

DISCUSSION PAPER SERIES

IZA DP No. 12839

**Understanding Program Complementarities:
Estimating the Dynamic Effects of a Training
Program with Multiple Alternatives**

Antonio Dalla-Zuanna
Kai Liu

DECEMBER 2019

DISCUSSION PAPER SERIES

IZA DP No. 12839

Understanding Program Complementarities: Estimating the Dynamic Effects of a Training Program with Multiple Alternatives

Antonio Dalla-Zuanna

Institute for Fiscal Studies and Norwegian School of Economics

Kai Liu

University of Cambridge and IZA

DECEMBER 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

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

Understanding Program Complementarities: Estimating the Dynamic Effects of a Training Program with Multiple Alternatives*

In this paper we estimate the causal effect of a training program for disadvantaged youths on their long-run labor market outcomes. Individuals receive lottery offers to participate in the program, but are allowed to choose when to leave the program and to participate in alternative programs. We consider a multistage decision setting, where individuals sequentially select which program to participate in at every stage. The standard IV estimator using initial random assignment as instrumental variable identifies a weighted average of the effects of the treatment for subgroups of individuals differing in terms of potential duration of participation and choice of alternative programs. We estimate a sequential choice model that allows us to estimate the effect of the treatment for these different subgroups separately. We use the estimated model to investigate the dynamic complementarity between different training programs and explore program targeting to improve the cost-effectiveness relative to the existing program.

JEL Classification: J0, H4

Keywords: training, program evaluation, dynamic treatment effects, experiment

Corresponding author:

Kai Liu
Faculty of Economics
University of Cambridge
Sidgwick Avenue
Cambridge, CB3 9DD
United Kingdom
E-mail: kai.liu@econ.cam.ac.uk

* We thank Marc Chan, Robert Moffitt, Magne Mogstad, the participants of the 2018 IRP Summer Workshop in Madison, Wisconsin and seminar participants at several institutions for valuable comments. Taehee Oh provided excellent research assistance.

1 Introduction

Many less-skilled workers do not accumulate skills from work experiences as they move from one low-wage job to another, often with long unemployment spells in between. For these individuals, government-sponsored training programs are important policy tools to promote skill accumulation and employment prospects. However, existing evidence suggests that returns from participating in these means-tested training programs are typically low, and, even among the few programs generating positive returns, still fail to pass a social cost–benefit test (Barnow and Smith, 2016).

Thanks to the availability of randomized experiments, we have credible evidence on the causal impact of several training programs. For instance, in the National Job Corps Study (NJCS), applicants for the Job Corps program were randomly assigned into either a treatment group in which individuals were offered the opportunity to enroll in Job Corps, or a control group, in which individuals were excluded from Job Corps. Arguably, the experimental feature of program assignment addresses the problem of endogenous self-selection into the program. Exploiting the randomization in assignment in the NJCS, Schochet, Burghardt, and McConnell (2008) find that program participation increases educational attainment, employment and earnings for several post-program years.¹

Yet, even with random assignment in program access, the estimated treatment effects are difficult to interpret in the presence of (i) multiple outside options and (ii) dynamic selection. Two empirical patterns from the NJCS data motivate our paper. First, 68.3% of individuals in the control group enrolled in other education or training programs during the 48 months after random assignment. Second, there are large variations in time spent in Job Corps in the treatment group, and a substantial percentage of program participants enrolled in other types of training programs after some experience with Job Corps. For instance, 20% of the treated individuals enrolled for less than 3 months (relative to the mean duration of 8 months). Among these early-leavers, more than half subsequently enrolled in alternative training programs. In addition, we find substantial heterogeneity in outcomes depending on both the time spent in Job Corps and participation in alternative training programs.

Using data from the NJCS, this paper provides a comprehensive analysis of the causal effect of the

¹Other training programs that have been evaluated using a randomized control trial include the National Supported Work Demonstration (LaLonde, 1986; Ham and LaLonde, 1996; Calónico and Smith, 2017), and the U.S. Job Training Partnership Act (JTPA) program (see e.g., Bloom, Orr, Bell, Cave, Doolittle, Lin, Bos, et al., 1997; Heckman, Smith, and Taber, 1998; Heckman, Hohmann, Smith, and Khoo, 2000).

Job Corps program in a multistage decision setting, where individuals sequentially select which type of program to participate in every stage. One central feature of our paper that differs from the rest of the literature is that we distinguish the outside options facing individuals at each decision stage. In particular, both Job Corps participants and nonparticipants can choose to enroll in alternative training programs. Exploiting the experimental feature, we begin by building a multistage potential outcome framework with multiple alternatives in each period. Relative to a static framework with mutually exclusive choices, our framework features a richer set of substitution patterns, including *intertemporal* substitution in alternative training (where an individual delays the entry into the alternative program because of Job Corps) and *complementarity* between the Job Corps experience and alternative training (where experience in Job Corps leads the individual who otherwise would not enroll in any training to pursue additional training). In terms of program returns, our framework allows us to study the dynamic treatment effects of Job Corps and dynamic complementarity between the Job Corps experience and alternative training programs. Dynamic complementarity may arise if the Job Corps experience raises human capital and increases the return from subsequent training (Cunha and Heckman, 2007).

Without any parametric restriction, the experimental variation from the NJCS only allows us to identify the local average treatment effect (LATE) of the Job Corps program. We show that this LATE estimator is a mixture of the causal effect among different types of compliers differing in terms of the time they spend in Job Corps and their next-best alternatives. Consistent with Kirkeboen, Leuven, and Mogstad (2016), we show that it is not possible to identify these parameters because we do not have information on individuals' rankings of the available alternatives. We discuss the limitation of using instrumental variables (IVs) in general to draw inference about the interaction effects between Job Corps and alternative training programs. In particular, when individuals self-select based on individual-specific gains, comparing the average treatment effects from different types of compliers provides biased evidence of program complementarities.

Following the empirical approach in Heckman and Navarro (2007) and Heckman, Humphries, and Veramendi (2016), we instead construct a sequential choice model to estimate the dynamic treatment effects of the Job Corps program and how the returns interact with alternative training programs. We assume that a low dimensional set of unobservables affects both selection over time and the final outcome of interest. Even though our assumption places restrictions on the joint distribution of outcomes

and choices, our sequential choice model allows for a variety of selection schemes, including selection into training on levels and gains. Identification of these class of models rely on the assumption that conditional on these latent factors the choices and outcomes are independent, and on a measurement system to identify the unobservable that determines selection into treatment types and affecting outcomes.² In our case, initial randomization of Job Corps offers is used as exclusion restriction to help secure identification.

The model is estimated using data from the NJCS. We find that a substantial percentage of the individuals (28%) engage in intertemporal substitution, where they postpone other types of training until after they have enrolled in Job Corps. Relatively few individuals (7%) complement their Job Corps experience with additional training, whereas a larger percentage (21%) substitute Job Corps for alternative training. By comparison, a static framework with mutually exclusive choices would significantly overestimate the percentage of individuals engaging in program substitution. In terms of labor market outcomes, we find evidence for dynamic complementarity between a spell in Job Corps lasting between 3 and 6 months and training obtained from other programs, whereas longer spells in Job Corps tend to reduce the return from training obtained outside Job Corps. We also present the dynamic treatment effects of Job Corps at each decision stage and decompose these into a direct effect of receiving the treatment in that period, and a continuation value of moving to the following period. For instance, in the first period, we find that the direct effect of Job Corps is negative, whereas the continuation value beyond the first period is large and positive. This suggests that the benefits of the treatment arise only after the completion of a sufficiently long training period. We also show the decomposition of the overall estimated LATE, which shows positive but small effects both on earnings and on employment. We find that this overall effect is a weighted average of the negative effect of the treatment for those individuals who stay in Job Corps only for short periods and the larger and positive effect for those who stay longer.

We evaluate potential policy with the aim of improving the cost-effectiveness of the Job Corps program. We consider a targeting policy, where instead of offering Job Corps to all or randomly to a subset of individuals, we offer Job Corps to a targeted subpopulation who have the highest predicted gain from receiving offers. Importantly, an individual's labor market and education history at the

²A similar approach to model sequential selection has been used by Fruehwirth, Navarro, and Takahashi (2016) to analyze the effect of grade retention, by Heckman, Humphries, and Veramendi (2018) to analyze the effect of school choices on earnings and health and by Rodríguez, Saltiel, and Urzúa (2018) to analyze the effect of on-the-job training.

time of randomization reveals important information about her unobserved heterogeneity. Exploiting these observed characteristics, we use our estimated model to compute the posterior distribution of unobserved heterogeneity and the associated expected potential gain from receiving a Job Corps offer for each individual. Based on the predicted gains, we then rank the population which a policy maker considers to offer Job Corps to. We show that program targeting can significantly improve the cost-effectiveness of the program. For instance, 1\$ spent in the program targeting the top 30% of the population would raise the expected after-tax lifetime income of participants by approximately 1\$. By comparison, 1\$ spent in the existing program only raises the expected after-tax income of participants by about 30 cents.

This paper is closely related to two papers analyzing the role of program substitution in program evaluation. Heckman, Hohmann, Smith, and Khoo (2000) shows the importance of considering substitution and dropout biases in estimating the effects of job training. They separate between the *effect of the program* that compares treated individuals to controls who are allowed to select into any alternative and the *effect of the treatment* that compares treated individuals to a counterfactual scenario where they do not obtain any training. In Heckman, Hohmann, Smith, and Khoo (2000), the dropout behavior refers to dropping out of the program prior to receiving training, whereas we consider the issue arising by individuals who drop out at different points after having received the training. More recently, Kline and Walters (2016) provided a framework for evaluating the Head Start program, taking into account that individuals may self-select into an alternative educational program that might be close substitute to Head Start. Both of these papers rely on a static model where choices are mutually exclusive, and hence, different programs must be either substitutes or independent from each other. This paper considers a sequential choice setting, that opens up the possibility for two programs to be complements. This allows us to test for dynamic complementarity in terms of program returns and understand the dynamic program effects in terms of direct effects and continuation value.

Our paper is also related to a strand of literature analyzing the dynamic selection and dropouts of training programs. We contribute to this literature by combining a model of dynamic selection, together with multiple outside options, that enable us to explain richer substitution patterns and understand how the returns to one program may be affected by participation in another program. In addition, we contribute to this literature by considering that the effect of training may vary with different duration

in the program. Ham and LaLonde (1996) and Eberwein, Ham, and LaLonde (1997) consider the effect of participating in training programs on the duration of subsequent employment and unemployment spells. They assume that the treatment effect is constant with respect to the time spent in the program, whereas we are interested also in understanding the effect of program’s duration.³ Abbring and Van den Berg (2003) allow for the effect to change with the duration of the spell in the program, but this is at the expenses of ruling out the heterogeneous effect of the treatment with respect to the unobservable characteristics of the individuals which determine their choices in terms of duration in the program. In our framework, we explicitly allow for this type of heterogeneity.

The effect of spending different periods in the training is the focus of Heckman, Smith, and Taber (1998), who show that the Wald estimator is biased when only the individuals who spend long periods in the programs are considered treated (whereas those dropping out earlier are considered as not receiving any treatment), in that it assumes that individuals treated for shorter periods have the same outcome as if they had not been treated. Exploiting a dynamic version of the matching assumptions, Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) estimate the effect of length of exposure to academic and vocational training (two of the different training programs offered) within Job Corps. They also allow for unobserved factors that are time invariant and whose influence can thus be netted out using a difference-in-difference estimator (where the first difference is taken with respect to pretreatment levels).⁴ More recently, Rodríguez, Saltiel, and Urzúa (2018) use a similar approach to ours to estimate the dynamic treatment effects of an on-the-job training program in Chile and investigate the issue of complementarity and substitutability between different spells of the *same* training program.⁵ Differently from our paper, they focus on individuals who are already employed (and mostly in the formal sector).

³These papers also show that the estimation of the program effect on earnings is complicated by the fact that the program also affects selection into employment. Lee (2009), Blanco, Flores, and Flores-Lagunes (2013a) and Blanco, Flores, and Flores-Lagunes (2013b) address this issue by focusing only on the individuals who would be employed in both treatment status. They estimate bounds for the effects of Job Corps on earnings and wages of this subgroup under a “weak monotonicity” assumption, which implies that treatment assignment can affect selection in only one direction (we do not make this assumption, and thus, in our setting, individuals who would be employed when assigned to the control group are allowed to be unemployed if assigned to the treatment group). Frumento, Mealli, Pacini, and Rubin (2012), instead, use a different approach and jointly estimate the probability of being employed in both group and the final outcome.

⁴Other papers in the literature that estimate the dynamic treatment effects of training programs by invoking the sequential conditional-independence assumptions (a dynamic version of the matching assumption in which treatment is independent from potential outcomes conditional on observables) include Lechner (2002), Fitzenberger, Osikominu, and Völter (2008), and Lechner and Miquel (2010). Note that, unlike in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), we allow the unobserved factor to affect the potential outcomes differently under different durations, and, as a result, they will not be differenced out.

⁵For instance, they find that obtaining a second year of training does not complement the training obtained in the first year.

Moreover, the complementarity patterns we consider are not restricted to different spells within the same program, but involve the different options available outside the program of interest. We also conduct a comprehensive cost–benefit analysis and explore the implications of selective targeting to improve the cost-effectiveness of Job Corps.

Although our paper focuses on the evaluation of job training programs, our framework can potentially be applied to many different contexts. For instance, many development randomized control trials (RCT) studies rely on “encouragement design” where the randomization is based on access to or offer of a particular program rather than the final take-up of the program (Duflo, Glennerster, and Kremer, 2007). In many papers, the intention-to-treat (ITT) effects are often divided by changes in the probability of treatment induced by the policy intervention. This type of scaled estimate of the ITT effects is interpreted in the same way as IV estimates, and, therefore, may be hard to interpret and relevant for policy in the presence of multiple outside options and dynamic selection.

The paper is organized as follows. Section 2 describes the NJCS and related data and conducts an impact evaluation of Job Corps. Section 3 presents the potential outcomes framework and interprets the IV estimates using experimental variation. It further specifies parameters of interest beyond the usual LATE. In Section 4, we specify the sequential selection model and discuss assumptions for identification and estimation strategies. Section 5 reports the empirical results. Section 6 performs a cost–benefit analysis and evaluates program targeting in improving the cost-effectiveness of the program. Section 7 concludes.

2 Experimental Evaluation of the Job Corps (NJCS)

2.1 The National Job Corps Study

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”.⁶ Job Corps offers a comprehensive array

⁶More specifically, there are 11 criteria that should be satisfied to be eligible to participate in Job Corps: age 16–24 years; a legal US resident; receiving welfare or food stamps or having an income less than 70% below the DOL’s “lower living standard income level”; living in an environment characterized by a disruptive home life, high crime rates, or limited job opportunities; needing additional education, training, or job skills; free of serious behavioral problems; a clean health

of education and training services, including the teaching of academic, vocational, and employability skills. Instruction is personalized depending on the initial level of preparation of the students and on the pace at which students progress. In the sample we analyze, about 42% of the participants in Job Corps obtained training to get general education development (GED), 86% obtained some sort of vocational training, and 3% obtained high school education. In addition to the training and education at the center, Job Corps provides placement services for finding a job or pursuing additional training.

One 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. Most academic and vocational training in centers requires participants who do not reside at the centers to stay at the campus for the whole day. In the period we study, 110 centers ranging in size between 200 and 2,600 slots were available throughout the country.

The data set used in this paper comes from the NJCS, an experimental evaluation of Job Corps that was conducted in the mid-1990s. Youths who applied to the program and were found eligible for participation in the period between November 1994 and December 1995 were randomized into either a control or treated group.⁷ Applicants assigned to the treatment group were able to enroll in Job Corps. Among them, the vast majority of those who enrolled in Job Corps (92%) did so within the first 3 months after random assignment. Applicants assigned to the control group were excluded from Job Corps for an embargo period of 3 years.

The initial research sample in the NJCS includes about 6,000 individuals in the control group and about 9,400 in the treated group. For cost reasons, data were collected from only a random subset of those randomly assigned to the treated group. 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

history; an adequate child care plan (for those with children); registered with the Selective Service Board (if applicable); having parental consent (for minors); and judged to have the capability and aspirations to participate in Job Corps (Schochet, Burghardt, and McConnell, 2008).

⁷The process of outreach and admission in Job Corps works as follows: to recruit participants to the program, local agencies in disadvantaged communities interact with other organizations working with youths (such as schools or employment services), providing them information and verifying the eligibility of the applicants. Eligible youths are then assigned to a center in the first month after eligibility is assessed (Schochet, Burghardt, and McConnell, 2008). Note that 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, Burghardt, and McConnell, 2008).

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.⁸

2.2 Sample Selection, Choice Set and Outcomes

Similar to Schochet, Burghardt, and McConnell (2008), we focus on the sample of individuals who completed the last survey (around 80%). In addition, we exclude individuals who had missing information regarding the duration of participation in Job Corps (around 3% of the sample completing the last survey) and for whom we did not observe educational attainment before applying to Job Corps (7%). Because the outcome of interest is earnings at the end of the fourth year after random assignment, we further exclude the individuals who did not report this measure (1.6%), the few individuals who were still in full-time education at the end of the fourth year after random assignment (3.7%), and the individuals who completed Job Corps more than 3 years after random assignment (1.7%).⁹ Among these individuals, 142 were assigned to the control group and enrolled in Job Corps after the embargo period (this is 3.2% of the control group). We also exclude the individuals who violated the embargo period because this is a very small proportion of the control group (1.3%). Therefore, in our sample, none of the individuals in the control group participated in Job Corps. Our final sample includes 9,429 individuals, 5,648 of whom belong to the treated group, and 3,781 to the control group. In Section 2.3, we show that our sample replicates the experimental impact of the Job Corps documented in Schochet, Burghardt, and McConnell (2008) well.

Around 72% of the treated group enrolled in Job Corps. Conditional on enrollment, the average enrollment duration in Job Corps is about 8 months, ranging between less than 1 month to more than 30 months. We categorize the duration of participation into three groups. The first group enrolled for

⁸The 12-, 30- and 48-month surveys ask information on start and end date of employment for every job the respondent held in the months between the last and the current survey. They also ask information on either hourly, daily, weekly, or monthly labor earnings, before taxes and deduction, including tips and regular overtime pay. This information, together with information on day and hours of work, allows to create weekly measures of labor earnings for each individual. Similarly, the surveys ask questions on enrollment dates in different education/training programs, which allow to track duration in Job Corps and subsequent enrollment in other programs. Only the 48-month interview did not contain information about Job Corps participation (since very few individuals are enrolled in Job Corps after the 30th month), but it still contains information on all the other education/training programs. Information about the months spent in Job Corps for the few individuals enrolling in Job Corps in the periods exceeding 30 months comes from Student Pay and Allotment Management Information System (SPAMIS).

⁹Individuals are considered to be in full-time education if they report having spent, on average, more than 30 hours per week in education in the quarter.

3 months or less (about 28% of those who enroll in Job Corps), the second group between 3 and 6 months (19%) and the last group enrolled for more than 6 months (53%).

We classify individuals as choosing alternative training if they had enrolled in some training or educational programs outside Job Corps at any time between the end of Job Corps (or at random assignment for those who did not enroll in Job Corps) and the 48th month after random assignment. The alternative training mostly takes place in the form of certain “education” programs such as high school, other vocational programs, or community colleges.¹⁰ As described by Schochet, Burghardt, and McConnell (2008), Job Corps differentiates itself from these programs for three main reasons: it is more comprehensive, offering different services (such as education and job-seeking assistance) all in the same place, it is more intensive (partly due to its residential feature) and, hence, more expensive and, finally, it is more homogeneous within the country because it is administered by the US Department of Labor, and not by local agencies.

Based on the time spent in Job Corps and the training choice after Job Corps, individuals in the treated group are divided into eight mutually exclusive subgroups. The proportions in each subgroup are reported in Table 1. 20.3% of the individuals in the treatment group enrolled in Job Corps only for a short period of less than 3 months. Among these early-leavers, nearly 60% of them eventually enrolled in alternative training. Interestingly, even among individuals with more than 7 months of experience with Job Corps, a substantial fraction of them (about 40%) participated in additional training after Job Corps. Overall, there are more individuals in the treatment group who enrolled in some alternative education program than individuals who did not (54.2% as compared to 45.8%). As for the control group, Table 1 shows that the number of individuals who enrolled in some alternative education program more than doubled the number of individuals without enrolling in any program (68.3% as compared to 31.7%).

Our main outcomes of interest are average weekly earnings and employment in the last quarter in the sample, i.e., the 16th quarter after random assignment, reported in the interview at 48 months.¹¹ Therefore, the outcome variables that we focus on in this paper are terminal outcomes—those that are

¹⁰See Appendix A for details about the different types of educational institution attended by individuals when not enrolled in Job Corps.

¹¹For some individuals, this information is missing because the last interview is conducted during the 16th quarter after random assignment: for these individuals, we use information about 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 and unemployment status instead.

measured after Job Corps and other training are completed.¹²

In defining the choice set outside Job Corps, we make a few simplifying assumptions. First, we ignore the duration spent in alternative training and, therefore, focus on the extensive margin of enrollment. We also do not differentiate further the specific type of alternative training. Expanding the choice set of alternative training will enrich the analysis, although this comes at the cost of an additional computational burden. For instance, if we distinguish different subcategories under alternative training, we can then study the complementarity (or substitutability) between Job Corps and a specific alternative program. In its current form, the alternative training should be interpreted as a mixture of different types of programs. Second, we treat employment as a terminal outcome. One alternative is to model employment as a third choice that is parallel to the existing choices of Job Corps, alternative training and no training. This allows us to estimate the returns to education against employment and unemployment separately. This difference may be substantial in the case of high returns to job experience: if individuals have high return to work experience, then the “no-training” group would be a mix of individuals with high earnings due to their working experience, and individuals with low earnings due to long unemployment. In our case, we believe returns to job experience are not a big concern. First, as documented in Appendix Table A2, there is little difference in terms of the average number of quarters of job experience between all the different subgroups selecting different duration and education treatment outside Job Corps. Second, there is evidence that returns to experience are quite low for young and low-skilled individuals (Altonji and Williams, 2005; Card and Hyslop, 2005), which suggests that there may not be big differences in terms of potential earnings due to differential working experience within the group of individuals who do not enroll in any education program.

2.3 Experimental Impact of Job Corps

In Appendix Table A3, we show that our sample replicates the ITT results in Table 2 and Table 3 of Schochet, Burghardt, and McConnell (2008). The individuals assigned to the treated group have around a 2.5-percentage point higher probability of being employed 4 years after the program is launched (about 69% of the control group is employed 4 years after the beginning of the Job Corps program). They have also a higher hourly wage (0.23\$ higher, where the average for the control group is 7.33\$) and higher

¹²As mentioned in Section 2.1, we dropped from our sample the few individuals who reported still being in full-time education in the last quarter, as well as those who spent some periods in Job Corps during the last year.

average weekly earnings (18\$ higher, where the average for the control group is about 200\$).¹³ The treatment increases receipt of any education and training by about 23 percentage points. In addition, approximately 72% of the treated groups enrolls in at least one period of Job Corps.

The ITT results in Table A3 compare the averages between individuals randomly assigned to the treated group and those assigned to the control group. In Table 2, we show the ordinary least square regression of the average weekly earnings and employment on Z , the variable indicating random assignment (Columns (1) and (2)), and on Z interacted with eight different dummies, one for each subgroup generated on the basis of duration in and education outside Job Corps (Columns (3) and (4)). Columns (1) and (2) are thus the standard ITT estimates, whereas Columns (3) and (4) show the differences between the average outcomes of individuals in each subgroup and the average outcome of the control group. The individuals dropping out of Job Corps after one period (hence, in the first 3 months) have, on average, the same earnings and the same employment probability as the control group, and lower earnings than any other subgroup. Differently, the individuals selecting Job Corps for a long period have the highest earnings and the highest probability of employment, in particular, those who obtain more education once they complete Job Corps. These differences show interesting heterogeneous patterns, but do not reveal whether there are program complementarities or whether heterogeneity is entirely driven by the unobserved characteristics of individuals, because they self-select into the different treatments.

