i$ enrolls in the focal program and 0 otherwise. We can think of $Z_i$ as an instrument that alters the relative cost of enrolling in the focal program. This gives a potential program choice for each individual if the instrument were externally set to $Z_i = z$, denoted by $\mathsf{d}_i^z$.

Throughout the paper, we assume that receiving a program offer weakly increases enrollment in the focal program, commonly referred to as the monotonicity assumption in the literature (Angrist et al., 1996). The monotonicity assumption seems natural in our empirical applications, ruling out the possibility that receiving an offer to access the program could reduce program enrollment for any individual.

**Assumption 1.** *Monotonicity:* $\mathsf{d}_i^1 \geq \mathsf{d}_i^0,\ \forall i.$

Let $\delta(x)$ be the probability of assigning treatment ($Z_i = 1$) to individuals with $X_i = x$. Normalizing the population size to 1 and letting $F(x)$ denote the marginal c.d.f of $X_i$, the objective of the planner is to choose a targeting rule $\delta : X \to [0, 1]$ in order to

$$\max_{\delta(\cdot)} \int \delta(x)E(Y_i \mid Z_i = 1, X_i = x) + (1 - \delta(x))E(Y_i \mid Z_i = 0, X_i = x)dF(x) \tag{1}$$

subject to a resource constraint

$$\int \delta(x)C(Z_i=1,X_i=x)+(1-\delta(x))C(Z_i=0,X_i=x)dF(x) \leq R \tag{2}$$

The objective function is the social benefit of the program, defined in terms of the mean outcome in the entire population $E[Y_i]$.[6] The resource constraint (equation (2)) states that the total cost of offering this program cannot exceed an exogenous fiscal requirement, denoted by $R$.

Let $\phi$ be the cost of administrating the program to an additional individual, which is assumed independent of individual characteristics.[7] For any given individual of $Z_i = z$ and $X_i = x$, the conditional cost function is

$$C(z,x) = \phi P(\mathsf{d}_i = 1 \mid z, x).$$

$C(z,x)$ must weakly increase in the treatment status $z$, because offering the program to an additional person (weakly) increases the cost of the program (due to the monotonicity assumption). Fixing $Z_i = z$, $C(z,x)$ differs by $x$ because program take-up rates may vary by individual.

Dividing equation (2) by $\phi$, the resource constraint can be rewritten as a capacity constraint:

$$\int \left[\delta(x)\left(P(\mathsf{d}_i=1|Z_i=1,X_i=x)-P(\mathsf{d}_i=1|Z_i=0,X_i=x)\right)+P(\mathsf{d}_i=1|Z_i=0,X_i=x)\right]dF(x) \leq \tilde{R} \tag{3}$$

where $\tilde{R} = \frac{R}{\phi}$ is the maximum proportion in the population that can be allowed to enroll in the focal program. Note that $P(\mathsf{d}_i = 1|Z_i = 0, X_i = x)$ captures the share of always-takers among subpopulation $x$, who always enroll in the program regardless of the assigned value of $Z_i$. Since

[6] In the basic set-up we have assumed that the social welfare weights are equal across individuals. In our empirical applications, the individuals who are eligible for treatment are relatively homogeneous (e.g., children and youth from disadvantaged backgrounds). We allow the social welfare weight to differ by individual when we extend the model to multiple treatments in Section 2.3.

[7] This assumption can be relaxed to allow the cost to vary by $x$. We focus on the variable cost and ignore the fixed cost of setting up the program because we can always adjust the external expenditure requirement ($R$) to accommodate any fixed cost.

always-takers take up resources even when not assigned to the treatment group, the *effective* capacity constraint excluding always-takers is $\hat{R} \equiv \tilde{R} - P(\mathsf{d}_i = 1|Z_i = 0)$. We assume that $\hat{R} > 0$, meaning that the program has excess capacity (resources) after admitting the always-takers.

Denote the conditional intention-to-treat effect by $\Delta_Y(x) \equiv E(Y_i \mid Z_i = 1, X_i = x) - E(Y_i \mid Z_i = 0, X_i = x)$; $\Delta_Y(x)$ measures the average benefit of expanding the program offer to individuals of $X_i = x$. Let $\Delta_C(x) \equiv C(Z_i = 1, X_i = x) - C(Z_i = 0, X_i = x)$ be the mean cost change from the assignment of the treatment. To ensure that the optimization problem is well posed and admits a meaningful solution, we impose that $\Delta_Y(x)$ should be bounded and not identical almost everywhere with respect to $F(x)$, and that $f(x)$ is bounded away from zero over the support of $\Delta_Y(x)$. Under these assumptions, the optimal targeting rule that solves the planner's problem described by equations (1) and (2) is

$$\delta^*(x) = \begin{cases} 1 \text{ if } \frac{E(Y_i \mid Z_i=1, X_i=x) - E(Y_i \mid Z_i=0, X_i=x)}{C(Z_i=1, X_i=x) - C(Z_i=0, X_i=x)} \geq \lambda \\ 0 \text{ elsewhere} \end{cases} \tag{4}$$

where $\lambda$ satisfies

$$\int \left[\delta^*(x) \left(C(Z_i = 1, X_i = x) - C(Z_i = 0, X_i = x)\right) + C(Z_i = 0, X_i = x)\right] dF(x) = R \tag{5}$$

and

$$\lambda \geq 0$$

**Discussion.** Equation (4) states that the optimal treatment assignment depends on two quantities: the average gain in social benefit ($\Delta_Y(x)$) relative to the average increase in the cost of assigning the treatment to a given subpopulation ($\Delta_C(x)$), and a minimum target threshold $\lambda$. The parameter $\lambda$ is determined *endogenously* by equation (5): it is the Lagrangian multiplier that balances the benefit of the program with the cost and ensures that the total cost does not exceed the budget. A small budget implies a high $\lambda$, meaning that only individuals with sufficiently large per dollar benefits from the treatment should be assigned to the treatment. A more generous budget is associated with a lower $\lambda$, allowing people with small per dollar benefits to access the program.

Because expanding the program does not decrease the total cost of the program, it is never optimal to target people with $\Delta_Y(x) < 0$. If the fiscal requirement $R$ is so large that the budget constraint becomes non-binding, then $\lambda = 0$ and the treatment should be extended to any individual as long as doing so improves social welfare. By comparison, when the constraint is binding, there is an "opportunity cost" of extending the treatment to an extra person – the opportunity cost being the foregone social benefit if the same resources were redirected to another untreated individual. The target threshold $\lambda$ reveals the marginal opportunity cost of treatment at a fixed budget level. By prioritizing individuals who yield the greatest benefit per dollar spent, the planner minimizes the opportunity cost of treatment given the budget constraint.

An alternative way of characterizing the planner's problem is to define the social welfare as the mean outcome net of program cost (if both are measured in monetary terms), without an explicit budget constraint. This formulation of the problem entails the planner to offer treatment based on $\Delta_Y(x) - \Delta_C(x)$, that is, if the individual's expected gain net of cost is high. If $\Delta_C(x)$ is constant across individuals, ranking by $\Delta_Y(x) - \Delta_C(x)$ is the same as ranking by $\Delta_Y(x)/\Delta_C(x)$. When $\Delta_C(x)$ vary, the two rules may differ from each other: net-benefit ranking may pick high $\Delta_Y(x)$ and high-$\Delta_C(x)$ individuals, while the budget-constrained planner wants to maximize total benefit given the resource available and hence favors individuals where per-dollar-spend yields the most benefit.[8]

**Linking the targeting rule to the MVPF.** Let $Y_i^{\mathsf{d}}$ denote the *potential* outcome if the individual $i$ is externally assigned to program $\mathsf{d}$. This notation incorporates the usual exclusion restriction that the causal effect of the treatment on $Y_i$ is only due to its effect on $\mathsf{d}_i$ (Angrist et al., 1996). Given that $Z_i$ is randomly assigned, $Z_i \perp (Y_i^0, Y_i^1, \mathsf{d}_i^0, \mathsf{d}_i^1)$.

[8] As an example, suppose that $\Delta_C(x)$ exceeds $\Delta_Y(x)$ for group $x$ and that $\Delta_Y(x) > 0$. Under the targeting rule based on $\Delta_Y(x) - \Delta_C(x)$, it is never optimal to offer treatment to this group even when the net gain from the treatment is positive for this group of individuals. In our setting, it will be optimal to target these individual under a generous budget.

Consider a marginal increase in $\delta(x)$, the probability of offering the treatment to a subgroup of individuals with $X_i = x$. An increase in $\delta(x)$ means that a random set of individuals in the subgroup who did not receive the treatment is now included in the treatment group. Under the monotonicity assumption above (assumption 1), this leads to additional individuals being drawn into the program. The marginal social benefit of program expansion in the subgroup $x$ is

$$MB(x) =: \frac{\partial E[Y_i|X_i = x]}{\partial \delta(x)} = \underbrace{E[Y_i^1 - Y_i^0 \mid \mathsf{d}_i^1 \neq \mathsf{d}_i^0, X_i = x]}_{\equiv LATE(x)} \underbrace{P(\mathsf{d}_i^1 \neq \mathsf{d}_i^0 \mid X_i = x)}_{\equiv \pi(x)} \tag{6}$$

which equals the average impact of the program on compliers ($LATE(x)$) times the measure of compliers ($\pi(x)$). This result, shown in Kline and Walters (2016) and reproduced in Appendix Section A, follows from the assumption that program offers are randomly assigned and excludable from potential outcomes.

The marginal cost of expanding the program for the subgroup $x$ is:

$$MC(x) =: \frac{\partial C(x)}{\partial \delta(x)} = \phi \frac{\partial P(\mathsf{d}_i = 1|X_i = x)}{\partial \delta(x)} = \phi \underbrace{P(\mathsf{d}_i^1 \neq \mathsf{d}_i^0 \mid X_i = x)}_{\equiv \pi(x)} \tag{7}$$

$MC(x)$ must be nonnegative, because, under the monotonicity assumption, the expansion draws additional individuals to the program ($\pi(x) \geq 0$) and hence increases program cost.

The Marginal Value of Public Funds measures the social value of an extra dollar spent on the program. The MVPF of subgroup $x$ is given by the ratio of the marginal social benefit to the marginal cost defined in equations (6)-(7):

$$MVPF(x) = \frac{MB(x)}{MC(x)} = \frac{LATE(x)}{\phi} = \frac{E(Y_i \mid Z_i = 1, X_i = x) - E(Y_i \mid Z_i = 0, X_i = x)}{C(Z_i = 1, X_i = x) - C(Z_i = 0, X_i = x)} \tag{8}$$

where the last equality uses the fact that $LATE(x)$ can be estimated by a Wald estimator using $Z_i$ as an instrument for $\mathsf{d}_i$: $LATE(x) = \frac{E(Y_i|Z_i=1,X_i=x)-E(Y_i|Z_i=0,X_i=x)}{P(\mathsf{d}_i=1|Z_i=1,X_i=x)-P(\mathsf{d}_i=1|Z_i=0,X_i=x)}$.

A comparison of equation (8) with (4) shows that the group-specific MVPF is the metric to infer optimal treatment assignment in the broader population:

$$\delta^*(x) = \begin{cases} 1 \text{ if } MVPF(x) \geq \lambda \\ 0 \text{ elsewhere} \end{cases}$$

Although $MVPF(x)$ is derived from a marginal change in the offer probability $\delta(x)$, the composition of compliers and $LATE(x)$ does not vary with $\delta(x)$. Therefore, the effects of a group-level change in offer probabilities (the MVPF) are equivalent to the effects of an individual-level randomized trial with a fixed offer probability (the metric in the optimal targeting rule).

## 2.2 Accounting for Fiscal Externalities from Competing Programs

Our analysis so far has assumed that the focal program is the only option for individuals to acquire services. In reality, competing subsidized programs may be available that offer similar services. For instance, Kline and Walters (2016) show that the Head Start program draws roughly a third of its participants from competing preschool programs, many of which receive public funds. Therefore, it is important to consider the cost savings that arise when Head Start draws children away from competing subsidized preschool programs.

Now suppose that an individual faces three program alternatives: the focal program ($h$), a competing program ($c$), and an outside option ($n$). Denote the program choice by $d \in \{h, c, n\}$. The planner continues to maximize the same objective function (as in equation (1)) and chooses the probability of treatment for program $h$. However, the existence of a competing program changes the total cost function, the resource constraint, and, consequently, the targeting threshold.

Following Kline and Walters (2016), we assume that an offer ($Z_i = 1$) only induces certain individuals who would otherwise choose $c$ or $n$ to enroll in $h$. This restriction leads to the following assumption, which extends the monotonicity assumption to a setting with multiple program alternatives:

**Assumption 2.** $\mathsf{d}_i^0 \neq \mathsf{d}_i^1 \implies \mathsf{d}_i^1 = h, \mathsf{d}_i^0 \neq h.$

Assumption 2 states that if individuals change their behavior as a consequence of being assigned to the treatment group, they do so by enrolling in the focal program.[9] Under this assumption, the compliers in the overall population can be partitioned into two mutually exclusive groups, depending on their potential program choices in the absence of the program offer: (1) $c$-compliers, with $\mathsf{d}_i^1 = h, \mathsf{d}_i^0 = c$; (2) $n$-compliers, with $\mathsf{d}_i^1 = h, \mathsf{d}_i^0 = n$.

