DISCUSSION PAPER SERIES

IZA DP No. 2846

Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps

Alfonso Flores-Lagunes
Arturo Gonzalez
Todd C. Neumann

June 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps

Alfonso Flores-Lagunes
University of Arizona
and Princeton University

Arturo Gonzalez
Public Policy Institute of California
and IZA

Todd C. Neumann
University of California, Merced

Discussion Paper No. 2846
June 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 2846
June 2007

ABSTRACT

Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps*

Most of the literature on the evaluation of training programs focuses on the effect of participation on a particular outcome (e.g. earnings). The "treatment" is generally represented by a binary variable equal to one if participation in the program occurs, and equal to zero if no participation occurs. While the use of a binary treatment indicator is attractive for ease of interpretation and estimation, it treats all exposure the same. The extent of exposure to the treatment, however, is potentially important in determining the outcome; particularly in training programs where a main feature is the varying length of the training spells of participating individuals. In this paper, we illustrate how recently developed methods for the estimation of causal effects from continuous treatments can be used to learn about the consequences of heterogeneous lengths of enrollment in the evaluation of training programs. We apply these methods to data on Job Corps (JC), America's largest and most comprehensive job training program for disadvantaged youth. The length of exposure is a significant source of heterogeneity in these data: while the average participation spell in JC is 28 weeks, its standard deviation and interdecile range are 27 and 62 weeks, respectively. We estimate average causal effects of different lengths of exposure to JC using the "generalized propensity score" under the assumption that the length of the individual's JC spell is randomly assigned, conditional on a rich set of covariates. Finally, using this approach, we document important differences across different spell lengths and across three racial and ethnic groups of participants (blacks, whites and Hispanics) that help understand why the benefits these groups receive from JC are so disparate from estimates derived using traditional methods.

JEL Classification: C21, J24, I38

Keywords: training programs, continuous treatments, generalized propensity score, dose-response function

Corresponding author:

Arturo Gonzalez
Public Policy Institute of California
500 Washington St
Suite 800
San Francisco, 94111
USA
E-mail: gonzalez@ppic.org

* We thank Kalena Cortes and David Green for useful discussions of the paper at the ASSA Meetings in January 2007 and the Society of Labor Economists (SOLE) Meetings in May 2007, respectively. We are grateful for useful comments provided at the Princeton Junior Faculty Presentation Series and by Carlos A. Flores. Flores-Lagunes gratefully acknowledges financial support from the Industrial Relations Section at Princeton University

I. Introduction

Interest in whether various types of labor market interventions are effective, such as publicly sponsored job training programs, has spawned methods to estimate the causal effect of the receipt of a treatment on an outcome of interest, generally earnings or employment. While this literature has paid particular attention to estimating the causal relationship between the treatment and outcome variable under different assumptions about how individuals select into the treatment, it has for the most part only considered the case of a binary treatment variable, e.g. whether an individual undertakes the training program or not (e.g. Heckman, LaLonde and Smith, 1999; Imbens, 2004). Yet, it is possible that limiting the treatment to the binary case masks important heterogeneous effects of the program under consideration. For instance, participants in a job training program are typically exposed to different levels of training, suggesting that within treatment-group members, the treatment is not the same.

The length of program participation potentially provides more information regarding the effectiveness of the program than an indicator of participation, particularly in training programs where a main feature is the varying length of the training spells of participating individuals. If individuals that take up training receive different levels of training, then the average treatment effect estimated by any of the conventional estimators in the literature is unlikely to capture the heterogeneity in effects arising from different dosages of the treatment. This is not a novel concern, as some authors have considered the estimation of different components of a treatment. For instance, Imbens (2000) and Lechner (2001) were among the first to consider estimation of multi-valued treatment effects; while Hirano and Imbens (2004) considered continuous treatments.[2] In this paper, we illustrate how the estimation of causal effects from continuous treatments can be used to learn about the consequences of heterogeneous lengths of enrollment in the evaluation of job training programs, highlighting the type of insights that can be learned about the effects of training when its continuous nature is considered.

Recent work has paid especial attention to the heterogeneity in program effects by analyzing the distribution of impacts. For example, Heckman, Smith and Clements (1997) examine the distribution of treatment effects of the US Job Training Partnership Act (JTPA)

[2] Empirical applications of multi-valued treatment effects are in Gerfin and Lechner (2002), Lechner (2002) and Larsson (2003). Applications of continuous treatment effects are difficult to find. One exception is Fryges and Wagner (2007).

program, and Bitler, Gelbach and Hoynes (2006, 2007) estimate quantile treatment effects of Connecticut's Jobs First welfare program and Canada's Self-Sufficiency Project, respectively. The present paper relates to this literature by examining one factor (different training exposures by participants) that results in treatment heterogeneity, while still concentrating on mean effects, that is, average treatment effects for individuals receiving the same amount of training.

To estimate causal effects from continuous treatments, we employ the methods by Hirano and Imbens (2004) with properties akin to propensity score methods for a binary treatment variable (Rosenbaum and Rubin, 1983).[3] Under the assumption that selection into levels of the treatment is random conditional on a rich set of observable covariates, we use the "generalized propensity score" (GPS) to estimate the causal effect of different lengths of exposure to academic and vocational training on earnings. Conditional on the GPS it is possible to estimate average treatment effects of receiving different levels of exposure to the training program, thereby constructing a "dose-response function" (DRF) and hence providing insight into the causal effects of the training program that otherwise might be ignored.

We apply these methods to data on Job Corps (JC), America's largest and most comprehensive job training program for disadvantaged youth. These data are ideal for various reasons. First, the JC program consists of several types of academic and vocational instruction leading to different weeks of exposure by participants within the program. Length of exposure is a significant source of heterogeneity among JC participants: while the average participation spell in JC is 28 weeks, its standard deviation and interdecile range are 27 and 63 weeks, respectively. Second, the data available to us contain very detailed information about participants in the program, such as expectations and motivations for applying as well as information about the specific training center attended, all of which strengthens the plausibility of the "selection-on-observables" assumption necessary to our methodology.

Finally, as documented in Flores-Lagunes, Gonzalez, and Neumann (2006), the National Job Corps Study found positive and significant average treatment effects of JC training on weekly earnings for white and black participants, but a negative and insignificant effect for Hispanics (Schochet, Burghardt and Glazerman, 2001). The lack of a treatment effect for Hispanics is robust to the use of different conventional estimators and specifications (see Flores-

[3] Behrman, Cheng and Todd (2004) consider the role of exposure using a related propensity score method. Their method, however, is not a natural extension of propensity score methods for binary treatments.

Lagunes, Gonzalez, and Neumann (2006) for further details). Within this context, estimating a DRF can help analyze the extent to which differences in the variation in the length of training exposure across demographic groups matters in the assessment of the effectiveness of JC. Ultimately, the hope is to develop useful guidelines from this analysis to inform policymakers about how to improve the efficacy of JC for different groups of the population.