3 The Dynamic Setting with Multiple Alternatives

3.1 Potential Outcomes and Potential Choices

We consider a sequential multistage decision framework where individuals select which education to obtain at every stage. The different stages (s) correspond to the three periods described in Section 2.2. The setting is shown in Figure 1. $s = 0$, thus, corresponds to the period immediately after randomization; $s = 1$ to a period lasting 3 months or less; $s = 2$ to a period lasting between 3 and 6 months; and $s = 3$ to a period lasting more than 6 months. As mentioned in Section 2.1, in our sample, individuals in the control group ($Z_i = 0$) can only choose among two alternatives in terms of education (k): either to enroll in a program different from Job Corps ($k = a$), or not to enroll in any program ($k = n$). In

¹³All the measures of these outcomes are taken as the average outcome in the 16th quarter after random assignment.

period $s = 0$, the treated individuals ($Z_i = 1$) have the additional alternative of enrolling in Job Corps ($k = j$). Selecting $k = a$ or $k = n$ at $s = 0$ does not allow individuals to choose different education in the following periods (hence, these are absorbing states). Individuals who select Job Corps in one period transit to the next period, where they face a new decision, comparing again three different options. Z_i may shift the choice at period $s = 0$, but has no direct effect on the choices of the following periods. When an individual leaves the Job Corps program, she cannot reenroll. In the last period ($s = 3$) Job Corps is not available, and individuals have to select between a and n . Each finishing node in the tree shown in Figure 1 for those with $Z_i = 1$ represents one of the eight subgroups described in Section 2.2.

The choice of an individual i is defined in terms of random variable S_i , i.e., the first period in which a person is observed not enrolled in Job Corps, and in terms of the random variable K_i , the choice made in period S_i . This definition implies that the duration in Job Corps for individual i is equal to S_i . Let $D_i = (s, k)$ indicate the observed choice of individual i (hence, $S_i = s$ and $K_i = k$) and $D_i(s, k)$ be an indicator function for each possible choice, where $D_i(s, k) = \mathbf{1}[D_i = (s, k)]$. The realized and potential choices for individual i are linked by:

$$D_i(s, k) = D_i^0(s, k) + (D_i^1(s, k) - D_i^0(s, k))Z_i$$

where $D_i^z(s, k)$ is $D_i(s, k)$ if $Z_i = z$.

Let Y_i be the observed outcome for the individual i , and $Y_i^{D_i}$ the potential outcome for the individual i if she selects D_i , then $Y_i^{D_i} \equiv Y_i^{s,k}$ if $D_i = (s, k)$.¹⁴ We can then link realized and potential outcomes:

$$Y_i = \sum_{s=0}^3 D_i(s, a)(Y_i^{s,a}) + \sum_{s=0}^3 D_i(s, n)(Y_i^{s,n}). \quad (1)$$

In what follows we maintain the following four assumptions:

Assumption 1. *Random assignment:* $Y_i^{D_i}, D_i^{Z_i} \perp\!\!\!\perp Z_i \quad \forall D_i, Z_i$,

Assumption 2. *Exclusion restriction:* $Y_i^{D_i^z} = Y_i^{D_i}, \quad \forall D_i, Z_i$,

Assumption 3. *Stable rank of next-best alternative at randomization:* $D_i^0 \neq D_i^1 \implies D_i^1 = (s, k), s > 0$,

¹⁴In our context, Y_i is realized only after D_i is chosen, and can be a vector of outcomes.

Assumption 4. *No Always-Takers:* $Z_i = 0 \implies S_i = 0$.

Assumption 1 and Assumption 2 are the usual assumptions that the instrument (Z_i) is randomly assigned and has no direct effect on the outcome.¹⁵ Assumption 3 implies that if individuals change their behavior as a consequence of being assigned to the treatment group, they do so by enrolling in at least one period of Job Corps. Hence, for example, this assumption rules out the possibility that individuals who select a when assigned to the control group would select n at period 0 when assigned to the treated group. This is the same as the “irrelevance” assumption invoked in the static framework examined by Kirkeboen, Leuven, and Mogstad (2016) and Kline and Walters (2016). Note that we impose this assumption only at period 0—beyond period 0 we allow the duration of Job Corps to affect the rank of the alternative program relative to no training. Assumption 4 is driven by the data given that we have excluded a small number of individuals in the control group who enrolled in Job Corps (see Section 2.1).

Note that here, we make no assumption about the individuals’ choice processes, their knowledge about the future outcomes, their utilities, or what they maximize. Hence, this framework allows for essential heterogeneity (Heckman, Urzua, and Vytlacil, 2006). For instance, individuals can select their education based on the gain they receive from it (where, in this case, individuals who select into an alternative education program at a point in time are those who benefit the most from alternative education as compared with no education or Job Corps at that point in time).

3.2 Substitution Patterns

Our potential outcome framework allows for a rich substitution patterns in the demand of training programs. Similar to the definition for market products, we define two programs to be complements or substitutes depending on the change in demand for one program in response to a change in price (or in availability) for the other (see, e.g., Gentzkow, 2007). An increase in the demand for programs other

¹⁵Ba, Ham, LaLonde, and Li (2017) point out that the random assignment may have a direct effect on potential outcomes. In our context, the knowledge of being in the treatment group at the time of randomization may imply that individuals anticipate the option of participating in Job Corps in the future, which could change her current behavior (e.g. by reducing job search efforts, or not accepting a job offer that he/she would otherwise accept). In the experimental setting considered here this problem is less severe because over 90% of enrollments in Job Corps take place within the first 3 months after randomization. Therefore, the time horizon that individuals can use to adjust their behavior is short. In addition, in this paper, we focus on long-term terminal outcomes, i.e., outcomes that are measured 4 years after random assignment when Job Corps is concluded for all treated individuals. This issue may be more significant if we estimate the program returns for short-term outcomes.

than Job Corps after the introduction of Job Corps implies that Job Corps and the other programs are complementary, whereas a decrease implies that they are substitute for each other.

We can investigate in more detail whether different times spent in Job Corps substitute or complement the demand for more education. Note that different periods in Job Corps can substitute other types of education for some individuals and complement for others. In our setting, we can characterize the proportion of individuals for whom the programs are substitutes or complements. In more detail, call $\pi_{k,sk'}$ the proportion of the population for whom $D_i^0 = (0, k)$, $D_i^1 = (s, k')$, i.e., the joint probability of selecting $k \in (a, n)$ if $Z_i = 0$ and s periods in Job Corps and $k' \in (a, n)$ if $Z_i = 1$. We can then define the proportion of individuals complementing $s \geq 1$ periods of Job Corps and a by:

$$\pi_{n,sa} \equiv P(D_i^0 = (0, n), D_i^1 = (s, a)),$$

which is the proportion of individuals who select more education only after having completed s periods in Job Corps. As the “price” of Job Corps declines from $Z = 0$ to $Z = 1$, these individuals increase their demand for both alternative education and Job Corps. Similarly, we can define the proportion of individuals who substitute the alternative education program with $s \geq 1$ periods in Job Corps as the proportion of individuals who would obtain some education in the absence of Job Corps, but do not enroll in more education once Job Corps becomes available, and they obtain s periods of Job Corps. This proportion is:

$$\pi_{a,sn} \equiv P(D_i^0 = (0, a), D_i^1 = (s, n)).$$

The dynamic framework introduces an additional type of substitutability, as individuals who receive the offer of Job Corps may postpone the decision to enroll in some already available education after having enrolled in Job Corps for some time. This would be evidence of “intertemporal substitution” where the overall demand for an alternative education does not change, but the timing does.¹⁶ The individuals who select some education when assigned to the control group and who select more education after $s \geq 1$ periods in Job Corps if assigned to the treatment group are those for whom Job Corps and other

¹⁶Our framework assumes that there is only one category of alternative education and, as a consequence, the only effect of the intertemporal substitution is to postpone enrollment in an alternative education. It is, however, possible that individuals who would have enrolled in one education program when assigned to the control group, and who instead receive Job Corps when assigned to the treatment group, then decide to enroll in a different type of education. As mentioned, we could expand the model and estimate complementarity/substitutability in demand for different types of educational programs; however in the current form, it is to be interpreted as a general definition of education relative to not getting education at all.

programs are intertemporal substitutes; hence, their proportion in the population is:

$$\pi_{a,sa} \equiv P(D_i^0 = (0, a), D_i^1 = (s, a)).$$

3.3 LATE interpretation

Schochet, Burghardt, and McConnell (2008) use random assignment as an instrument for program participation and the Wald estimator to recover the causal effect of the program. Under Assumptions 1 and 2, this allows us to estimate the LATE, i.e. the effect of Job Corps for the population of compliers. In this specification, the compliers are the individuals who enroll in at least one period of Job Corps when assigned to the treatment group and do not enroll in Job Corps when assigned to the control group. In particular, the Wald estimator is used to estimate the LATE parameter:

$$\begin{aligned} LATE_{Overall} &= \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{P(S_i > 0|Z_i = 1) - P(S_i > 0|Z_i = 0)} \\ &= E[Y_i^{D^1} - Y_i^{D^0} | D_i^0 = 0, D_i^1 = S_i], \quad S_i > 0. \end{aligned} \tag{2}$$

Equation (2) implies that the estimate obtained by instrumenting Job Corps participation using initial random assignment is an average effect of Job Corps for all the individuals who enroll in the program for at least one period relative to their own counterfactual education choices. We call $LATE_{Overall}$ the average effect of Job Corps for all compliers. The denominator of the fraction in equation (2) is the proportion of compliers in the population. Note that, by assumption 4 (no Always Takers), all the individuals who enroll in the program when assigned to the treatment group are compliers and, as a consequence, the second term in the denominator is zero.

Given Assumptions 3 and 4 and the dynamic setting with three periods and two potential choices when not enrolled in Job Corps (for both the treatment and the control group), we can define 12 different subpopulations of compliers. Each subpopulation differs in terms of the combinations of duration and education selected when assigned to the treatment group ($D_i^1 = (s, k)$, $s \in (1, 2, 3)$, $k \in (a, n)$) and in terms of education selected when assigned to the control group ($D_i^0 = (0, k)$, $k \in (a, n)$). Kline and Walters (2016) show, in a static framework, that, in the presence of subpopulation of compliers differing in terms of potential choices, the overall LATE estimated through the Wald estimator is an

average of the “subLATEs” for the different subgroups. Proposition 1 extends the result decomposition by Kline and Walters (2016) to a dynamic framework.¹⁷

Proposition 1. *The LATE identified by the Wald estimator is a weighted average of the LATE the subpopulations of the compliers:*

$$LATE_{Overall} = \sum_{s=1}^3 \sum_{k \in (a,n)} \sum_{k' \in (a,n)} \frac{\pi_{k,sk'}}{\pi_C} E[Y_i^{s,k'} - Y_i^{0,k} | D_i^0 = (0, k), D_i^1 = (s, k')] \quad (3)$$

where $\pi_C \equiv \sum_{s=1}^3 \sum_{k \in (a,n)} \sum_{k' \in (a,n)} (\pi_{k,sk'})$. In our framework, the proportions of individuals in each complier sub-population and the subLATEs are not nonparametrically identified exploiting only initial random assignment as an exclusion restriction.

Proof. See Appendix B.

Each subLATE is the LATE of the given combination of duration and education for the subgroup selecting that combination. For example, the subLATE for the individuals who select no education when assigned to the control group and enroll in three periods in Job Corps and then obtain more education when assigned to the treatment group is $E[Y_i^{3,a} - Y_i^{0,n} | D_i^1 = (3, a), D_i^0 = (0, n)]$. The intuition behind the nonidentification result is that as individuals face multiple choices at each node, one instrument exogenously shifting choices of individuals at time zero only is insufficient to identify the effect of Job Corps for all the different combinations of observed and counterfactual educational decisions. As shown in Kirkeboen, Leuven, and Mogstad (2016), even under Assumption 3 in a static framework, it is necessary to know the next best alternative of each individual to identify the causal effects on the different types of compliers in the presence of multiple alternatives. Proposition 1 extends their findings to a multiperiod setting.¹⁸

¹⁷Proposition 1 is similar to the decomposition of the Wald estimator in Heckman, Humphries, and Veramendi (2016). The difference between their case and Proposition 1 is that Heckman, Humphries, and Veramendi (2016) only consider the choice of the duration in the treatment, not the multiple options for the treatment and control groups. In addition, in our case, the instrument shifts the choices of the individuals only at period 0, whereas the instrument considered by Heckman, Humphries, and Veramendi (2016) can have an effect in shifting the choices also at later periods. Heckman, Humphries, and Veramendi (2016) also show that, in their case, the proportions of compliers in the different subgroups are nonparametrically identified, whereas the subLATEs are not. Proposition 1 clarifies that neither proportions nor subLATEs are nonparametrically identified in the presence of multiple alternatives. In Appendix B we show that, in the static framework described by Kline and Walters (2016), the proportions in the different subgroups are nonparametrically identified, whereas the subLATEs are not. A general decomposition of the IV estimator in the presence of multiple but nonsequential choices is reported in Heckman and Urzua (2010) who show that, in their case, neither the proportions nor the subLATEs are identified. Our case can be regarded as a special case of their decomposition, where our instrument is binary (and thus the IV estimator we consider is the Wald estimator) and our choice structure is sequential.

¹⁸One way to identify the separate effect of the different duration/education choice on earnings would be to conduct

Note that certain proportions of the compliers’ subpopulation correspond to the proportions of individuals complementing/substituting Job Corps and alternative education programs described in the previous section. For instance, $\pi_{a,sa}$ corresponds to the proportion of individuals for whom s periods in Job Corps and a are intertemporal substitutes. Thus, estimating the proportion of compliers in each subgroup not only is informative on the LATE decomposition, but also reveals the patterns of complementarity/substitutability in the demand for different education programs.

Note also that, as the earnings are only observed conditional on employment, the overall LATE for earnings estimated using the Wald estimator are biased due to selection into employment (see, e.g., Ham and LaLonde, 1996). In Appendix C we show the decomposition of the overall LATE taking into account endogenous selection in employment.¹⁹

3.4 Beyond LATE

Our framework allows Job Corps participant to spend multiple periods in Job Corps and to complement Job Corps training with additional education after Job Corps is concluded. Hence, Job Corps may have different effect depending on the duration of participation (“dynamic treatment effect”) and choice of additional education (“dynamic program interaction”). In this section, we formally define these additional treatment effects. We show that many of these economically important parameters are different from the LATE and subLATEs discussed in Section 3.3, and that using the subLATEs alone may produce misleading estimates of program complementarities if individuals self-select into treatment types based on gains. This further motivates us to develop an econometric model in Section 4 that can be used to identify these treatment effects.

two-stage least squares, where each $D_i(s, k)$ is a different endogenous variable. However, this requires different exclusion restrictions (different instruments) for any time point that affects selection into each education level available at that time point, but not the outcome (Taber, 2000).

¹⁹Appendix C shows that the overall LATE estimated using the Wald estimator is the sum of the causal effect of the treatment on individuals, who are employed when assigned to the treatment group, and selection bias due to the fact that some of the individuals who are employed when assigned to the treatment group are not employed when assigned to the control group. In this context, we can identify two decompositions of interest: the causal part of the estimated LATE, and the effect of the treatment on the whole population, including the unemployed individuals for whom we compute the latent earnings. The decomposition of the causal part of the LATE has different weights for the different subLATEs, as compared with the decomposition of the LATE for the whole population.

3.4.1 Dynamic Program Interactions

We define the alternative education programs and the Job Corps program to be dynamically complementary in *outcome* if the returns to the alternative education programs are higher after having enrolled in Job Corps. At the individual level and with continuous treatments, this condition can be written as $\frac{\partial \frac{\partial Y_i}{\partial a}}{\partial JC} > 0$. The individual-level complementarity between s periods in Job Corps and alternative programs with discrete treatments is defined as:

$$(Y_i^{s,a} - Y_i^{s,n}) - (Y_i^{0,a} - Y_i^{0,n}) > 0. \quad (4)$$

A negative value of the difference in equation (4) implies that returns to a are lower after having enrolled in Job Corps, indicating that the skills acquired during the s periods in Job Corps training substitute for the skills obtained in alternative educational programs. Note that this definition of complementarity in outcome is different from that of complementarity in the *demand* between Job Corps and a , as defined in section 3.2.

At the population level, the complementarity in outcome between a and s periods in Job Corps is then defined as:

$$E[(Y_i^{s,a} - Y_i^{s,n}) - (Y_i^{0,a} - Y_i^{0,n})] > 0. \quad (5a)$$

The difference in equation (5a) can be rewritten as:

$$E[Y_i^{s,a} - Y_i^{0,a}] - E[Y_i^{s,n} - Y_i^{0,n}], \quad (5b)$$

where the two terms of the difference in equation (5b) are the ATE of completing s periods in Job Corps and then selecting a (n) as compared with enrolling in a (n) without Job Corps. The average differences between the outcome of individuals selecting (s, k) and the outcome of the same individuals

selecting $(0, k)$ are related to the subLATEs described in Section 3.3 by

$$\begin{aligned}
& \underbrace{E[Y_i^{s,k} - Y_i^{0,k} | D_i^0 = (0, k), D_i^1 = (s, k)]}_{\text{subLATE}} = \\
& \underbrace{E[Y_i^{s,k} - Y_i^{0,k}]}_{\text{Pairwise ATE}} + \underbrace{E[Y_i^{s,k} - Y_i^{0,k} | D_i^0 = (0, k), D_i^1 = (s, k)] - E[Y_i^{s,k} - Y_i^{0,k}]}_{\text{Sorting on Gains}}, \forall k \in (a, n).
\end{aligned} \tag{6}$$

Therefore, if individuals select s and k based on gains, the subLATEs do not correspond to the ATE required to identify complementarities. As a consequence, estimates of program complementarities cannot be recovered by only comparing the subLATEs for individuals selecting $D_i^0 = (0, a), D_i^1 = (s, a)$ and those selecting $D_i^0 = (0, n), D_i^1 = (s, n)$, because sorting on gains would confound the result:

$$\begin{aligned}
& \underbrace{E[Y_i^{s,a} - Y_i^{0,a} | D_i^0 = (0, a), D_i^1 = (s, a)]}_{\text{subLATE}(D_i^0 = (0, a), D_i^1(s, a))} - \underbrace{E[Y_i^{s,n} - Y_i^{0,n} | D_i^0 = (0, n), D_i^1 = (s, n)]}_{\text{subLATE}(D_i^0 = (0, n), D_i^1(s, n))} = \\
& \underbrace{E[Y_i^{s,a} - Y_i^{0,a}] - E[Y_i^{s,n} - Y_i^{0,n}]}_{\text{Complementarity / Substitutability}} + \\
& \underbrace{E[Y_i^{s,a} - Y_i^{0,a} | D_i^0 = (0, a), D_i^1 = (s, a)] - E[Y_i^{s,a} - Y_i^{0,a}]}_{\text{Sorting on Gains in } a} - \\
& \underbrace{E[Y_i^{s,n} - Y_i^{0,n} | D_i^0 = (0, n), D_i^1 = (s, n)] - E[Y_i^{s,n} - Y_i^{0,n}]}_{\text{Sorting on Gains in } n}.
\end{aligned} \tag{7}$$

Hence, if individuals select into s and k based on gains, the subLATEs estimate the causal effect of s periods in Job Corps and k education for individuals selecting s and k , but they do not inform on the population-level average effect of obtaining s and k .

Note that these parameters are unique to a dynamic framework with multiple alternatives. In a static framework, the different educational programs are necessarily mutually exclusive. In our case, this would imply that there are two subgroups of compliers: those who select n when in the control group and Job Corps when treated, and those who select a when not treated and Job Corps when treated. This implies that a static model does not permit estimating dynamic complementarity or substitutability.²⁰ In addition, it does not permit estimating any dynamic treatment effect of the

²⁰A similar static framework is the one adopted by Kline and Walters (2016), who do not consider any dynamic interaction between Head Start and other preschool programs in generating the final outcome.

program, as defined in the next section.

3.4.2 Dynamic Treatment Effect

Given the sequential structure, we can estimate the value of transitioning to the next period, i.e., the dynamic treatment effects of sequential training decisions (Rodríguez, Saltiel, and Urzúa, 2018), at every period. Thus, we can define three different parameters, the dynamic average treatment effect on the treated (DATT) at any of the three periods:

$$DATT^{\bar{s}} = E[Y_i^{S_i, K_i} - Y_i^{\bar{s}-1, k} | S_i \geq \bar{s}] \quad (8)$$

with $\bar{s} \in (1, 2, 3)$. $Y_i^{S_i, K_i}$ is the outcome of individual i when selecting to spend the preferred S_i periods in Job Corps and then choosing K_i as education afterwards.²¹ $Y_i^{\bar{s}-1, k}$ is the outcome for the same individual i when enrolling in $\bar{s} - 1$ periods of Job Corps and then selecting k . We define k as the optimal choice of an individual in terms of a and n if forced to leave Job Corps at period $\bar{s} - 1$. Note that $DATT^1$ is the effect of enrolling in at least one period of Job Corps for those who select at least one period of Job Corps, which then corresponds to the overall LATE.

Equation (8) can be decomposed as follows:

$$DATT^{\bar{s}} = \underbrace{E[Y_i^{\bar{s}, k} - Y_i^{\bar{s}-1, k} | S_i \geq \bar{s}]}_{\text{Direct Effect}} + \underbrace{E[Y_i^{S_i, K_i} - Y_i^{\bar{s}, k} | S_i \geq \bar{s}]}_{\text{Continuation Value}} \quad (9)$$

where the direct effect is the returns of obtaining exactly \bar{s} periods in Job Corps as compared with stopping at the period before, whereas the continuation value is the additional gain of moving forward to the next periods for those individuals who select to obtain additional Job Corps. Because, in our framework, it is possible to obtain training in period \bar{s} only by obtaining training at period $\bar{s} - 1$, it is possible that the real gain of training at period $\bar{s} - 1$ in terms of outcome Y is just the fact of offering the possibility of additional training in the following periods.²² Thus, this decomposition

²¹We define these parameters as treatments on the treated because we condition on individuals selecting their optimal $S_i > \bar{s}$ and making their optimal choice in the periods following \bar{s} . A population-level treatment effect (i.e. the dynamic average treatment effect, DATE) can be calculated if we impose a specific choice at $s > \bar{s}$ for the individuals.

²²Because Job Corps stops at period $s = 3$, by construction, $DATT^3$ has no continuation value. Note that none of the terms in the decomposition correspond to the subLATEs. As for the subLATEs, we only compare the different choices of individuals when treated and when in the control group, whereas here we compare individuals when treated with their outcome in the next-best choice when forced to stay in Job Corps for different amounts of time.

allows us to investigate the source of the dynamic effect in Job Corps and is related to the results on dynamic complementarity. If, for example, one period in Job Corps is strongly complementary with more education, and individuals select on the basis of gains, we expect the direct effect of one period in Job Corps to be quite high, because most individuals would select more education after one period in Job Corps and have high gains from it. The continuation value would then be low, because individuals can get the same high returns from one period in Job Corps by dropping out and selecting an alternative education program and by continuing in Job Corps for one more period. If, instead, the returns to one period in Job Corps are generally low, irrespective of the next-best alternative after $s = 1$, then the direct effect is low and an eventually positive effect of transitioning to period 1 would be driven by the continuation value.