Let $\phi_h$ and $\phi_c$ be the administration costs of providing the focal and competing program to an additional participant, respectively. For a given individual with $Z_i = z, X_i = x$, the conditional cost function becomes

$$\tilde{C}(z,x) = \phi_h P(\mathsf{d}_i = h|z,x) + \phi_c P(\mathsf{d}_i = c|z,x) \tag{9}$$

where $P(\mathsf{d}_i = h|z,x)$ and $P(\mathsf{d}_i = c|z,x)$ capture the proportions of individuals who have chosen the focal program $h$ and the competing program $c$, respectively. The marginal cost from an increase in $\delta(x)$ is:[10]

$$\begin{aligned}\frac{\partial \tilde{C}(x)}{\partial \delta(x)} &= \phi_h \frac{\partial P(\mathsf{d}_i = h|X_i = x)}{\partial \delta(x)} + \phi_c \frac{\partial P(\mathsf{d}_i = c|X_i = x)}{\partial \delta(x)} \\ &= \phi_h \underbrace{P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 \neq h|X_i = x)}_{\equiv \pi_h(x)} - \phi_c \underbrace{P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 = c|X_i = x)}_{\equiv \pi_c(x)} \end{aligned} \tag{10}$$

Within the subgroup $x$, the effect of an increase in $\delta(x)$ on the enrollment fraction of $h$ is equivalent to the share of compliers who switch to $h$ from $c$ or $n$, and the impact on the proportion of individuals who enroll in $c$ is the share of compliers substituting $c$ for $h$. Both compliance shares can be estimated directly from data.[11] A larger $\pi_c$ implies that the experimental offer induces more individuals to shift from $c$ to $h$ (but not vice versa and not from $c$ to $n$). This reduces the fiscal burden of expanding the focal program, creating a positive fiscal externality.[12]

[9] Similar assumptions are made in Kirkeboen et al. (2016). This assumption extends the monotonicity assumption of Imbens and Angrist (1994) to a setting with multiple counterfactual treatments.

[10] See Appendix A.1 for a discussion of the derivation.

[11] $\pi_h = P(\mathsf{d}_i = h|Z_i = 0) - P(\mathsf{d}_i = h|Z_i = 1)$ and $\pi_c = P(\mathsf{d}_i = c|Z_i = 0) - P(\mathsf{d}_i = c|Z_i = 1)$.

[12] We assume that $\phi_c$ is not too large relative to $\phi_h$ so that targeting an additional person weakly increases the total program cost even after accounting for the fiscal externality. Otherwise, it is possible that expanding the focal program $h$

The marginal social benefit of the program expansion is

$$\frac{\partial E[Y_i|X_i=x]}{\partial \delta} = \underbrace{E[Y_i^h - Y_i^{\mathsf{d}_i^0} \mid \mathsf{d}_i^1 \neq \mathsf{d}_i^0, X_i = x]}_{\equiv LATE_h(x)} \underbrace{P(\mathsf{d}_i^1 \neq \mathsf{d}_i^0 | X_i = x)}_{\equiv \pi_h(x)}, \tag{11}$$

which equals the average effect of enrolling in program $h$ for compliers (relative to a mix of program alternatives) times the measure of compliers in the subgroup. $LATE_h(x) \times \pi_h(x)$ can be recovered directly from the intention-to-treat effect conditional on $x$.

Dividing (11) by (10), the group-specific MVPF is given by

$$MVPF(x) = \frac{\pi_h(x) LATE_h(x)}{\phi_h \pi_h(x) - \phi_c \pi_c(x)} = \frac{E(Y_i \mid Z_i = 1, X_i = x) - E(Y_i \mid Z_i = 0, X_i = x)}{\tilde{C}(Z_i = 1, X_i = x) - \tilde{C}(Z_i = 0, X_i = x)} \tag{12}$$

which shows again that the effects of a group-level change in offer probabilities (the MVPF) are equivalent to the within-group average benefit-to-cost ratio that is used for targeting. Given this result, the optimal targeting rule compares MVPF(x) with a threshold $\tilde{\lambda}$ that satisfies

$$\int \left[\delta^*(x) \left(\tilde{C}(Z_i = 1, X_i = x) - \tilde{C}(Z_i = 0, X_i = x)\right) + \tilde{C}(Z_i = 0, X_i = x)\right] dF(x) = R \tag{13}$$

where $\delta^*(x) = \mathbf{1}(MVPF(x) > \tilde{\lambda})$.

In general, we expect the maximized social welfare to be higher after accounting for fiscal externality from the competing program. By expanding the scale of program $h$, we can reduce the enrollment in the competing program $c$ hence save cost. The savings relax the resource constraint, meaning that additional individuals can be targeted to improve social welfare.

**Fiscal externality from additional tax revenue.** Another type of fiscal externality is the additional tax revenue generated by higher earnings associated with the social benefit of the program. For example, increased earnings and the associated tax revenue are part of Hendren and Sprung-Keyser (2020) MVPF calculations for the Job Corps program. Accounting for tax revenue relaxes the government

pays for itself, leading to an infinite MVPF (Hendren and Sprung-Keyser, 2020).

budget constraint as the additional revenue can be used to extend the program offering to more people. At the same time, by allowing the revenue generated by taxes to reduce the net financing costs to the government, we need to assume that the government is able to borrow against the additional tax revenues. This may not be feasible under borrowing constraints, especially when the additional tax revenues can only be realized years after program assignment (such as subsidized childcare programs). For this reason, we do not take into account the additional tax revenue in our main analysis. Nevertheless, for completeness, in Section 4.3 we show how incorporating fiscal externality from tax revenues can affect the optimal assignment of treatment.

## 2.3 Optimal Assignment of Multiple Treatments

Now suppose that the planner needs to assign two treatments to individuals. Let $Z_{1i} = 1$ if individual $i$ receives an offer to access the first program (and 0 otherwise), and $Z_{2i} = 1$ if the individual receives an offer to access the second program. Assume that program $1$ is for a population with the marginal c.d.f. of characteristics denoted by $F(\cdot)$, and program 2 is for a different population with the marginal c.d.f. denoted by $G(\cdot)$. An example is the assignment of individuals to different skill investment programs, one targeting young children and the other targeting youth. Assume that there are no crossovers between the two programs (due to eligibility rules), but there is imperfect compliance to treatment in each program.

When considering multiple treatments, the planner now faces an additional tradeoff: expanding one program may come with an opportunity cost of having to reduce the size of the other program. Normalizing the total population size to 1, let $p_n$ denote the fraction of the total population eligible for program $n \in \{1, 2\}$. Let $\eta_n$ denote the social welfare weights associated with the eligible population for program $n$, which measures the impact on social welfare from a 1-unit improvement in their mean outcomes. The planner's problem is to choose two targeting probabilities, $\delta : X \to [0, 1]$ and $\gamma : W \to$

$[0, 1]$, in order to maximize the overall social welfare

$$p_1\eta_1 \int (\delta(x)E(Y_i|Z_{1i}=1, X_i=x) + (1-\delta(x))E(Y_i|Z_{1i}=0, X_i=x))\, dF(x)$$
$$+ (1-p_1)\eta_2 \int (\gamma(w)E(Y_i|Z_{2i}=1, W_i=w) + (1-\gamma(w))E(Y_i|Z_{2i}=0, W_i=w))\, dG(w),$$

subject to an aggregate budget constraint

$$p_1 \int (\delta(x)C(Z_{1i}=1, X_i=x) + (1-\delta(x))C(Z_{1i}=0, X_i=x))\, dF(x)$$
$$+ (1-p_1) \int (\gamma(w)C(Z_{2i}=1, W_i=w) + (1-\gamma(w))C(Z_{2i}=0, W_i=w))\, dG(w) \leq R.$$

The solution to the problem is an optimal targeting rule for the first program

$$\delta^*(x) = \begin{cases} 1, & \text{if } \eta_1 MVPF_1(x) \geq \lambda, \\ 0, & \text{otherwise.} \end{cases}$$

and an optimal targeting rule for the second program

$$\gamma^*(w) = \begin{cases} 1, & \text{if } \eta_2 MVPF_2(w) \geq \lambda, \\ 0, & \text{otherwise.} \end{cases}$$

The optimal assignment rule is to compare the impact on social welfare per dollar of government expenditure on subgroup $x$ in program $j$ ($\eta_j \times MVPF_j(x)$) with the threshold $\lambda$ that is endogenously determined by the aggregate budget constraint. The assignment rules imply the welfare-maximizing sizes of different social programs: given an aggregate budget constraint, the optimal size of a program is simply given by the number of people whose weight-adjusted MVPF exceeds the threshold. A high social welfare weight increases the probability of targeting and the size of the program.

# 3 Applications: Background and Data

## 3.1 The Head Start Impact Study

**Background.** Launched in 1965, Head Start is a U.S. federal program that offers year-long care to children between three and five years of age, with the aim of fostering their early reading and math skills to be ready for school. It is funded by the Department of Health and Human Services, and on average it enrolled approximately 900,000 children per year between 2000 and 2020, throughout the U.S. territory (DHHS, 2021). Head Start is administered by local agencies (both public and private), which compete for funding and are required to adhere to national standards. The vast majority of Head Start providers offer center-based care. It is targeted to poor families (with income below the federal poverty line), but agencies are allowed to admit up to ten percent of children from wealthier families. Children who attend Head Start are not charged any fee.

**Data.** We use data from the Head Start Impact Study (HSIS), an experimental evaluation of Head Start's impact on children. In fall 2002, more than 4,000 newly entering children aged 3 and 4 were randomly assigned to a treatment group, which was offered Head Start, or a control group, which had no access to Head Start and were free to select any other type of care available to them.[13] Children are tracked over time, collecting data on the type of childcare they attend and their performance on selected cognitive assessments. Participants with missing information at any time point are excluded from the analysis. The final sample comprises 3,574 children; 63% of them were randomly assigned to the treatment group, and the remaining 37% were assigned to a control group who did not have access to Head Start in that year.[14]

Panel A of Table 1 reports the program take-up rates in our sample. Focusing on the full sample (column 1), 79.8% of the children in the treatment group enrolled in Head Start in the year they received the offer, while the enrollment rate is 13.7% among the children in the control group.[15] This

[13] The randomization was implemented at the center level – at each Head Start center, eligible applicants were randomly assigned to the treatment and control groups.

[14] We provide additional information on the sample we use in Appendix B.

[15] Despite efforts to maintain the integrity of the control group, some children assigned to the control group enrolled

implies an overall compliance share of about 66%(=$\pi_h$). Children who did not choose Head Start had other program options aside from home care, including private care or other federal or state-funded programs. In the control group, more than 1 in 4 children received other center-based care. The Head Start offer reduces the share of children in other center-based care to 6.7%, implying that $19.3\%(=\pi_c)$ of children switch from competing programs to Head Start when receiving an Head Start offer. By substituting Head Start for other center-based care, this group of children reduces the fiscal burden of a Head Start expansion.

Panel B reports the experimental impact of the HSIS on children's test scores one year after random assignment (Spring 2003).[16] Overall, children in the treatment group gained 0.19 of a standard deviation (sd) of the test scores in the control group. Given the proportion of experimental compliers (66%; see above), this implies that Head Start participation increases the test score among compliers by 0.28 sd.

In columns (2)–(4), we investigate the heterogeneity in the proportion of compliers and in the experimental effect by children's baseline skills and language use. The shares of overall compliers are fairly similar across these subgroups, although program substitution between Head Start and other-center care is more common among Spanish speakers and those with high baseline test scores. The experimental effect is larger for children with lower baseline score (consistent with Bitler et al., 2014) and for non-Spanish speakers.

in Head Start in the same program year ("crossovers"). These crossovers occur because local staff intentionally enrolled control group children into Head Start, and more commonly, parents applied to another nearby Head Start program as information on HSIS was not shared with programs not involved in the study (Puma et al., 2010).

[16] Following Kline and Walters (2016), we use the average of the scores obtained in the Woodcock Johnson III (WJIII) test scores and in the Peabody Picture and Vocabulary Test (PPVT) scores. This measure is normalized to have mean zero and variance one in the respective cohort's control group. Similar to Kline and Walters (2016), we control for household size, number of siblings, dummies for whether the child is female, black, hispanic, uses English as home language, is living in urban area, is living with both parents, is in need of special education, is the child of a teen mother, is the child of a mother who never married or is separated, is the child of a mother with high school or more than high school. Our point estimates are similar if these baseline characteristics are excluded from the regression.

## 3.2 The National Job Corps Study

**Background.** Administered by the Department of Labor (DOL), Job Corps is the largest vocationally focused education and training program for disadvantaged youths in the US. The program targets young individuals (between 16 and 24 years of age) who are receiving welfare or food stamps or have an income less than 70% below the DOL's "lower living standard income level".[17] Job Corps offers a comprehensive array of education and training services, including the teaching of academic, vocational, and employability skills. A unique characteristic of Job Corps is its residential nature—in the majority of the training centers (about 87%) participants are required to reside at a center while training. The remaining participants who do not reside at the centers are usually required to stay at the campus for the whole day. Participants are not charged any fee and receive a weekly cash payment, free meals and clothing.

**Data.** We use data from the National Job Corps Studies (NJCS), an experimental evaluation of Job Corps conducted in the mid-1990s. Youths who applied to the program and were eligible for participation between November 1994 and December 1995 were randomized into either a treatment group or a control group.[18] Applicants assigned to the treatment group were able to enroll in Job Corps. Applicants assigned to the control group were excluded from Job Corps for an embargo period of 3 years.

Our sample includes individuals who completed the evaluation survey conducted 48 months after random assignment.[19] 60% of the 11,094 individuals in our sample were assigned to the treatment group. We use the sample weights provided in Schochet et al. (2008) to adjust for the sample design and survey design; Job Corps used to enroll 60,000 new attendees every year when the NJCS was conducted.

[17]Besides age and income, there are 11 criteria that should be satisfied to be eligible to participate in Job Corps. See Schochet et al. (2008) for details.

[18]Applicants were randomized into the treatment or control group after eligibility was assessed but before applicants were assigned a training center. This was implemented to balance the comparability of the treated and control groups and minimize any direct negative impact that the randomization may have for the control group (Schochet et al., 2008).

[19]We further exclude few (219) individuals for whom baseline information is missing, thus our sample is slightly smaller than Schochet et al. (2008). We provide additional information on the sample we use in Appendix C.