Our results suggest that the estimation of a (causal) DRF is indeed informative about the heterogeneity in average treatment effects, both across different lengths of exposure and for the racial and ethnic groups considered. In particular, we find that the estimated (marginal) effects of an additional week of training decline with the total length of enrollment in the program, such that the estimated DRF for JC participants is not uniform with weeks of training. Moreover, compared to whites and blacks, Hispanics' estimated impacts are larger, and their magnitude persists over longer training spells.

The rest of this paper is organized as follows. Section II presents an overview of the estimation method and the empirical strategy used to estimate the DRF. Section III discusses the JC and the National Job Corps Study, including a description of how heterogeneity in the length of enrollment in JC arises. Section IV presents the results of estimating the generalized propensity score, while section V describes the estimated average dose-response function for the different samples. Section VI concludes with a discussion and implications of the results.

II. Estimating the Dose-Response Function of Length of Enrollment in Job Corps Training

In the case of a binary treatment (e.g. participation on a job training program or not), the propensity score is commonly used to estimate average treatment effects. In particular, Rosenbaum and Rubin (1983) show that adjusting for differences in the conditional probability of receipt of the treatment given pre-treatment covariates (the propensity score) eliminates selection bias between treated and untreated individuals, if selection into treatment is purely based on observable factors. The propensity score simplifies the estimation of the average treatment effect by reducing the dimensionality of the conditioning set to one, avoiding the need to adjust for all pre-treatment variables simultaneously. For this reason, a large number of studies employ propensity score methods for the estimation of average (binary) treatment effects (see, e.g. the review by Imbens, 2004). Recently, methods that extend this framework to multi-valued and continuous treatments have been introduced.

A. The Generalized Propensity Score and the Dose-Response Function

Imbens (2000) extends Rosenbaum and Rubin's (1983) conditions for the validity of the propensity score to multi-valued treatments, while Hirano and Imbens (2004) extend the results to continuous treatments. Both of these papers employ the concept of a "generalized propensity score" (GPS) that is used to adjust for selection bias in the estimation of the average "dose-response" function (DRF) in a similar way that the usual propensity score does. In this context, the DRF is defined as the average effect of the multi-valued or continuous treatment on the outcome of interest.[4]

Borrowing notation from Hirano and Imbens (2004), let $Y_i(t)$ be the potential outcome of a treatment $t \in \Im$, where $\Im$ may be an interval (i.e. a continuous treatment). $Y_i(t)$ may also be thought of as the individual dose-response function. The observed variables for each unit *i* are a vector of pre-treatment covariates X_i, the level of the treatment received, T_i, and the observed outcome for the level of the treatment actually received $Y_i(T_i)$. Interest lies on the estimation of the average dose-response function (DRF): $\mu(t) = \mathrm{E}[Y_i(t)]$.

Similar to the case of estimation of binary treatment effects (Rosenbaum and Rubin 1983), an unconfoundedness assumption is needed. A key insight from Imbens (2000) is a weak version of unconfoundedness:[5]

$$Y_i(t) \perp T_i \mid X_i \text{ for all } t \in \Im .$$

This assumption that, conditional on observed covariates, the level of the treatment received (T_i) is independent of the potential outcome $Y_i(t)$, is at the heart of the literature of selection-on-observables. Particularly, this assumption rules out any systematic "selection" into levels of the treatment based on unobservable characteristics not captured by observable ones.

Under the weak unconfoundedness assumption, the average DRF could be derived by estimating average outcomes in subpopulations defined by pre-treatment covariates and different levels of the treatment. However, as the number of pre-treatment covariates increases, it becomes

[4] Imai and van Dyk (2004) introduce a similar concept to the GPS, the "propensity function", but propose a slightly different way to control for it in order to remove bias.

[5] It is referred to as weak unconfoundedness since it does not require joint independence of all potential outcomes, but instead requires conditional independence to hold for each value of the treatment.

difficult to simultaneously adjust for all covariates in X. In analogy to the binary treatment case in Rosenbaum and Rubin (1983), this dimensionality problem is solved by employing the generalized propensity score (GPS).

The GPS can be defined as follows. Let the conditional (on pre-treatment covariates) density of the treatment be given by

$$r(t,x) = f_{T|X}(t \mid X = x). \quad (1)$$

Then, the GPS is the conditional density of receiving a particular level of the treatment, $t = T$:

$$R = r(T, X). \quad (2)$$

Note the subtlety of the notation. The function $r(\cdot,\cdot)$ defines both the GPS, which is a single random variable at level T of the treatment and X, $r(T, X)$, and a family of random variables indexed by t, $r(t, X)$.

Similarly to the binary treatment case, the GPS has the "balancing property" in that $X \perp 1\{t = T\} \mid r(T, X)$. In other words, within strata defined by values of the GPS, the probability that $t = T$ does not depend on the value of X. This property, combined with the assumption of weak unconfoundedness above, has the important implication that *assignment* to the *level of* treatment is unconfounded *given* the GPS (Theorem 1 in Hirano and Imbens, 2004). If f_T is the conditional probability of receiving T, then, for every t:

$$f_T(t \mid r(t, X), Y(t)) = f_T(t \mid r(t, X)).$$

This result allows the estimation of the average dose-response function by using the GPS to remove selection bias. Bias-removal under the weak unconfoundedness assumption is achieved in two steps (Imbens, 2000, and Hirano and Imbens, 2004). The first step is to estimate the conditional expectation of the outcome as a function of the observed treatment level (T_i) and the GPS (R_i):

$$\beta(t,r) \equiv \mathrm{E}[Y(t) \mid r(t, X) = r] = \mathrm{E}[Y \mid t = T, r(T, X) = r].$$

$\beta(t,r)$ is the conditional mean of the outcome Y given the *observed* value of the treatment and the probability of receiving that value.

The second step is to estimate a value of the dose-response function by averaging $\beta(t,r)$ *over the values of the GPS* (R_i) at that particular level of the treatment:

$$\mu(t) = \mathrm{E}[Y(t)] = \mathrm{E}[\beta(t, r(t, X))].$$

Imbens (2000) and Hirano and Imbens (2004) demonstrate that, under the weak unconfoundedness assumption, estimating values of the DRF adjusting for the GPS in this way removes all selection bias (Theorem 2 of Hirano and Imbens, 2004).

To provide some intuition about this result, consider the following. The function $\beta(t,r)$ represents the average potential outcome for the strata defined by $r(T,X)=r$; however, it does not allow causal comparisons across different levels of the treatment since for other treatment levels the strata will be different, say $r(T',X)=s$ for treatment level T'. In other words, $\beta(t,s)$ defines the conditional expectation outcome for a different strata than $\beta(t,r)$, and hence, directly comparing these values *does not* yield a causal difference in the outcome of receiving treatment level *t* versus *s*. Therefore, the second step is needed for causal comparisons, which consists of *averaging* the conditional means $\beta(t,r)$ over the distribution of the GPS $r(t,X)$ (i.e., the "family" of random variables mentioned above). Computing the average DRF in this way yields values whose comparisons can be given causal interpretation.