4 Sequential Choice Model

Thus far, we have not imposed any restrictions on how individuals sort into different potential choices. We have shown that the compositions of the program compliers and their associated returns are not identified, nor are the parameters characterizing the complementarity/substitutability between different programs. If we are interested in evaluating the impact of any reforms to Job Corps, we need to have a framework to predict how selection into Job Corps is likely to change, and how this impacts the program's rate of return. A model that specifies the dynamic sorting of individuals into potential outcomes also allows us to identify the mix of program compliers and the associated individual-specific gains.

In this section, we build and estimate a sequential choice model that predicts an individual's potential choices and combine it with potential outcomes. Building on Heckman and Navarro (2007) and Heckman, Humphries, and Veramendi (2016), we characterize the decision process at each decision node by a threshold model and parameterize decision rules and potential outcomes using an unobserved factor. It is worth noting that even the heavily parameterized model is still quite flexible, allowing for treatment effect heterogeneity and for selection into training based on levels and gains. As in Heckman and Navarro (2007) and Heckman, Humphries, and Veramendi (2016), the model should be interpreted as being midway between dynamic discrete choice models in the structural econometrics literature, and the reduced-form estimation of the treatment effects literature. Unlike the literature on

discrete dynamic choice models, the information sets, preferences, and budget constraints governing agent choices at different nodes are left unspecified. This implies that we are only able to estimate the ex-post returns to Job Corps, not the ex-ante returns (i.e. returns perceived by the agents prior to their choices) (Heckman, Humphries, and Veramendi, 2016, 2018).

4.1 The Setup

We impose the same choice structure as discussed in the previous section (shown in Figure 1). Time is discrete, and all individuals are assumed to start in period 0 in an initial state that is randomly determined by the experimental variation Z . If $Z = 1$, individuals will have access to Job Corps. Participation in Job Corps is dynamic and sequential, and stopping times are nonrecurrent (i.e., there is no recall when an individual decides to drop out of the Job Corps). The sequential feature of the model means that new information arrives each period and anticipations of individuals as to when they will stop and the consequences of alternative stopping times can be revised sequentially.

We model the decision process at each choice node by a threshold model. Define $I_{ik}(s)$ as indicator function for whether an individual i chooses option k in period s . Time spent in Job Corps (treatment time) is the first period at which a person exits from Job Corps. We can link the individuals' potential choices ($D_i^Z(s, k)$) (defined in Section 3.1) with their sequential realization of $I_{ik}(s)$. For instance, $D_i^Z(s, k) = \mathbf{1}[I_{ik}(s) = 1, I_{ij}(s-1) = 1, I_{ij}(s-2) = 1, \dots]$, where the treatment time s can only be chosen if Job Corps option is taken up to period $s-1$.

We specify the process that determines $I_{ik}(s)$ using a reduced-form threshold model that characterizes choices as a function of observables and unobservables. In general, conditional on reaching the decision node in period s (where $s \in [0, 3]$), utilities from each choice in the current period are given by:

$$U_{is}^n = 0 \tag{10}$$

$$U_{is}^a = \beta_x^{s,a} \tilde{X}_i + \beta_\theta^{s,a} \theta_i + u_{is}^a \tag{11}$$

$$U_{is}^j = \beta_x^{s,j} X_i + \beta_\theta^{s,j} \theta_i + u_{is}^j \tag{12}$$

where U_{is}^a is the value of alternative training for the individual i in period s , U_{is}^j is his valuation of Job

Corps, and the value of no training in any period (U_{is}^n) is normalized to zero. The unobserved factor, θ_i , is the key source of dynamic selection in this model, and, together with the observed variables, determines the dependence between selection into different duration/education bundles and outcomes. Given the normalization assumption, we make in Section 4.2, θ_i captures unobserved differences in individual demand for alternative training relative to no training. This includes both any unobserved preferences and any unobserved constraints such as credit constraints for paying for alternative training. The factor loadings, $\beta_\theta^{s,a}$ and $\beta_\theta^{s,j}$, are common across the population, but evolve flexibly over time, thereby allowing the effects of unobserved heterogeneity in demand to vary over time. The observable characteristics, X_i and \tilde{X}_i , affect the utilities. In both X_i and \tilde{X}_i , we include a constant, an age indicator for older individuals, and a female indicator variable. In \tilde{X}_i , we also include an additional variable indicating the availability of alternative education programs in the state of residence, that serves as an exclusion restriction.²³ The choice structure imposes additional exclusion restrictions. If an individual is not offered to join Job Corps ($Z = 0$), U_{is}^j is excluded from their choice set, and they only choose between a and n . Similarly, if an individual is in the last period of Job Corps ($s = 3$), U_{is}^j is also excluded from their choice set.

The choice-specific idiosyncratic shocks, u_{is}^j and u_{is}^a , are uncorrelated with the factor θ_i and other covariates in the model. They reflect unobserved shocks to demand for Job Corps and alternative training, respectively, relative to no training. They are assumed to be independent over time and may be correlated with each other within a given period. u_{is}^a and u_{is}^j are jointly normally distributed with means zero, variances normalized to 1 and correlation coefficient ρ .

All parameters in the threshold model may vary according to the potential duration in Job Corps. Therefore, even for a given individual, their valuation of Job Corps and alternative programs may change over time. This could capture, for instance, that experience under the Job Corps program makes alternative programs more attractive to individuals in the current period. Following the stable-rank assumption, discussed in section 3.1 (Assumption 3), U_{i0}^a and U_{i0}^n are invariant to the instrument Z . However, in a dynamic setting, U_{is}^a (where $s \geq 1$) may indirectly be affected by Z because Z affects

²³This variable is equal to one if the number of community colleges and public post-secondary institutions in the state of residence is above the national median level, and equal to zero otherwise. We use the number of community colleges and public post-secondary institutions by state in the academic year 1997/98 (published by The National Center for Education Statistics). In the data that we have access to, the state of residence of each individual is available only at the 48th month interview, which implies that we assign each individual to the post-treatment state of residence. However, almost 90% of the sample we consider reports living in the same state at the 48th interview as at the random assignment.

the duration of Job Corps, and the duration of Job Corps affects the relative value of the alternative program.

Individuals make choices $k \in (a, n, j)$ by comparing the potential utility of all available alternatives conditional upon reaching a specific decision node. Therefore:

$$I_{ik}(s) = \begin{cases} 1 & \text{if and only if } U_{is}^k \geq U_{is}^m, \forall m, \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

$I_{ik}(s)$ are determined sequentially. Because individuals proceed to the next period only conditional on obtaining Job Corps in the previous period, $I_{ik}(s)$ is revealed only if $I_{ij}(s-1) = 1, \forall s \geq 1$. One can think of the potential utilities as approximating the choice-specific value function in each period s (conditional on having reached this decision node) as a function of observables, permanent unobservables, and shocks.²⁴ However, contrary to dynamic discrete-choice structural models, the information and preference structure at each node is not specified, and with the model, only ex-post returns, not ex-ante returns, are identified. In our model, agents face stage-specific costs and/or preference shocks that they may or may not anticipate before they reach the stage (Heckman, Humphries, and Veramendi, 2016).

We parameterize the potential outcomes associated with the treatment of time in Job Corps (s) and the alternative choice ($k \in (a, n)$) using the same set of factors and observable characteristics:

$$Y_i^{s,k} = \alpha^{s,k} + \gamma_y^{s,k} X_i + \gamma_\theta^{s,k} \theta_i + \varepsilon_i^{s,k} \quad (14)$$

where $Y_i^{s,k}$ is a terminal outcome (measured beyond s) associated with s periods of Job Corps experience and choosing option k outside Job Corps. The factor loading, $\gamma_\theta^{s,k}$ captures how potential outcomes vary with the permanent unobservable factor. The choice-specific idiosyncratic error terms are $\varepsilon_i^{s,k}$. They are assumed to be independent from each other, the covariates X , the unobserved factor, and the choice-specific demand shocks (u_{is}^a and u_{is}^j).²⁵ The factor loadings in the outcome equation reveal

²⁴Individuals can separate from Job Corps either voluntarily (quits) or involuntarily (asked to leave the program). As argued by Borjas and Rosen (2012), regardless of who initiates it, separation represents the same underlying phenomenon, that of individuals' marginal product being higher elsewhere.

²⁵The independence between $\varepsilon_i^{s,k}$ and u_{is}^a and u_{is}^j is an identifying assumption (see Section 4.2 for details). In our case, this assumption is less stringent because of a timing restriction: the outcome is measured in the medium term beyond period s , after the individuals have selected their training decisions. Therefore, if the shocks to the outcomes are temporary and mean-reverting, they are unlikely to be correlated with shocks to the potential choices that took place previously.

important information about a variety of selection schemes. For instance, if $\gamma_\theta^{s,k} = 0 \forall s, k$, then potential outcomes are independent of the unobserved factor, implying no selection. If $\gamma_\theta^{S,a} = \gamma_\theta^{S,n}$, then there is no selection on gains into alternative training. If $\gamma_\theta^{s,k} = \gamma_\theta^{s',k} \forall s \neq s'$, then there is no dynamic selection into Job Corps on gains.

As mentioned in Section 3.3, we need to correct for selection bias given that earnings are only realized conditional on employment. Without controlling for employment, the estimated effect of the treatment on earnings is the sum of the causal effect of the treatment and the bias due to the different composition of the sample of employed individuals in the different treatment statuses. We define potential employment in the terminal period as

$$H_i = H_i^{0,n} + \sum_{s=0}^3 D_i(s, a)(H_i^{s,a} - H_i^{0,n}) + \sum_{s=1}^3 D_i(s, n)(H_i^{s,n} - H_i^{0,n}).$$

We model $H_i^{s,k}$ as an index function that depends on the same set of observables and unobservables as above:

$$H_i^{s,k} = \mathbf{1}[\alpha_H^{s,k} + \gamma_H^{s,k} X_i + \gamma_{H,\theta}^{s,k} \theta_i + \varepsilon_{i,H}^{s,k}] \quad (15)$$

where $\varepsilon_{i,H}^{s,k}$ are independently and identically distributed shocks that are uncorrelated with $\varepsilon_i^{s,k}$, the unobserved factor and the demand shocks. Potential earnings $Y_i^{s,k}$ are observed only conditional on $D_i(s, k) = 1$ and $H_i^{s,k} = 1$.

4.2 Identification

Identification of the sequential choice model relies on the conditional independence assumption and the exclusion restrictions. The conditional independence assumption implies that, conditional on the unobserved factors, choices and outcomes are independent from each other. This assumption is analogous to the conditional independence assumption in the matching literature, except that here, factors are unobserved to the econometrician. Following the literature on dynamic treatment effects (e.g., Heckman, Humphries, and Veramendi (2016)), we supplement the outcomes with a measurement system to

proxy the unobserved factors and correct for the effects of measurement error in the proxy:

$$C_i = \alpha_c + \gamma_c X_i + \gamma_{c,\theta} \theta_i + \varepsilon_{c,i} \quad (16)$$

$$M_i^{s,k} = \alpha_m^{s,k} + \gamma_m^{s,k} X_i + \gamma_{m,\theta}^{s,k} \theta_i + \varepsilon_{m,i}^{s,k} \quad (17)$$

where C_i is a vector of measurements that are observed at the time of randomization, $M_i^{s,k}$ is a vector of measurements that are evaluated after the individuals' choices and $\varepsilon_{c,i}$ and $\varepsilon_{m,i}^{s,k}$ are the respective classical measurement errors that are independent from outcomes and choices. In C_i , we include the following three measurements, all measured at the time of randomization using the baseline survey: the fractions of time spent in an educational/training program in the previous year, the fractions of time spent in employment in the previous year, and whether the individual had any high school credentials (including a GED).²⁶ Given that these measurements are observed at the time of randomization, the parameters affecting C_i are common across different treatment values.

In $M_i^{s,k}$, we include literacy test scores available from a random subsample of individuals. The test score sample contains 1,117 program members and 1,156 control group members who completed the literacy test. The test is on three dimensions: ability to understand a text (prose), ability to understand a table/graph (document), and knowledge of arithmetic operations (quantitative). We use the log quantitative scores and the log average of the prose and document scores, given the high correlation between the latter two test scores.²⁷ The tests were implemented in the 30th month after the time of randomization, and, therefore, the parameters in the measurement system of $M_i^{s,k}$ vary with the potential treatment. In this respect, these test score measurements are used as additional outcomes. The parameters are of interest by themselves because they are informative of the treatment effects on cognitive ability.

Location and scale of the factor are not identified, so normalizations on the factor are applied. Specifically, the mean of the factor is normalized to zero (to fix the location), and, to fix the scale, the factor loading in the selection equation for a in period 0 is fixed to one ($\beta_a^0=1$). The current

²⁶The range of work and educational history in the year before randomization is between 0 and 1. As a result, instead of a linear model, we use a fractional outcome regression model for these two measurements. Similarly, for the high school credential measurement, we use a probit model.

²⁷The literacy data set includes five measurements for each of the three dimensions. These were computed using the answers to the tests and the characteristics of individuals to build a distribution of the individual's ability, and then five plausible values are simulated from each individual's distribution in each domain. For each test dimension, we use the averages of these five measurements.

normalization assumption implies that θ_i captures unobserved differences in individual demand for alternative training in period 0 relative to no training. Alternative normalizations can also rationalize the given set of measurements. By linking the unobservable to measurements, we are also able to make the unobservable factor θ_i interpretable.

In Appendix D, we utilize identification results from Heckman and Navarro (2007) and show that, under the conditional independence assumption implied by the factor structure, all parameters of the model are non-parametrically identified using covariances restrictions provided by the outcomes, the choices and the measurements. In addition, the conditional independence assumption is combined with exclusion restrictions. More specifically, the instrumental variable Z from the experimental variation provides exclusion restrictions on the choice of Job Corps in period 0. As mentioned in Section 4.1, as additional exclusion restriction for the choice of alternative training, we use the availability of alternative education programs in the state in which the individual resides.

4.3 Estimation Strategy

The model is estimated using maximum likelihood method. The complete likelihood function consists of products over workers:

$$L = \prod_{i \in B_1} L^{(1)}(D_i, Y_i, H_i, M_i | X_i) \prod_{i \in B_2} L^{(2)}(D_i, Y_i, H_i, M_i | X_i) \quad (18)$$

where $i \in B_m$ denotes the set of workers who belong to the m th case of the likelihood function. The first case includes individuals who do not receive an offer of Job Corps ($Z = 0$). Because Job Corps is not an option for these individuals, the likelihood contribution is defined with respect to the choice between a and n , the associated outcomes and measurements. The second case includes all individuals who received an offer of Job Corps ($Z = 1$). In this case, the likelihood function is defined over the duration of participation, alternative education following Job Corps, and the associated outcomes and measurements.²⁸

²⁸ M_i includes all measurements available to individual i , including cognitive test scores and the individual's observables at the time of randomization.

Conditional on the individual's type θ_i , the likelihood function of individuals with $Z = 0$ is

$$\begin{aligned} L^{(1)}(D_i = k, Y_i, H_i, M_i \mid X_i, \theta_i) \\ = P(U_{i0}^k > U_{i0}^{k-} \mid X_i, \theta_i)h(Y_i, H_i, M_i \mid D_i = k, X_i, \theta_i) \end{aligned} \quad (19)$$

where $k \in (a, n)$ and $k-$ denote alternative choices other than k . When $Z = 0$, we know that $s_i = 0$, so the potential outcomes are defined based only on $k \in (a, n)$.

Conditional on the individual's type θ_i , the likelihood function of individuals with $Z = 1$ is

$$\begin{aligned} L^{(2)}(D_i = (s, k), Y_i, H_i, M_i \mid X_i, \theta_i) \\ = P(U_{i0}^j > \max(U_{i0}^{j-}), \dots, U_{is-1}^j > \max(U_{is-1}^{j-}), U_{is}^k > \max(U_{is}^{k-}) \mid X_i, \theta_i)h(Y_i, H_i, M_i \mid D_i = (s, k), X_i, \theta_i) \end{aligned} \quad (20)$$

where $j-$ denotes all the available options other than the Job Corps (j). When $Z = 1$, an individual may choose either Job Corps or one of the outside options in period 0. In period 1, she can only choose Job Corps provided that she chose Job Corps in period 0.²⁹

To form the likelihood contribution for the individual, we need to average out over all possible individual types (using the Gauss-Hermite quadrature):

$$L^{(m)}(D_i, Y_i, H_i, M_i \mid X_i) = \int L^{(m)}(D_i, Y_i, H_i, M_i \mid X_i, \theta_i)f(\theta_i)d\theta_i, \quad m = \{1, 2\}$$

The choice probabilities are evaluated using GHK simulator. The distribution of the unobservables $(\theta_i, u_{is}^a, u_{is}^j, \varepsilon_i^{s,k}, \varepsilon_{i,H}^{s,k})$ are nonparametrically identified, as shown in Appendix Section D. However, for estimation purpose, we impose additional distributional assumptions. The unobserved factor (θ_i) and $\varepsilon_{i,H}^{s,k}$ each follows a normal distribution. The utility shocks follow a bivariate normal distribution as already specified in Section 4.1. Shocks to latent earnings, $\varepsilon_i^{s,k}$, follow a mixture of normal distribution with two components.³⁰ The standard errors are computed using the Berndt, Hall, Hall, and Hausman

²⁹As mentioned in Section 3.1, we model potential outcomes at each potential duration of participation in Job Corps, given that individuals cannot return to Job Corps after exiting.

³⁰We use a mixture of normal distribution because the earnings distribution in the data has a thick left-tail. Our estimated model is able to fit the earnings distribution quite well (see Figure A1 and Section 5.1). We also experimented with a two-component mixture of normal distribution for the unobserved factor, which did not lead to large improvement in terms the goodness of fit.

(1974) (BHHH) algorithm. When implementing the maximum likelihood estimator, we follow Hansen, Heckman, and Mullen (2004) and Heckman, Humphries, and Veramendi (2016) to perform the maximum likelihood estimation in two steps. In the first step, we correct the estimated factor distributions for the causal effect of choices on measurements by jointly estimating the choice and measurement equations. The outcome equations are estimated in the second stage using estimates from the first stage. Details of the estimation procedure are presented in Appendix Section E.

5 Empirical Results

5.1 Model Parameters and Selection Patterns

The estimates and standard errors of the estimators of the parameters of the selection and the outcome equations are reported in Tables 3 and 4. The normalization we impose implies that θ positively affects the probability of enrolling in some alternative education program at period $s = 0$. Larger values of θ also imply higher probability of enrolling in Job Corps at period $s = 0$, as compared with not enrolling in any education. In the following periods, individuals with high values of θ who have enrolled in Job Corps in period 0 are more likely to drop out of Job Corps to enroll in an alternative education program, but θ also has a (smaller) positive effect on the probability of continuation in Job Corps, and, hence, the latent factor is, in general, positively correlated with the probability of enrolling in any education program. In Appendix Table A6, we report the estimates for the measurement equations and, not surprisingly, we find that θ is also positively correlated with the months of schooling in the year before application, negatively correlated with the probability of having already obtained high school certification, and negatively (but not significantly) correlated with the months spent in job in the year before application.

The estimates of the selection equations also show that higher availability of alternative education programs, as expected, has a positive impact on the desire for alternative education, especially in period 3, when individuals have to select between a and n only. In every period, older and male individuals are less likely to enroll in any education, either Job Corps or another available program. This is true despite the fact that older individuals have consistently higher returns when they obtain an alternative education program, whereas the difference in returns to a between male and female participants is quite

small at every period.

We test for three different types of selection of unobservables: selection on *levels* implies that individuals with different desires for one type of education have different returns from that education (formally, this implies that $\gamma_{\theta}^{s,k} \neq 0$), selection on *gains* implies that, at every period s , individuals who prefer a to n or vice versa have different returns by enrolling in a rather than in n (formally $\gamma_{\theta}^{s,a} \neq \gamma_{\theta}^{s,n}$), *dynamic* selection on gains implies that individuals postpone their enrollment in k on the basis of the returns they can get by staying for longer periods in the training (formally $\gamma_{\theta}^{s,k} \neq \gamma_{\theta}^{s',k} \forall s \neq s'$). The bottom of Table 4 report the p-values of different tests for the null hypotheses of no selection on levels, no selection on gains and no dynamic selection on gains. We reject the null of no selection on levels and on gains, while we do not reject the null of no dynamic selection on gains.³¹ Note that the negative selection on observables is reversed when we analyse selection on unobservables: given the normalization we imposed and given the results in Table 3 the unobserved factor is positively correlated with the proportion of individuals enrolling in a at every period. Table 4 shows that individuals with higher value of θ have also higher returns in terms of earnings if they select a as compared to selecting n . This implies that individuals engage in “Roy-style” selection patterns. Note also that the individuals who are more likely to select an alternative education program based on the unobserved characteristics perform better in the cognitive test taken 30 months after initial random assignment if they enroll in alternative education after a Job Corps spell of any duration (see Appendix Table A5). This suggests that they may select into the treatment also based on the gain they receive in terms of cognitive ability.

Figure 2 plots the estimates and the 90% confidence intervals of the average potential log-earnings (Panel (a)) and employment (Panel(b)) if the whole population is assigned to each (s, k) choice.³² As expected, without Job Corps ($s = 0$) getting some type of education gives higher earnings and employment probability than not getting any education. The average outcomes, however, do not increase monotonically with duration in Job Corps, and only three periods of Job Corps clearly give higher returns than no Job Corps at all, if Job Corps is not aided by additional education. Earnings and employment probability after three periods in Job Corps are roughly the same with and without additional education, suggesting that, overall, the returns to education after three periods in Job Corps are small. The dashed line in the figure shows the average log-earnings and employment proportion

³¹We used the delta method to compute the standard errors of the test statistics.

³²Because we normalize the covariates to have mean zero, the constant term in the outcome equation estimates the average potential outcome, i.e. $E[Y_i^{s,k}] = \alpha^{s,k}$ in the case of earnings and $E[H_i^{s,k}] = \Phi(\alpha^{s,k})$ for employment.

of the individuals who in the data are observed in each (s, k) couple (for earnings, we report only the earnings of those who are employed). The comparison between the observed and potential earnings is an indication of the selection in the different (s, k) and in employment, showing that, on average, individuals select in the treatment that is more beneficial to them, because the realized outcome is always above the potential one.

To assess the fit of the model, in Table 5, we compare the sample moments for outcomes and choices to the moments predicted by the model, separating between the moments of the treated and control groups. The left panel of Table 5 compares the choice probabilities, and shows that the model matches the actual proportion well for both the treatment and control groups. Similarly, the model matches the employment proportion for the different subgroups. For log-earnings, there are some differences between the predictions of the models and the actual moments, but in the order of no more than 0.04 log-points. Furthermore, in Figure A1, we show that the density of the predicted and actual distribution of log-earnings overlap well.

5.2 Substitution Pattern and subLATEs

Using the estimated parameters of the sequential choice model, we compute the proportion of different types of compliers and the subLATEs, i.e. the ATE for the different types of compliers, by computing the difference in the predicted choices and the outcomes when the individuals are assigned to the treatment or the control group.³³ Column (1) of Table 6 shows the percentage of the whole population selecting each specific node; Column (2) shows the employment effect of the treatment for each compliance group; and Column (3) shows the effect of the treatment on earnings. In Table 6, earnings correspond to the “latent” productivity of the workers, i.e., the earnings predicted if everyone is employed. The first row shows the estimate of the overall LATE, when all the individuals who enroll in at least one period of Job Corps are considered as one unique compliance group. The estimates compare quite well with the nonparametric estimates of the effect of the treatment in the main sample, where the percentage of compliers is 71.8 (the model predicts 72.2) and the LATE on employment is

³³In more detail, with the estimates of the model, we simulate 30 paths of choices and associated outcomes for each individual in the sample starting from period $s = 0$. Each simulation consists of random draws of the unobserved factor and utility shocks. For each simulation, we simulate each individual’s choices when treated ($Z=1$) and when not treated ($Z=0$).