Panel A of Table 2 reports the program take-up rates. Focusing on the full sample (column 1), 73% of individuals in the treatment group enrolled in Job Corps, while 4.2% of the control group individuals crossed over to Job Corps. This implies an overall compliance share of about 68.8%(=$\pi_h$). Other programs were available for disadvantage youths, including high school, community colleges, and other publicly funded programs (Schochet et al., 2008). The following row reports the proportion of children who enrolled in a competing program but not in Job Corps, separately for the treatment and control groups. The Job Corps offer reduces the share of individuals enrolling in alternative programs from 66.5% to 21.2%, implying that the proportion of compliers switching from competing programs to Job Corps is 45.4%. In panel B of Table 2 we report the reduced form effect of being treated on weekly earnings 4 years after random assignment; in the overall sample these increase by 17.5$ (an increase by 8.8%). Dividing by the compliance share we obtain a treatment effect of 25.5$.

Columns (2)–(4) report the compliance shares and experimental impacts for observationally different groups. We focus on age (separating between those younger than 20 and those 20 and above) and past education experience because previous studies have found heterogeneous effect of the program along these dimensions (Schochet et al., 2008; Blanco et al., 2013). We find that the overall compliance rates are similar across these groups, but the fraction of individuals substituting Job Corps for competing training programs is higher for the young. As in Blanco et al. (2013), we find that the treatment effect is larger for older individuals, both in absolute and relative terms: weekly earnings increase by almost 42$ (14%) for older participants compared to 14.3$ (5%) for younger ones. In contrast, differences by prior educational attainment are less pronounced.

# 4 Results

## 4.1 Empirical Implementation

In Section 2, we assumed that the distribution of covariates $X$ is smooth and well-behaved. In practice, given finite sample sizes, conditioning tends to decrease statistical precision of the sample estimate of the conditional treatment effect and compliance shares. Therefore, it may be preferable to use a subset of observed covariates to estimate the parameters of interest.

**Grouping strategies.** We partition the population into groups based on observed characteristics, with the goal of identifying subpopulations for which the treatment yields a higher MVPF.

We choose the groups based on empirical evidence from studies that have examined the heterogeneity of treatment effects in our applications. For Head Start, we follow Bitler et al. (2014) and use the test scores at the time of randomization and language status (Spanish vs. non-Spanish speakers) as grouping variables. For Job Corps, we draw on the findings from Blanco et al. (2013), using the educational history, age, and race of the individuals. Note that all these characteristics are observed before randomization, so they could not possibly have been affected by treatment.[20]

We implement two grouping strategies. The first involves manually constructing groups by applying thresholds to continuous characteristics. To balance the precision of treatment effect estimation with the ability to capture meaningful heterogeneity, we define 6 groups for Head Start and 8 groups for Job Corps, the latter benefiting from a larger experimental sample size. For Head Start, we divide the distribution of pre-assignment test scores (a simple average of PPVT and WJIII pre-academic skills test scores) into three equal-sized groups and then further split each group by language status, resulting in six subgroups. For Job Corps, we define groups based on three binary characteristics: being older or younger than 20, being white or non-white, and having completed high school or not. This yields eight distinct groups corresponding to all possible combinations of these attributes.

As an alternative strategy, we employ the generalized random forests (GRF) algorithm to estimate individual-level treatment effects based on pre-determined covariates, following the framework for conditional average treatment effects (CATE) outlined in Athey et al. (2019).[21] The covariates used in estimating the CATE closely mirror the grouping variables used above.[22] We then use recursive

[20]Abadie et al. (2018) discuss potential biases from endogenous stratification, when individuals' outcomes are used to define subgroups.

[21]Our implementation uses the GRF package developed by Athey et al. (2019), which has been applied in recent work such as Haushofer et al. (2025) to predict causal treatment effects and assess the impact of targeted interventions. We predict the effect of Head Start on children's test score one year after assignment and the effect of Job Corps on weekly earnings 4 years after assignment.

[22]For Head Start, we use the continuous measures of the two pre-enrollment test scores (WJIII and PPVT) and an indicator for whether the child is a Spanish speaker. For Job Corps, we use a continuous measure of age, categorical race

partitioning to group individuals based on their characteristics, identifying patterns in how these traits relate to the predicted effect of the treatment.[23] Compared to manual splitting this strategy allows us to consider more variables and their relationship with predicted treatment effect. It freely selects which variables and which categories/cutoffs within each variable are the relevant ones, thus also being more agnostic on the number of groups. It generates 7 groups for Head Start and 14 groups for Job Corps, several of which diverge substantially from those produced with manual grouping. For example, in the Head Start case, while the majority of the generated groups involve children with either relatively high or relatively low grades in both pre-enrollment tests, two of them, covering about 22% of the population, involve children with a high PPVT score but low WJIII score. The manual grouping strategy aggregates these children together with those displaying average performance in both tests into a single, undifferentiated group. For Job Corps, participants below age 20 are further split among those above and below 18, but only for non-white participants with little job experience. Those above 20 are instead split between those younger and older than 22 only for non-high school graduates. Figures A1 and A2 illustrate the partitioning trees and the resulting characteristics of the different groups.

**Estimation steps.** We estimate the values of $\lambda$ for a given budget constraint using the following procedure. First, we compute the MVPF for each group. As shown in equation (8), the group-level MVPF is defined as the ratio of the group-specific intention-to-treat (ITT) effect to the marginal cost of offering the program to an additional individual in that group, the latter of which depends on the compliance shares (see Equation (7)). We estimate ITT effects and compliance shares using the experimental data. Next, we rank the groups by their MVPF and sequentially expand the offer of the program to the group with the highest MVPF among untreated individuals. We continue this process until adding the next group would violate the budget constraint. The $\lambda$ associated with each budget is

(white, black, Hispanic, native American), the share of months in the previous year spent in employment and in education, and whether the individual held a GED or completed high school at the time of application.

[23]Recursive partitioning is a data-driven method used to segment a population into subgroups that are internally homogeneous with respect to a target variable—in this case, the predicted treatment effect. The algorithm iteratively splits the sample based on covariates, selecting the partitioning rules that best explain variation in the outcome. Each split aims to maximize within-group similarity and between-group differences, resulting in a hierarchical structure of subpopulations. To prevent overfitting, we set a maximum tree depth of 5, which limits the number of recursive splits from the root node.

then given by the MVPF of the last target group — i.e., the group treated at the margin.[24]

**Social benefit.** We measure $Y_i$ using lifetime earnings, predicted from child achievement scores in Head Start (as in Kline and Walters (2016)) and earnings 4 years after random assignment in Job Corps (as in McConnell and Glazerman (2001)). Details are included in Appendices B and C. As pointed out in Kline and Walters (2016), focusing only on lifetime earnings ignores other potential benefits of the program (such as reduced criminal activity, or improved health), thus the social benefit imputed from test score gains is likely a lower bound.

## 4.2 Results: Optimal Assignment of a Single Treatment

Table 3 reports the estimated parameter $\lambda$ under different budget constraints, separately for Head Start (columns (1)-(4)) and Job Corps (columns (5)-(8)). We consider four budget levels that would allow enrolling up to 25, 50, 75, or 100 percent of applicants; the final level implies a non-binding budget constraint. To calculate the budget levels, we consider a maximum number of 900,000 participants for Head Start and 60,000 for Job Corps (see Section 3).[25] As explained in Section 2.1, the effective capacity ($\hat{R}$) is below the overall program capacity ($\tilde{R}$) because of always takers who crossover from the control group to enroll in the program. For Head Start, the share of always takers is relatively high (=0.14) because local staff intentionally enrolled control group children into their center and parents applied to another nearby program centers as information on HSIS was not shared with centers not involved in the study (Puma et al., 2010). In Job Corps, the share of always takers is small (=0.03).

The first two rows of Panel A report the MVPF threshold ($\lambda$) and the implied share of individuals offered program access ($P(Z = 1)$), assuming no fiscal externalities and manually constructed groups. As expected, tighter budget constraints lead to higher values of $\lambda$. When the budget expands

[24] This procedure implies that if there is some budget left after treating one subgroup, but the cost of treating the next-ranked group exceeds the remaining budget, then the remaining budget is unused. This can be relaxed, for example by allowing the remaining budget to be targeted to a random sample of the next subgroup up until the budget is exhausted. This latter procedure is used in computing the regret analysis in Table 4.

[25] Recall from equation (3) that, in the case of a single treatment without fiscal externalities, budget constraint is equivalent to a capacity constraint of the program. See Appendices B and C for the specifications of program costs of Head Start and Job Corps, respectively.

and more applicants can be offered enrollment, $\lambda$ declines monotonically, admitting individuals from lower-MVPF groups. When the budget becomes sufficiently large, it may be optimal to target groups with MVPFs below one—that is, even when the policy does not yield a "bang for the buck." This occurs because policies remain welfare-improving as long as the MVPF is non-negative. In Job Corps, some groups exhibit negative marginal benefits and therefore negative MVPFs; the optimal strategy excludes these individuals from program access. Consequently, even in the absence of binding budget constraints, the share of individuals assigned to treatment remains below one (=0.76).

Imperfect compliance implies that the share of individuals targeted for treatment ($P(Z=1)$) may exceed the effective capacity constraint, provided that the targeted groups have non-negative MVPFs. For example, with an effective constraint of 11%, the optimal Head Start strategy targets 17% of the population (column (1)), reflecting the presence of never-takers within targeted groups. Whether $P(Z=1)$ exceeds or falls below the overall constraint depends on the relative shares of always-takers and never-takers.

Figure 1 shows the change in social welfare as we expand the program budget. The welfare change from optimal treatment assignment is given by $\int \delta^*(x)(E(Y_i \mid Z_i = 1, X_i = x) - E(Y_i \mid Z_i = 0, X_i = x))\, dF(x)$, which is simply the maximized social welfare function in equation (1) subtracting $E(Y_i \mid Z_i = 0)$, the average welfare in the control group.[26] Correspondingly, the budget (x-axis) measures the cost of enrolling the compliers in each group, net of the costs incurred by the always takers. To ensure comparability with the overall budget constraints in Table 3, we scale costs and benefits by the number of potential participants (900,000 for Head Start; 60,000 for Job Corps).

Our discussion centers on panels (a) and (b), ignoring for now any fiscal externality. Under optimal targeting, social benefits (solid lines) are a concave function of program costs. The slope of the curve corresponds to the ratio between the benefit of treating the next marginal group and the cost of expansion to that group.[27] As shown in equations (6)–(8), this ratio corresponds to the MVPF for the

[26]The figure also plots the 95% confidence interval of the estimated welfare when the program is offered to the individuals in all the targeted groups. This is calculated based on the s.e. from a single ITT estimate that jointly considers all individuals across the targeted groups, and then scaling it by the proportion of these groups in the overall population ($\int \delta^*(x)\, dF(x)$).

[27]Formally, at each point the gain of targeting the next marginal group (call it $x'$) is $E[Y_i|Z_i = 1, X_i = x'] - E[Y_i|Z_i = 0, X_i = x']P(x')$, while the cost is $\phi[P(\mathrm{d}_i = 1|Z_1 = 1, X_i = x') - P(\mathrm{d}_i = 1|Z_1 = 0, X_i = x')]P(x')$, where $P(x')$ is

next-targeted group. With tight budgets the optimal strategy involves treating groups with high MVPF. As the budget constraint is relaxed, treatment is extended to groups with progressively lower MVPF, resulting in a diminishing slope of the benefits-cost curve. The vertical dotted line marks the point at which program expansion should cease under an optimal targeting rule. In the case of Job Corps, the presence of groups with negative MVPF implies that, even with a relaxed budget constraint, one group would remain untreated. Consequently, it would not be optimal to expand the program beyond a total cost of approximately 516 mln $.

We contrast the social benefit under optimal targeting (solid line) to the benefit when the same amount is spent on the program but targeting individuals at random (dashed line). Under random assignment, expanding the program means giving offers to a random set of individuals in the population. This means that the marginal benefit and cost of a program expansion remain constant regardless of the budget constraint. Thus, the social benefit curve is linear, with a slope equal to the population-level MVPF. The gap between the two lines represents the difference in social benefits between the two program assignment strategies. By prioritizing groups with the highest MVPF, optimal targeting yields significantly greater social returns, especially under tight budget constraints. For example, in the Job Corps program, directing resources to the top five MVPF groups produces benefits approximately 1.5 times greater than random allocation (320 mln $ as compared to 130 mln $). Achieving comparable outcomes through random assignment would require treating nearly the entire eligible population at more than double the cost.

Table 4 reports the gain in social benefits generated by optimal targeting relative to that from random assignment at different budget levels (a regret analysis). We also test for the null hypothesis that the gains are equal between the two approaches, treating the gain from random assignment as fixed.[28] We find that optimal targeting generates substantially greater social benefits, particularly under strin-

the proportion of group $x'$ within the population. The ratio of these two terms, which defines the slope of the curve when $P(x')$ is small, corresponds to the MVPF of the marginal group $x'$.

[28] In Table 4 (and Table 3), "budget" refers to the total budget required to treat certain proportions of the population, *including* always takers. This is different from Figure 1, where the x-axis reflects the budget *excluding* the cost associated with always takers. As a result, the same nominal budget allows to offer the program to a different proportion of the population in the tables versus the figure. In Table 4 social benefits for the targeted groups at each budget level, along with their s.e., derived from a single estimation using experimental variation across all targeted groups jointly, and then scaled by their population shares -the same procedure used in Figure 1.

gent budget constraints. For Head Start, when the budget allows to treat 25% of the applicants, optimal treatment assignment increases social welfare by about 1 bn $ compared to randomly assignment (column 1, panel A). This corresponds to doubling the social benefits (under random assignment the total benefits are estimated to be around 990 mln $) and the gain is more than half of the total budget spent to target this group. For Job Corps, the relative improvement is even more pronounced under a tight budget, because the population-average MVPF is relatively low and the MVPFs of certain groups are significantly higher than the average (e.g., Hendren and Sprung-Keyser (2020) report an MVPF of 0.18).[29] With a budget sufficient to cover 25% of the population, random targeting would generate benefits of approximately 103 mln $. In contrast, optimal targeting yields benefits nearly three times as large, with an incremental gain of about 193 mln $ (column 5, panel A).