B. Estimation Strategy

The empirical implementation of these concepts entails making a number of decisions and assumptions (e.g., such as parameterizations and functional forms) to sensibly estimate the objects defined above. In this paper we follow the implementation outlined in Hirano and Imbens (2004), paying special attention to assessing the validity of the assumptions made.

First, a lognormal distribution is used to model the conditional distribution of the treatment T_i (weeks spent in academic and vocational JC training) given the covariates. That is, we estimate $\ln(T_i)\,|\,X_i \sim N(\gamma_0+\gamma_1'X_i,\sigma^2)$. The lognormal distributional assumption is predicated based on the empirical distribution of the treatment for each of the samples considered (see Figure 1). Thus, the estimated GPS based on this model is simply

$$\hat{R}_i=\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\exp\left(-\frac{1}{2\hat{\sigma}^2}(\ln(T_i)-\hat{\gamma}_0-\hat{\gamma}_1'X_i)^2\right), \tag{3}$$

where $\hat{\gamma}_0$, $\hat{\gamma}_1$, and $\hat{\sigma}^2$ are estimated by ordinary least squares (OLS).

In the second step, the conditional expectation of the outcome given the observed treatment level (T_i) and the estimated GPS (e.g. $\hat{R}_i$) is modeled with a flexible linear specification and estimated with OLS:

$$E[Y_i \mid T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 \hat{R}_i + \alpha_4 \hat{R}_i^2 + \alpha_5 T_i \cdot \hat{R}_i . \quad (4)$$

Finally, in the third step, we estimate the value of the dose-response function at treatment level t by averaging the above regression function over the distribution of the GPS (holding constant the treatment level *t*):

$$\widehat{E[Y(t)]} = \frac{1}{N} \sum_{i=1}^{N} [\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t \cdot \hat{r}(t, X_i)] . \quad (5)$$

We can estimate values of the DRF corresponding to different values of the treatment repeating this last step.

III. Data: National Job Corps Study

A. The Job Corps Program

JC was created in 1964 as part of the War on Poverty under the Economic Opportunity Act, and has served over 2 million young persons ages 16-24.[6] JC provides academic, vocational, and social skills training at over 120 centers throughout the country where most students reside during training. In addition to education and vocational training, JC also provides health services and a stipend during program enrollment (Schochet, Burghardt and Glazerman, 2001).

Individuals are eligible based on several criteria, including age (16-24), poverty status, residence in a disruptive environment, not on parole, being a high school dropout or in need of additional training or education, and citizen or permanent resident. Approximately 60,000 new students participate every year at a cost of about $1 billion, and the typical JC student is a minority (70% of all students), 18 years of age, who has dropped out of high school (75%) and reads at a seventh grade level (U.S. Department of Labor, 2005).The motivation for applying to JC varies with age. In particular, the younger the applicant, the more likely he or she is interested in completing high school or GED degrees. Older applicants are less interested in general

[6] Job Corps operated under the Job Training Partnership Act from 1982 to July 2000, when it was replaced by Title I of the 1998 Workforce Investment Act (WIA).

training, and instead want job training. Above all, they see JC training as a means of finding employment since the majority has never held a full-time job (Schochet, 1998).

B. The National Job Corps Study

The data collected and used for this paper come from the National Job Corps Study (NJCS), a randomized experiment carried out during the mid- to late-1990s. The sampling frame for the NJCS consisted of first-time JC applicants in the 48 contiguous states and the District of Columbia. Since all JC training centers open in 1995 were part of the study, the NJCS (and the data used there) is based on a fully national sample. All pre-screened eligible applications (80,833) were randomly assigned into control, treatment, and program non-research groups between November 1994 and February 1996. Approximately 7% of the eligible applicants was assigned to the control group (N = 5,977) while 12% was assigned to the treatment group (N = 9,409). The remaining 65,497 eligible applicants were assigned to a group permitted to enroll in JC but were not part of the research sample.

Randomization took place before assignment to a JC center. As a result, not all of those randomized into the research treatment group enrolled in JC (73% of the treatment group enrolled in JC). Meanwhile control group members were barred from enrolling in JC for a period of three years. They could, however, enroll in other programs, some of which also offer job training and vocational opportunities which might be similar in nature or content as some of the JC training. The control and treatment groups were tracked with a baseline interview immediately after randomization and continuing 12, 30, and 48 months after randomization. Flores-Lagunes, Gonzalez and Neumann (2006) discuss other features of the NJCS.

C. NJCS Findings and Beyond

The original NJCS program evaluation is mostly based on a difference-in-means (or cross-section) estimator, modified to account for non-compliance: individuals in the treatment group who never enroll in JC, and individuals in the control group that enroll in JC before the three-year embargo (Schochet, 2001). This estimator identifies the local average treatment effect (LATE) of Imbens and Angrist (1994) on those individuals that comply with their treatment

assignment, since it is a Wald estimator where random assignment is used as an instrumental variable for the actual receipt of treatment.

The NJCS estimates imply an overall (full sample) gain of $22.1 in average weekly earnings at the 48-month after randomization, although it is not uniform across demographic groups: whites and blacks gain $46.2 and $22.8 per week, respectively, both statistically significant, while Hispanics show a statistically insignificant loss of $15.1.[7]

Flores-Lagunes, Gonzalez, and Neumann (2006) present evidence that a plausible explanation for this puzzling outcome is that Hispanics in the control group earn a significant amount of labor market experience during the study compared to treated Hispanics (and also control-group blacks and whites), resulting in an earnings advantage that treated Hispanics are not able to overcome by the end of the study. In addition, they show that Hispanics benefit from JC in the form of higher earnings growth relative to both control-group Hispanics and treated blacks and whites. The analysis to be presented below is consistent with and complements the findings in Flores-Lagunes, Gonzalez, and Neumann (2006) by shedding new light based on the heterogeneity in dose-responses across these samples.

D. Institutional Details of the Job Corps Program

Before providing summary statistics about the samples to be employed, this section describes relevant institutional details of the JC program, which is important in understanding the source of variability in our continuous treatment variable (weeks of training), and thus the selection mechanism. As will be clear shortly, even though most of the variability in length of enrollment is determined individually by the student, it is possible that JC staff influences it as well. Fortunately, our data is rich enough to allow controlling for the specific center each individual attends, accounting for most, if not all of the institutional factors determining selection.[8]

[7] All the NJCS estimates for the entire sample are based on average weekly earnings in quarter 16; however, the estimates by race and ethnic group in the NJCS report employ average weekly earnings in year 4. Throughout this paper we employ earnings in quarter 16 as our measure since it is the most recent measure.