0.03 (the model predicts 0.04).³⁴

5.2.1 Substitution Pattern

The proportions in Column (1) show the patterns in complementarity and substitutability in terms of the demand for education. These estimates point at low levels of complementarities (participants who increase their demand for alternative education because of the increased availability of Job Corps—hence $\pi_{a,1a} + \pi_{a,2a} + \pi_{a,3a}$). Approximately 7% of the population complements Job Corps with additional education over the different periods, mostly due to individuals who spend a long period in Job Corps before receiving additional education (more than half of those who complement Job Corps and additional education stay longer than 6 months in the program).

The substitution rate is much higher: for 21.7% of the population the availability of Job Corps decreases the demand for other types of education. The largest part of individuals who substitute other programs with Job Corps are also those who spend more than 6 months in Job Corps. For the majority of the compliers who stay in Job Corps less than 6 months, Job Corps and the other education programs are intertemporal substitutes. Overall, the degree of intertemporal substitutability is thus quite high, about 28%. This can be due to different reasons: for example, participants may regret and drop out from Job Corps if they realize that they can be better off in another program, or they may see the returns from other education being higher after Job Corps.

In the last two rows of Table 6, we show the decomposition based on a *static* model, where we divide the compliers on the basis of their potential choice when assigned to the control groups, but not on the basis of the dynamic choice they make when assigned to the treatment group. The static

³⁴The LATE for earnings estimated nonparametrically is different from the LATE on the latent productivity, as reported in Table 6, because the LATE estimated nonparametrically is conditional on employment (see Appendix C). The overall LATE estimated nonparametrically is 0.05, while the model predicts a LATE of 0.02 for the latent productivity (the equivalent of the nonparametric LATE as predicted by our model is also 0.02). As mentioned in Section 3.3 the LATE estimated nonparametrically is the sum of the causal effect of Job Corps for the compliers employed when assigned to the treatment group and selection into employment. In Table A7 and in Appendix C.1, we show the decomposition of the LATE for individuals who are employed when assigned to the treatment group and we report the exact weights for that decomposition. We also checked whether our results are in line with the ones in Lee (2009) and Blanco, Flores, and Flores-Lagunes (2013a) who estimate upper and lower bounds for the effect of Job Corps on the wage rate of individuals who are employed both in the treatment and control group. Our estimate of this effect is again 0.02, which is inside the bounds estimated by Lee (2009) (-0.04, 0.11), but slightly outside the bounds in Blanco, Flores, and Flores-Lagunes (2013a) (0.04 and 0.10). Note that, unlike our model and crucial for their bounds estimation, both Lee (2009) and Blanco, Flores, and Flores-Lagunes (2013a) assume that receiving Job Corps cannot have a negative effect on the employment probability of the compliers. Frumento, Mealli, Pacini, and Rubin (2012) point estimates the effect for the same subgroup of compliers by imposing structure on the probability of being employed when assigned to the treatment or control group and finds an effect on hourly wage in the range 4-5% hence not far from the 2% estimates of our model.

framework would indicate that 49.8% of the population substitutes an alternative education program with Job Corps, because it considers a and jc mutually exclusive (see Section 3.2) and does not allow for intertemporal substitution. Thus, a more comprehensive dynamic framework is necessary not only to identify complementarities in demand (which are not accounted for in a static framework), but also to show the correct substitution rate.

5.2.2 subLATEs

Columns (2) and (3) in Table 6 show the ATE on employment and earnings for the whole population (in the first row) and for each subgroup of compliers (the following 12 rows). As mentioned previously, and similar to the findings in Schochet, Burghardt, and McConnell (2008), our model predicts a small positive effect of the treatment when considering the average effect for all the compliers. The decomposition of this overall effect in the effect for the 12 subgroups shows a quite heterogeneous pattern.

The effect of the treatment on earnings and employment is negative or quite small for individuals selecting only one period in Job Corps. It is particularly negative for the group for whom one period in Job Corps and the other education programs are intertemporal substitutes. This finding is in line with a learning model where individuals have imperfect information about the outcome of the treatment and thus enroll in the program in period one, but decide to drop out once they realize they would be better off selecting a different education. The negative effect could then be explained by the delay in starting the alternative education program.

The effect for three periods in Job Corps are instead positive and much larger than the overall LATE as estimated in the first row. The proportion of participants who stay more than six months in the program is larger than the proportion of those who stay for a shorter period; thus, the overall effect is still estimated to be positive. This evidence confirms the relevance of accounting for the dynamic framework when evaluating the effect of a treatment.

Another general pattern that emerges from the results on earnings and employment in Table 6 is that the effect of the treatment is smaller for the individuals whose next-best alternative in the control group is to enroll in some type of education. This result confirms that the small overall effect shown in the first row is partly due to individuals in the control group enrolling in alternative programs.

In Section 5.1, we show evidence that individuals select into education based on gains. In Table 7,

we show further evidence that selection into education plays a role in determining the final outcome of individuals by showing the characteristics of the different subgroups of compliers. These different characteristics affect selection and, ultimately, the effect of the treatment on the subgroup. For example, in Table 6 we have reported that the effect of three periods in Job Corps among those whose next-best option in the control group is a is smaller for those who substitute a (5 percentage points in employment, 3% in earnings) than for those who obtain more education afterwards (7 percentage points in employment, 7% in earnings). This difference is not informative on the actual effect of substituting education with Job Corps, because Table 7 shows that these groups are quite different in terms of both observables and unobservables (compliers in the latter subgroup are younger, more likely to be female, and have a stronger preference for a), suggesting that part of the difference in outcome is likely due to differences in the selection pattern.³⁵

5.3 Beyond LATE

5.3.1 Dynamic Program Interaction

Table 8 reports the estimates of the levels of dynamic complementarity or substitutability between Job Corps and the other available programs. In Panels A and B, we show the results for earnings and employment, respectively. The numbers in parentheses are the p-values of tests for the null hypothesis that the interaction is equal to zero. Different columns show the results for different subsamples of the population in terms of observable characteristics. In order to estimate the degree of complementarity or substitutability between each duration in Job Corps and a , we predict the potential outcomes in every combination of s and k for each individual in the simulated sample. We then compute the difference between $E[Y_i^{s,a} - Y_i^{s,n}]$ and $E[Y_i^{0,a} - Y_i^{0,n}]$ for the whole population, as described in Section 3.4.1. Thus, population-level estimates are not affected by the selection in the different combinations of duration in Job Corps and education, and allows us to compute the dynamic interaction between programs in generating the final outcome.

Overall, a short period in Job Corps seems to reduce the returns to additional education, but the

³⁵The same issue on the interpretation of the effect for the different subgroup of compliers when using the Wald estimator is raised in Heckman, Humphries, and Veramendi (2016), who analyze the interpretation of IV estimates in the case of an instrument shifting different individuals into different levels of education. They show that the effect may differ between subgroups shifted to different education levels because of differences between the subgroups or in the effect of different education levels.

effect is positive for women and, in general, not significant, for either earnings or employment. Obtaining between 3 and 6 months of Job Corps, instead, complements the other education programs available. The effect is large and significant for earnings, and implies that Job Corps increases the returns to different education programs by almost 40%. The magnitude of the complementarity is different for observationally different groups, but is always positive and large. The results on employment point in the same direction (complementarity between two periods in Job Corps and alternative education programs), despite not being statistically significant.

Table 8 shows no evidence of complementarities between long periods in Job Corps and the other available programs. In fact, the returns to alternative education programs are lower after long periods in Job Corps, especially for employment. This result points at some degree of substitutability between long enough periods in Job Corps and the other programs available and is mostly driven by the fact that, on average, the earnings and the employment rate after three periods in Job Corps are quite high even without obtaining more education (see also the graph in Figure 2). This is an indication that more than 6 months in Job Corps provides skills similar to the other programs and, thus, acquiring additional skills after Job Corps has smaller returns than acquiring them without long periods in Job Corps. These results are also quite heterogeneous and seem stronger for young women.

In addition to heterogeneity on observable characteristics, because individuals select on gains at any period, it is reasonable to expect that the returns to an alternative education program after Job Corps are higher for individuals with strong preference for a and that there may be heterogeneity in terms of complementarity conditional on the unobserved factor. In Figure 3, we plot the complementarity parameter separately for five different bins of the distribution of θ (where 1 is the bin with the lowest value and 5 is the highest). For individuals with a high value of θ , one period of Job Corps does complement a for earnings, and the complementarity between two periods in Job Corps and a is stronger than that for the rest of the population. The complementarity for employment between two periods in Job Corps and an alternative education program is also higher for individuals with stronger preference for a , but for them, the substitutability between long periods in Job Corps and other education programs is also higher, confirming that the selection on gains is more likely to happen in terms of earnings rather than employment opportunities.

Table 9 compares the differences in $E[Y_i^{s,k} - Y_i^{0,k}]$ for $k = a$ (Column (1)) and $k = n$ (Column (3))

when estimated over the whole population, i.e., the ATE of the program for different durations, and when estimated over the subgroup of compliers selecting $D^0 = (0, k), D^1 = (s, k)$ (Columns (2) and (4)), i.e., the subLATEs. As shown in equation (7), the difference between the effect for the whole population and the effect for the subgroup shows the degree of selection on gains. As expected, the subLATEs are in general larger than the ATE and hence there is positive selection on gain. If there were no selection, the difference between the subLATEs for individuals selecting $D_i^1 = (s, a), D_i^0 = (0, a)$ and those selecting $D_i^1 = (s, n), D_i^0 = (0, n)$ would be equivalent to the dynamic complementarity. Overall, looking only at the difference between subLATEs would predict smaller degree of complementarity/substitutability: for example, the returns to additional education increase by 40% after two periods in Job Corps, but the comparison between subLATEs predicts an increase by only 5%. The difference between the degree of program interaction, as predicted by the subLATEs, and the one in the population, depends on the difference in selection on gain into the subgroup selecting $D^0 = (0, a), D^1 = (s, a)$ and the one selecting $D^0 = (0, n), D^1 = (s, n)$. There is no a priori reason to assume that the selection into the different subgroups must be the same (and, in fact, we find that this is not the case). This implies that, even in the presence of IVs that allow us to identify the subLATEs, it is not possible to estimate the degree of complementarity or substitutability between programs.

5.3.2 Dynamic Treatment Effect

Table 10 shows the DATTE and its decomposition for earnings (Panel A) and employment (Panel B). Following Section 3.4.2, we refer to $DATTE^s$ as the effect of transitioning to s from period $s - 1$. We compare the returns to the next best choice in terms of a and n at $s - 1$ to the returns to the optimal choice for all the individuals who select transition to s periods in Job Corps. Using our estimates, we predict the next-best alternative in terms of a and n for all the individuals in every period and their outcomes when selecting their next-best alternative. We then decompose the $DATTE^s$ in the *direct effect* of obtaining s periods of training instead of $s - 1$ and the *continuation value* of being able to select additional Job Corps the following periods.

The DATTE is positive for both earnings and employment at any period, but the magnitudes and the mechanisms producing these effects are different. At period 1, the DATTE corresponds to the LATE estimated using the the Wald estimator (see Section 3.4.2). The direct effect of Job Corps is

negative for both outcomes (-9% for earnings and -3 percentage points for employment), but this is compensated by a high continuation value. In the second period the direct effect is high, especially for earnings (+10%, 3 percentage points for employment). Because two periods in Job Corps and an alternative education program are complementary in the outcome, this result is not surprising: most individuals who reach period 2 would select a if they were forced to stop at period 2, and, given the high degree of complementarity between two periods in Job Corps and a , would have quite high gain from selecting $D = (2, a)$. The continuation value is however also positive, suggesting that moving to period 3 gives, on average, even higher returns than stopping at period 2 and selecting a . The overall value of transitioning from period 1 to period 2 is thus quite high (increase by 14% in earnings and 9 percentage points in employment). Because period 3 is the terminating period, it only involves a direct effect, which is positive for both outcomes (6% higher earnings and 9 percentage points increase in employment).³⁶

In Figure 4, we show that the effect is quite heterogeneous across the distribution of the unobservables. We divide the distribution of θ into five bins and plot the overall DATT and the direct effect. The difference between the DATT and the direct effect is the continuation value. Having only one period in Job Corps has a positive direct effect for the individuals with weaker preferences for a , who are then more likely to select n when not enrolled in Job Corps. This finding aligns well with the result that obtaining one period in Job Corps does not complement alternative education, especially for individuals with weaker preference for a . Similarly, the direct effect of two periods in Job Corps is high for individuals who are more likely to later enrol in additional education, possibly because of the high degree of complementarity between two periods in Job Corps and the other educational programs. Panel (b) shows more evidence of selection on gains. Recall from Section 5.1 that high values of θ are strongly linked with the probability of selecting a after 2 periods in Job Corps. Individuals in the three highest bins of θ have similar values of $DATT^2$ for employment, but this is caused mostly by the direct effect for those individuals who are more likely to stop at $s = 2$ and then select a (i.e., those with the highest value of θ), whereas it is driven by the continuation value for individuals who are more likely to proceed to period 3 (i.e., those with lowest value of θ in the third bin).

³⁶There is a mechanical link between the continuation value at period 2 and the DATT period 3, when the program terminates. $DATT^3$ is the continuation value at $s = 2$ divided by the proportion of individuals who continued to $s = 3$ among those who reached period 2.

6 Selective Targeting and Cost–Benefit Analysis

The heterogeneity in the effect of Job Corps shown in the previous section, suggests that an expansion of the program targeting those individuals who benefit the most from it may increase the overall welfare more than would a random expansion in the offer. In this section, we begin by defining the benefit from access to Job Corps, considering the benefits for individuals in terms of the net lifetime earnings for their life-span after the program is finished. We then show how we can use our estimated model to predict the expected gain from access to Job Corps at *individual* level, which form the basis for program targeting. Our exercise may be particularly useful and relevant if a policy maker would like to expand Job Corps to a pool of new individuals but face a budget constraint. We show that the selective targeting sample would outperform in terms of cost-effectiveness relative to the benchmark of offering Job Corps to a random set of individuals.

We designate $\bar{Y}_i^{D_i^1}$ and $\bar{Y}_i^{D_i^0}$ the lifetime earnings of individual i when assigned to the treatment and control group, respectively, and τ is the tax rate.³⁷ The individual net benefit of receiving the treatment is:

$$B_i = (1 - \tau)(\bar{Y}_i^{D_i^1} - \bar{Y}_i^{D_i^0}). \quad (21)$$

Given that there is a direct mapping of the program effect in terms of predicted log-earnings on the benefit, the results in the previous sections suggest that there should also be significant heterogeneity in the individual benefit based on both the observable and unobservable characteristics. In Figure 5 we plot the average benefit for 10 different bins of the distribution of θ (where 1 is the smallest and 10 the highest) separately for different age groups and gender. The average of the individual benefit from the program in the population is estimated to be around 3,300\$ over the life cycle. The pattern for young individuals is monotonic decreasing, indicating that among younger participants, those who have higher propensity to enroll in an alternative education program gain less from the program, whereas those who are less likely to take some education when in the control group can gain up to three times the average individual if assigned to the treatment group. Differently, for older participants those with

³⁷We predict the evolution of earnings over the lifetime following the discussion in McConnell and Glazerman (2001), and we transform the estimates of log-earnings predicted by our model in the present discount value of lifetime earnings (in 1995 dollars terms), assuming that the positive effect of Job Corps on earnings decreases over time. We use the tax rate as defined in McConnell and Glazerman (2001). Details about the computation of benefits are in the Appendix F.1.

high values of θ (and thus higher taste for more education) can gain equally as much.

The heterogeneity shown in Figure 5 suggests that ideal targeting should account for both the observable and unobservable characteristics of the individuals. Although θ is not observed, we estimate its density function and can thus recover its posterior distribution, conditional on observed characteristics, choices, and measurements by exploiting Bayes rule:

$$P(\theta_i | \mathbf{M}_i, \mathbf{D}_i, \mathbf{x}_i) = \frac{P(\mathbf{M}_i, \mathbf{D}_i | \theta_i, \mathbf{x}_i) P(\theta_i | \mathbf{x}_i)}{\int P(\mathbf{M}_i, \mathbf{D}_i, | \theta_i, \mathbf{x}_i) P(\theta_i | \mathbf{x}_i) d\theta} \quad (22)$$

Using the posterior distribution of θ , we can predict the benefit for the individual i (EB_i) as a function of the measurements and choices:

$$EB_i \equiv E(B_i | M_i, D_i, x_i) = \int B_i(\theta_i, \mathbf{x}_i) P(\theta_i | \mathbf{M}_i, \mathbf{D}_i, \mathbf{x}_i) d\theta \quad (23)$$

Suppose that the policy maker wants to expand the treatment to 60% of the control group. If the expansion is done at random, the expected benefit for the newly treated would be the same as that for the whole population. The expected benefit of a random expansion is then the same as that for the group randomly treated in the NJCS, which is about 3,300\$. However, in Figure 6 we show that the variance of the distribution of the individual benefit, as computed using the formula in equation (23), is large; hence, targeting the top 60% in terms of benefit instead of a random allocation can lead to significant improvements in the overall benefits.³⁸ Call $F_B(b) = P(EB_i \leq b)$ as the cumulative distribution function of the (posterior) empirical distribution of EB_i . Its inverse $F_B^{-1}(p)$ gives the value b such that $P(EB_i \leq b) = p$. Hence, to target the proportion p of the individuals with the highest B_i , we target the individuals for whom $EB_i \geq F_B^{-1}(1 - p)$. The expected benefits from an expansion of the treatment targeting the proportion p with the highest EB_i is thus:

$$E[B_i | EB_i \geq F_B^{-1}(1 - p)] = (1 - \tau) E[\bar{Y}_i^{D_i^1} - \bar{Y}_i^{D_i^0} | EB_i \geq F_B^{-1}(1 - p)]. \quad (24)$$

In the first row of Table 11, we report the expected benefit from a random expansion to 60% of

³⁸An alternative to targeting on the basis of the benefit computed using the conditional posterior distribution of θ would be to target individuals only on the basis of their observable characteristics (Frumento, Mealli, Pacini, and Rubin, 2012). However, in Figure A3, we show that the predicted benefit is very heterogeneous, even within the subgroups of individuals with similar observable characteristics, and thus, targeting based on unobservable characteristics should outperform that based on the observable characteristics.

the control group (3,265\$) and the expected benefit of an expansion targeting the individuals in the top 60% of the distribution of the individual benefit (6,533\$). The expected benefit from a targeted expansion is almost double that for a randomized expansion. However, this increase in the total individual benefits may not be cost-effective, because by targeting the individuals with the highest benefit, we may also increase the overall cost of training. To compute the costs of trainings, we follow McConnell and Glazerman (2001). Details about the computation of benefits and costs are in the Appendix F.1. When computing the cost of the expansion we consider that the substitution of other education programs with Job Corps represents a saving for the public finance, and additional savings may come from the increased tax revenues due to higher earnings of the the individuals. The overall cost is then the sum of costs and savings.³⁹

We can, thus, compare the expected benefit and the expected costs of a random and a targeted expansion. In Appendix F.2, we show that the expected benefit and the expected cost in our definition correspond to the benefit and cost of a marginal expansion of the program to a random individual in the target population (i.e., either the overall population or the population with the highest individual benefit). The ratio between the benefit of marginally expanding the offer and the cost of the marginal expansion (known as the marginal value of public fund [MVPF]) measures the value of an extra dollar spent in the policy net of fiscal externalities. Comparing the MVPF of different policies informs whether moving resources from one to the other improves social welfare (Hendren, 2016; Kline and Walters, 2016). In other words, by comparing the MVPF of an expansion for the whole population and of an expansion for the group with the highest EB_i , we can investigate whether social welfare decreases moving from one policy to the other.

In the first row of Table 11, we report the MVPF for a random expansion of the program (Column (3)) and for a targeted expansion to the 60% with the highest individual benefit (Column (4)). The MVPF is smaller than 1, indicating that the costs of administrating the program are larger than the

³⁹In this analysis, we assume that the alternative education programs are not rationed, implying that the expansion in Job Corps does not encourage a larger proportion of individuals in the control group to enroll in alternative education. If we allowed for this, there is another category of individuals affected by the reform, i.e., those in the control group who fill the slots in alternative educational programs left by those enrolling in Job Corps. This would also reduce the savings because of individuals substituting existing education with Job Corps (Kline and Walters, 2016). In a dynamic framework this may also imply that some individuals who select Job Corps may not be able to reenroll in education after Job Corps if the available slots in an alternative education program have already been filled by control group members. These effects are not nonparametrically identified, but our model can account for them, for example, by assuming that the treatment increases the utility of a for the individuals in the control group who currently select n , and that these shifts are permanent, thus reducing the probability of receiving a after Job Corps for the individuals in the treatment group who enroll in Job Corps.

benefits, as has been shown in the cost–benefit analysis conducted by McConnell and Glazerman (2001). A targeted expansion that increases individual benefit doubles the MVPF, implying that the change in costs due to expanding the treatment to targeted individuals is not larger than the increase in benefits, and also makes the targeted expansion preferable from a cost-effectiveness perspective (in fact, the marginal costs decrease because of the increase in tax revenues).

To gain a better understanding of the importance of an alternative education program in determining the costs and benefits of Job Corps, in the second and the third row of Table 11, we compare a random with a targeted expansion in different subgroups of the population, depending on the availability of alternative education programs. When the availability of alternative programs is high, the expected benefits are lower than that for the overall sample, because fewer people enroll in long periods of Job Corps, and the MVPF is also lower than that for the whole population. This suggests that the availability of alternative education programs plays an important role in determining the cost-effectiveness of Job Corps.⁴⁰

In Table 12, we show how the average characteristics of the overall sample (Column (1)) and the top 60% sample (the “Targeted Sample”, Column (2)) differ. In Column (3), we show the p-value of a t-test for the equality in the average of these variables between the top 60% and the bottom 40% of the distribution of individual benefits. As expected from Figure 5, the majority of the targeted individuals are men and older. They are also more likely to live in states with lower availability of alternative education programs. Given these differences, the proportions selecting into different subgroups of compliers change, and, due to heterogeneity in the treatment effect, so do the subLATEs. As shown in Appendix Table A8, the targeted group includes less individuals who stay for only one period in Job Corps and more who stay three periods. The targeting excludes the individuals who stay a short period in Job Corps and receive the lowest returns from this treatment (because the subLATEs for those selecting 1 period in Job Corps increase), and includes the individuals who benefit the most from three periods in Job Corps (the subLATEs for those selecting three periods in Job Corps also increase).

⁴⁰In Table A9 we show the difference in the shares of compliers and subLATEs separately for the sample with high and low availability of alternative education programs (here we do not consider the targeted sample, but only the whole sample). When availability is high, the proportion of compliers enrolling in a after Job Corps is higher, but the returns to that choice seem smaller (for example the effect on earnings and education for the subgroups selecting $D^0 = (0, a)$, $D^1 = (3, a)$ and those selecting $D^0 = (0, n)$, $D^1 = (3, a)$ is smaller in the sample with high availability). In general, the high availability of a increases the probability of enrolling in a after Job Corps, but the marginal individual induced to take more education by the fact of having higher availability is negatively selected as compared with the participants who also select a when availability is low.