As the budget constraint becomes less binding, the relative difference between the two assignment strategies diminishes and becomes statistically insignificant. For Job Corps, even without constraints, the estimated difference remains positive (albeit statistically insignificant), because the group with negative program returns is never treated under optimal targeting.

**Groups predicted by GRF and recursive partitioning.** The rest of Panel A in Table 3 reports the estimated $\lambda$ and the corresponding share of targeted individuals, when groups are defined by recursive partitioning based on the CATE (GRF-recursive). In general, this approach produces substantially different groups compared to manual grouping and, consequently, a different distribution of the MVPF. Figure 2 shows that the distribution of MVPF under GRF-recursive grouping is more spread than the distribution of MVPF under manual grouping. For Head Start, the GRF-recursive approach is especially successful in uncovering groups with high MVPFs. For Job Corps, the recursive partitioning approach more effectively identifies subpopulations with negative program returns. This suggests that this alternative grouping strategy captures more complex interactions among observable characteristics, potentially revealing more nuanced patterns in treatment effects that manual grouping may

[29] Notably, Hendren and Sprung-Keyser (2020) also acknowledge that certain subgroups exhibit MVPF estimates above one. Their calculations, which incorporate benefits from reduced disability insurance receipt, report an MVPF of 1.18 for older Job Corps participants. In line with this, three of the four groups with the highest MVPF consist of older individuals.

overlook.

Because the recursive partitioning approach more effectively identifies subpopulations with heterogeneous program returns, it enables a more efficient targeting strategy. For Head Start, due to the ability of the algorithm in identifying large MVPF groups, the estimated $\lambda$ under a tight budget constraint is substantially higher than the one estimated with manual grouping. This translates into a steeper social benefit curve at the smallest cost level (panel (a) in Figure A3) and a larger welfare gain relative to random assignment (Panel B in Table 4): when the budget allows treatment of 25% of the population, regret increases by nearly 50% compared to manual grouping.

Yet the efficiency gain from the GRF-recursive grouping may not persist as the budget becomes more generous. Table 4 shows that the difference in regret between GRF-recursive grouping and manual grouping becomes small and statistically insignificant for Head Start. This is because the GRF-recursive strategy also identifies more groups with small (and positive) MVPFs.[30] Indeed, panel A in Table 3 shows that the estimated $\lambda$ with more relaxed budget constraints are much lower under the GRF-recursive approach as compared to manual grouping (for budgets allowing to treat 75% and 100% of the population the estimated $\lambda$s are 0.20 and 0.16 with recursive partitioning grouping as compared to 0.80 and 0.68 with manual grouping).

In contrast, for Job Corps, the increase in regret extends to more generous budget levels. Specifically, when the budget can afford treating half of the population, the difference with random assignment rises by over 70% and becomes statistically significant. This is because the GRF-recursive approach identifies more groups with negative MVPFs, implying that the program should stop expanding at a smaller budget level than before.[31]

**Limitations.** As already mentioned above, we measure welfare by individual lifetime earnings and ignore other potential social benefits of the program such as reduced criminal activity or improved

[30] Because the population-level MVPF does not depend on the grouping strategy, a relatively high estimated value of $\lambda$ under a constrained budget is necessarily compensated by a sharper decline in $\lambda$ as the budget increases.

[31] Panel (b) in Figure A3 shows that the curve for Job Corps declines more sharply, with the maximum level of social benefits is achieved with approximately 10 mln $ less in spending.

health. Therefore, our estimated overall social welfare gains from targeting are likely lower bounds.[32] In addition, we have made the standard assumption in program evaluation that the potential outcome for each individual is only a function of the individual's own program choice (the SUTVA assumption in Angrist et al. (1996)). This assumption allows us to conveniently draw causal inference on the MVPF from the experimental variation. At the same time, it abstracts from peer effects, which might be important as targeting individuals at a large scale could change the composition of program peers.

We have illustrated the welfare gain of optimal targeting using a set of covariates that are predetermined and informed by prior studies. The chosen covariates are easily observed by a policymaker, yet they are by no means optimal–there may well be other covariates that one can use to deliver even larger welfare improvement. However, given this set of covariates, we have used both an ad-hoc approach as well as a machine learning approach to characterize heterogeneity in the averaged treatment effect, with the latter approach building on recent advances optimized specifically for the study of heterogeneous treatment effects. Our goal is to showcase how MVPF identified from experimental variations and grouping of individuals by both basic and modern approaches can be combined to improve social welfare.

## 4.3 Results: Optimal Assignment of a Single Treatment Accounting for Fiscal Externalities

As shown in Tables 1 and 2, a substantial share of compliers would have enrolled in a competing program if they had not been offered the focal program ($c$-compliers). Thus, the marginal cost of expanding the focal program should account for savings from reduced participation in competing programs. This cost adjustment has two key implications: (1) it increases the MVPF by reducing the marginal cost, particularly for groups with a higher share of $c$-compliers; and (2) it relaxes the budget constraint as the cost savings from competing programs can be reallocated to expand the focal program further.

[32] An interesting question is whether the degree of understatement likely varies across groups, and the resulting effect on the targeting rule and subgroup ordering. We leave this for future work.

Panel B of Table 3 reports the estimated $\lambda$ after accounting for fiscal externality created by the $c$-compliers. For Head Start, under tight budget constraints (columns (1) and (2)) incorporating fiscal externalities raises the estimated value of $\lambda$ (relative to panel A) but affects neither the size nor the composition of targeted individuals. Considering that the estimated MVPF is higher in all groups after accounting for fiscal externality, this leads to a mechanical increase in the target threshold. By comparison, when the budget is large to accommodate 75% of the eligible population, the fiscal savings from $c$-compliers become substantial for the planner to offer the program to the entire population (column (3), panel B). Without accounting for these savings in program costs, only 78% (89% in the case of the GRF grouping strategy) of the population could be targeted (column (3), panel A).

Accounting for fiscal externalities can also alter the MVPF ranks and hence the ordering of the groups to be targeted, especially when the shares of $c$-compliers vary significantly across groups. For instance, when the maximum overall capacity of Job Corps is 0.25, the share of targeted individuals when grouped manually *decreases* from 30% to 26% when fiscal externality is accounted for. This occurs because the fiscal externality disproportionately benefits a relatively small group, making it optimal to treat that group first. However, doing so exhausts more of the budget upfront, leaving insufficient resources to treat larger groups that would have been included in the absence of fiscal adjustments. This reflects a crowding-out effect driven by changes in the MVPF ranks.

Panels (c) and (d) of Figure 1 show the social benefits as a function of costs. Comparing them to the corresponding figures without fiscal externality (panels (a) and (b)), we find that the cost savings allow the same level of social benefit to be reached with a lower budget overall. For example, targeting the three groups with the highest MVPF in the Head Start case (with an overall benefit of about 2,800 mln \$) would cost 1,070 mln \$ after taking externalities in consideration, a reduction of about 18% compared to the cost without externalities (1,300 mln \$). Similarly, treating the five groups with the largest MVPF in the Job Corps case costs 17% less, when fiscal externalities are taken into account (215 vs. 260 mln \$).

**Fiscal externality from tax revenue.** Finally, we incorporate the additional fiscal externalities arising from the tax revenue in the computation of the MVPF, assuming that the government is able to borrow against the additional tax revenues.[33] This externality enables further program expansion at any budget level, since the additional tax revenues are treated as a reduction in program costs. Notably, when both externalities—stemming from reduced participation in competing programs and increased tax revenues—are considered, the cost of the program becomes negative for one Head Start group (10% of the population).[34] This implies that the tax revenue generated by treating this group exceeds the cost of treatment. As explained in Hendren and Sprung-Keyser (2020), the MVPF of this group is infinite, meaning it is always optimal to offer the program to its members, as it effectively pays for itself. In a constrained optimization framework, treating this group has an additional benefit: it expands the effective budget, since the surplus tax revenue can be used to fund treatment for other individuals.

Panel C of Table 3 illustrates this effect. With a nominal budget sufficient to treat only 25% of the Head Start population, accounting for the tax externality allows the program to be offered to manually generated groups covering more than half of the population. This expansion necessarily lowers the estimated $\lambda$ compared to optimal assignment without the tax externality, as it includes groups with lower MVPF. In contrast, the GRF-recursive strategy does not expand coverage as much, due to the presence of a large group—covering over 45% of the population—that is only targeted under more generous budgets. For both Head Start and Job Corps, the combined effect of the two externalities frees up sufficient resources to offer the program to the entire applicant population, even when the nominal budget would only cover half.

[33] In line with the approach adopted by Kline and Walters (2016), we assume a marginal tax rate of 35% for participants in both the Head Start and Job Corps program, leading to the following MVPF formula that extends equation 12 in Section 2.2: $\frac{\pi_h(x)LATE_h(x)(1-0.35)}{\phi_h\pi_h(x)-\phi_c\pi_c(x)-0.35\pi_h(x)LATE_h(x)}$. The effect of the tax externality is to mechanically increases the MVPFs which exceed one when tax effects are not accounted for, while it decreases the MVPF for groups with values below one; this results from the following inequality $\frac{\pi_h(x)LATE_h(x)(1-0.35)}{\phi_h\pi_h(x)-\phi_c\pi_c(x)-0.35\pi_h(x)LATE_h(x)} > \frac{\pi_h(x)LATE_h(x)}{\phi_h\pi_h(x)-\phi_c\pi_c(x)}$ which corresponds to $\pi_h(x)LATE_h(x) > \phi_h\pi_h(x) - \phi_c\pi_c(x)$.

[34] This group includes non-Spanish-speaking children with low PPVT baseline scores. Both manual and GRF-recursive grouping consistently identify this group.

## 4.4 Results: Optimal Assignment of Multiple Treatments

In this section, we examine the joint allocation of Head Start and Job Corps treatment under a single aggregate budget constraint. Advocates of early childhood investment have argued that early skill investment programs generate larger returns than investment programs at later ages, due to the hypothesis that skill begets skills (Cunha et al., 2010). Our results thus add to this discussion by showing the optimal treatment assignment of an early skill investment program (Head Start) and a late skill investment program (Job Corps), under a realistic budget constraint and different social welfare functions.

We maintain the same groupings as in the single-program analyses and assume no overlap in participation between the two programs. It is important to note that Head Start is substantially larger in scale, serving up to 900,000 participants compared to Job Corps's 60,000, implying that Head Start involves significantly higher total costs and potentially higher overall benefits due to its scale. We present results from the basic setting without fiscal externalities by competing programs, but the main insights hold after accounting for fiscal externalities.

Figure 3 illustrates the optimal joint allocation given different budget constraints for manual grouping. Panel (a) depicts program benefits against program costs in which potential participants across both programs are assigned equal social welfare weights, with distinct markers indicating whether the marginal group served at each budget level receives Head Start or Job Corps.[35] On average, Head Start exhibits a higher MVPF than Job Corps, which explains why more Head Start groups are prioritized in the optimal allocation. In particular, three groups of individuals in Job Corps have very low or even negative MVPF, meaning that they will be treated only at higher budget levels or excluded entirely. Nevertheless, even at low and intermediate budget levels, an optimal allocation involves offering both programs. For instance, consider a relatively modest budget of approximately 1,500 mln \$. Because some groups of potential Job Corps participants exhibit especially high returns, the optimal strategy includes 6.6% of Job Corps participants among those who are offered the treatment, which is more than the fraction of the overall Job Corps population (6.2% of the total population). Hence, despite Head Start having a higher average MVPF, optimal targeting under a fixed budget leverages program

[35] As in Figure 1, Figure 3 considers only the incremental welfare generated by program participation for compliers.

heterogeneity: it does not imply allocating all resources toward early childhood programs. In contrast, when directed toward specific subgroups of older youth, a training program like Job Corps can yield high returns per dollar spent and thus becomes preferable for certain segments of the population.

The analysis in panel (a) assumes that the policymaker values outcomes for children and young adults equally. Panels (b) and (c) relax this assumption and illustrate how the optimal treatment ordering shifts when differential welfare weights are applied. In panel (b), we assign a welfare weight of 1.5 to potential Head Start participants and of 1 to Job Corps participants, while in panel (c), we give a weight of 1.5 to potential Job Corps participants instead, leaving a weight of 1 to Head Start participants. A high welfare weight raises the social welfare per dollar of government expenditure, thereby increasing the likelihood of receiving treatment (at the expense of less-weighted groups). As shown in panel (b), under a 1,500 mln $ budget, the majority of resources are allocated to Head Start participants, with Job Corps recipients comprising only 2.4% of those offered treatment. In contrast, when Job Corps participants receive the higher welfare weight (Panel (c)), the optimal allocation includes 8.4% of the treated population from Job Corps, despite the same overall budget constraint.

Finally, we calculate the minimum relative welfare weight a social planner would need to assign to Head Start participants for it to become optimal to allocate the entire budget exclusively to Head Start, assuming that the budget is fully exhausted when all Head Start-eligible subgroups are treated. This requires the weighted MVPF of all Head Start subgroups to be larger than the MVPFs of all the Job Corps subgroups. We find that the planner must place a relative social welfare weight of 2.7 for Head Start participants compared to Job Corps participants to justify this reallocation. Although early childhood programs typically generate higher average returns, prioritizing investments in young children over disadvantaged youths is only warranted when there is a strong preference for the former group, once treatment effect heterogeneity is taken into account.

# 5 Conclusion

This paper shows that the MVPF can be used as a metric to determine treatment allocation and improve social welfare. There are two pillars in our analysis. One is heterogeneity in the treatment effects (and program costs) across policies and/or subgroups in the population, which implies heterogeneity in the MVPF. This gives scope for the use of selective targeting to improve social welfare. The second pillar of our analysis is the government budget constraint, which is very common in reality but largely ignored by the existing literature. Importantly, the budget constraint means that there is an opportunity cost in offering treatment to additional individuals. The optimal decision is to offer the treatment to those with MVPF above a cost-effectiveness threshold that is endogenously determined by the available resources.