[8] Larsson (2003) and Gerfin and Lechner (2002) motivate the use of variables related to the caseworker or placement officer to help control for idiosyncratic systematic differences in assigning individuals into types of training. Even though in our case (as explained below) the Job Corps counselor is supposed to play a passive role, it is still important to account for their potential effect into the determination of training lengths. While we do not have individual counselor information, we are able to adjust for Job Corps center "fixed effects". In addition to

From the point of view of the student, the JC program consists of four stages (U.S. Department of Labor, 2005): (1) outreach and admission leading to the decision to enroll in JC, (2) the career preparation period shortly after enrollment, (3) career development during the training portion of participation, and (4) transition into the labor market. Students play an integral part in each stage and they determine which course of action to take after counselors provide information and advice. The JC staff plays a significant role in helping students successfully transition out of each stage, but the program is streamlined and formalized to minimize any potentially subjective role by JC counselors.

In the first stage, counselors determine the eligibility of applicants using a standardized form based on the objective criteria outlined in section III.A above. Counselors also determine eligibility based on whether the applicant demonstrates a desire to gain from academic and career technical training and are judged capable of getting along with others in a group setting (U.S. Department of Labor, 2005). While this last criterion is subjective, it should have no bearing on our results, since inclusion into the sample (randomization in the original NJCS) took place after applicants were deemed eligible for JC. Subsequently, and along with the counselor, students choose a vocational program after the counselor informs them of labor market trends and available options for vocational training through JC. Counselors also provide details about the rules, expectations, and graduation requirements at this stage, including the expectation that students commit to 8 to 12 months of training (U.S. Department of Labor, 2006, Appendix 102).[9] In principle, if the vocational program of choice is not available at the closest JC center, students may choose to attend one that does offer that particular program.[10]

In stage two, which takes place within the first 60 days in JC, students and the JC counselors establish a "career preparation plan" tailored to each student's needs. This plan helps students acclimate to center life, assesses their skills and interests, helps them choose a career, and an academic, social and vocational training plan. Students that score below the threshold for

---

institutional factors, these fixed effects should help account for differences in local labor market conditions across center locations.

[9] This expectation is merely a recommendation, since Job Corps emphasizes to students that: "Job Corps is a self-paced program. That means you learn at your own pace. Depending on the career area you choose and the learning pace you set for yourself, training can take from eight months to two years to complete. Job Corps recommends that you remain on-center for at least one year to gain the knowledge and social skills needed for your new career." (see, e.g. http://jobcorps.dol.gov/faq.htm#stay).

[10] For applicants younger than 18, students and counselors are constrained with regards to the choice of Job Corps center as regulations require them to be assigned to the center closest to the applicants' residence, unless parents request a different assignment.

the reading instruction requirement must continue receiving such instruction until they score above this threshold.[11] Students not proficient in English may thus have more difficulty meeting this requirement in the same period of time as other students.

In the career development period (third stage), students undertake all training needed to achieve the goals of the career preparation plan, culminating with the search for jobs. Evaluation of the students' progress takes place every 60 days as part of this stage. Students carry on training at their own pace and counselors are not expected to discriminate between students by their length of stay (U.S. Department of Labor, 2006). This means that counselors are unlikely to want to either minimize or maximize the students' length of stay. In the fourth stage, students obtain their first job and find living accommodations. Importantly, while they can use placement services after "graduation" from JC if needed, this is not counted as part of their training spell.

These institutional details highlight the fact that the length of enrollment in JC training is determined mainly by student's choices and to a lesser extent by the JC counselors. Given our maintained assumption that conditional on observed covariates the enrollment length is random, it is crucial that we effectively control for all factors determining enrollment spells. Fortunately, the NJCS provides very rich data that (in our view) makes plausible this selection-on-observables assumption, such as center attended while in JC and a myriad of variables reflecting the individual expectations and motivation upon applying to JC. We explain these variables in detail below.

E. Summary Statistics of the Data Employed

The pre-treatment covariates and labor market earnings of interest for this study are taken from the baseline and 48th month surveys, respectively, of the NJCS. We concentrate mainly on those individuals who enrolled in JC in order to compare the effect of length of enrollment in the program on their weekly earnings 48 months after randomization took place. This sample consists of 3,406 individuals who report being white, black or Hispanic. For comparison, we also employ a sample of non-participants available in the NJCS. The long list of pre-treatment covariates used in the GPS model can be classified into demographic, education, health, and

[11] Centers test all students at the beginning of their enrollment in JC and provide them with reading instruction if they test below 567 on the Reading subtest of the Tests of Adult Basic Education (TABE). Students continue to receive reading instruction as a part of their overall academic and vocational training programs and are not exempt from follow-up TABE testing until they achieve the required reading score.

economic variables, pre-treatment expectations about and motivations to enroll in JC, and geographical variables such as state of residence and the JC center attended.

Table 1 presents means (first column) and standard deviations (second column) of selected pre-treatment variables for our four samples, arranged in vertical panels. The majority of JC participants in our sample is black (54%), 18% is Hispanic, and 28% is white. We measure time of enrollment in JC employing an item in the NJCS survey that measures the hours spent in either vocational or academic training while in the program, rescaling them into weeks by assuming a 40-hour workweek. JC is a time-intensive program, with the average participant enrolling in 28 weeks of vocational and academic training. This is equivalent to almost 60% of a full-time year-long job. Whites and blacks completed similar levels of training, 26 and 27 weeks respectively, but significantly less than Hispanics, who enrolled for over 34 weeks of training. The values of the interdecile range are 62 (full sample), 58 (whites), 61 (blacks), and 73 (Hispanics), reflecting the large variation in enrollment spells among participants.

Prior to enrollment, JC participants have an average age of 18.7, completed 10 years of school on average and just over 20% have completed high school or the GED, are predominantly unmarried (99%), and are more likely to be male (57%). About 80% of JC enrollees have ever worked and have average weekly earnings of a little over $115, while about 18% of them still live with their parents. Their self-reported health indicators imply that 87% are in excellent or good health; although 51, 54 and 31% report to ever have smoked cigarettes, pot, or drank alcohol. Additionally, about 23% have been arrested and most live in an urban area (79%). Even though most respondents are not married, 10% are the head of the household, and nearly one in five (17%) have a child.

A number of average characteristics vary significantly across the samples by race and ethnicity. For example, whites are less likely to be female (34%), have about $30 higher weekly earnings at baseline than the other two groups, are more likely to have been arrested (29%) and ever smoked or consumed alcohol. Both blacks and Hispanics are more likely to have children (20%), and to be assigned to non-residential training (17%); while blacks are less likely to have a GED degree (2%). Hispanics are particularly less likely to speak English fluently, 52% compared to almost 100% for whites and blacks. They also undertake about 8 more weeks of training, are more likely to live in a PMSA (43%) and to live with their parents (27%).