We can further ask whether we can make the program cost-effective in absolute terms (i.e., with $MVPF > 1$) by targeting a different population. In Table 13, we show the change in expected benefit and in the MVPF when different proportions of the population are targeted in terms of their individual benefit. The MVPF is inversely related to the proportion of targeted individuals, suggesting that the change in expected cost is smaller than that in expected benefit, because we target a population that has higher gains from the treatment. In particular, a treatment that targets at most the top 30% of the applicants in terms of their individual benefit would be cost-effective, as 1\$ spent in a program targeting the top 30% would raise the expected after-tax income of participants by approximately 1\$.

7 Conclusion

In this paper, we used a sequential choice model to analyze the dynamic treatment effects of the Job Corps program on labor market outcomes among disadvantaged youths. Our model features sequential selection in training duration and accounts for the additional training opportunities facing both program participants and nonparticipants. These features are important empirically, because a substantial fraction of program participants and nonparticipants enrolled in training/education programs outside Job Corps. Using data from the NJCS, we estimated the model by exploiting the randomization of Job Corps offer as a source of identification and imposing a factor structure on the unobservables in characterizing the joint distribution of choices and outcomes.

We found that enrolling in Job Corps for a period between 3 and 6 months increases the returns to other types of education, but shorter and longer periods do not. If anything, longer periods in Job Corps substitute for the skills eventually gained if enrolled in other education programs. In line with this finding, we also showed that the direct effect of the first period in Job Corps on both earnings and employment is negative. The overall benefit of enrolling in the first period of Job Corps (which corresponds to the LATE estimated using the standard IV framework) thus lies in the opportunity of obtaining additional education once the first period is over.

We also showed that the treatment effect is heterogeneous at the individual level, depending on both observable characteristics and the predicted unobservable factor. This implies that targeting the individuals with the highest benefit should be a way to improve the effectiveness of the program. For example, we showed that targeting the 60% of the population with the highest predicted individual

returns doubles the expected benefits of the program as compared to a random allocation to 60% of the population. Comparing the expected benefits with the expected costs, we found that, overall, the program is not cost-effective, which is consistent with the findings of McConnell and Glazerman (2001). The proposed targeting strategy can improve the cost-effectiveness of Job Corps, and we show that, if the program targeted only the 30% of the population with the highest returns, the expected benefits would eventually exceed the expected costs.

Our dynamic potential outcome framework can be applied in many different contexts, especially in the analysis of RCTs that rely on “encouragement design”. Like the NJCS, in these experiments, the duration of participation may be endogenous, and participation in the program of interest may provide incentives for enrolling in additional programs. Our results suggest that analyzing the selection patterns into alternative programs seems fruitful, because it is informative in terms of understanding the nature of complementarity (or substitutability) between different programs. We also believe that the selective targeting exercise conducted in this paper can be useful for a policy maker considering to expand a given program in other contexts. Researchers often collect rich sets of baseline information from experiment participants, and this paper shows that such information can be better exploited to predict their unobserved heterogeneity and potential gains from program participation. In this respect, combining the experimental variation with a choice-theoretic framework is useful because it can provide useful implications for policymakers in regard to how to change the design of programs in order to improve their cost effectiveness.

References

- ABBRING, J. H., AND G. J. VAN DEN BERG (2003): “The nonparametric identification of treatment effects in duration models,” *Econometrica*, 71(5), 1491–1517.
- ALTONJI, J., AND N. WILLIAMS (2005): “Do Wages Rise with Job Seniority? A Reassessment,” *Industrial and Labor Relations Review*, 58(3), 370–397.
- BA, B. A., J. C. HAM, R. J. LALONDE, AND X. LI (2017): “Estimating (easily interpreted) dynamic training effects from experimental data,” *Journal of Labor Economics*, 35(S1), S149–S200.
- BARNOW, B. S., AND J. SMITH (2016): “Employment and training programs,” in *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, pp. 127–234. University of Chicago Press.
- BERNDT, E. K., B. H. HALL, R. E. HALL, AND J. HAUSMAN (1974): “Estimation and inference in nonlinear structural models,” *Annals of Economic and Social Measurement*, 3(4), 653–665.
- BLANCO, G., C. A. FLORES, AND A. FLORES-LAGUNES (2013a): “Bounds on average and quantile treatment effects of Job Corps training on wages,” *Journal of Human Resources*, 48(3), 659–701.
- (2013b): “The effects of Job Corps training on wages of adolescents and young adults,” *American Economic Review: Papers & Proceedings*, 103(3), 418–22.
- BLOOM, H. S., L. L. ORR, S. H. BELL, G. CAVE, F. DOOLITTLE, W. LIN, J. M. BOS, ET AL. (1997): “The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study,” *Journal of human resources*, 32(3).
- BORJAS, G. J., AND S. ROSEN (2012): “Income prospects and job mobility of younger men,” in *35th Anniversary Retrospective*, pp. 441–463. Emerald Group Publishing Limited.
- CALÓNICO, S., AND J. SMITH (2017): “The women of the national supported work demonstration,” *Journal of Labor Economics*, 35(S1), S65–S97.
- CARD, D., AND D. R. HYSLOP (2005): “Estimating the effects of a time-limited earnings subsidy for welfare-leavers,” *Econometrica*, 73(6), 1723–1770.
- CUNHA, F., AND J. HECKMAN (2007): “The technology of skill formation,” *American Economic Review*, 97(2), 31–47.
- DUFLO, E., R. GLENNERSTER, AND M. KREMER (2007): “Using randomization in development economics research: A toolkit,” *Handbook of development economics*, 4, 3895–3962.
- EBERWEIN, C., J. C. HAM, AND R. J. LALONDE (1997): “The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data,” *The Review of Economic Studies*, 64(4), 655–682.

- FITZENBERGER, B., A. OSIKOMINU, AND R. VÖLTER (2008): “Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany,” *Annales d’Economie et de Statistique*, pp. 321–355.
- FLORES, C. A., A. FLORES-LAGUNES, A. GONZALEZ, AND T. C. NEUMANN (2012): “Estimating the effects of length of exposure to instruction in a training program: the case of job corps,” *Review of Economics and Statistics*, 94(1), 153–171.
- FRUEHWIRTH, J. C., S. NAVARRO, AND Y. TAKAHASHI (2016): “How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects,” *Journal of Labor Economics*, 34(4), 979–1021.
- FRUMENTO, P., F. MEALLI, B. PACINI, AND D. B. RUBIN (2012): “Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data,” *Journal of the American Statistical Association*, 107(498), 450–466.
- GENTZKOW, M. (2007): “Valuing new goods in a model with complementarity: Online newspapers,” *American Economic Review*, 97(3), 713–744.
- HAM, J. C., AND R. J. LALONDE (1996): “The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training,” *Econometrica*, 64, 175–206.
- HANSEN, K. T., J. J. HECKMAN, AND K. J. MULLEN (2004): “The effect of schooling and ability on achievement test scores,” *Journal of econometrics*, 121(1-2), 39–98.
- HECKMAN, J., N. HOHMANN, J. SMITH, AND M. KHOO (2000): “Substitution and dropout bias in social experiments: A study of an influential social experiment,” *The Quarterly Journal of Economics*, 115(2), 651–694.
- HECKMAN, J., J. SMITH, AND C. TABER (1998): “Accounting for dropouts in evaluations of social programs,” *Review of Economics and Statistics*, 80(1), 1–14.
- HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2016): “Dynamic treatment effects,” *Journal of econometrics*, 191(2), 276–292.
- (2018): “Returns to education: The causal effects of education on earnings, health, and smoking,” *Journal of Political Economy*, 126(S1), S197–S246.
- HECKMAN, J. J., AND S. NAVARRO (2007): “Dynamic discrete choice and dynamic treatment effects,” *Journal of Econometrics*, 136(2), 341–396.
- HECKMAN, J. J., AND S. URZUA (2010): “Comparing IV with structural models: What simple IV can and cannot identify,” *Journal of Econometrics*, 156(1), 27–37.
- HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2006): “Understanding instrumental variables in models with essential heterogeneity,” *The Review of Economics and Statistics*, 88(3), 389–432.

- HENDREN, N. (2016): “The policy elasticity,” *Tax Policy and the Economy*, 30(1), 51–89.
- KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): “Field of study, earnings, and self-selection,” *The Quarterly Journal of Economics*, 131(3), 1057–1111.
- KLINE, P., AND C. R. WALTERS (2016): “Evaluating public programs with close substitutes: The case of Head Start,” *The Quarterly Journal of Economics*, 131(4), 1795–1848.
- LALONDE, R. J. (1986): “Evaluating the econometric evaluations of training programs with experimental data,” *The American economic review*, pp. 604–620.
- LECHNER, M. (2002): “Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies,” *Review of Economics and Statistics*, 84(2), 205–220.
- LECHNER, M., AND R. MIQUEL (2010): “Identification of the effects of dynamic treatments by sequential conditional independence assumptions,” *Empirical Economics*, 39(1), 111–137.
- LEE, D. S. (2009): “Training, wages, and sample selection: Estimating sharp bounds on treatment effects,” *The Review of Economic Studies*, 76(3), 1071–1102.
- MCCONNELL, S., AND S. GLAZERMAN (2001): “National Job Corps Study: The Benefits and Costs of Job Corps.,” *Princeton, NJ: Mathematica Policy Research, Inc.*
- RODRÍGUEZ, J., F. SALTIEL, AND S. S. URZÚA (2018): “Dynamic Treatment Effects of Job Training,” Discussion paper, National Bureau of Economic Research.
- SCHOCHET, P. Z., J. BURGHARDT, AND S. MCCONNELL (2008): “Does job corps work? Impact findings from the National Job Corps Study,” *The American economic review*, 98(5), 1864–1886.
- TABER, C. R. (2000): “Semiparametric identification and heterogeneity in discrete choice dynamic programming models,” *Journal of econometrics*, 96(2), 201–229.

Table 1: Percentages in Each Duration/Education Bundle

JC Duration in the (Months)	No Training (n) (1)	Alternative Training (a) (2)	Total (3)	Ratio (a/n) (4)
<i>control (Z=0)</i>				
No Job Corps	31.7	68.3	100	2.1
<i>treatment (Z=1)</i>				
No Job Corps	8.9	19.2	28.1	2.2
1-3	8.2	12.1	20.3	1.5
4-6	6.3	8.0	14.3	1.5
7+	22.4	15.0	37.4	0.7
Treatment Total	45.8	54.2	100.0	1.2

Note: Cells in Columns (1) and (2) show the percentages of individuals selecting different combinations of periods in Job Corps and education when not enrolled in Job Corps. Column (3) shows the sum of cells in Columns (1) and (2), i.e., the percentages of individuals selecting each period in Job Corps, irrespective of the education choice after Job Corps.

Table 2: Empirical Estimates for Ordinary Least Squares Regressions

	Earnings (1)	Probability of being Employed (2)	Earnings (3)	Probability of being Employed (4)
Z	18.0*** (4.5)	2.5** (1.0)		
$Z \times n$ without JC			2.8 (9.9)	-0.4 (2.3)
$Z \times a$ without JC			16.6** (7.9)	1.2 (1.6)
$Z \times n$ after 1 Period in JC			0.4 (10.6)	-1.8 (2.4)
$Z \times a$ after 1 Period in JC			-13.2 (8.3)	-2.8 (2.0)
$Z \times n$ after 2 Periods in JC			16.5 (12.3)	2.9 (2.6)
$Z \times a$ after 2 Periods in JC			20.5* (11.7)	1.9 (2.3)
$Z \times n$ after 3 Periods in JC			31.6*** (7.2)	5.8*** (1.5)
$Z \times a$ after 3 Period in JC			43.0*** (8.7)	7.8*** (1.7)
Constant	200.6*** (3.4)	68.1*** (7.8)	200.6*** (3.4)	68.1*** (7.8)
N	9,429	9,429	9,429	9,429

Note: $Z = 1$ if the person is assigned to treatment, $a =$ alternative education, $n =$ no education. Each column is from a separate regression of the outcome variable on a dummy variable for having been randomly assigned to Job Corps (columns (1) and (2)) and its interaction with dummy variables for the different duration/education combinations (columns (3) and (4)). As in Schochet, Burghardt, and McConnell (2008), we use sample weights in these regressions to adjust for the sample and survey design. Standard errors are in parentheses. * indicates that the estimates are significant at 10% level, ** indicates that the estimates are significant at 5% level, *** indicates that the estimates are significant at 1% level.

Table 3: Empirical Estimates for Selection Models

Periods in JC	Parameter	Alternative Education	Job Corps
0	Constant	0.6*** (0.05)	1.45*** (0.07)
	Old	-0.5*** (0.05)	-0.52*** (0.05)
	Female	0.33*** (0.05)	0.06 (0.05)
	Availability	0.08** (0.03)	
	$\beta_{\theta}^{0,k}$	1 ^[a] -	0.75*** (0.07)
1	Constant	0.03 (0.22)	1.12*** (0.13)
	Old	-0.39*** (0.12)	-0.06 (0.08)
	Female	0.29*** (0.09)	0.29*** (0.07)
	Availability	0.03 (0.05)	
	$\beta_{\theta}^{1,k}$	1.18** (0.54)	0.32 (0.41)
2	Constant	0.94** (0.42)	1.92*** (0.39)
	Old	-0.63*** (0.2)	-0.25* (0.13)
	Female	0.65*** (0.18)	0.51*** (0.15)
	Availability	0.03 (0.06)	
	$\beta_{\theta}^{2,k}$	2.19*** (0.82)	1.74*** (0.57)
3	Constant	-0.41*** (0.09)	
	Old	-0.26*** (0.1)	
	Female	0.34*** (0.09)	
	Availability	0.26*** (0.07)	
	$\beta_{\theta}^{3,k}$	0.65 (0.45)	

^[a] The factor loading on θ_i in the selection equation for a at period 0 (β_a^0) is normalized to 1 to fix the scale.

Note: “Old” is a dummy variable for being more than 20 years of age, “Female” is a dummy variable for being female. These two dummy variables are normed to have mean 0. “Availability” is a dummy variable for living in a state where the availability of educational programs other than Job Corps (community colleges and postsecondary public institutions) is above the median availability in the cross-states distribution. Standard errors are in parentheses. Significance level (t-test for testing if each parameter=0): *** 1%, ** 5%, * 10%.

Table 4: Empirical Estimates for Earnings Model

Periods in JC	Parameter	Alternative Education	No Education
0	Constant	5.46*** (0.02)	5.36*** (0.05)
	Old	0.11*** (0.02)	0.07** (0.03)
	Female	-0.22*** (0.02)	-0.25*** (0.04)
	$\gamma_{\theta}^{0,k}$	0.08** (0.03)	-0.13** (0.05)
1	Constant	5.28*** (0.08)	5.26*** (0.11)
	Old	0.16** (0.07)	0.07 (0.07)
	Female	-0.2*** (0.06)	-0.47*** (0.07)
	$\gamma_{\theta}^{1,k}$	0.23*** (0.08)	-0.27*** (0.1)
2	Constant	5.55*** (0.08)	5.04*** (0.18)
	Old	0.2** (0.08)	0.09 (0.09)
	Female	-0.28*** (0.06)	-0.33*** (0.09)
	$\gamma_{\theta}^{2,k}$	-0.04 (0.19)	-0.38** (0.15)
3	Constant	5.5*** (0.05)	5.49*** (0.03)
	Old	0.1* (0.05)	0.04 (0.04)
	Female	-0.26*** (0.05)	-0.24*** (0.04)
	$\gamma_{\theta}^{3,k}$	0.21*** (0.07)	-0.03 (0.09)

Testing for selection (p-values):

No selection on levels ($H_0 : \gamma_{\theta}^{s,k} = 0 \forall s, k$) 0.00

No selection on gains ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s,n} \forall s$) 0.00

No dynamic selection on gains in a ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s',a} \forall s \neq s'$) 0.52

No dynamic selection on gains in n ($H_0 : \gamma_{\theta}^{s,n} = \gamma_{\theta}^{s',n} \forall s \neq s'$) 0.42

Note: “Old” is a dummy variable for being more than 20 years of age and “Female” is a dummy variable for being female. These two dummy variables are normed to have mean 0, which allows the intercept in the outcome equation to be interpreted as the average potential outcome in the population. Standard errors are in parentheses. Significance level (t-test for testing if each parameter=0): *** 1%, ** 5%, * 10%.

Table 5: Goodness of Fit of the Models

JC duration	Proportions		Log-Earnings		Employment	
	No Education	Alternative	No Education	Alternative	No Education	Alternative
	<i>control group</i>		<i>control group</i>		<i>control group</i>	
0 (predicted)	31.1	68.9	5.48	5.49	65.5	68.8
0 (data)	<i>31.7</i>	<i>68.3</i>	<i>5.49</i>	<i>5.45</i>	<i>66.5</i>	<i>69.1</i>
	<i>treated group</i>		<i>treated group</i>		<i>treated group</i>	
0 (predicted)	8.7	19.3	5.50	5.46	68.0	69.6
0 (data)	<i>8.9</i>	<i>19.2</i>	<i>5.46</i>	<i>5.48</i>	<i>67.5</i>	<i>69.4</i>
1 (predicted)	8.1	12.3	5.46	5.44	66.6	65.5
1 (data)	<i>8.2</i>	<i>12.1</i>	<i>5.47</i>	<i>5.45</i>	<i>66.2</i>	<i>65.1</i>
2 (predicted)	6.0	7.9	5.49	5.49	69.8	69.4
2 (data)	<i>6.3</i>	<i>8.0</i>	<i>5.48</i>	<i>5.51</i>	<i>70.3</i>	<i>69.4</i>
3 (predicted)	22.6	15.1	5.50	5.52	73.7	76.1
3 (data)	<i>22.4</i>	<i>15.0</i>	<i>5.54</i>	<i>5.54</i>	<i>74.0</i>	<i>75.7</i>

Note: Numbers in italics are the population equivalents (computed directly from the data).

Table 6: Empirical Estimates of Subgroups' Shares and subLATEs

D^0	D^1	Share	subLATEs	
		(1)	Employment (2)	Earnings (3)
Overall LATE				
no jc	jc	72.2	0.04	0.02
Dynamic Model				
a	(1,a)	10.4	-0.07	-0.07
n	(1,a)	1.7	0.10	-0.01
n	(1,n)	3.6	0.02	0.04
a	(1,n)	4.6	-0.03	-0.11
a	(2,a)	6.5	0.00	0.02
n	(2,a)	1.6	0.11	0.17
n	(2,n)	4.0	0.01	-0.03
a	(2,n)	2.4	0.02	-0.03
a	(3,a)	11.2	0.07	0.07
n	(3,a)	3.6	0.15	0.09
n	(3,n)	7.9	0.13	0.09
a	(3,n)	14.7	0.05	0.03
Static Model				
n	jc	22.4	0.09	0.06
a	jc	49.8	0.02	0.00

Note: D^0 and D^1 refer to the potential choices in terms of periods in Job Corps (s) and of education (k) for the individuals when assigned to the control and treatment groups, respectively. Because the individuals in the control group cannot enroll in Job Corps, D^0 is expressed only in terms of $k \in (a, n)$. As in Section 3.3, “subLATEs” refer to average treatment effect for the subgroup of individuals selecting each (D^0, D^1) bundle. The “Overall LATE” specification includes in one unique group all individuals for whom $s > 0$ under D^1 , not separating between different k under D^0 . This is the LATE computed by the standard Wald estimator. The “Dynamic Model” separates between different s and k under D^1 and different k under D^0 . The “Static Model” includes in one unique group all individuals for whom $s > 0$ under D^1 , but separates between different k under D^0 .

Table 7: Difference in Compliers Characteristics

D^0	D^1	Female (1)	Old (2)	θ (3)
a	(1,a)	-0.02	-0.11	0.77
n	(1,a)	-0.10	0.00	0.22
n	(1,n)	-0.17	0.04	-0.69
a	(1,n)	-0.10	-0.04	-0.18
a	(2,a)	0.05	-0.11	0.42
n	(2,a)	0.00	-0.04	-0.05
n	(2,n)	-0.13	0.03	-1.09
a	(2,n)	-0.09	-0.04	-0.81
a	(3,a)	0.08	-0.04	0.31
n	(3,a)	0.03	0.05	-0.14
n	(3,n)	-0.07	0.09	-0.40
a	(3,n)	-0.01	0.00	-0.01

Note: D^0 and D^1 refer to the potential choices of the individuals when assigned to the control and treatment groups, respectively. Columns (1)-(3) show the difference between the average characteristics in the subgroup of compliers and those in the population.

Table 8: Dynamic Complementarities between Different Duration in Job Corps and Alternative Education Programs

Periods in Job Corps	Overall	Men		Women	
		Young	Old	Young	Old
Panel A: Earnings					
1	-0.07 (0.60)	-0.19 (0.18)	-0.14 (0.36)	0.05 (0.76)	0.10 (0.58)
2	0.40 (0.04)	0.37 (0.05)	0.44 (0.04)	0.40 (0.08)	0.46 (0.05)
3	-0.09 (0.19)	-0.07 (0.36)	-0.06 (0.50)	-0.13 (0.16)	-0.11 (0.22)
Panel B: Employment					
1	-0.10 (0.31)	-0.12 (0.30)	-0.01 (0.90)	-0.15 (0.15)	-0.05 (0.67)
2	0.16 (0.42)	0.13 (0.50)	0.14 (0.41)	0.17 (0.42)	0.20 (0.35)
3	-0.09 (0.03)	-0.09 (0.10)	-0.03 (0.66)	-0.14 (0.01)	-0.08 (0.20)

Note: Each column shows the difference between $E[Y_i^{s,a} - Y_i^{s,n}]$ and $E[Y_i^{0,a} - Y_i^{0,n}]$ (see equation (5a)). In parentheses we report the p-values for tests of the null hypotheses for which the complementarities are 0. As these estimated effect for the average in the population are a combination of the parameters estimated in the model, we use the delta method to conduct inference on different combinations of parameters which would imply the estimated effect to be zero. Different rows correspond to different periods s . Different columns correspond to different subsamples.

Table 9: Complementarity Decomposition

Periods in Job Corps	Alternative Education (a)		No Education (n)		Complementarity	
	subLATE (1)	Pairwise ATE (2)	subLATE (3)	Pairwise ATE (4)	subLATE (5)	Pairwise ATE (6)
Panel A: Earnings						
1	-0.07	-0.18	0.04	-0.10	-0.11	-0.08
2	0.02	0.09	-0.03	-0.32	0.05	0.40
3	0.07	0.04	0.09	0.13	-0.01	-0.09
Panel B: Employment						
1	-0.07	-0.05	0.02	0.05	-0.08	-0.10
2	0.00	0.01	-0.01	-0.13	0.01	0.16
3	0.07	0.06	0.13	0.16	-0.06	-0.09

Note: Column (1) reports the subLATE for $D^0 = (0, a)$ and $D^1 = (s, a)$, and Column (3) reports the subLATE for $D^0 = (0, n)$ and $D^1 = (s, n)$. Column (2) reports the ATE $E[Y_i^{s,a} - Y_i^{0,a}]$, and Column (4) reports the ATE $E[Y_i^{s,n} - Y_i^{0,n}]$. The difference between the subLATE and the pairwise ATE is the selection on gain, as shown in equation (6). Column (5) reports the differences between Columns (1) and (3), and Column (6) reports the differences between Columns (2) and (4), i.e., the complementarities net of selection on gains, as shown in equation (7). Different rows correspond to different periods s .