Using experimental variations from Head Start and Job Corps, we draw causal inference on the heterogeneity of MVPF across subgroups in the population. The subgroups are classified using both an ad-hoc simple approach and a recent machine learning approach. By prioritizing groups with markedly high MVPF, our analysis suggests that optimal treatment targeting can double the social benefits of Head Start compared to the randomized treatment assignment used in the experiment when the government budget is tight. For Job Corps, the benefit of targeting is even more pronounced, particularly given that the population-level average MVPF is relatively low despite significant heterogeneity. For example, providing Job Corps randomly would cost twice as much as targeting the five highest-ranking groups to achieve the same level of benefits.

We also analyze the joint treatment assignment of Head Start (an early skill investment program) and Job Corps (a late skill investment program), contributing to the debate whether resources should be reallocated to early investment programs given the larger program returns than late investment programs. When the policymaker places equal social weights on the beneficiaries, we show that it is suboptimal to allocate all resources toward early childhood programs, because Job Corps can yield high returns per dollar spent and thus becomes preferable for certain groups of the population. Prioritizing investments in young children over disadvantaged youth is only warranted when a significantly larger social welfare weight is placed for the former group than the latter.

We believe our results are highly relevant in today's world with many governments facing budget deficit and fiscal austerity. Our cost-benefit calculations are based on estimates in the literature to convert the test score effects of Head Start into earnings gains. Our estimates of program cost use reported average cost from the literature, ignoring any heterogeneity in the program cost per person. These calculations are necessarily speculative and may limit the MVPF heterogeneity that we are able to recover from the data.

# References

**Abadie, Alberto, Matthew M Chingos, and Martin R West**, "Endogenous stratification in randomized experiments," *Review of Economics and Statistics*, 2018, *100* (4), 567–580.

**Aizer, Anna, Nancy Early, Shari Eli, Guido Imbens, Keyoung Lee, Adriana Lleras-Muney, and Alexander Strand**, "The Lifetime Impacts of the New Deal's Youth Employment Program," *The Quarterly Journal of Economics*, 2024, *139* (4), 2579–2635.

**Angrist, Joshua D, Guido W Imbens, and Donald B Rubin**, "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 1996, *91* (434), 444–455.

**Athey, Susan, Julie Tibshirani, and Stefan Wager**, "GENERALIZED RANDOM FORESTS," *The Annals of Statistics*, 2019, *47* (2), pp. 1148–1178.

**Berger, Mark C, Dan Black, and Jeffrey A Smith**, "Evaluating profiling as a means of allocating government services," *Econometric Evaluation of Labour Market Policies*, 2001, *13*, 59.

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

**Bitler, Marianne P, Hilary W Hoynes, and Thurston Domina**, "Experimental evidence on distributional effects of Head Start," Technical Report, National Bureau of Economic Research 2014.

**Blanco, German, Carlos A Flores, and Alfonso Flores-Lagunes**, "The effects of Job Corps training on wages of adolescents and young adults," *American Economic Review: Papers & Proceedings*, 2013, *103* (3), 418–22.

**Cascio, Elizabeth U**, "Does universal preschool hit the target?: Program access and preschool impacts," *Journal of Human Resources*, 2023, *58* (1), 1–42.

**Chan, Marc K, Antonio Dalla-Zuanna, and Kai Liu**, "Understanding Program Complementarities: Estimating the Dynamic Effects of Head Start with Multiple Alternatives," Technical Report, IZA Discussion Papers 2024.

**Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan**, "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly journal of economics*, 2011, *126* (4), 1593–1660.

**Cook, Jason B and Chloe N East**, "The effect of means-tested transfers on work: Evidence from quasi-randomly assigned SNAP caseworkers," Technical Report, National Bureau of Economic Research 2023.

**Cunha, F., J. Heckman, and S. Schennach**, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78(3)*, 883–931.

**Dehejia, Rajeev H**, "Program evaluation as a decision problem," *Journal of Econometrics*, 2005, *125* (1-2), 141–173.

**DHHS**, "Head Start program facts fiscal year 2013," 2013. Available at: https://headstart.gov/sites/default/files/pdf/head-start-fact-sheet-fy-2013.pdf.

__ , "Head Start Federal Funding and Funded Enrollment History," 2021. Available at: https://headstart.gov/sites/default/files/pdf/head-start-federal-funding-funded-enrollment-history-eng.pdf.

**Finkelstein, Amy and Nathaniel Hendren**, "Welfare analysis meets causal inference," *Journal of Economic Perspectives*, 2020, *34* (4), 146–167.

**__ , __ , and Erzo FP Luttmer**, "The value of medicaid: Interpreting results from the oregon health insurance experiment," *Journal of Political Economy*, 2019, *127* (6), 2836–2874.

**Frölich, Markus**, "Statistical treatment choice: an application to active labor market programs," *Journal of the American Statistical Association*, 2008, *103* (482), 547–558.

**García, Jorge Luis and James J Heckman**, "On criteria for evaluating social programs," Technical Report, National Bureau of Economic Research 2022.

**Hahn, Robert W, Nathaniel Hendren, Robert D Metcalfe, and Ben Sprung-Keyser**, "A welfare analysis of policies impacting climate change," Technical Report, National Bureau of Economic Research 2024.

**Haushofer, Johannes, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael Walker**, "Targeting impact versus deprivation," *American Economic Review*, 2025, *115* (6), 1936–1974.

**Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo**, "Substitution and dropout bias in social experiments: A study of an influential social experiment," *The Quarterly Journal of Economics*, 2000, *115* (2), 651–694.

**Hendren, Nathaniel and Ben Sprung-Keyser**, "A unified welfare analysis of government policies," *The Quarterly journal of economics*, 2020, *135* (3), 1209–1318.

**Imbens, Guido W and Joshua D Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

**Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad**, "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

**Kitagawa, Toru and Aleksey Tetenov**, "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 2018, *86* (2), 591–616.

**Kleven, Henrik Jacobsen and Claus Thustrup Kreiner**, "The marginal cost of public funds: Hours of work versus labor force participation," *Journal of Public Economics*, 2006, *90* (10-11), 1955–1973.

**Kline, Patrick and Christopher R Walters**, "Evaluating public programs with close substitutes: The case of Head Start," *The Quarterly Journal of Economics*, 2016, *131* (4), 1795–1848.

**Lechner, Michael and Jeffrey Smith**, "What is the value added by caseworkers?," *Labour economics*, 2007, *14* (2), 135–151.

**Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan**, "The unreasonable effectiveness of algorithms," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2024, pp. 623–627.

**Manski, Charles F**, "Statistical treatment rules for heterogeneous populations," *Econometrica*, 2004, *72* (4), 1221–1246.

**Mayshar, Joram**, "On measures of excess burden and their application," *Journal of Public Economics*, 1990, *43* (3), 263–289.

**McConnell, Sheena and Steven Glazerman**, "National Job Corps Study: The Benefits and Costs of Job Corps.," *Princeton, NJ: Mathematica Policy Research, Inc.*, 2001.

**Puma, Michael et al.**, "Head Start Impact Study. Final Report.," *Administration for Children & Families, U.S. Department of Health and Human Services*, 2010.

**Schochet, Peter Z, John Burghardt, and Sheena McConnell**, "Does job corps work? Impact findings from the National Job Corps Study," *The American economic review*, 2008, *98* (5), 1864–1886.

Table 1: Compliance Share and Experimental Impact, Head Start

| | | (1) Overall | (2) Low Base Score | (3) High Base Score | (4) Non-Spanish Speakers | (5) Spanish Speakers |
|---|---|---|---|---|---|---|
| **Panel A: Program Participation** | | | | | | |
| Head Start | | | | | | |
| | Treatment | 79.8 | 81.3 | 78.2 | 80.6 | 79.5 |
| | Control | 13.7 | 14.2 | 13.1 | 16.5 | 12.6 |
| | $\pi_h$ | 66.1 | 67.1 | 65.1 | 64.1 | 66.9 |
| Competing Program | | | | | | |
| | Treatment | 6.7 | 4.7 | 8.7 | 6.6 | 6.7 |
| | Control | 26.0 | 21.6 | 30.3 | 22.3 | 27.5 |
| | $\pi_c$ | 19.3 | 16.9 | 21.6 | 15.7 | 20.7 |
| **Panel B: Experimental Impact** | | | | | | |
| ITT | | 0.19 | 0.22 | 0.15 | 0.32 | 0.14 |
| | | (0.025) | (0.036) | (0.032) | (0.047) | (0.026) |
| 2SLS | | 0.28 | 0.33 | 0.23 | 0.50 | 0.21 |
| | | (0.038) | (0.055) | (0.049) | (0.075) | (0.039) |
| N | | 3,574 | 1,787 | 1,787 | 1,017 | 2,557 |

$Notes$ : $\pi_h$ is the difference in the mean participation rate in Head Start between the treatment and the control group. $\pi_c$ is the difference in the mean participation rate in a competing program between the control and the treatment group. ITT estimates come from an OLS regressions of the outcome (simple average of WJIII and PPVT test one year after random assignment) on the randomly assigned treatment variable ($Z$). 2SLS estimates comes from 2SLS model where program participation is instrumented with the treatment variable. Both models control for household size, number of siblings, dummies for whether the child is female, black, hispanic, uses English as home language, is living in urban area, is living with both parents, is in need of special education, is the child of a teen mother, is the child of a mother who never married or is separated, is the child of a mother with high school or more than high school. Standard errors are clustered at the level of the Head Start center to which both treated and control individuals are assigned. “Low Base Score” and “High Base Score” indicate whether the child scores below or above the median grade in baseline test scores (average of WJIII and PPVT test completed before random assignment).

Table 2: Compliance Share and Experimental Impact, Job Corps

| | | (1) Overall | (2) Young | (3) Old | (4) Dropouts | (5) Non-Dropouts |
|---|---|---|---|---|---|---|
| **Panel A: Program Participation** | | | | | | |
| Job Corps | | | | | | |
| | Treated | 73.0 | 75.5 | 69.4 | 73.4 | 72.6 |
| | Control | 4.2 | 5.3 | 2.6 | 4.5 | 3.9 |
| | $\pi_h$ | 68.8 | 70.2 | 66.8 | 68.9 | 68.8 |
| Competing Program | | | | | | |
| | Treated | 21.2 | 21.5 | 20.7 | 20.4 | 21.9 |
| | Control | 66.5 | 73.0 | 56.8 | 64.2 | 68.9 |
| | $\pi_c$ | 45.4 | 51.5 | 36.1 | 43.7 | 47.0 |
| **Panel B: Experimental Impact** | | | | | | |
| ITT | | 17.5 | 10.1 | 27.9 | 18.8 | 16.3 |
| | | (4.191) | (5.307) | (6.752) | (5.94) | (5.909) |
| 2SLS | | 25.5 | 14.3 | 41.9 | 27.2 | 23.7 |
| | | (6.101) | (7.55) | (10.168) | (8.631) | (8.613) |
| N | | 11,094 | 6,575 | 4,519 | 8,155 | 2,939 |

$Notes$ : $\pi_h$ is the difference in the mean participation rate in Job Corps between the treatment and the control group. $\pi_c$ is the difference in the mean participation rate in a competing program between the control and the treatment group. ITT estimates come from an OLS regressions of the outcome (weekly earnings during the last quarter of the fourth year after random assignment) on the randomly assigned treatment variable ($Z$). 2SLS estimates comes from 2SLS model where program participation is instrumented with the treatment variable. All the statistics and estimates use sample weights to adjust for the sample design and survey design. “Young” and “Old” indicate whether the participant is below or above 19 years of age. “Dropouts” refer to participants dropping out of high school before applying for Job Corps.

Table 3: Values of $\lambda$ for Different Constraints

| | Head Start | | | | Job Corps | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Constraint** | | | | | | | | |
| Capacity: Overall | 0.25 | 0.5 | 0.75 | 1 | 0.25 | 0.5 | 0.75 | 1 |
| Capacity: Effective | 0.11 | 0.36 | 0.61 | 0.86 | 0.22 | 0.47 | 0.72 | 0.97 |
| Budget | 1800 | 3600 | 5400 | 7200 | 252 | 505 | 758 | 1011 |
| **Panel A: No Externalities** | | | | | | | | |
| *Manually constructed groups* | | | | | | | | |
| $\lambda$ | 1.53 | 1.04 | 0.80 | 0.68 | 1.05 | 0.24 | 0.18 | 0.18 |
| P(Z=1) | 0.17 | 0.56 | 0.78 | 1 | 0.30 | 0.57 | 0.76 | 0.76 |
| *GRF with recursive partitioning* | | | | | | | | |
| $\lambda$ | 3.38 | 1.60 | 0.20 | 0.16 | 1.16 | 0.29 | 0.13 | 0.13 |
| P(Z=1) | 0.10 | 0.28 | 0.89 | 1 | 0.28 | 0.64 | 0.74 | 0.74 |
| **Panel B: Fiscal Externalities from Program Substitution** | | | | | | | | |
| *Manually constructed groups* | | | | | | | | |
| $\lambda$ | 2.09 | 1.41 | 0.86 | 0.86 | 1.25 | 0.22 | 0.22 | 0.22 |
| P(Z=1) | 0.17 | 0.56 | 1 | 1 | 0.26 | 0.76 | 0.76 | 0.76 |
| *GRF with recursive partitioning* | | | | | | | | |
| $\lambda$ | 3.69 | 2.10 | 0.18 | 0.18 | 0.69 | 0.15 | 0.15 | 0.15 |
| P(Z=1) | 0.10 | 0.28 | 1 | 1 | 0.36 | 0.74 | 0.74 | 0.74 |
| **Panel C: Fiscal Externalities from Program Substitution Considering Taxes** | | | | | | | | |
| *Manually constructed groups* | | | | | | | | |
| $\lambda$ | 1.80 | 0.80 | 0.80 | 0.80 | 1.41 | 0.16 | 0.16 | 0.16 |
| P(Z=1) | 0.56 | 1 | 1 | 1 | 0.39 | 0.76 | 0.76 | 0.76 |
| *GRF with recursive partitioning* | | | | | | | | |
| $\lambda$ | 5.19 | 0.12 | 0.12 | 0.12 | 0.37 | 0.10 | 0.10 | 0.10 |
| P(Z=1) | 0.28 | 1 | 1 | 1 | 0.41 | 0.74 | 0.74 | 0.74 |

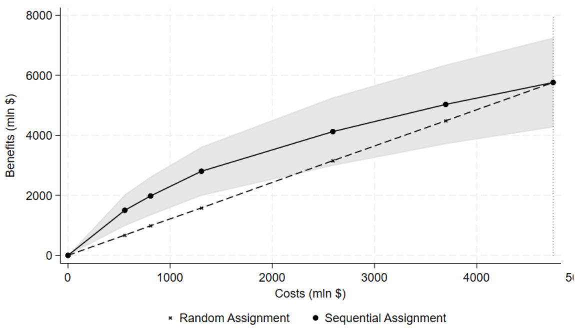
$Notes$ : Overall capacity constraint refers to the maximum capacity of the program. We consider four budget levels that would allow enrolling up to 25, 50, 75, or 100 percent of applicants; the final level implies a non-binding budget constraint. Effective capacity constraint is the capacity constraint excluding Always Takers. Budget constraint is expressed in mln dollars. We estimate $\lambda$ for each budget level by ranking groups by their MVPF and allocating the program to groups in descending order of MVPF. The program is expanded until the next group's cost would exceed the remaining budget. The $\lambda$ for each budget corresponds to the MVPF of the marginal (last-treated) group.