Variables pertaining to the expectation from and motivation for enrolling in JC are important since these variables are intended to help control for unobserved characteristics that may be related to both the outcome (earnings) and length of enrollment in JC. Such variables include: whether the individual had any worries about attending JC, knew the type of job he/she would like to train for at JC, knew what center wished to attend, whether the individual joined JC to improve math skills, reading skills, to help get along better with people, to improve self-control, to improve self-esteem, to find a specific job, to help find friends, whether individual heard about JC from parents, knew somebody who took JC in the past, whether the individual joined JC to get away from home, to get away from a community problem, to get trained, to attain a career goal, to get a HS or GED degree, to find work, or joined JC for other reason, and a prediction by the interviewing counselor about whether the individual would enroll in residential or nonresidential training. We believe that including this set of variables in the GPS model strengthen the argument for the validity of the selection-on-observables assumption.

Finally, the state and JC center-attended indicators (not listed in Table 1 to save space) are intended to control for local labor market dynamics and also for the potential role played by counselors in the case of the latter indicators. We note that the number of centers represented by these indicators is 109 for the full sample, with no center having more than 5.2% of individuals, so that these indicators are likely correlated with local labor market conditions. Overall, in our view, the richness of our data makes a strong case for the validity of the weak unconfoundedness assumption. In the next sections, we describe the role played by these variables in the estimation of the GPS and some exercises undertaken to evaluate the specification of the GPS model. In addition, we document some differences in the estimated coefficients among the three racial and ethnic groups that, together with the differences in characteristics described above, argue in favor of considering them separately.

IV. Estimation of the Generalized Propensity Score

A. Estimates of the GPS

The third and fourth columns for each group in Table 1 present the estimated coefficients of the GPS model and their estimated standard errors. The bottom panel of the table shows the R^2 of the model along with p-values of F-tests for the joint significance of different sets of variables.

Recall that the estimated GPS model is the basis for controlling for selection bias into different lengths of training undertaken. The GPS model is estimated using least squares under the log-normal distribution assumption described in the previous section.

For the full sample, several individual estimated GPS coefficients are statistically significant. Variables related to higher length of enrollment are indicators for being black or Hispanic (relative to whites), female, living in a PMSA, and knowing someone who attended JC; while variables negatively related to (log) weeks in training are indicators for having a child and ever smoking or drinking. The bottom panel shows that the GPS model has an R^2 of 9.4% and that the demographic, health, expectations and motivation, and JC center indicators are each statistically significant as groups of variables at the 7% level or better.

The GPS model corresponding to subsamples defined by racial and ethnic groups are similarly specified with the exception of whites. For this group, dropping the variable "Joined JC to achieve career goal" (because it was perfectly collinear with another motivation variable) and specifying the expectations in joining JC variables and the reasons to join JC as percentages of the total number of expectations and reasons result in an improved GPS model with better balancing properties. Overall, the three groups show the common pattern of achieving a higher model R^2 relative to the full sample, but having just a few statistically significant estimated coefficients (except whites). In particular, the R^2 of the models range from 25% for whites and Hispanics to 12% for blacks; while the groups of statistically significant variables are the JC center indicators for whites and blacks, the demographic and health variables for blacks, and the economic variables for Hispanics. Nevertheless, given the predictive purpose of the GPS model, we do not consider this a serious concern, which likely arises as a consequence of high collinearity among the variables and relative smaller samples.

The estimated GPS model also reveals important differences among the subsamples. The indicator for female significantly lengthens enrollment for Hispanics, while being the head of household shortens spells for blacks. Being in good health is relevant for both whites and blacks, although they have the opposite effect on length of enrollment. Ever smoking shortens the spells for whites and blacks but it is not statistically significant for Hispanics, while ever drank alcohol is only relevant for blacks. For Hispanics, residing in a PMSA is significantly correlated with larger spells (and most Hispanics reside in those areas), as well as enrolling in JC expecting to get training for a specific job. Indicators that significantly lengthen the enrollment of blacks but

not other groups are having heard about JC from parents, knowing someone who attended JC and joining JC to achieve a career goal. Finally, several of the expectations and motivation variables are statistically significant for whites, all of them increasing their JC enrollment length. In summary, these differences in estimated GPS coefficients argue in favor of considering these groups separately.

B. Balancing Properties of the GPS

Recall that an important property of the GPS is that it "balances" the covariates within strata defined by the values of the GPS, such that, within strata, the probability that $t = T$ does not depend on the value of X. More formally, $X \perp 1\{T = t\} \mid r(t, X)$. This balancing property can be employed to empirically assess the adequacy of our chosen functional form to estimate the GPS in a similar spirit in which it is done in the binary treatment case with the propensity score (e.g., Dehejia and Wahba, 2002; Smith and Todd, 2005). In the case of a continuous treatment, one approach to check the balance of each covariate consists of running a regression of each covariate on the log of the treatment and the GPS (Imai and van Dijk, 2004). If the covariate is balanced, then the treatment variable should have no predictive power conditional on the GPS. A comparison of this coefficient to the corresponding coefficient of a regression that does not include the GPS can be used to gauge the balancing provided by the GPS.

We perform this exercise to check the balance of each covariate that was included in the GPS model. We summarize the results of this balancing check using standard normal quantile plots of the *t*-statistics for the coefficient on the log of the treatment variable. Figures 2.1 to 2.4 show the standard normal quantile plots for the full, white, black and Hispanic sample, respectively. Panel (a) of each figure shows the normal quantile plot for the *t*-statistics on the regressions that do not include the GPS. For each covariate we use OLS or a logit specification if the covariate is binary, and we also include the square of the treatment variable in the specification. This panel in each figure conveys an idea about how "unbalanced" the covariates are in each of our four samples when not controlling for any covariates or the GPS. Consequently, Figure 2.1(a) shows that the full sample has 11 covariates with *t*-statistics greater than 2 in absolute value, while Figure 2.2 (a) shows 3 such *t*-statistics for whites, Figure 2.3 (a) 8 for blacks and lastly Figure 2.4 (a) shows 2 statistically significant *t*-statistics for Hispanics.