Table 10: Decomposition of Dynamic Average Treatment Effect on the Treated (DATTT)

Periods in Job Corps	Direct Effect	Continuation Value	Overall DATTT
Panel A: Earnings			
1	-0.09	0.10	0.02
2	0.10	0.04	0.14
3	0.06		0.06
Panel B: Employment			
1	-0.03	0.07	0.04
2	0.03	0.06	0.09
3	0.09		0.09

Note: The direct effect is the difference $E[Y_i^{\bar{s},k} - Y_i^{\bar{s}-1,k} | S_i \geq \bar{s}]$; the continuation value is the difference $E[Y_i^{S_i, T_i} - Y_i^{\bar{s},k} | S_i \geq \bar{s}]$ (see equation (9)). Different rows correspond to different periods \bar{s}

Table 11: Marginal Value of Public Fund

	Expected Benefit		Marginal Value of Public Fund	
	Baseline (1)	Targeted (2)	Baseline (3)	Targeted (4)
Overall	3,265	6,533	0.30	0.61
High Availability	2,965	6,201	0.28	0.59
Low Availability	3,607	6,893	0.33	0.64

Note: The table shows the expected benefit and marginal value of public funds (MVPF; the ratio between expected benefits and expected costs) of the program. Columns (1) and (3) refer to the policy when the treatment is assigned randomly, whereas Columns (2) and (4) are for when the 60% individuals with the highest individual benefits are targeted. In the first row, we include all the individuals in the control group. In the second row, we include only individuals living in the states where the availability of alternative education programs is above the median of the availability in the cross-states distribution. In the last row, we include individuals living in states where the availability of alternative education programs is below the median of the availability in the cross-states distribution.

Table 12: Characteristics of the Overall Control and Targeted Samples

	Overall Sample	Targeted Sample	T-test for difference (p-value)
	(1)	(2)	(3)
% Female	38.16	31.42	0.00
% Old	27.43	28.12	0.00
Prose Test (s.d.)	0.00	-0.15	0.00
Quantitative Test (s.d.)	0.00	-0.16	0.00
Education in past year (months)	4.48	4.42	0.00
Job in past year (months)	3.64	3.67	0.01
% Has HS at randomization	24.33	24.72	0.00
% Availability	53.24	50.46	0.00

Note: The “Overall Sample” includes all the individuals in the control group from our main sample. The “Targeted Sample” includes the individuals in the control group with the highest 60% individual benefits. Column (3) reports the p-value of a t-test for whether the average characteristics of the top 60% in terms of benefit are different from those of the bottom 40%. “Female” is a dummy variable taking value 1 for females. “Old” is a dummy variable taking value 1 for individuals more than 20 years of age. “Prose” and “Quantitative Test” are the scores in cognitive tests taken in the 30th month after random assignment (standardized to mean 0 and s.d. 1). “Education” and “Jobs” in the past years are the numbers of months spent in education and jobs in the year before random assignment, respectively. “HS” at randomization is a dummy variable equal to 1 if the person has a high school qualification at random assignment. “Availability” is a dummy variable taking the value of 1 for the states where the availability of alternative education programs is above the median of the availability in the cross-states distribution.

Table 13: Marginal Value of Public Fund Targeting on the Basis of Individual Benefit

Targeted Percentages with Highest Benefit	Expected Benefit	MVPF
10%	16,112	1.65
20%	12,146	1.21
30%	10,044	0.97
40%	8,595	0.82
50%	7,468	0.70
60%	6,533	0.61
70%	5,715	0.53
80%	4,987	0.47
90%	4,287	0.40
100%	3,265	0.30

Note: The table shows the expected benefit and marginal value of public funds (MVPF, the ratio between marginal benefits and marginal costs) of a marginal expansion of the program targeting different proportions of the population on the basis of the predicted individual benefit.

Figure 1: The Sequential Multistage Decision Framework with Multiple Alternatives

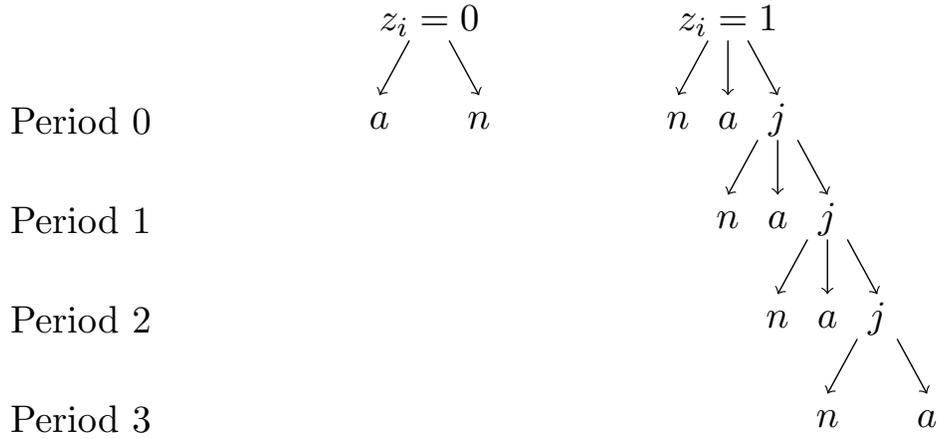
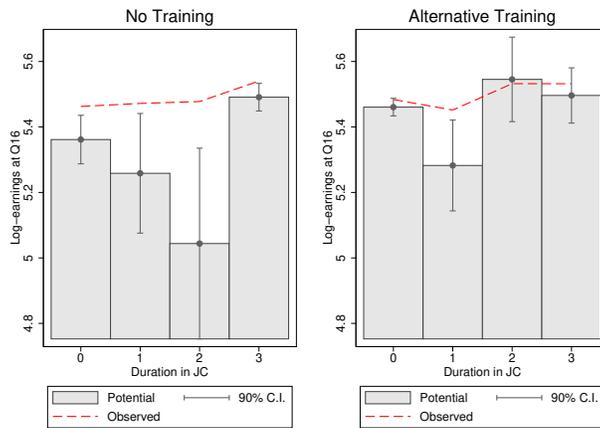
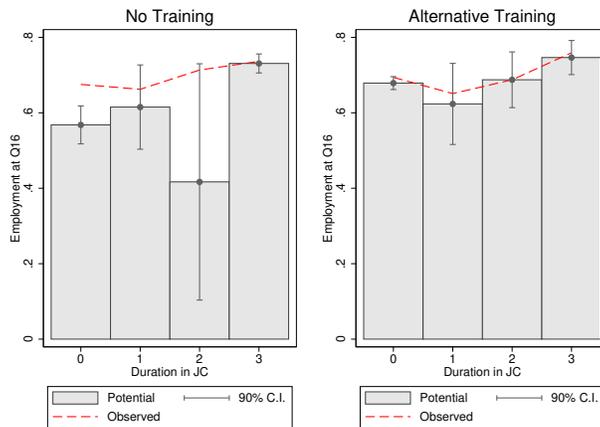


Figure 2: Empirical Estimates and 90% Confidence Intervals of the Average Potential Outcomes

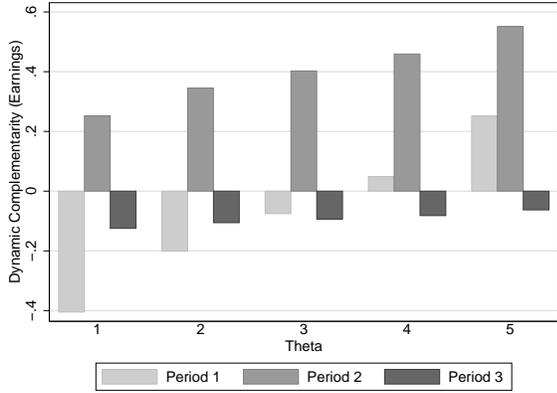


(a) Earnings

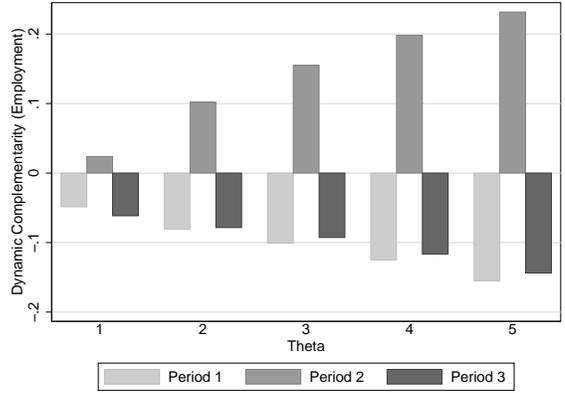


(b) Employment

Figure 3: Heterogeneity in Dynamic Interaction between Job Corps and a



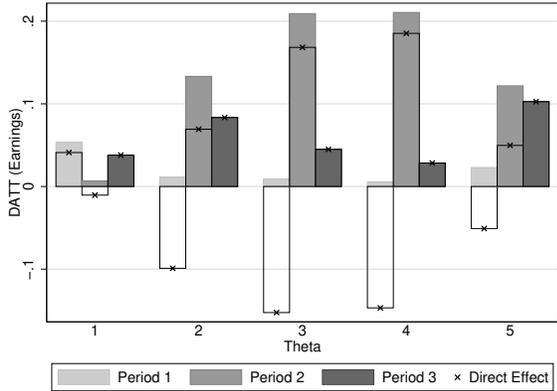
(a) Earnings



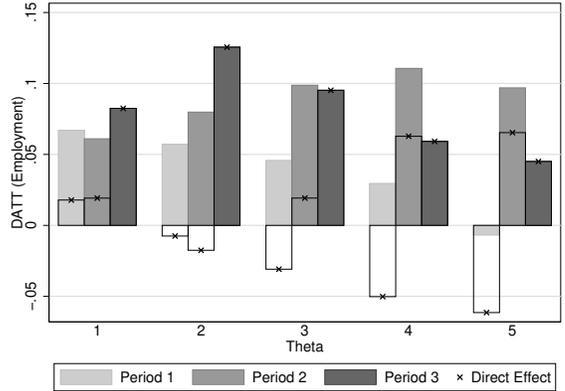
(b) Employment

Notes: Each bar shows the difference between $E[Y_i^{s,a} - Y_i^{s,n}]$ and $E[Y_i^{0,a} - Y_i^{0,n}]$ (see equation (5a)) for different bins of the distribution of θ . We split the distribution of θ in 5 bins where 1 is the lowest and 5 is the highest.

Figure 4: Heterogeneity in the DATT



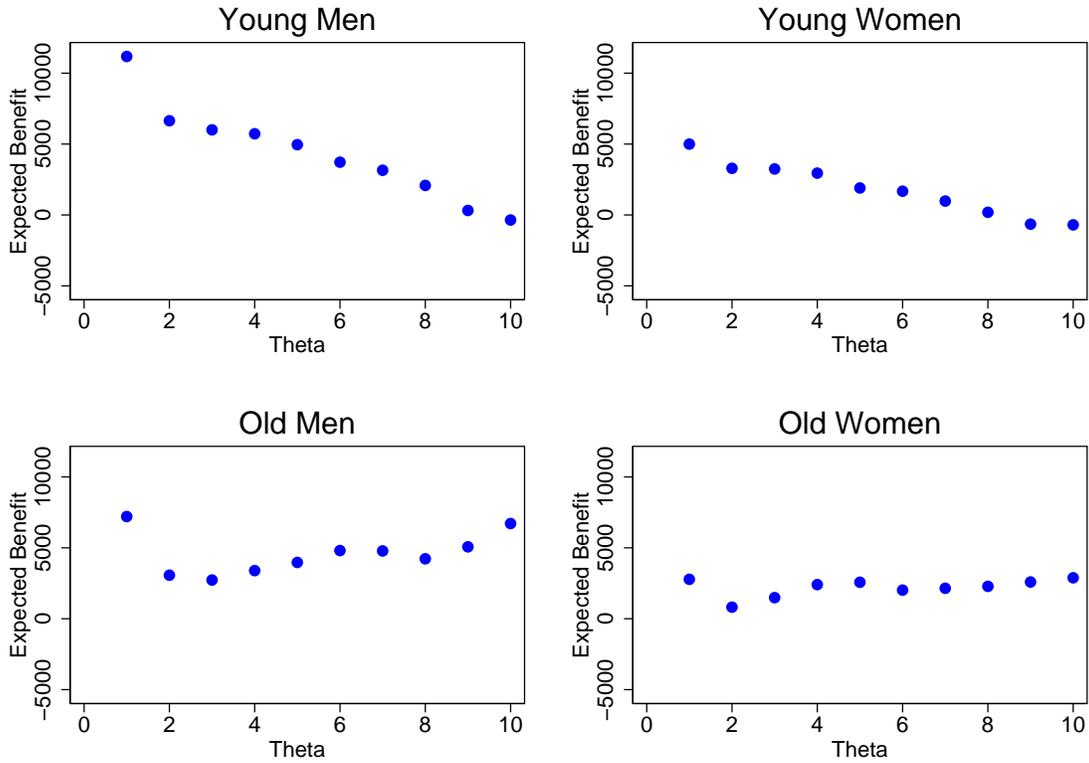
(a) Earnings



(b) Employment

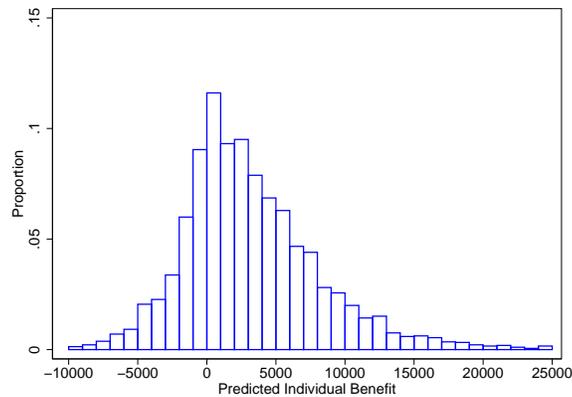
Notes: Each bar shows the DATT (i.e. the difference $E[Y_i^{S_i, K_i} - Y_i^{\bar{s}-1, k} | S_i \geq \bar{s}]$, where S_i is the optimal choice of the individual in terms of s , K_i is the optimal choice between a and n at period S_i and k is the optimal choice at periods $\bar{s} - 1$) for different periods \bar{s} and for different bins of the distribution of θ . We split the distribution of θ in five bins where 1 is the lowest and 5 is the highest. The figure also shows the decomposition in terms of the direct effect and the continuation value, where the direct effect is the difference $E[Y_i^{\bar{s}, k} - Y_i^{\bar{s}-1, k} | S_i \geq \bar{s}]$, and the continuation value is the difference $E[Y_i^{S_i, K_i} - Y_i^{\bar{s}, k} | S_i \geq \bar{s}]$ (see equation (9)). Period 3 has no continuation value.

Figure 5: Predicted Expected Benefit for Different Values of θ_i



Notes: Each dot shows the average of the predicted individual benefits among individuals with the same value of observed characteristics (age and gender) and in the same bin of the distribution of θ , where we divided the distribution of θ in 10 bins, 1 has the lowest value and 10 the highest.

Figure 6: Histogram of Individual Expected Benefits in the Control Group



Notes: In this graph the individual benefits are computed based on the conditional posterior distribution of θ (see equation (23))

ONLINE APPENDIX

A Characteristics of Alternative Education

In Table A1, we report the proportions of individuals enrolling in the most relevant categories of education available to individuals when not enrolled in Job Corps: “High School” refers to grade 9–12 in the education system; “Vocational” refers to vocational, technical, or trade schools; “GED” refers to programs preparing for the GED test; “2-year College” refers to community or junior colleges, while “4-year College” refers to college programs lasting 4 years. These different programs are offered by different types of institutions; some are public (as it is the case for most community colleges), whereas others are private. In the different interviews, respondents are asked the source of funding for the training they obtained. Despite the fact that few individuals answered the question (among those who claimed having obtained some training only 20% answered the question about funding in the 12th and 30th month interviews and about 30% in the 48th month interview), in general, we find that between 25 and 30% claim that the training was provided by some government program and a similar proportion claim it was provided by the employer. However, even among the individuals who claim that the training was offered by the employer, very few reported having obtained on-the-job training: among the 4,580 individuals who claim to have obtained some education between random assignment and the 12-month interview, only 4 claimed having obtained “on-the-job training”. Similarly, among the 3,153 individuals who obtained education between the 30th and 48th month, three claim having obtained “on-the-job training” (but two of them also obtained publicly provided education—GED and college—before finding a job). Hence, our definition of education mostly includes education offered by some established institution.

Most of the education obtained in the first few months after random assignment is full-time education. On average, students assigned to the control group reported spending 5.3 hours per day in the first education they obtained between the random assignment and the 12-month interview. Similarly, the individuals who enrolled in Job Corps and then obtain additional education before the 12-month interview (hence, mostly those spending between one and two periods in Job Corps, according to our definition) report spending, on average, 5.4 hours per day in education. The average hours spent in education decreased in the most recent periods both for the treated and the control groups, suggesting that more individuals enrolled in part-time education over time.

Note that the alternative education program selected by individuals after Job Corps is different conditional on the duration in the program, with a decrease in the proportions enrolling in high school or GED and an increase in 2-year college enrollment for individuals who stay in Job Corps for more than 6 months. This may be because these individuals who stay longer in Job Corps are more likely to obtain a GED within the program that then substitutes for GED or high school investment outside the program.

As mentioned in Section 2.2, the training obtained under the supervision of Job Corps differentiates itself from these programs for three main reasons: first, it is more comprehensive, offering different

services (such as education and job-seeking assistance) all in the same place; second it is more intensive (partly due to its residential feature) and, hence, more expensive; and, third, it is more homogeneous within the country because it is administered by US DOL and not by local agencies.

B LATE Decomposition

The LATE is estimated using the Wald estimator:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{P(S_i > 0|Z_i = 1) - P(S_i > 0|Z_i = 0)} \quad (\text{B.1})$$

The first term in the numerator is the average outcome Y_i for all the individuals who have received the treatment. Given that we group them in eight groups on the basis of the number of periods spent in Job Corps and of their choice when not enrolled in Job Corps (see Section 2.1), this term can be rewritten as the weighted average of the outcome for eight different subgroups, where the weights are given by the proportion of individuals selecting in each subgroup (note that we suppressed the subscript i for convenience):

$$\begin{aligned} E[Y|Z = 1] = & P(D^1 = (0, a))E[Y^{0,a}|D^1 = (0, a)] + P(D^1 = (0, n))E[Y^{0,n}|D^1 = (0, n)] \\ & + P(D^1 = (1, a))E[Y^{1,a}|D^1 = (1, a)] + P(D^1 = (1, n))E[Y^{1,n}|D^1 = (1, n)] \\ & + P(D^1 = (2, a))E[Y^{2,a}|D^1 = (2, a)] + P(D^1 = (2, n))E[Y^{2,n}|D^1 = (2, n)] \\ & + P(D^1 = (3, a))E[Y^{3,a}|D^1 = (3, a)] + P(D^1 = (3, n))E[Y^{3,n}|D^1 = (3, n)]. \end{aligned} \quad (\text{B.2})$$

$E[Y^{3,a}|D^1 = (3, a)]$, for example, is the average outcome for those individuals who select $s = 3$ and $k = a$ when assigned to the treatment group. We can rewrite this in compact notation as

$$E[Y|Z = 1] = \sum_{s=0}^3 P(D^1 = (s, a))E[Y^{s,a}|D^1 = (s, a)] + \sum_{s=0}^3 P(D^1 = (s, n))E[Y^{s,n}|D^1 = (s, n)] \quad (\text{B.3})$$

Given Assumptions 1–4 described in Section 3.1 and calling k' the treatment that is the opposite of k (hence, if $k = a$, then $k' = n$), the last six of these eight averages can be decomposed in:⁴¹

$$\begin{aligned} E[Y^{s,k}|D^1 = (s, k)] = & \frac{P(D^0 = (0, k), D^1 = (s, k))}{P(D^1 = (s, k))} E[Y^{s,k}|D^0 = (0, k), D^1 = (s, k)] \\ & + \frac{P(D^0 = (0, k'), D^1 = (s, k))}{P(D^1 = (s, k))} E[Y^{s,k}|D^0 = (0, k'), D^1 = (s, k)]. \end{aligned} \quad (\text{B.4})$$

Hence, the first term in the numerator of Equation B.1 can be decomposed as:

⁴¹Because, by Assumption 3, the treatment does not change the choice when $s = 0$, the outcome for those who select $s = 0$ is the same irrespective of the value of Z

$$\begin{aligned}
E[Y|Z = 1] &= \sum_{s=0}^3 P(D^0 = (0, a), D^1 = (s, a))E[Y^{s,a}|D^0 = (0, a), D^1 = (s, a)] \\
&+ \sum_{s=1}^3 P(D^0 = (0, n), D^1 = (s, a))E[Y^{s,a}|D^0 = (0, n), D^1 = (s, a)] \\
&+ \sum_{s=1}^3 P(D^0 = (0, a), D^1 = (s, n))E[Y^{s,n}|D^0 = (0, a), D^1 = (s, n)] \\
&+ \sum_{s=0}^3 P(D^0 = (0, n), D^1 = (s, n))E[Y^{s,n}|D^0 = (0, n), D^1 = (s, n)].
\end{aligned} \tag{B.5}$$

Similarly, the second term in the numerator of Equation B.1 can be decomposed as the weighted average of the outcome of two subgroups (given the absence of Always Takers):

$$E[Y|Z = 0] = P(D^0 = (0, a))E[Y^{0,a}|D^0 = (0, a)] + P(D^0 = (0, n))E[Y^{0,n}|D^0 = (0, n)], \tag{B.6}$$

and each of these weighted averages can be decomposed as:

$$\begin{aligned}
E[Y^{0,k}|D^0 = (0, k)] &= \frac{P(D^1 = (0, k))}{P(D^0 = (0, k))}E[Y|D^0(0, k), D^1 = (0, k)] \\
&+ \sum_{s=1}^3 \frac{P(D^1 = (s, k))}{P(D^0 = (0, k))}E[Y^{0,k}|D^0(0, k), D^1 = (s, k)] \\
&+ \sum_{s=1}^3 \frac{P(D^1 = (s, k'))}{P(D^0 = (0, k))}E[Y^{0,k}|D^0(0, k), D^1 = (s, k')].
\end{aligned} \tag{B.7}$$

The second term in the numerator of Equation B.1 can thus be rewritten as:

$$\begin{aligned}
E[Y|Z = 0] &= \sum_{s=0}^3 P(D^0 = (0, a), D^1 = (s, a))E[Y^{0,a}|D^0 = (0, a), D^1 = (s, a)] \\
&+ \sum_{s=1}^3 P(D^0 = (0, n), D^1 = (s, a))E[Y^{0,n}|D^0 = (0, n), D^1 = (s, a)] \\
&+ \sum_{s=1}^3 P(D^0 = (0, a), D^1 = (s, n))E[Y^{0,a}|D^0 = (0, a), D^1 = (s, n)] \\
&+ \sum_{s=0}^3 P(D^0 = (0, n), D^1 = (s, n))E[Y^{0,n}|D^0 = (0, n), D^1 = (s, n)].
\end{aligned} \tag{B.8}$$

Combining Equation B.5 and Equation B.8, we can rewrite the numerator of the Wald estimator as:

$$\begin{aligned}
E[Y|Z = 1] - E[Y|Z = 0] = & \\
& \sum_{s=1}^3 P(D^0 = (0, a), D^1 = (s, a))(E[Y^{s,a} - Y^{0,a}|D^0 = (0, a), D^1 = (s, a)]) \\
& + \sum_{s=1}^3 P(D^0 = (0, n), D^1 = (s, a))(E[Y^{s,a} - Y^{0,n}|D^0 = (0, n), D^1 = (s, a)]) \\
& + \sum_{s=1}^3 P(D^0 = (0, a), D^1 = (s, n))(E[Y^{s,n} - Y^{0,a}|D^0 = (0, a), D^1 = (s, n)]) \tag{B.9} \\
& + \sum_{s=1}^3 P(D^0 = (0, n), D^1 = (s, n))(E[Y^{s,n} - Y^{0,n}|D^0 = (0, n), D^1 = (s, n)]) \\
& + P(D^0 = (0, a), D^1 = (0, a))(E[Y^{0,a} - Y^{0,a}|D^0 = (0, a), D^1 = (0, a)]) \\
& + P(D^0 = (0, n), D^1 = (0, n))(E[Y^{0,n} - Y^{0,n}|D^0 = (0, n), D^1 = (0, n)]),
\end{aligned}$$

where the last two terms are zero. This equation shows that, in the case of three periods and two alternative treatments, the numerator of the Wald estimator can be decomposed into the average effect for 12 different subpopulations.