Table 4: Regret Analysis Comparing Targeting to Random Assignment

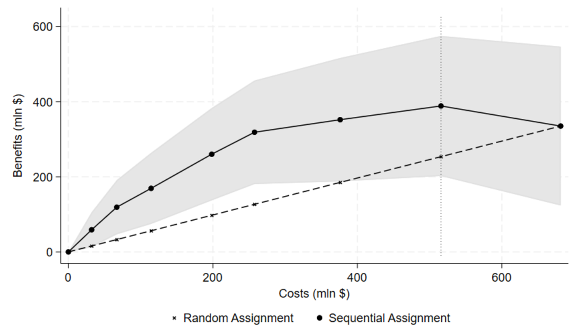
| | Head Start | | | | Job Corps | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Constraint** | | | | | | | | |
| Capacity: Overall | 0.25 | 0.5 | 0.75 | 1 | 0.25 | 0.5 | 0.75 | 1 |
| Capacity: Effective | 0.11 | 0.36 | 0.61 | 0.86 | 0.22 | 0.47 | 0.72 | 0.97 |
| Budget | 1800 | 3600 | 5400 | 7200 | 252 | 505 | 758 | 1011 |
| **Panel A: Targeting manually constructed groups** | | | | | | | | |
| Difference vs. random assignment | 1,029 | 932 | 279 | 0 | 193 | 127 | 53 | 53 |
| p-value | 0.002 | 0.110 | 0.701 | - | 0.003 | 0.164 | 0.576 | 0.576 |
| **Panel B: Targeting groups generated via GRF and recursive partitioning** | | | | | | | | |
| Difference vs. random assignment | 1,518 | 879 | 296 | 0 | 279 | 215 | 110 | 110 |
| p-value | 0.000 | 0.151 | 0.694 | - | 0.000 | 0.016 | 0.236 | 0.236 |

$Notes$ : Overall capacity constraint refers to the maximum capacity of the program. We consider four budget levels that would allow enrolling up to 25, 50, 75, or 100 percent of applicants; the final level implies a non-binding budget constraint. Effective capacity constraint is the capacity constraint excluding Always Takers. Budget constraint and the difference in aggregate social benefits are expressed in mln dollars. With targeting, for each constraint we sequentially expand the treatment offer to different groups, ranked on the basis of their MVPF. If, after targeting one group, there is some budget left to only targeting a part of the following group, we select some individuals at random from this marginal group (see also footnote 24). We estimate the social benefits of offering the program to different combinations of groups as the ITT effect on lifetime income, aggregated across all individuals in the targeted groups. These estimates are scaled by the number of targeted individuals to reflect total benefits. S.e. are derived from the ITT estimates and similarly scaled, allowing us to formally test (chi-squared test) for statistically significant differences in aggregate social benefits across alternative assignment rules. We report the p-value of these tests in the table.

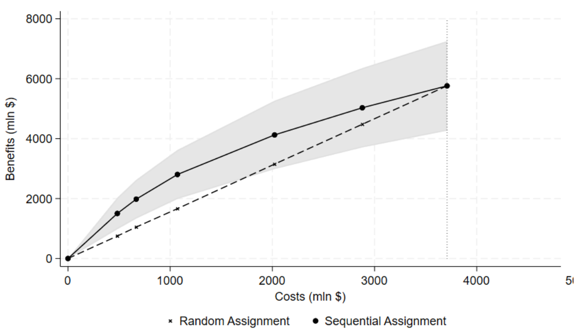
Figure 1: Benefit as a Function of Costs, Manually Constructed Groups



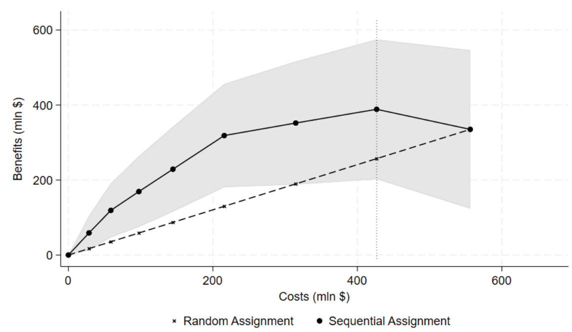

(a) Head Start




(b) Job Corps



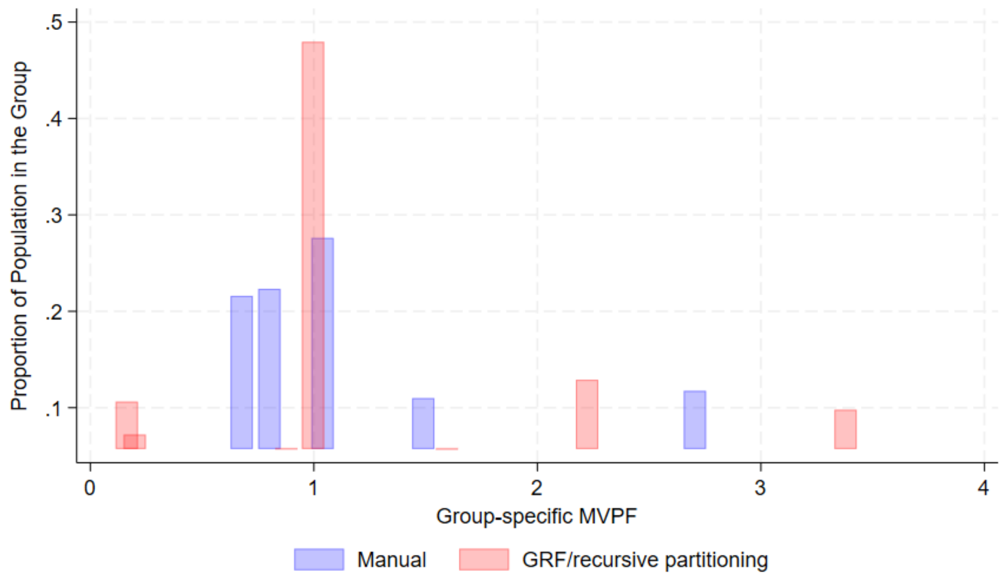

(c) Head Start - With Fiscal Externalities



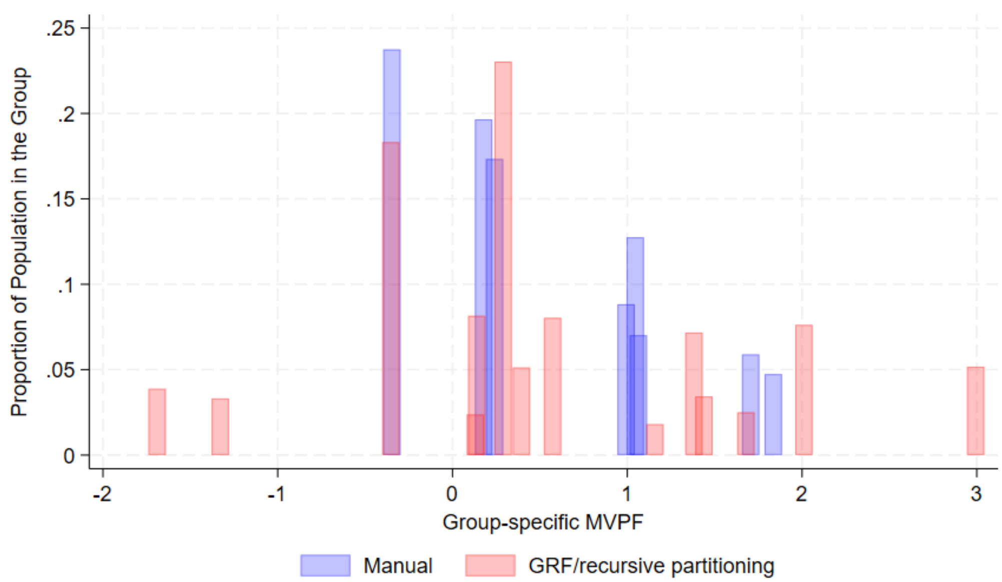

(d) Job Corps - With Fiscal Externalities

*Notes* : Each point represents the benefits of being treated for individuals in the subgroups targeted under a given budget constraint. Benefits are estimated as the ITT in terms of lifetime income aggregating all the individuals in the targeted subgroups, scaled by the total number of individuals in these subgroups. The shaded gray area denotes the 95% confidence interval for these scaled ITT estimates. The standard errors for the Head Start estimates are clustered at the level of the Head Start center to which both treated and control individuals are assigned.

Figure 2: MVPF Distribution across Different Grouping Strategies



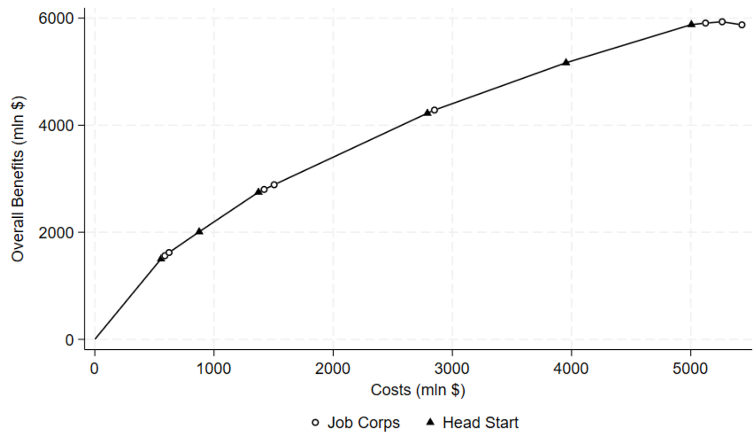

(a) Head Start



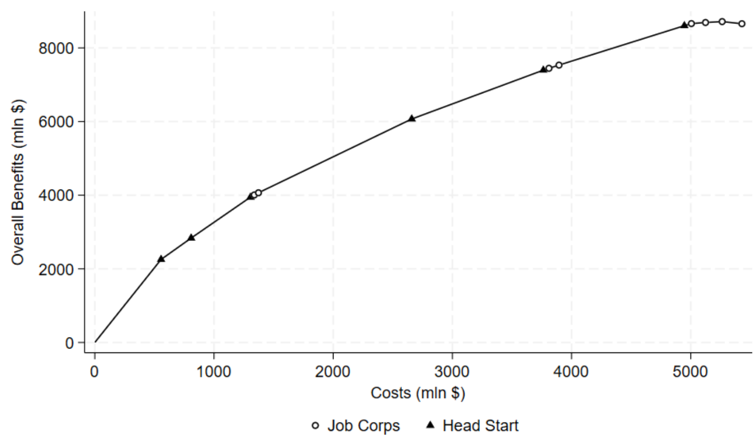

(b) Job Corps

*Notes* : The plot displays, for each subgroup—defined either through manual splitting or via the GRF-recursive strategy—the estimated MVPF on the x-axis and the subgroup's share in the population on the y-axis.

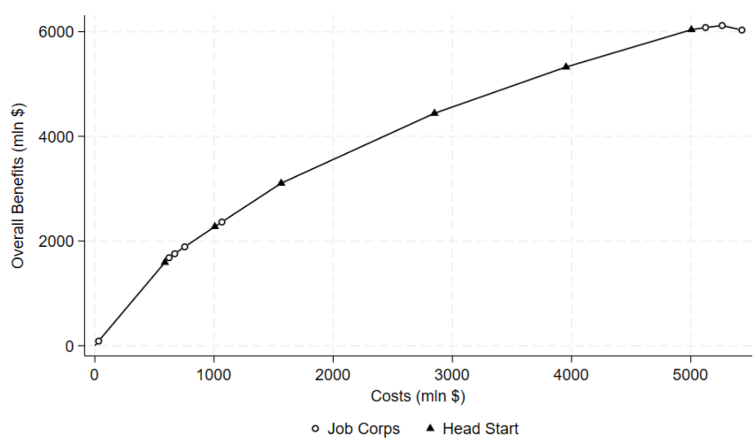
Figure 3: Benefit as a Function of Costs, Assigning Multiple Treatments, Manually constructed Groups




(a) Same Weights




(b) Larger Weights on Head Start Participants




(c) Larger Weights on Job Corps Participants

$Notes$ : Each point represents the benefits of being treated for individuals in the subgroups targeted (either with Head Start or with Job Corps) under a given budget constraint. Benefits are estimated as the sum of the subgroup-specific ITT in terms of lifetime income, scaled by the number of individuals in each subgroup.