Panel (b) in each of the figures shows that, once the GPS is included in the regressions,

the number of statistically significant *t*-statistics declines in all cases—that is, the balance of the covariates improves. These regressions are specified as before but adding the GPS in level, square, and cube. In addition, an indicator for gender is explicitly included in the specification to improve the balance, as well as a race and ethnicity indicator in the full sample.[12] The best balancing occurs in the full sample, for which no covariate regression shows a statistically significant (greater than two) *t*-statistic on the coefficient on the treatment variable in Figure 2.1 (b). For whites, Figure 2.2 (b) reveals that 2 *t*-statistic remain statistically significant, while for blacks Figure 2.3 (b) shows 3 significant *t*-statistics, down from 8. Finally, Hispanics have only one *t*-statistic which is greater than two on Figure 2.4 (b). Overall, the fact that the GPS achieves a better balance in all samples suggests that its specification is adequate.[13]

C. Assessing the Support Overlap Condition

In the binary treatment literature, it is well known that methods that adjust for pre-treatment observable variables are likely to work poorly if there is not enough overlap in the distribution of covariates by treatment status. In that literature, it is common to gauge the overlap by looking at the distribution of the propensity score among treated and non-treated individuals, sometimes restricting estimation to the common support region. However, controversy still exists about what the best way to gauge overlap is, and how best to tackle the issue of lack of overlap (Imbens, 2004).

While in the case of multi-valued or continuous treatments it is also true that inference may be poor if there is no sufficient overlap in the distribution of covariates across different levels of the treatment, it is considerably more difficult to gauge this condition. The main reason for this is that there are many levels of the treatment and consequently multiple parameters of interest, each of them requiring a potentially different support condition. In the case of multi-valued treatments, for example, Gerfin and Lechner (2002) consider estimation of effects from

[12] We control explicitly for gender (and race/ethnicity) since these are important variables that the GPS may not weight enough. Imai and van Dijk (2004) take a similar approach in their application, including explicitly age and its square. In the estimation of the DRF we also control explicitly for gender (and race/ethnicity), so that the application is internally consistent.

[13] We also undertook another exercise to assess the balancing of covariates obtaining similar results indicating that the specification of the GPS is adequate. This exercise follows Hirano and Imbens (2004) and consists of dividing the levels of the treatment into three intervals. Then, within those intervals, we stratify individuals into five values of the GPS evaluated at the median value of the treatment of the corresponding interval. Finally, we test whether the observed covariates are "balanced" among these strata. For details on this procedure, see Hirano and Imbens (2004).

nine different subprograms with interest in pair wise comparisons among them. In estimation, they restrict the sample to those individuals that have the possibility of participating in all states (according to their estimated model). The case of a continuous treatment is even more complicated since there is a continuum of treatment levels by definition. To our knowledge, there are no concrete suggestions in the literature on how to gauge the common support condition in this context. Consequently, we present here an exercise to examine the extent to which the support condition is satisfied, but we do not restrict our sample in any way based on this analysis. Fortunately, the evidence suggests that the support condition is likely to be satisfied in our data.

To informally gauge the extent of overlap in the supports of different levels of the treatment, we divide these values into five quintiles.[14] For each quintile, we compute the value of the GPS for each individual at the median level of the treatment for the quintile. Subsequently, we compute the value of the GPS at the same median level of the treatment for all individuals that are not part of the quintile in question. Finally, we compare the supports of the values of the GPS for these two groups (individuals in the quintile in question and the rest) by superimposing their histograms. This is similar to what Dehejia and Wahba (2002) do in the binary treatment case.

This exercise is repeated for each quintile in turn, resulting in five plots for each of our samples, which are shown in Figure 3(a) through 3(e) for the full sample. These figures show that the overlap in the support of the estimated GPS across quintiles is very good in general, with only a few instances at the tails in which the support condition fails. Similar conclusions can be drawn from the corresponding figures for the other samples (available upon request), although, as expected, the overlap deteriorates slightly due to the smaller sample sizes. Overall, we conclude that the overlap support condition is not a serious concern in our estimated model and samples, although the evidence upon which this conclusion is drawn is only suggestive.

V. Estimates and Plots of the Dose Response Function (DRF)

Recall that the second step towards the estimation of values of the DRF, after the estimation of the GPS, is to estimate the conditional mean of the outcome given the observed

[14] We also undertook the same exercise described here using deciles. The results are very similar as those reported here, although, as expected, the overlap deteriorates slightly.

treatment level (T_i) and the estimated GPS ($\hat{R}_i$). We report in Table 2 the results of this step employing a flexible linear specification estimated with OLS:[15]

$$E[Y_i \mid T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 \hat{R}_i + \alpha_4 R_i^2 + \alpha_5 T_i \cdot \hat{R}_i \ .$$

It is important to note that the estimated coefficients presented in Table 2 *do not* have any causal interpretation, as causality requires averaging this conditional mean over the distribution of the GPS. However, as Hirano and Imbens (2004) note, a test of joint significance of the coefficients related to the GPS "can be interpreted as a test of whether the covariates introduce any bias." Test statistics and p-values of a test of joint significance of the GPS coefficients (and covariates explicitly included) are reported at the bottom of Table 2 for each of the samples under analysis. They show that, for all samples, the null hypothesis that the coefficients are jointly zero is strongly rejected.

Once the conditional expectation of the outcome given T_i and $\hat{R}_i$ has been computed, we can obtain average effects for different values of the treatment in order to construct the dose-response function (DRF). We present the entire DRF for the full sample and each racial and ethnic group in Figures 4 – 7, providing a general overview of how participants benefit from additional JC training. The DRF plots are obtained with 99 different values of the treatment that correspond to the 99 percentiles of the corresponding empirical distribution. We also present a plot of a function that predicts earnings using OLS of weekly earnings on a quadratic function of the length of training and the full set of covariates used in the estimation of the GPS. The difference between the two is the more flexible specification provided by the GPS method. For each of them (DRF and OLS), we provide 90% (point-wise) confidence bands obtained with 1,000 bootstrap replications. Finally, each figure also shows a covariate-adjusted non-participants mean earnings at month 48 (horizontal line) and the length of training that correspond to the 25th, 50th, and 75th percentiles (vertical lines).[16]

[15] In addition to the treatment level and the estimated GPS variables, we also include an indicator for gender and indicators for race and ethnicity for the full sample. These variables are important to control for explicitly, as explained in section IV.B.

[16] The covariate-adjusted non-participants mean earnings at month 48 are obtained in the following way. First, we fit a linear model for earnings as a function of all covariates for JC non-participants (control-group members plus treatment-group members that never enrolled in JC). Then, we use the estimated coefficients of this model on the sample of JC participants to predict their earnings (as a function of their observable characteristics) "had they been non-participants". In other words, we construct a counterfactual non-participant mean earnings for a group with the same observable characteristics as the JC participants.

Figures 4 - 7 present the results for the full sample of JC trainees, whites, blacks and Hispanics, respectively. There are a few general observations to make about all figures. First, is the fact that for all groups the (marginal) impact of an extra week of training declines with the length of training. Second, there are important differences between the shapes of the OLS function and the DRF, likely arising by the flexibility allowed by the DRF. However, the confidence bands indicate that they are not statistically different from one another. Finally, we note that the covariate-adjusted non-participants mean earnings for all groups nicely aligns with the origin of the DRF, providing an informal validation of the GPS methodology, as by definition non-participants have zero weeks of training.