The denominator in the Wald estimator is the difference between the probability of enrolling in at least one period of Job Corps when assigned to the treatment group, and the same probability when assigned to the control group. Given Assumption 4 (no Always Takers), the second term of the denominator is zero. This is the proportion of compliers, which is π_C , described in Section 3.3.

B.1 Non-identifiability of the Proportions

Given Assumptions 3 and 4, the following equations hold, where the population proportions are equalized to the proportions of the 14 different subgroups (12 subgroups of compliers and two Never Takers), which follow from Assumptions 3 and 4:

$$\begin{aligned}
P(D_i = (0, k)|Z_i = 1) &= \pi_{k,0k} \quad \forall k \in (a, n) \\
P(D_i = (s, k)|Z_i = 1) &= \pi_{k',sk} + \pi_{k,sk} \quad \forall s \in (1, 2, 3), \forall k \in (a, n), k \neq k' \\
P(D_i = (0, k)|Z_i = 0) &= \pi_{k,0k} + \pi_{k,1k'} + \pi_{k,1k} + \pi_{k,2k'} + \pi_{k,2k} + \pi_{k,3k'} + \pi_{k,3k} \quad \forall k \in (a, n), k \neq k'
\end{aligned}$$

which is a system of 10 equations and 14 unknowns, implying that these proportions are not identifiable.

Note that this case is different from the static case described in Kline and Walters (2016), where the proportions are nonparametrically identifiable. Consider the case where we do not account for the dynamic setting; hence, we are only interested in the choice individuals make at time 0, where they can select n , a , or j if assigned to the treatment group, and a or n when assigned to the control group. We thus have four proportions: two subgroups of compliers ($\pi_{n,j}$ and $\pi_{a,j}$) and two subgroups of Never

Takers ($\pi_{a,a}$ and $\pi_{n,n}$). From the data, the following equations hold:

$$\begin{aligned}
P(D_i = (a)|Z_i = 1) &= \pi_{a,a} \\
P(D_i = (n)|Z_i = 1) &= \pi_{n,n} \\
P(D_i = (jc)|Z_i = 1) &= \pi_{j,a} + \pi_{j,n} \\
P(D_i = (a)|Z_i = 0) &= \pi_{a,a} + \pi_{j,a} \\
P(D_i = (n)|Z_i = 0) &= \pi_{n,n} + \pi_{j,n}
\end{aligned}$$

being five equations with four unknowns, which has a solution.

B.2 Non-identifiability of the subLATEs

In terms of outcome, given Assumptions 3 and 4, the following equations hold, linking the population moments to the subgroup potential outcomes:

$$E[Y|Z_i = 1, D_i = (0, k)] = E[Y_i^1|D_i^0 = (0, k), D_i^1 = (0, k)] = E[Y_i^0|D_i^0 = (0, k), D_i^1 = (0, k)] \quad \forall k \in (a, n)$$

$$\begin{aligned}
E[Y|Z_i = 1, D_i = (1, k)] &= \frac{\pi_{k,sk}}{\pi_{k,sk} + \pi_{k',sk}} E[Y_i^1|D_i^0 = (0, k), D_i^1 = (s, k)] \\
&+ \frac{\pi_{k,sk}}{\pi_{k,sk} + \pi_{k',sk}} E[Y_i^1|D_i^0 = (0, k'), D_i^1 = (s, k)] \quad \forall s \in (1, 2, 3), \forall k \in (a, n), k \neq k'
\end{aligned}$$

$$\begin{aligned}
E[Y|Z_i = 0, D_i = (0, k)] &= \left(\sum_{s=1}^3 \frac{\pi_{k,sk}}{\sum_{k'' \in (a,n)} \sum_{s''=0}^3 \pi_{k,s''k''}} E[Y_i^0|D_i^0 = (0, k), D_i^1 = (s, k)] \right. \\
&+ \left. \frac{\pi_{k,sk'}}{\sum_{k'' \in (a,n)} \sum_{s''=0}^3 \pi_{k,s''k''}} E[Y_i^0|D_i^0 = (0, a), D_i^1 = (s, n)] \right) \\
&+ \frac{\pi_{k,0k}}{\sum_{k'' \in (a,n)} \sum_{s''=0}^3 \pi_{k,s''k''}} E[Y_i^0|D_i^0 = (0, k), D_i^1 = (0, k)] \quad \forall k \in (a, n)
\end{aligned}$$

where the first equations identify the average potential outcomes for the two types of Never Takers (the same no matter which treatment group they are assigned to), but the system with 10 equations and 24 unknowns does not allow the effect for the different subgroups to be identified.

In the static case with multiple alternatives (where individuals can select $t \in (a, n, j)$ if $z_i = 1$ and $t \in (a, n)$ if $z_i = 0$), the subLATEs are also not nonparametrically identified, as mentioned in Kline and Walters (2016). In this case, to estimate the subLATEs, we should be able to identify six average potential outcomes: four for the compliers ($E[Y_i^1|D_i^1 = j, D_i^0 = a]$, $E[Y_i^1|D_i^1 = j, D_i^0 = n]$, $E[Y_i^0|D_i^1 = j, D_i^0 = a]$, $E[Y_i^0|D_i^1 = j, D_i^0 = n]$) and two for the Never Takers ($E[Y_i^1|D_i^1 = a, D_i^0 = a] = E[Y_i^0|D_i^1 = a, D_i^0 = a]$ and $E[Y_i^1|D_i^1 = n, D_i^0 = n] = E[Y_i^0|D_i^1 = n, D_i^0 = n]$). However, because we only observe five sample moments ($E[Y_i|Z_i = 1, D_i = j]$, $E[Y_i|Z_i = 1, D_i = n]$, $E[Y_i|Z_i = 1, D_i = a]$,

$E[Y_i|Z_i = 0, D_i = n], E[Y_i|Z_i = 0, D_i = a]$) the parameters of the model are not identified.

As is known, the static case with only one alternative (as the standard LATE framework) is instead identified, because it requires the estimation of three average potential outcomes when Always Takers are excluded (call the only alternative n , the three average potential outcomes needed to estimate the LATE are $E[Y_i^1|D_i^0 = n, D_i^1 = n] = E[Y_i^0|D_i^0 = n, D_i^1 = n]$, $E[Y_i^1|D_i^0 = n, D_i^1 = j]$, $E[Y_i^0|D_i^0 = n, D_i^1 = j]$) and we observe three sample moments ($E[Y_i|Z_i = 1, D_i = j]$, $E[Y_i|Z_i = 1, D_i = n]$, $E[Y_i|Z_i = 0, D_i = n]$).

C Selection into Employment

Earnings are observed only conditional on employment; hence, when estimating the causal effect of the treatment on earnings using the Wald estimator, it is possible to incur bias due to selection into employment. In particular, the Wald estimator of the causal effect on log-earnings can be written as:

$$\frac{E[Y_i|H_i = 1, Z_i = 1] - E[Y_i|H_i = 1, Z_i = 0]}{\sum_{s=1}^3 \sum_{k \in (a,n)} P(D_i(s, k) = 1|H_i = 1, Z_i = 1) - \sum_{s=1}^3 \sum_{k \in (a,n)} P(D_i(s, k) = 0|H_i = 1, Z_i = 0)} \quad (\text{C.1})$$

Given Assumptions 1–4 we can rewrite this as:

$$\begin{aligned} & \frac{1}{P(D_i^1 \neq (0, k)|H_i^1 = 1)} (E[Y_i^1|H_i^1 = 1] - E[Y_i^0|H_i^0 = 1]) = \\ & \frac{1}{P(D_i^1 \neq (0, k)|H_i^1 = 1)} (E[Y_i^1 - Y_i^0|H_i^1 = 1] + E[Y_i^0|H_i^1 = 1] - E[Y_i^0|H_i^0 = 1]) \end{aligned} \quad (\text{C.2})$$

where $P(D_i^1 \neq (0, k)|H_i^1 = 1)$ is the proportion of individuals among those employed after being assigned to the treatment group whose potential choice when $Z_i = 1$ is to get at least one period of Job Corps.⁴² The first term in the numerator is the causal effect of the treatment on the individuals who are employed when assigned to the treatment group, whereas the difference between the last two terms is the selection bias due to selection into employment (i.e. the fact that some of the individuals who are employed when assigned to the treatment group are not employed when assigned to the control group and vice versa). For some individuals, the earnings are not observed in the counterfactual scenario; hence, if they are selected, the difference in the composition of employed individuals biases the causal estimates.

C.1 subLATEs for the LATE Conditional on Employment

As an alternative to the decomposition of the LATE for the potential earnings, we can decompose the causal part of the LATE described above, i.e., the LATE conditional on employment when $Z_i = 1$. The decomposition follows the same steps as the decomposition of the overall LATE, shown in Appendix B. The weights for this case are the proportion of compliers conditional on employment when assigned to

⁴²Note that here we consider the LATE for log-earnings, which implies that the individuals who report having zero earnings are excluded from the estimation.

the treatment group. As a consequence, if, for example, the proportion of employed individuals who get one period in Job Corps and then no education is low and the proportion of employed individuals who enrol for three periods in Job Corps and then does not get education is high, the first group is weighted less in the computation of the causal impact for employed individuals even if the unconditional proportion of individuals in the two groups is the same.

In Table A7, we report the conditional proportion, together with the results of the decomposition of the causal part of the LATE. As expected, these results show that the subgroups for which the treatment has a positive effect on employment contribute more to the overall LATE as compared to the unconditional case. For example, in Table 6 we reported that for the subgroup $D_i^0 = (0, n), D_i^1 = (3, n)$ the effect of the treatment on employment is 13 percentage point. This group is 7.9% of the population, but conditional on employment its share is 8.6%. Symmetrically, the group selecting $D_i^0 = (0, a), D_i^1 = (1, a)$, for which the effect of the treatment on employment is negative, reduces its share from 10.4 in the overall population to 9.6 when conditioning on employment. Hence there are two reasons why the overall conditional LATE is larger than the overall unconditional LATE (the first rows in Table A7 and Table 6). The first is that the the potential earnings of the employed individuals are higher. The second is that among the employed individuals there is a higher proportion of individuals selecting the combinations of education and duration which have higher returns (following the previous example, the returns to $D_i^0 = (0, n), D_i^1 = (3, n)$ is 8%, while the returns to $D_i^0 = (0, a), D_i^1 = (1, a)$ is negative 6%).

D Identification of the Sequential Choice Model

Define the unobservable component of equations (10)–(12) in every period s as η_{is}^k , where $\eta_{is}^a \equiv \beta_\theta^{s,a} \theta_i + u_{is}^a$ and $\eta_{is}^j \equiv \beta_\theta^{s,j} \theta_i + u_{is}^j$. Similarly, define the unobserved component of the outcome equation (14) as ω_{is}^k , hence $\omega_{is}^k \equiv \gamma_\theta^{s,k} \theta_i + \varepsilon_i^{s,k}$ (for simplicity, here we assume away any employment selection). In period 0, individuals face three alternative choices $\{n, a, j\}$. Suppose we have one measurement denoted by C_i which is free from selection. Define the unobserved component of the measurement by $\kappa_i \equiv \gamma_{c,\theta} \theta_i + \varepsilon_{c,i}$. Because the measurements C_i are independent from the choices of the individuals, we can identify the joint distribution of the κ_i and $(\eta_{i0}^a, \eta_{i0}^j, \omega_{i0}^a, \omega_{i0}^n)$ and, given the exclusion from program randomization, we can also identify the joint distribution of η_{i0}^a and ω_{i0}^a (e.g., as shown in Theorem B.1 in Heckman and Navarro (2007)). Therefore, given the conditional independence assumption and the factor structure from Section 4, the following equations hold:

$$Cov(\eta_{i0}^a, \omega_{i0}^a) = \gamma_\theta^{0,a} \sigma_\theta \tag{D.1}$$

$$Cov(\eta_{i0}^j, \kappa_i) = \beta_\theta^{0,j} \gamma_{c,\theta} \sigma_\theta \tag{D.2}$$

$$Cov(\eta_{i0}^a, \kappa_i) = \gamma_{c,\theta} \sigma_\theta, \tag{D.3}$$

$$Cov(\omega_{i0}^n, \kappa_i) = \gamma_\theta^{0,n} \gamma_{c,\theta} \sigma_\theta \tag{D.4}$$

$$Cov(\omega_{i0}^a, \kappa_i) = \gamma_\theta^{0,a} \gamma_{c,\theta} \sigma_\theta, \tag{D.5}$$

where σ_θ is the variance of the unobserved factor θ and we have normalized $\beta_\theta^{0,a}$ to 1 (see Section 4). The ratio between equation (D.2) and (D.3) identifies $\beta_\theta^{0,j}$. Then, combining equation (D.2) with equation (D.4) identifies $\gamma_\theta^{0,n}$ and combining equation (D.2) with equation (D.5) identifies $\gamma_\theta^{0,a}$. Then, σ_θ , the variance of the unobserved factor, is identified using equation (D.1). Finally, $\gamma_{c,\theta}$ is identified from equation (D.3).

Next, we can use the choice probabilities provided by the data in the period 0 to identify the remaining parameters in the choice equations: $P(D = a|Z = 1), P(D = a|Z = 0), P(D = jc|Z = 1)$. In this case, there are three remaining parameters: $\beta^{0,a}, \beta^{0,j}, \rho$, where $\beta^{0,a}$ is the intercept of the choice equation for a , $\beta^{0,j}$ is the intercept of the choice equation for j , and ρ is the correlation between utility shocks. These parameters are just-identified and perfectly fit the three independent conditional choice probabilities.⁴³ Given that the choice equations are identified and the correlation between the outcome and the choice is identified, we can construct selection correction terms (which are generalizations of the standard inverse Mills ratio correction term used in the literature) and identify the parameters of average potential outcomes.

Moving beyond period 0 and repeating the above steps, we can rely on the covariance restrictions between measurement and choices and between measurement and outcomes (similar to equations (D.2) to (D.5)) to identify the factor loadings in the selection and outcome equations. The remaining parameters in the choice equations (containing $\beta^{s,a}, \beta^{s,j}$) are just identified from the two choice probabilities in period s (of choosing a and j).

E Estimation Details: The Likelihood Function

Following Hansen, Heckman, and Mullen (2004) and Heckman, Humphries, and Veramendi (2016), the likelihood function is constructed in two steps. In the first step, the distribution of the latent factors is estimated using only data on educational choices and measurements. Therefore, we estimate factor distributions by jointly estimating the choice and measurement equations in the first stage, without using any information from outcomes. The parameters of the outcome equations are estimated in the second step.

Using the notations defined in Section 4.3, the conditional likelihood for individual i is:

$$L_i = L_i^{(1)}(D_i, Y_i, H_i, M_i | X_i)Z_i + L_i^{(2)}(D_i, Y_i, H_i, M_i | X_i)(1 - Z_i) \quad (\text{E.1})$$

where

$$\begin{aligned} L_i^{(m)}(D_i, Y_i, H_i, M_i | X_i) &= \int L^{(m)}(D_i, Y_i, H_i, M_i | X_i, \theta_i) f(\theta_i) d\theta_i \\ &= \int f^{(m)}(Y_i, H_i | D_i, X_i, \theta_i) f^{(m)}(D_i, M_i | X_i, \theta_i) f(\theta_i) d\theta_i, \quad m = \{1, 2\} \end{aligned} \quad (\text{E.2})$$

⁴³The discussion here omits the role of any covariate. Suppose we include a covariate X that is binary. Then we have six choice probabilities (three probabilities as specified above, each of which conditional on $X=0$ and $X=1$) and five unknown parameters (assuming ρ does not vary by X). The model would be over-identified.

In the first step, the individual-specific conditional likelihood function is:

$$\begin{aligned} & \int f^{(m)}(D_i, M_i | X_i, \theta_i) f(\theta_i) d\theta_i \\ & = \int f^{(m)}(M_i | D_i, X_i, \theta_i) f^{(m)}(D_i | X_i, \theta_i) f(\theta_i) d\theta_i, \quad m = \{1, 2\} \end{aligned} \quad (\text{E.3})$$

where $f^{(m)}(M_i | D_i, X_i, \theta_i) = \prod_{q=1}^Q f^{(m)}(M_{iq} | D_i, X_i, \theta_i)^{Q_i^q}$ where Q_i^q is an indicator function that is equal to 1 if the q th measurement is available for the individual i . We assume that $f(\theta_i)$ follows a normal distribution. The first step yields estimates of all the parameters in the measurement and choice equations. It also estimates the parameters that characterize the distribution of unobserved factor, γ_θ . The vector γ_θ contains: μ_θ (mean of the factor) and σ_θ^2 (variance of the factor). Because of the normalization assumption, μ_θ is fixed at zero.

In the second step of the procedure, information from the outcomes is added together with the estimated distribution of unobserved factor to form the complete likelihood function for individual i :

$$L_i^{(m)}(D_i, Y_i, H_i, M_i | X_i) = \int f^{(m)}(Y_i, H_i | D_i, X_i, \theta_i) \hat{f}^{(m)}(D_i, M_i | X_i, \theta_i) \hat{f}(\theta_i) d\theta_i, \quad m = \{1, 2\} \quad (\text{E.4})$$

where $\hat{f}^{(m)}(D_i, M_i | X_i, \theta_i)$ and $\hat{f}(\theta_i)$ are the density functions from the first-step estimation (using the estimated vector γ_θ from the first-step). Because outcomes are independent from the measurements conditional on X and θ , we can obtain consistent estimates of the parameters for the outcomes by maximizing the likelihood function.

F Costs and Benefits

F.1 Definitions of Benefits and Costs

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. To compute the lifetime gain of Job Corps, we rely on the information on the weekly earnings in the last quarter of the fourth year after random assignment. In our main analysis we estimated that the increase in potential wage for the treatment group is about 2%. To compute the lifetime earnings gain, we follow McConnell and Glazerman (2001) and assume that the earnings of all individuals increase with worker age, and that the dollar value of the increase in income 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 exponential of the potential log-weekly earnings in the last quarter individuals are observed in the survey.⁴⁴ 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. The average effect on lifetime earnings for those who enrolled in Job Corps for at least one period is estimated to be 3,300\$. We then consider a tax rate of 23.5%. This includes the federal income tax rate at the end of the 1990s (15%), the payroll tax (7.65%) and the average of the state taxes (0.8%).

In terms of costs, although we show the formulas allowing ϕ_a to differ depending on the period spent in Job Corps before obtaining a , for simplicity, in our calculations, we assume that the costs per student of alternative education program is the same no matter whether the individuals enrolled in Job Corps before starting education (i.e., $\phi_a^0 = \phi_a^s = \phi_a$). McConnell and Glazerman (2001) compute that, overall, the average cost of enrolling in some other education program for a student who, instead, enrolls in Job Corps, would be around 2,100\$. We keep this as an indication also of the costs of attending additional education after Job Corps. This assumption may not hold, because the educational choices of individuals after they have enrolled in Job Corps are quite different as compared with the educational choices for the control group or for the Never Takers. Most of the difference lies in the smaller proportion of individuals enrolling in high school or obtaining a GED and the higher proportions of individuals enrolling in a 2-year College for the treated who attend Job Corps. McConnell and Glazerman (2001) estimate a cost of 6.41\$ per hour for high school and 181\$ per week for 2-year colleges, suggesting that the cost of these two should not differ much. However, this assumption, while relevant for our calculation, can be easily removed, and we can assign different costs of alternative education programs depending on the duration in Job Corps. The cost of Job Corps for the average student is about 16,489\$. The average student spends 7.5 months in Job Corps, so Job Corps costs about 2,200\$ per month. We thus assign a Job Corps cost of 4,400\$ for those who dropout during period $s = 1$, 11,000\$ for those who stay $s = 2$ periods, and 26,400\$ for individuals with $s = 3$ (this corresponds to the cost of 1 year of Job Corps, about the average duration for those who have $s = 3$).

F.2 Derivation of Marginal Benefits and Marginal Costs

Following the discussion in Kline and Walters (2016), we define the overall net expected lifetime earnings for all the individuals potentially treated by a program as $\bar{Y} = (1 - \tau)E[\bar{Y}_i]$ (where τ is the tax rate). Consider the case where a fraction δ of the population is treated. \bar{Y} can be decomposed as:

$$\bar{Y} = (1 - \tau)E[\bar{Y}_i^{D^1} | Z_i = 1]\delta + E[\bar{Y}_i^{D^0} | Z_i = 0](1 - \delta) \quad (\text{F.1})$$

⁴⁴Note that Schochet, Burghardt, and McConnell (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.

If treatment is randomly allocated, then we can rewrite this as

$$\bar{Y} = (1 - \tau)E[\bar{Y}_i^{D_i^1}]\delta + E[\bar{Y}_i^{D_i^0}](1 - \delta) \quad (\text{F.2})$$

A marginal expansion in the probability of being randomly assigned to the treatment group can then be defined as:

$$\begin{aligned} \frac{\partial \bar{Y}}{\partial \delta} &= (1 - \tau)E[\bar{Y}_i^{D_i^1}] - E[\bar{Y}_i^{D_i^0}] \\ &= (1 - \tau)E[\bar{Y}_i^{D_i^1} - \bar{Y}_i^{D_i^0}] \end{aligned} \quad (\text{F.3})$$

which corresponds to the expected benefit in the population among which the treatment is randomly allocated.

The costs of training for the population depend on the proportions selecting into different training programs. To get a measure of “net” costs, we discount the taxes paid by the individual from the cost function. Let ϕ_{jc}^s indicate the administrative cost of providing s periods in Job Corps to a participant, ϕ_a^0 the cost of providing alternative education to students who have not enrolled in Job Corps before, and ϕ_a^s the cost of providing alternative education programs after the students enrolled in Job Corps for s periods. The net cost of the training is C , defined as:

$$C = \sum_{s=1}^3 \sum_{k \in (a,n)} \phi_{jc}^s P(D_i = (s, k)) + \sum_{s=0}^3 \phi_a^s P(D_i = (s, a)) - \tau E[\bar{Y}_i]. \quad (\text{F.4})$$

where $P(D_i = (s, k))$ is the proportion of individuals for whom the realized potential choice is s periods in Job Corps and k when not enrolled in Job Corps. To simplify the calculation, we assume that the costs of alternative education programs do not change with the duration in Job Corps; hence, $\phi_a^0 = \phi_a^1 = \dots = \phi_a$ (see Appendix F.1). Each proportion $P(D_i = s, k)$ can be decomposed as:

$$P(D_i = s, k) = P(D_i^1 = s, k)\delta + P(D_i^0 = s, k)(1 - \delta). \quad (\text{F.5})$$

Given the assumption of no Always Takers, then $P(D_i^0 = s, k) = 0$ for $s > 0$. Therefore, the cost of a marginal increase in the fraction δ of randomly treated individuals is:

$$\begin{aligned} \frac{\partial C}{\partial \delta} &= \sum_{s=1}^3 \sum_{k \in (a,n)} \phi_{jc}^s P(D_i^1 = (s, k)) + \sum_{s=0}^3 \phi_a^s P(D_i^1(s, a)) \\ &\quad - \phi_a^0 P(D_i^0(0, a)) - \tau E[\bar{Y}_i^{D_i^1} - \bar{Y}_i^{D_i^0}]. \end{aligned} \quad (\text{F.6})$$

which is the expected cost in the population that is randomly allocated to the treatment.