# ONLINE APPENDIX

# A Linking the Targeting Rule to the MVPF: Derivation

**Marginal Benefits** The average benefit for the subgroup defined by $X = x$ is $E[Y_i|X_i = x]$. Consider a random allocation of the offer to individuals *within* this subgroup, where a fraction $\delta(x)$ of individuals in this subgroup receive the offer. We can write the average benefit as (see also Kline and Walters, 2016):

$$E[Y_i|X_i = x] = E[Y_i^1|Z_i = 1, X_i = x]\delta(x) + E[Y_i^0|Z_i = 0, X_i = x](1 - \delta(x)) \tag{A.1}$$

and because of random allocation within the group this becomes:

$$E[Y_i|X_i = x] = E[Y_i^1|X_i = x]\delta(x) + E[Y_i^0|X_i = x](1 - \delta(x)). \tag{A.2}$$

The effect of a marginal expansion in the probability of randomly receiving the offer on the average benefits is then:

$$\frac{\partial E[Y_i|X_i = x]}{\partial \delta(x)} = E[Y_i^1|X_i = x] - E[Y_i^0|X_i = x] = E[Y_i^1 - Y_i^0|X_i = x]. \tag{A.3}$$

This corresponds to the expected benefit in the subsample among which the treatment is randomly allocated. It can be rewritten as:

$$E[Y_i^1 - Y_i^0|X_i = x] = \underbrace{E[Y_i^1 - Y_i^0 \mid \mathsf{d}_i^1 \neq \mathsf{d}_i^0, X_i = x]}_{\equiv LATE(x)} \underbrace{P(\mathsf{d}_i^1 \neq \mathsf{d}_i^0 \mid X_i = x)}_{\equiv \pi(x)}. \tag{A.4}$$

**Marginal Costs** The average cost of treating a fraction $\delta(x)$ within the subgroup with $X_i = x$ corresponds to the per person cost $\phi$ times the proportion of individuals enrolling in the program. These can be both individuals who receive the offer ($\mathsf{d}_i^1 = 1$) and individuals who do not receive the offer ($\mathsf{d}_i^0 = 1$):

$$\begin{aligned}\phi P(\mathsf{d}_i = 1|X_i = x) =& \phi(P(\mathsf{d}_i^1 = 1|Z_i = 1, X_i = x)\delta(x) \\ &+ P(\mathsf{d}_i^0 = 1|Z_i = 0, X_i = x)(1 - \delta(x))).\end{aligned} \tag{A.5}$$

Because $Z_i$ is randomly assigned, this can be rewritten as

$$\phi P(\mathsf{d}_i = 1|X_i = x) = \phi(P(\mathsf{d}_i^1 = 1|X_i = x)\delta(x) + P(\mathsf{d}_i^0 = 1|X_i = x)(1 - \delta(x))). \tag{A.6}$$

The cost of a marginal expansion in the probability of randomly receiving the offer is then:

$$\frac{\partial \phi P(\mathsf{d}_i = 1|X_i = x)}{\partial \delta(x)} = \phi(P(\mathsf{d}_i^1 = 1|X_i = x) - P(\mathsf{d}_i^0 = 1|X_i = x)) = \phi\pi(x). \tag{A.7}$$

## A.1 Accounting for Fiscal Externalities

**Marginal Costs** The average cost of treating a fraction $\delta(x)$ within the subgroup with $X_i = x$ corresponds to the per person cost $\phi$ times the proportion of individuals enrolling in the program. These can be both individuals who receive the offer ($\mathsf{d}_i^1 = j, j \in \{h, c\}$) and individuals who do not receive the offer ($\mathsf{d}_i^0 = j$):

$$\begin{aligned} \phi_h P(\mathsf{d}_i = h|X_i = x) + &\phi_c P(\mathsf{d}_i = c|X_i = x) = \\ &\phi_h(P(\mathsf{d}_i^1 = h|Z_i = 1, X_i = x)\delta(x) + P(\mathsf{d}_i^0 = h|Z_i = 0, X_i = x)(1 - \delta(x))) + \\ &\phi_c(P(\mathsf{d}_i^1 = c|Z_i = 1, X_i = x)\delta(x) + P(\mathsf{d}_i^0 = c|Z_i = 0, X_i = x)(1 - \delta(x))). \end{aligned} \tag{A.8}$$

Because of random assignment this can be rewritten as

$$\begin{aligned} \phi_h P(\mathsf{d}_i = h|X_i = x) + &\phi_c P(\mathsf{d}_i = c|X_i = x) = \\ &\phi_h(P(\mathsf{d}_i^1 = h|X_i = x)\delta(x) + P(\mathsf{d}_i^0 = h|X_i = x)(1 - \delta(x))) \\ &+ \phi_c(P(\mathsf{d}_i^1 = c|X_i = x)\delta(x) + P(\mathsf{d}_i^0 = c|X_i = x)(1 - \delta(x))). \end{aligned} \tag{A.9}$$

The cost of a marginal expansion in the probability of randomly receiving the offer is then:

$$\begin{aligned} \frac{\partial(\phi_h P(\mathsf{d}_i = h|X_i = x) + \phi_c P(\mathsf{d}_i = c|X_i = x))}{\partial \delta(x)} = &\phi_h(P(\mathsf{d}_i^1 = h|X_i = x) - P(\mathsf{d}_i^0 = h|X_i = x)) \\ &- \phi_c(P(\mathsf{d}_i^0 = c|X_i = x) - P(\mathsf{d}_i^1 = c|X_i = x)). \end{aligned} \tag{A.10}$$

Due to assumption 2, the first term on the right hand side corresponds to the product of the cost of the focal program times the proportion of compliers $P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 \neq h)$, i.e. all participants switching to $h$ either from $c$ or from $n$ when receiving the offer. The second term corresponds to the product of the cost of the competing program and the proportion of individuals who switch from $c$ to an alternative program when $Z_i = 1$. Given assumption 2, they can only switch to $h$. Hence, this quantity corresponds to $P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 = c)$. The latter equation can thus be written as:

$$\phi_h \frac{\partial P(\mathsf{d}_i = h|X_i = x)}{\partial \delta(x)} + \phi_c \frac{\partial P(\mathsf{d}_i = c|X_i = x)}{\partial \delta(x)}$$
$$= \phi_h \underbrace{P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 \neq h|X_i = x)}_{\equiv \pi_h(x)} - \phi_c \underbrace{P(\mathsf{d}_i^1 = h, \mathsf{d}_i^0 = c|X_i = x)}_{\equiv \pi_c(x)}. \tag{A.11}$$

# B Head Start: Sample, Program Benefits and Costs

**Sample.** We use data from the HSIS which follows children from the time of randomization until the third grade. We use both the cohort which is 3 years old and the one which is 4 years old at the time of randomization. We combine information from two surveys. From the baseline survey conducted in Fall 2002 (at the time of randomization), we collect information including the random assignment and baseline characteristics of the children. From subsequent follow-up survey (Spring 2003), we collect information on childcare use and cognitive test scores. To investigate the extent of within-year changes in program participation, we cross tabulate program participation using the Fall 2002 survey (at the start of the first Head Start year) and the Spring 2003 survey (at the end of the first Head Start year). We find a high degree of overlap in childcare use – over 90% of children report identical childcare use in the two surveys. Similar to Kline and Walters (2016) we include in the sample all applicants with nonmissing baseline characteristics and spring 2003 test scores (3,574 individuals).

Descriptive statistics from our final sample indicate that the treatment and control groups are very similar across many dimensions – in demographic characteristics, baseline scores (taken in the baseline survey), and characteristics of the Head Start centers they applied to (see Table A1). The statistically significant differences are minimal in economic terms, except for the treatment group having a higher proportion of children in need of special education, but the two groups are fairly similar along any other dimension which can indicate children in the treatment group being more disadvantaged. If anything, they do slightly better in the PPVT baseline score (but no differences in WJIII test score).

**Benefits.** We measure Head Start' benefits starting from the estimated impact of the program on children's test scores. As test scores, we use an average of the scores obtained by the children in the Peabody Picture Vocabulary Test (PPVT) and in the Woodcock-Johnson III (WJIII) Pre-Academic Skills test. This average is then normalized to have mean zero and variance one in the control group, where normalization is done separately for age 3 and age 4 cohort. PPVT measures the comprehension of spoken words in standard English, and its levels and scales are comparable across years. The Woodcock-Johnson III (WJIII) Pre-Academic Skills comprises three tests, i.e., Letter-Word Identification, Spelling, and Applied Problems. Both PPVT and WJIII scores obtained from HSIS data have been transformed using item response theory to make them comparable across ages. We measure test scores one year after enrollment (Spring 2003) because age-3 cohort children in the control group could eventually be admitted to Head Start in the year they turn 4.

These test score gains are then converted in impact on lifetime earnings. As pointed out in Kline and Walters (2016), focusing only on lifetime earnings ignores a bunch of additional potential benefits of the program (such as reduced criminal activity, improved health or parental labor supply decisions), thus the effect is to be interpreted as a lower bound.

We follow Kline and Walters (2016) to recover lifetime earnings starting from measured test scores. Their approach is to attribute a price to human capital and multiply observed test scores times this price. Tennessee STAR experiments suggest that program's induced early childhood gains in test scores predict earnings gains despite the fadeout of test scores impact in the medium term (Chetty et al., 2011). We exploit this finding and estimates from the STAR study to argue that a 1 s.d. increase in test scores at age 3 increases lifetime earnings by 10 percent (this is to be conservative, as other studies find much larger effects of test scores increases on earnings, see Appendix C in Kline and Walters (2016) for a discussion). As a baseline for the discounted lifetime earnings for the HSIS population at age 3, we assign the value of 343,492 USD. This follows Chetty et al. (2011) who estimate the present value of lifetime earnings at age 3 to be 438,000 USD.[36] This is an an average for the U.S. and is then scaled by a term which takes into account that household who participate in HSIS earn 46% of the national average and that intergenerational income elasticity in the U.S. is estimated to be about 0.4. Scaling 438,000 by (1-(1-0.46)×0.4) we obtain 343,492 USD. Hence, the overall effect of Head Start on lifetime earnings is computed as $0.1 \times 343,492 \times LATE$, where $LATE$ is the impact of Head Start in terms of test scores s.d. The measure we exploit to compute benefits is $pre - tax$ lifetime earnings.

**Costs.** Per-pupil expenditure in Head Start is about 8,000$ (DHHS, 2013) and we use this as the cost of one year of Head Start participation. The public expenditure on alternative programs depends on their costs and who finances them. Following the discussion in Kline and Walters (2016), we assume alternative programs have the same cost of Head Start (hence 8,000$ per year). However, not all this cost is paid by public funding: in fact about 25% of the alternative programs attended by HSIS participants are primarily funded by parents' fees. Hence, we assign a cost of 0.75×8,000=6,000$ to the other childcare programs. This is admittedly an upper bound; because 39% of the childcare centers which were asked about their source of funding did not answer the question, if we assume they are all primarily paid by parents' fees we end up with the government paying only for 50% of the alternative programs.

To compute the *total* cost of the program, which needs to be compared to the budget constraint, we need to multiply the per-person cost to the number of participants. In this application we consider a

[36] Chetty et al. (2011) predicts 522,000 USD at age 12, which is then discounted at a 3-percentage discounted rate to recover the present value at age 3.4, the average application age for Head Start at $t = 1$.

population of 900,000 potentially targeted children, as this is the number of children who enrolled in Head Start in 2013 (between 2000 and 2020 the number of children enrolled in Head Start fluctuated around an average of approximately 900,000 children DHHS, 2021). In our application, this is the maximum number of children who may be *offered* a place, while the total number of attendees depends on the compliance rate. Because of imperfect compliance we end up considering a lower number of participants as compared to actual participants in each year. As a consequence, the total cost when the program is offered to the whole population, which we estimate being about $5.7 billions, is lower than the yearly cost of Head Start in 2013 (about $7.6 billions in 2013).

# C Job Corps: Sample, Program Benefits and Costs

**Sample.** Four surveys were conducted during the NJCS to allow for program evaluation: one at baseline (at or shortly after the random assignment), and the others at 12, 30 and 48 months after random assignment. For all interviews in the survey and for both the treatment and the control group, detailed information was collected on the duration of participation in Job Corps, training choices other than Job Corps, individual characteristics, and labor market outcomes.[37] Sample weights are also provided to adjust for the sample and survey designs.

Following Schochet et al. (2008), we focus on the sample of individuals who completed the last survey (around 80%). In addition, we exclude 219 for whom baseline information is missing (less than 2% of the original sample in Schochet et al. (2008)). Descriptive statistics shows that individuals in the treatment and control group on average are similar along different demographic and past-experience dimension (Table A2).

**Benefit** McConnell and Glazerman (2001) conduct an extensive cost–benefit analysis of Job Corps exploiting information from the NJCS. We rely on their calculations for our definitions of benefits and costs, although we consider only the private benefit of Job Corps, where we define as private benefits only the increase in the lifetime earnings of individuals. We thus ignore other potential benefits such as reduced crime activty (as reported in Schochet et al., 2008).

To compute the lifetime gain of Job Corps, we start from the observed earnings in the last quarter of observation and then impose realistic assumptions on their lifetime evolution. In particular, we rely on the information on the weekly earnings in the last quarter of the fourth year after random assignment. For some individuals information about earnings in the last quarter is missing because the last interview is conducted $during$ the 16th quarter after random assignment. For them, we use information about

[37] The 12-, 30- and 48-month surveys contain information on enrollment dates in different education/training programs, which allow to track enrollment in Job Corps and other programs.

the reported average weekly income in the 15th quarter after random assignment. For individuals who have weekly earnings below 10$ we assign them zero earnings.