Analyzing each group in turn, the returns to training (slope of the DRF) for the full sample in Figure 4 decrease across the entire distribution but remain positive through 57 weeks of training, where earnings are maximized and after which additional training lowers participants' earnings. The returns to training are higher than the covariate-adjusted non-participants mean throughout all levels of training. OLS seems to overstate the returns to training in the first few weeks up until the 25^{th} percentile, where the two lines cross. Between the 25^{th} and 75^{th} percentile the two lines are similar although the DRF estimates higher returns. Finally, OLS overestimates returns again beyond the 75^{th} percentile of training lengths.

An evident difference in Figure 5, which shows the results for whites, is that the DRF is considerably flatter than that of the full sample. For this group, the returns to longer training spells also decrease across the entire distribution but remain positive through 52 weeks of training. The returns to training are higher than the covariate-adjusted non-participants mean after only the first week of training. In this sample, the difference between OLS and the DRF is more evident, with OLS underestimating the effects from the second week of training up until the 50^{th} percentile of training lengths, after which OLS overestimates the impacts. In addition, the DRF confidence bands are tighter than those of OLS.

The corresponding graph for blacks in Figure 6 shows that their DRF is even flatter than that of whites, although it achieves a maximum in the 62^{nd} week of training. For this sample, the difference between OLS and the DRF is minimal, although a similar pattern emerges in which OLS first underestimates, then overestimates the impacts. Finally, the OLS confidence bands for this group cannot rule out zero effects throughout the distribution of training spells, while the DRF provides much tighter bands.

Figure 7 presents the results for Hispanics, revealing particular features not present for the other groups. First, it is evident from the figure that Hispanics show the steepest DRF of all, implying that additional JC training is highly beneficial for them in the first few weeks of training since their weekly earnings increase rapidly; nevertheless, as in the other groups, the marginal impacts decrease throughout the graph. Second, this is the only group for which weekly earnings decrease for the first few weeks of training, after which the returns become positive and remain so up until the 58th week of training. Third, Hispanic participants show higher weekly earnings than the covariate-adjusted non-participants mean only after the 8th week of training, and not immediately like the other groups. Fourth, for Hispanics the difference between OLS and the DRF is largest, with OLS overestimating the impacts through the 50th percentile of the distribution of spells, underestimating them between the 50th and 90th percentiles and overestimating them thereafter. Finally, just as for whites and blacks, the OLS results are more imprecise than the DRF, as judged by the width of the corresponding confidence bands.

To further examine the results implied by the DRF, Table 3 shows values of the DRF for the 25th, 50th and 75th percentiles of the empirical distribution of treatment levels along with its estimated derivative (DRF-Diff). The DRF-Diff is computed as the "forward" change of one additional week of training at that particular percentile level of the treatment. This derivative is informative about the extent to which further time spent in vocational and academic training in JC is predicted to increase weekly earnings at the 48th month after randomization.[17]

For the full sample, the 25th percentile corresponds to 7.5 weeks of training and results in 48-month post-randomization weekly earnings of $209. For comparison, the covariate-adjusted mean earnings of non-participants at 48-month is $199, reflecting a gain of $10. Looking at the derivative estimate at the 25th percentile, it is clear that an extra week of training is beneficial as it yields an extra $0.94 per week. Consistent with this notion, the estimate at the 50th percentile (20.6 weeks of training) is $221 with corresponding derivative of $0.77, reflecting positive but decreasing returns to training. Finally, the DRF estimate at the 75th percentile (40 weeks of training) is $230 with a positive derivative of $0.45. Therefore, according to our estimates, a randomly drawn individual is expected to benefit from additional JC vocational and academic

[17] Note that the estimate of an additional week of training (DRF-Diff) does not necessary correspond to the slope of the DRF at a given point in Figures 4 - 7. This is so because DRF-Diff is computed by averaging over the distribution of the GPS after obtaining the difference between a corresponding value of the treatment and the one-week change in such value, while the slope of the DRF is only an extrapolation between two estimated values of the DRF.

training even at the 75th percentile of the empirical distribution of training intensities, although the marginal gain in earnings is decreasing.

Looking at particular racial and ethnic groups, however, important differences emerge. For the white sample, the 25th percentile corresponds to 6.9 weeks of training and yields a DRF estimate of $266, compared to $242 covariate-adjusted mean earnings for non-participants, for a gain of $24. The corresponding derivative at this 25th percentile is $1.06 which shows that whites benefit greatly from more JC training at this level of training intensity. The DRF estimate for the 50th percentile is $276 with positive derivative of $0.73. However, in contrast to the full sample, whites experience almost half the returns from additional JC training at the 75th percentile of their empirical distribution of training intensity since, while the DRF estimate is $286, the predicted return from an additional week of JC training is only $0.29.

For blacks, the value of the DRF at the 25th percentile (7.1 weeks of training) is $189 with a derivative of $0.53, the smallest derivative value among the three groups for this percentile, while at the 50th percentile (19.8 weeks of training) the DRF has a value of $195 (derivative of $0.46) and at the 75th percentile is just $201 (derivative of $0.30). The values of these derivatives are consistent with the observed flatness in the DRF for this group. For comparison, the covariate-adjusted mean weekly earnings for black's non-participants is $171, implying positive gains for trainees relative to no training at each percentile.

Consistent with the DRF plots, Hispanics show the highest magnitude of the derivatives at each of the percentiles of their treatment distribution. This is the case despite Hispanics being the group having the longest average spells in JC, which means that they indeed benefit more from those longer spells.[18] Table 3 shows that mean earnings for Hispanics are $188, $212, and $229 for the 25th (10 weeks of training), 50th (26.2 weeks of training) and 75th (50 weeks of training) percentiles of their treatment empirical distribution, respectively. The corresponding benefits from an additional week of training are $1.62, $1.32 and $0.63, surpassing by much those of the other two groups. Conversely, the covariate-adjusted mean earnings of Hispanic's control group is $184, implying an overtaking point at a level of training (8 weeks) that corresponds to the 20th percentile of the empirical distribution, which stands in stark contrast to

[18] This is a factor consistent with the argument in Flores-Lagunes, Gonzalez and Neumann (2006) that Hispanics do in fact benefit from JC training, but that 48 months after randomization is not enough time for the average benefits from participation to be seen due to the relatively high mean earnings of the control group.

whites and blacks, who earn more than the corresponding control group essentially right after the start of JC.