The expected benefit and the expected cost for treated individuals thus correspond to the marginal benefit and the marginal cost of a random expansion of the program to the population of interest. Hence, their ratio corresponds to the Marginal value of Public Fund:

$$MVPF = \frac{\partial \bar{Y} / \partial \delta}{\partial C / \partial \delta} \quad (\text{F.7})$$

The MVPF is a measure of the economic efficiency of the policy and measures the value of an extra dollar spent on training for the population that is targeted by the policy, net of fiscal externality (Kline and Walters, 2016; Hendren, 2016).

Appendix Tables and Figures

Table A1: Proportions Enrolling in Different Types of Education, for the Treatment and Control Groups

Type of Education	Treated Group				Control Group
	$s = 0$	$s = 1$	$s = 2$	$s = 3$	
High School	0.23	0.21	0.14	0.07	0.33
Vocational	0.42	0.36	0.37	0.44	0.38
GED	0.42	0.46	0.40	0.23	0.45
2-year College	0.17	0.14	0.20	0.26	0.17
4-year College	0.05	0.04	0.05	0.07	0.04
Observations	1,085	682	432	864	2,582

Note: Proportions of individuals enrolling in different types of education, separately for individuals enrolling in Job Corps for different periods (s) in the treated and control groups. The table only includes individuals who obtain education when not enrolled in Job Corps (hence, it excludes the individuals who select n when not enrolled in Job Corps. For this reason, the total number in each group is smaller than the overall numbers in the treatment and control groups). The proportions do not sum to one because individuals may enroll in more than one education program.

Table A2: Quarters of Job Experience in Quarter 16

Duration in Job Corps (s)	Z=0		Z=1	
	n	a	n	a
0	9.40 (5.15)	9.24 (4.74)	9.38 (5.11)	9.39 (4.88)
1			8.91 (4.89)	8.72 (4.75)
2			9.54 (4.52)	8.82 (4.73)
3			8.66 (4.08)	8.88 (4.13)

Note: The means and standard deviations (in parentheses) for the number of quarters of job experience, separately for individuals enrolling in Job Corps for a different number of periods and making different educational choices when not enrolled in Job Corps. An individual is considered as employed in one specific quarter if she reports having worked during at least 1 week in that quarter. $z = 1$ if the person is assigned to the treatment group and 0 otherwise. a means alternative education when not enrolled in Job Corps, n for no education.

Table A3: Experimental Impact of Job Corps: Intention-to-Treat Effects

	Our Sample	Schochet et al. (2008)
Number of Treated	5,648	6,828
Number of Controls	3,781	4,485
Percentage of Treated Enrolling in JC	71.8*** (0.7)	73.0*** -
Panel A: Educational Outcomes		
Percentage Enrolling in any Education	22.8*** (0.9)	20.8*** (0.7)
Percentage Receiving GED	17.0*** (1.2)	15.0*** (1.0)
Percentage Receiving High School Diploma	-2.2*** (0.6)	-2.2*** (0.5)
Percentage Receiving Vocational Certificate	24.1*** (0.9)	22.3*** (0.9)
Percentage Receiving College Degree	-0.4 (0.3)	-0.2 (0.2)
Panel B: Labor Market Outcomes		
Employment in Q16	2.5** (1.0)	2.4*** (0.9)
Average Hourly Wage Q16	0.23*** (0.09)	0.22*** (0.08)
Average Weekly Earnings Q16	18.0*** (4.5)	18.1*** (4.1)

Note: The ITTs are computed as the difference between the weighted mean of the different outcomes for the treated and control groups; hence, for the results on percentage enrolling in different education these are to be interpreted as differences in percentage points. Results in Panel A are shown in Table 2 in Schochet, Burghardt, and McConnell (2008). Results in Panel B are shown in Table 3 in Schochet, Burghardt, and McConnell (2008). Results for average weekly earnings are shown in Table A1 in Schochet, Burghardt, and McConnell (2008). As in Schochet, Burghardt, and McConnell (2008), we use sample weights in these regressions to adjust for the sample and survey design. * indicates significance at the 10% level; ** indicates significance at the 5% level; and *** indicates significance at the 1% level.

Table A4: Empirical Estimate:s Employment Model

Periods in JC	Parameter	Alternative Education	No Education
0	Constant	0.46*** (0.03)	0.17** (0.08)
	Old	0.14*** (0.05)	0.27*** (0.07)
	Female	-0.04 (0.05)	-0.32*** (0.07)
	$\gamma_{\theta}^{0,k}$	0.18** (0.08)	-0.35*** (0.11)
1	Constant	0.31* (0.17)	0.29 (0.18)
	Old	0.28** (0.13)	0.14 (0.14)
	Female	-0.08 (0.1)	-0.27* (0.14)
	$\gamma_{\theta}^{1,k}$	0.14 (0.24)	-0.25 (0.4)
2	Constant	0.49*** (0.13)	-0.21 (0.49)
	Old	0.2 (0.18)	0.27 (0.2)
	Female	0.01 (0.13)	-0.39** (0.2)
	$\gamma_{\theta}^{2,k}$	0.12 (0.32)	-0.73 (0.51)
3	Constant	0.66*** (0.09)	0.62*** (0.05)
	Old	0.15 (0.12)	0.12 (0.09)
	Female	-0.09 (0.11)	-0.24*** (0.08)
	$\gamma_{\theta}^{3,k}$	0.28 (0.3)	-0.21 (0.19)

Testing for selection (p-values):

No selection on levels ($H_0 : \gamma_{\theta}^{s,k} = 0 \forall s, k$) 0.05

No selection on gains ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s,n} \forall s$) 0.00

No dynamic selection on gains in a ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s',a} \forall s \neq s'$) 0.99

No dynamic selection on gains in n ($H_0 : \gamma_{\theta}^{s,n} = \gamma_{\theta}^{s',n} \forall s \neq s'$) 0.98

Note: “Old” is a dummy variable for being more than 20 years of age; “Female” is a dummy variable for being female; and “Availability” is a dummy variable for living in a state where the availability of educational programs other than Job Corps (community colleges and postsecondary public institutions) in the academic year 1997/98 is above the median availability in the cross-states distribution. All of these variables are normed to have mean 0, which allows the intercept in the outcome equation to be interpreted as the average potential outcome in the population. Significance level (t-test for testing if each parameter=0): *** 1%, ** 5%, * 10%.

Table A5: Empirical Estimates: Cognitive Test Models

Periods in JC	Parameters	Prose Test		Quantitative Test	
		Alternative Education	No Education	Alternative Education	No Education
0	Constant	5.37*** (0.01)	5.22*** (0.02)	5.47*** (0.01)	5.37*** (0.02)
	Old	0.01 (0.02)	0.1*** (0.03)	-0.01 (0.01)	0.07*** (0.02)
	Female	0.05*** (0.01)	-0.04* (0.02)	0.05*** (0.01)	-0.02 (0.02)
	$\gamma_{\theta}^{0,k}$	0.26*** (0.03)	-0.3*** (0.03)	0.2*** (0.02)	-0.23*** (0.03)
1	Constant	5.22*** (0.07)	5.27*** (0.11)	5.35*** (0.06)	5.42*** (0.08)
	Old	0.08 (0.06)	-0.01 (0.08)	0.05 (0.04)	0 (0.05)
	Female	0.08* (0.04)	-0.01 (0.09)	0.1*** (0.03)	0 (0.07)
	$\gamma_{\theta}^{1,k}$	0.28*** (0.05)	-0.34*** (0.06)	0.22*** (0.04)	-0.23*** (0.04)
2	Constant	5.31*** (0.1)	5.00*** (0.07)	5.45*** (0.07)	5.24*** (0.07)
	Old	0.05 (0.06)	0.23** (0.12)	0.04 (0.05)	0.14* (0.07)
	Female	0.1* (0.05)	-0.18* (0.05)	0.06 (0.04)	-0.1** (0.05)
	$\gamma_{\theta}^{2,k}$	0.31*** (0.06)	-0.38*** (0.06)	0.21*** (0.04)	-0.27*** (0.06)
3	Constant	5.39*** (0.06)	5.41*** (0.03)	5.5*** (0.04)	5.5*** (0.03)
	Old	0.00 (0.04)	0.06* (0.03)	0.00 (0.03)	0.05* (0.03)
	Female	0.02 (0.04)	-0.04 (0.03)	0.04 (0.03)	-0.02 (0.03)
	$\gamma_{\theta}^{3,k}$	0.32*** (0.05)	-0.29*** (0.04)	0.22*** (0.04)	-0.22*** (0.04)
Testing for Selection (p-values):				Prose	Quantitative
No selection on levels ($H_0 : \gamma_{\theta}^{s,k} = 0 \forall s, k$)				0.00	0.00
No selection on gains ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s,n} \forall s$)				0.00	0.00
No dynamic selection on gains in a ($H_0 : \gamma_{\theta}^{s,a} = \gamma_{\theta}^{s',a} \forall s \neq s'$)				0.68	0.96
No dynamic selection on gains in n ($H_0 : \gamma_{\theta}^{s,n} = \gamma_{\theta}^{s',n} \forall s \neq s'$)				0.45	0.99

Note: “Old” is a dummy variable for being more than 20 years of age; “Female” is a dummy variable for being female. These variables are normed to have mean 0, which allows the intercept to be interpreted as the average potential outcome in the population. Significance level (t-test for testing if each parameter=0): *** 1%, ** 5%, * 10%.

Table A6: Empirical Estimate: Measurement Models

	Months in School	Months in Job	High School Credentials
Constant	-0.4*** (0.02)	-0.52*** (0.02)	-0.78*** (0.02)
Old	-1.04*** (0.05)	0.37*** (0.04)	1.1*** (0.03)
Female	-0.12*** (0.04)	-0.08** (0.04)	0.27*** (0.03)
θ	0.15*** (0.04)	-0.02 (0.03)	-0.13*** (0.03)

Note: Significance level (t-test for testing if each parameter=0): *** 1%, ** 5%, * 10%

Table A7: Empirical Estimates of the Subgroups' Shares and subLATEs Conditional on Employment

D^0	D^1	Shares $ H_i^1 = 1$	Earnings $ H_i^1 = 1$
Overall LATE			
no jc	jc	72.8	0.03
Dynamic Model			
a	(1,a)	9.6	-0.06
n	(1,a)	1.6	0.01
n	(1,n)	3.5	0.05
a	(1,n)	4.1	-0.09
a	(2,a)	6.4	0.02
n	(2,a)	1.6	0.18
n	(2,n)	4.2	-0.01
a	(2,n)	2.2	0.00
a	(3,a)	12.1	0.08
n	(3,a)	3.7	0.10
n	(3,n)	8.6	0.08
a	(3,n)	15.2	0.03
Static Model			
n	jc	23.1	0.07
a	jc	49.7	0.01

Note: Shares and subLATEs decomposition for the causal part of the estimated LATE on observed earnings. These are the estimates of the effects only for the individuals who are employed when assigned to the treatment group (see Section C.1). D^0 and D^1 refer to the potential choices in terms of periods in Job Corps (s) and of education (k) for the individuals when assigned to the control and treatment groups, respectively. Because individuals in the control group cannot enroll in Job Corps, D^0 is expressed only in terms of $k \in (a, n)$. As in Section 3.3, “subLATEs” refer to average treatment effect for the subgroup of individuals selecting each (D^0, D^1) bundle. The “Overall LATE” specification includes in one unique group all individuals for whom $s > 0$ under D^1 , not separating between different k under D^0 . The “Static Model” also includes in one unique group all individuals for whom $s > 0$ under D^1 , but separates between different k under D^0 . The “Dynamic Model” separates between different s and k under D^1 and different k under D^0 .

Table A8: Proportion and subLATEs, Overall Control Sample and Targeted Sample

D^0	D^1	Share		subLATE - Employment		subLATE - Earnings	
		Overall	Targeted	Overall	Targeted	Overall	Targeted
Overall LATE							
no jc	jc	72.40	72.90	0.04	0.06	0.02	0.03
Dynamic Model							
a	(1,a)	10.46	9.74	-0.07	-0.04	-0.07	-0.07
n	(1,a)	1.78	1.87	0.09	0.12	-0.01	0.00
n	(1,n)	3.67	3.94	0.01	0.02	0.04	0.05
a	(1,n)	4.62	4.45	-0.02	0.01	-0.10	-0.07
a	(2,a)	6.52	6.42	0.00	0.00	0.02	0.03
n	(2,a)	1.65	1.77	0.12	0.14	0.18	0.18
n	(2,n)	4.10	4.16	0.01	0.02	-0.03	-0.02
a	(2,n)	2.43	2.45	0.03	0.07	-0.03	-0.01
a	(3,a)	11.06	11.14	0.07	0.10	0.08	0.08
n	(3,a)	3.53	3.61	0.15	0.18	0.09	0.11
n	(3,n)	7.94	8.53	0.13	0.16	0.09	0.09
a	(3,n)	14.64	14.83	0.05	0.08	0.03	0.03
Static Model							
n	jc	22.66	23.88	0.09	0.11	0.06	0.07
a	jc	49.74	49.03	0.02	0.04	0.00	0.01

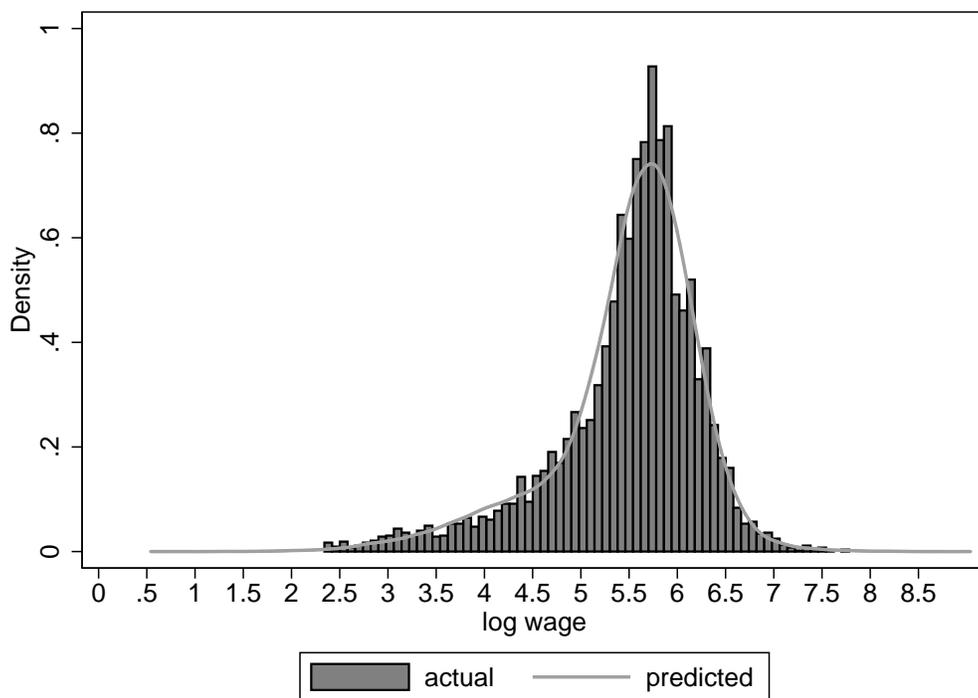
Note: The “Overall” columns include all the individuals in the control group. The “Targeted” columns include only the individuals in the control group with the highest 60% individual predicted benefit. D^0 and D^1 refer to the potential choices in terms of periods in Job Corps (s) and of education (k) for the individuals when assigned to the control and treatment groups, respectively. Because individuals in the control group cannot enroll in Job Corps, D^0 is expressed only in terms of $k \in (a, n)$. As in Section 3.3, “subLATEs” refer to average treatment effect for the subgroup of individuals selecting each (D^0, D^1) bundle. The “Overall LATE” specification includes in one unique group all individuals for whom $s > 0$ under D^1 , not separating between different k under D^0 . This is the LATE computed by the standard Wald estimator. The “Static Model” also includes in one unique group all individuals for whom $s > 0$ under D^1 , but separates between different k under D^0 . The “Dynamic Model” separates between different s and k under D^1 and different k under D^0 .

Table A9: Predicted Subgroup Shares and subLATEs with Different Availability of Further Education

D^0	D^1	Share		Employment Rate		Earnings	
		Low Availability	High Availability	Low Availability	High Availability	Low Availability	High Availability
Overall LATE							
no jc	jc	73.8	71.2	0.04	0.04	0.02	0.02
Dynamic Model							
a	(1,a)	10.3	10.6	-0.06	-0.07	-0.07	-0.07
n	(1,a)	1.9	1.7	0.10	0.08	-0.01	-0.01
n	(1,n)	3.9	3.4	0.02	0.01	0.05	0.03
a	(1,n)	4.8	4.5	-0.01	-0.03	-0.09	-0.12
a	(2,a)	6.5	6.5	0.00	0.00	0.02	0.02
n	(2,a)	1.7	1.6	0.12	0.11	0.18	0.18
n	(2,n)	4.2	4.0	0.00	0.01	-0.03	-0.02
a	(2,n)	2.5	2.4	0.03	0.02	-0.04	-0.02
a	(3,a)	9.9	12.1	0.08	0.07	0.08	0.07
n	(3,a)	3.2	3.8	0.15	0.14	0.11	0.07
n	(3,n)	9.1	7.0	0.13	0.13	0.09	0.08
a	(3,n)	15.9	13.5	0.05	0.05	0.03	0.03
Static Model							
n	jc	23.9	21.5	0.09	0.08	0.06	0.05
a	jc	49.8	49.7	0.02	0.01	0.00	0.00

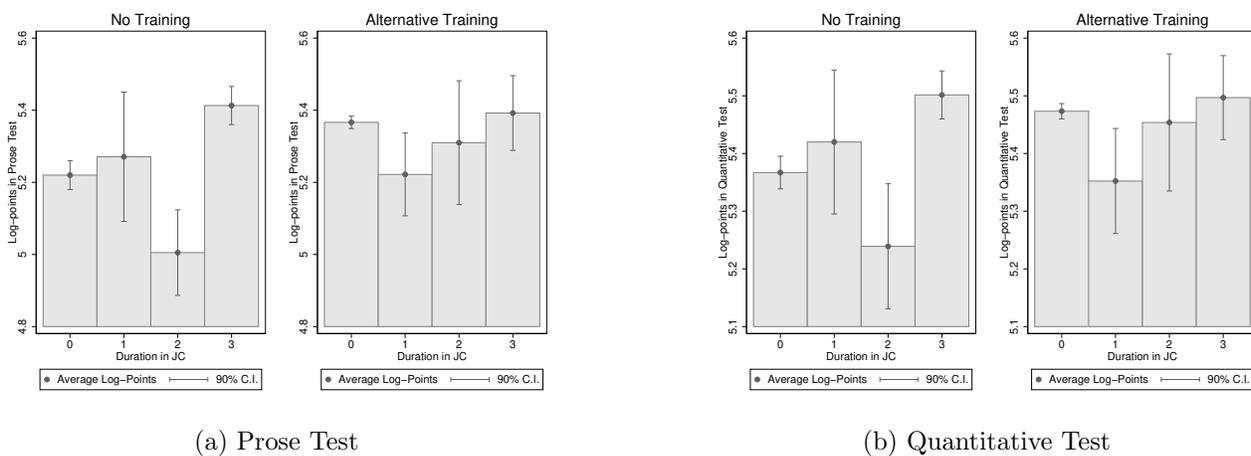
Note: D^0 and D^1 refer to the potential choices of the individuals when assigned to the control and treatment groups, respectively. As in Section 3.3, “subLATEs” refer to average treatment effect for the subgroup of individuals selecting each (D^0, D^1) bundle. “Static Model” refers to a model where we do not separate between the post-Job Corps choices (hence, all the individuals who have the same outside option when assigned to the control group and who enroll in at least one period of Job Corps are part of the same subgroup of compliers). The columns reporting results for “Low Availability” show the subLATE decomposition for the sample of individuals living in states where the availability of public education other than Job Corps is below the median of the availability in the cross-states distribution. The columns reporting results for “High Availability” show the results for the same sample of individuals, when we simulate the case that the availability is the same as that in the states where availability is above the median.

Figure A1: Goodness of Fit, Earnings Distribution



Notes: Comparison between the actual distribution of log-earnings as in the whole population and the distribution of log-earnings as predicted by our model.

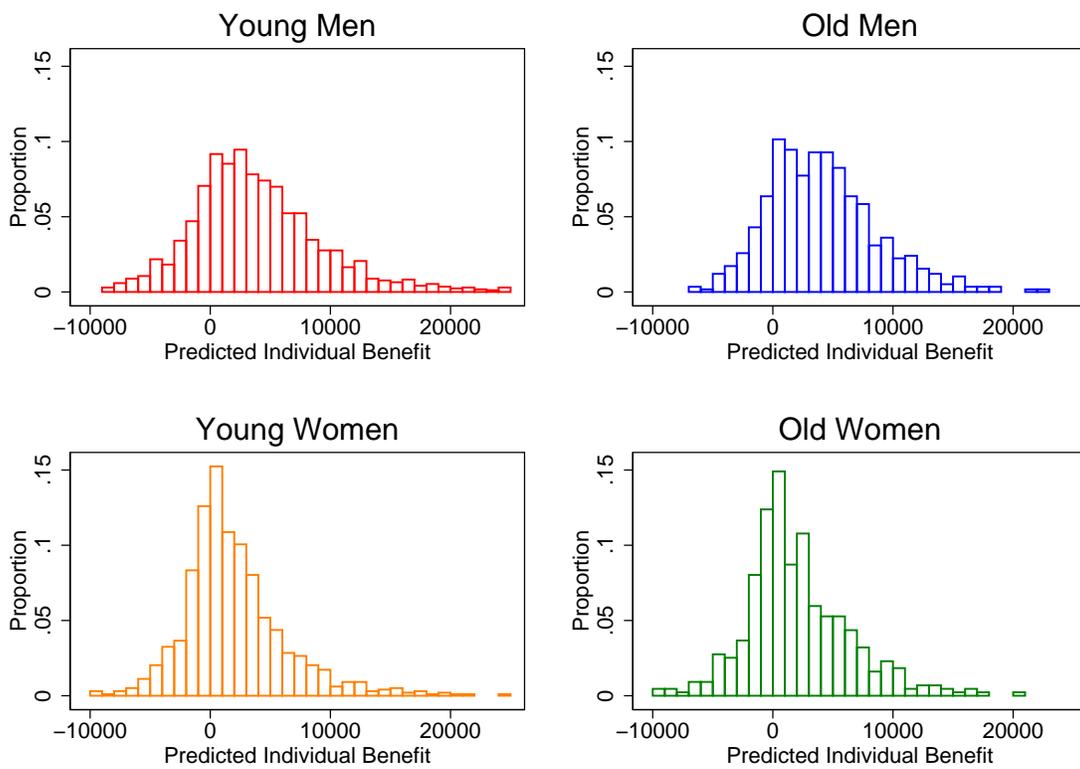
Figure A2: Empirical Estimates and 90% Confidence Intervals of the Constants in the Test Scores Models



(a) Prose Test

(b) Quantitative Test

Figure A3: Histograms of the Individual Benefits for Observationally Different Subgroups



Notes: In these graphs, the individual benefits for individuals are computed based on the conditional posterior distribution of θ , as explained in Section 6