As in McConnell and Glazerman (2001) we assume that the earnings of all individuals increase with worker age, and that the dollar value of the gain in income due to Job Corps persists. This means that the difference in income between individuals is constant over time, but the proportionate difference is decreasing, because overall income is increasing. In fact, with this assumption, the proportionate gain decreased quite substantially, being halved within 10 years. We assume that the control group's earnings increase with worker age by $8.1\%$ in the first year, and then at a rate that decays by 0.24 percentage points annually. The treated group's income increases accordingly to maintain the dollar value of the difference constant. We assume 40 years of expected working life for the individuals, and we consider as first-year income the year-equivalent of the weekly earnings in the last quarter individuals are observed in the survey.[38] To calculate the present value, we allow for a discount rate of $4\%$, which is about the average real rate of return on 30-year treasury bonds for the period 1990s-2000s. Also for Job Corps we consider $pre - tax$ lifetime earnings.

The cost of Job Corps is estimated to be about 16,489$ per capita. As a cost of alternative education we consider the cost of 6 months of high school (4,615$), because high school together with GED is the most common alternative program in our data.

**Costs** The cost of Job Corps for the average student is about 16,489$. This number includes the reported program costs (incurred either by the centers operating the program or by national and regional Job Corps offices, which amount to about 14,901$), the unreported program costs (which include he costs of *operating* the national and regional offices and the cost of donated items and services, 551$), and the capital costs (depreciation and opportunity costs for land, buildings, equipment and furniture, 1,037$) (McConnell and Glazerman, 2001). As a cost for the alternative education obtained when not in Job Corps we consider the cost of spending 6 months in high school. High school, together with GED preparation is the most common alternative to Job Corps in the control group. On average Job Corps participants spend 7.5 months in the program. We consider 6 months in high school instead of 7.5 to compute the cost of competing programs to take into account the cheaper GED training alternative. McConnell and Glazerman (2001) estimate a cost of 6.41$ per hour for high school. We consider 30 hours per week, 4 weeks per month. Overall, we thus consider a cost of 4,615$ for competing programs.

Similar to Head Start, we need to define the cost in terms of the total number of participants. We

[38]Note that Schochet et al. (2008) find no difference in the average earnings of the two groups when looking at official earnings records. However, these data underestimate the effect of the treatment on earnings, according to the surveys, suggesting that they may miss some earnings from casual jobs because nonrespondent bias cannot explain all the differences between the survey and the official earnings record. In addition, our assumption on the persistence of the difference in dollar values, not in proportionate value, allows for a fading out of the proportionate gap between the two groups.

consider 60,000 potential participants (as Job Corps used to enroll 60,000 new attendees every year when the NJCS was conducted), which need then to be scaled by the compliance rate to obtain the number of participants.

# Appendix Tables and Figures

Table A1: Descriptive Statistics in the Baseline Survey – Head Start

| | Control Group | Treatment Group | Difference |
|---|---|---|---|
| Household Size | 4.54 | 4.58 | -0.03 |
| Female=1 | 0.50 | 0.51 | -0.01 |
| Black=1 | 0.29 | 0.31 | -0.01 |
| Hispanic=1 | 0.37 | 0.37 | -0.00 |
| English as Home Language=1 | 0.72 | 0.71 | 0.01 |
| In Need of Special Education=1 | 0.10 | 0.13 | -0.02** |
| PPVT Score at Baseline | 2.50 | 2.47 | 0.02* |
| WJIII Score at Baseline | 3.46 | 3.47 | -0.00 |
| Number of Siblings | 1.36 | 1.43 | -0.07* |
| Both Parents in the Household=1 | 0.49 | 0.49 | -0.00 |
| Teen Mother=1 | 0.17 | 0.16 | 0.01 |
| Mother not Married=1 | 0.39 | 0.40 | -0.00 |
| Mother Separated=1 | 0.15 | 0.16 | -0.00 |
| Mother Education | | | |
| High School Diploma=1 | 0.32 | 0.33 | -0.01 |
| More than High School=1 | 0.27 | 0.29 | -0.01 |
| Urban Area=1 | 0.83 | 0.83 | -0.00 |
| HS Center Quality Index | 0.70 | 0.70 | -0.00 |
| HS Center Transportation Available | 0.64 | 0.65 | -0.00 |
| Other Center Quality Index | 0.58 | 0.58 | 0.00 |
| Other Center Transportation Available | 0.50 | 0.49 | 0.00 |
| N | 1,318 | 2,256 | |

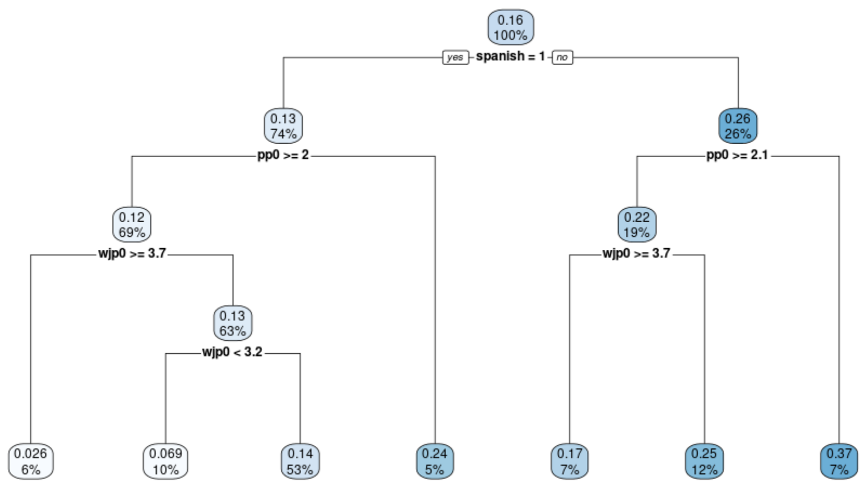
*Notes* : This table reports the mean value of different individual characteristics comparing the control to the treatment group. The center quality index (obtained directly from the HSIS data) and the availability of transportation refer to the center of random assignment. Significance level (t-test for the difference between the average for the treatment group and the one for the control group=0): *** 1%, ** 5%, * 10%

Table A2: Descriptive Statistics in the Baseline Survey – Job Corps

| | Control Group | Treatment Group | Difference |
|---|---|---|---|
| Female | 0.40 | 0.41 | -0.00 |
| Age | 18.36 | 18.39 | -0.03 |
| Education in past year (months) | 4.48 | 4.44 | 0.04 |
| Job in past year (months) | 3.61 | 3.65 | -0.03 |
| Has High School at randomization | 0.23 | 0.22 | 0.00 |
| N | 4,395 | 6,699 | |

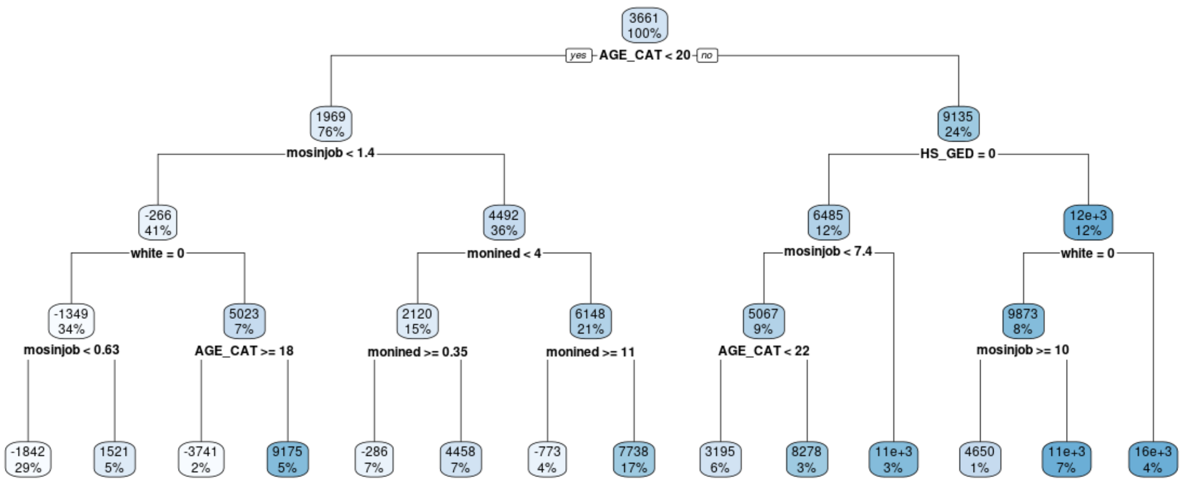
*Notes* : This table reports the mean value of different individual characteristics comparing the control to the treatment group. "Education" and "Jobs" in the past year are the numbers of months spent in education and jobs in the year before random assignment, respectively. Significance level (t-test for the difference between the average for the treatment group and the one for the control group=0): *** 1%, ** 5%, * 10%

Figure A1: Recursive Partitioning of Predicted CATE by Covariates, Head Start



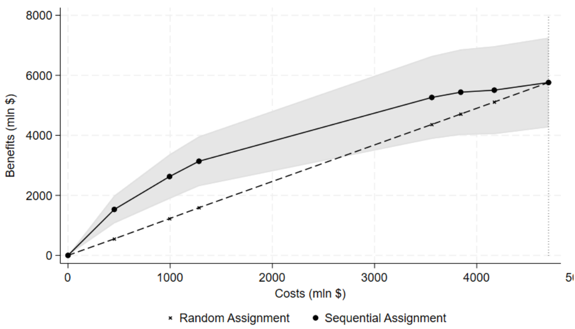

$Notes$ : The tree illustrates how the sample is split via recursive partitioning into subgroups to capture heterogeneity in the predicted conditional average treatment effect (CATE), which was computed via generalized random forest. The CATE we consider is in terms of the standardized average of the PPVT and WJIII scores one year after enrollment, and the average of all individual CATE for each groups at each node is reported on the top of each node in the figure. We specified a maximum depth of 5, to avoid overfitting. In the figure, at each node a condition is specified; the branches to the bottom-left of the node comply with the condition, the branches to the bottom-right do not comply with that condition. The darkest colors at each node correspond to higher values of the average predicted CATE within the node. “spanish” refers to whether the children are spanish speakers. “pp0” is the score in the baseline PPVT test (mean 3.47, s.d. 0.23). “wjp0” is the score in the baseline WJIII test (mean 2.48, s.d. 0.43).

Figure A2: Recursive Partitioning of Predicted CATE by Covariates, Job Corps

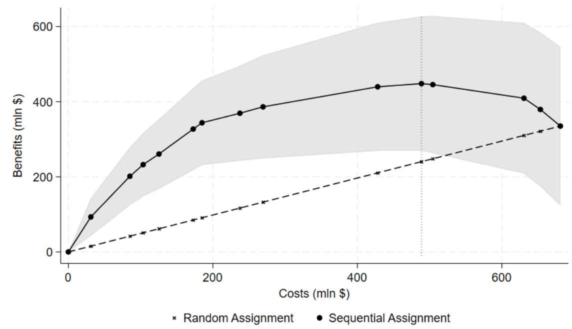



*Notes* : The tree illustrates how the sample is split via recursive partitioning into subgroups to capture heterogeneity in the predicted conditional average treatment effect (CATE), computed via generalized random forest. The average of all individual CATE for each groups at each node is reported on the top of each node in the figure. We specified a maximum depth of 5, to avoid overfitting. In the figure, at each node a condition is specified; the branches to the bottom-left of the node comply with the condition, the branches to the bottom-right do not comply with that condition. The darkest colors at each node correspond to higher values of the average predicted CATE within the node. "AGE_CAT" refers to age of the participant (ranging between 16 and 24, mean 18.4, s.d. 2.1). "mosinjob" refers to the number of months spent in employment in the last month (mean 3.6, s.d. 4.3). "monined" refers to the number of months spent in education in the last month (mean 4.5, s.d. 4.5). "HS_GED" is a dummy for whether the person completed high school or GED at the time of application in Job Corps. "white" is a dummy for whether the person is white (where 0 includes all the alternatives: Black, Hispanic, native American).

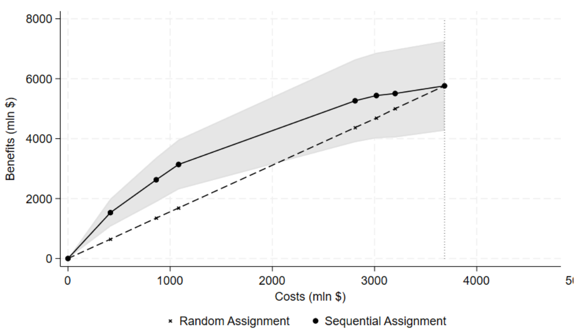
Figure A3: Benefit as a Function of Costs, GRF Grouping



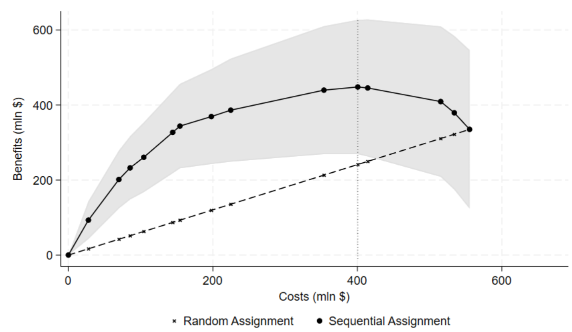

(a) Head Start




(b) Job Corps




(c) Head Start - With Fiscal Externalities




(d) Job Corps - With Fiscal Externalities

*Notes* : Each point represents the benefits of being treated for individuals in the groups targeted under a given budget constraint. Benefits are estimated as the ITT in terms of lifetime income aggregating all the individuals in the targeted subgroups, scaled by the total number of individuals in these subgroups. The shaded gray area denotes the 95% confidence interval for these scaled ITT estimates. The standard errors for the Head Start estimates are clustered at the level of the Head Start center to which both treated and control individuals are assigned.