These observations about the differences among racial and ethnic groups are strengthened when we compare the DRF and DRF-Diff across groups holding constant the level of the treatment, as opposed to using the own subgroup's empirical distribution of training intensities. We fix the levels of the treatment at the 25^{th}, 50^{th} and 75^{th} percentiles for the full sample (i.e. 7.5, 20.6 and 40 weeks of training, respectively) and obtain the corresponding values of the DRF and DRF-Diff for all three groups. The bottom panel of Table 3 shows that the key insights from the top panel prevail for whites and blacks. For Hispanics it is now even more evident that they experience higher returns from increased time spent in JC training relative to the other two subgroups, as judged by the derivative DRF-Diff, which is between .6 and 4.38 times larger at the percentiles shown.

Consequently, while Hispanics spend more time in JC training relative to the other two subgroups, by our estimates it is rational for them to do so (conditional on a sufficiently low discount factor and budget constraint) given that their future average earnings are predicted to be higher relative to having a shorter JC spell. This insight weakens the originally estimated lack of average effects for Hispanics in the NJCS, given that Hispanics need more time in training to maximize their benefits from the program and the fixed point in time at which the outcome is measured (48 months after randomization).

It is of interest to speculate on why Hispanics show a DRF that is different from that of whites and blacks. One plausible explanation is as follows. Recall that fluency in English is lowest among Hispanics relative to whites and blacks, while JC emphasizes proficiency in English in order to complete their academic or vocational training (see section III.D). Therefore, the returns to longer enrollment in JC and the steeper DRF among Hispanics can reflect returns to increased English proficiency, if such attribute is valued in the labor market over other skills (Bleakley and Chinn, 2004; Gonzalez, 2000; McManus, Gould and Welch, 1983). It may also be the case that Hispanics need more time in JC to attain a given level of proficiency in a particular academic or vocational track due to the necessary longer enrollment because of lower English proficiency. If true, an implication is that, given the estimated returns from additional training, Hispanics as a group would benefit from longer JC spells, so that retention efforts targeted at them may be beneficial. Retention efforts may also be beneficial for whites and blacks, as it is

clear from their estimated DRFs that most participants do not stay in the program long enough to maximize their benefits, assuming that a higher number of participants staying in JC longer does not change the estimated DRF pattern.

VI. Conclusion

This study is one of the first in the program evaluation literature to estimate the causal impact of the length of enrollment in training on an outcome. By employing recently developed methods for the estimation of dose-response functions (DRFs), we estimate the average effect of the length of enrollment in Job Corps (JC) training under the assumption that, conditional on observable characteristics, the length of enrollment is randomly assigned. In addition, we estimate the DRF of JC for three different racial and ethnic groups (whites, blacks, and Hispanics), finding important differences among them.

The results in this paper show that estimation of the causal effects of the length of enrollment on a training program under the selection-on-observables assumption is feasible when a sufficiently rich dataset that allows embracing this assumption is available. Furthermore, they show that differences in the estimated DRFs between Hispanics and whites and blacks can be important in explaining previous results based on average impacts from receipt of treatment pointing out a lack of positive effects of JC on Hispanics. In particular, our results suggest that Hispanics benefit from a longer enrollment time in JC relative to blacks and whites. Finally, our estimates show that the heterogeneity in lengths of enrollment is an important dimension to investigate when evaluating active labor market programs, which is typically missed by conventional program evaluation methods.

References

Behrman, Jere R., Cheng, Yingmei and Todd, Petra E. (2004), "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics*, 86(1), 108-132.

Bitler, Marianne P., Gelbach, Jonah B., and Hoynes, Hillary W. (2006), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review*, 96 (4), 988-1012.

Bitler, Marianne P., Gelbach, Jonah B., and Hoynes, Hillary W. (2007), "Distributional Impacts of the Self-Sufficiency Project." University of California-Davis Working Paper.

Bleakley, Hoyt and Chinn, Aimee (2004), "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics*, 86(2), 481-496.

Dehejia, Rajeev H. and Wahba, Sadek (2002), "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics*, 84(1), 151-161.

Flores-Lagunes, Alfonso, Gonzalez, Arturo and Neumann, Todd (2006), "Learning but Not Earning? The Value of Job Corps Training for Hispanic Youths." University of Arizona Working Paper (http://www.u.arizona.edu/~afl/JC-FGN-7-27-06.pdf).

Fryges, Helmut and Wagner, Joachim (2007), "Exports and Productivity Growth: First Evidence from a Continuous Treatment Approach." IZA Discussion Paper No. 2782.

Gerfin, Michael and Lechner, Michael (2002), "A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland." *Economic Journal*, 112(482), 854-893.

Gonzalez, Arturo (2000), "The Acquisition and Labor Market Value of Four English Skills: New Evidence from Nals." *Contemporary Economic Policy*, 18(3), 259-269.

Heckman, James J., LaLonde, Robert J. and Smith, Jeffrey A. (1999), "The Economics and Econometrics of Active Labor Market Programs." In Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, 3A. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 1865-2097.

Heckman, James J., Smith, Jeffrey A., and Clements, Nancy (1997), "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*, 64, 487-535.

Hirano, Keisuke and Imbens, Guido W. (2004), "The Propensity Score with Continuous Treatments." In Andrew Gelman and Xiao-Li Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. West Sussex: John Wiley and Sons, 73-84.

Imai, Kosuke and van Dijk, David A. (2004), "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association*, 99, 854-866.

Imbens, Guido W. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika*, 87(3), 706-710.

Imbens, Guido W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics*, 86(1), 4-29.

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

Larsson (2003), "Evaluation of Swedish Youth Labor Market Programs." *Journal of Human Resources*,, 38(4), 891-927.

Lechner, Michael (2001), "Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption." In Michael Lechner and Friedhelm Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*, Heidelberg; New York: Physica-Verlag, 43-58.

Lechner, Michael (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.

McManus, Walter, Gould, William and Welch, Finis (1983), "Earnings of Hispanic Men: The Role of English Language Proficiency." *Journal of Labor Economics*, 1(2), 101-130.

Rosenbaum, Paul R. and Rubin, Donald B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1), 41-55.

Schochet, Peter Z. (1998), *National Job Corps Study: Eligible Applicants' Perspectives on Job Corps Outreach and Admissions*. Mathematica Policy Research, Inc., Princeton, NJ.

Schochet, Peter Z. (2001), "National Job Coprs Study: Methodological Appendixes on the Impact Analysis." 8140-530. Mathematica Policy Research, Inc., Princeton, NJ.

Schochet, Peter Z., Burghardt, John and Glazerman, Steven (2001), *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*. 8140-530. Mathematica Policy Research, Inc., Princeton, NJ.

Smith, Jeffrey A. and Todd, Petra E. (2005), "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125(1-2), 305-353.

U.S. Department of Labor (2005), "Job Corps Fact Sheet." http://www.doleta.gov/Programs/factsht/jobcorps.cfm (December 24, 2006).

U.S. Department of Labor (2006), *Policy and Requirements Handbook*. Washington, DC.

Figure 1. Kernel Density Estimates of the Length of Enrollment in JC

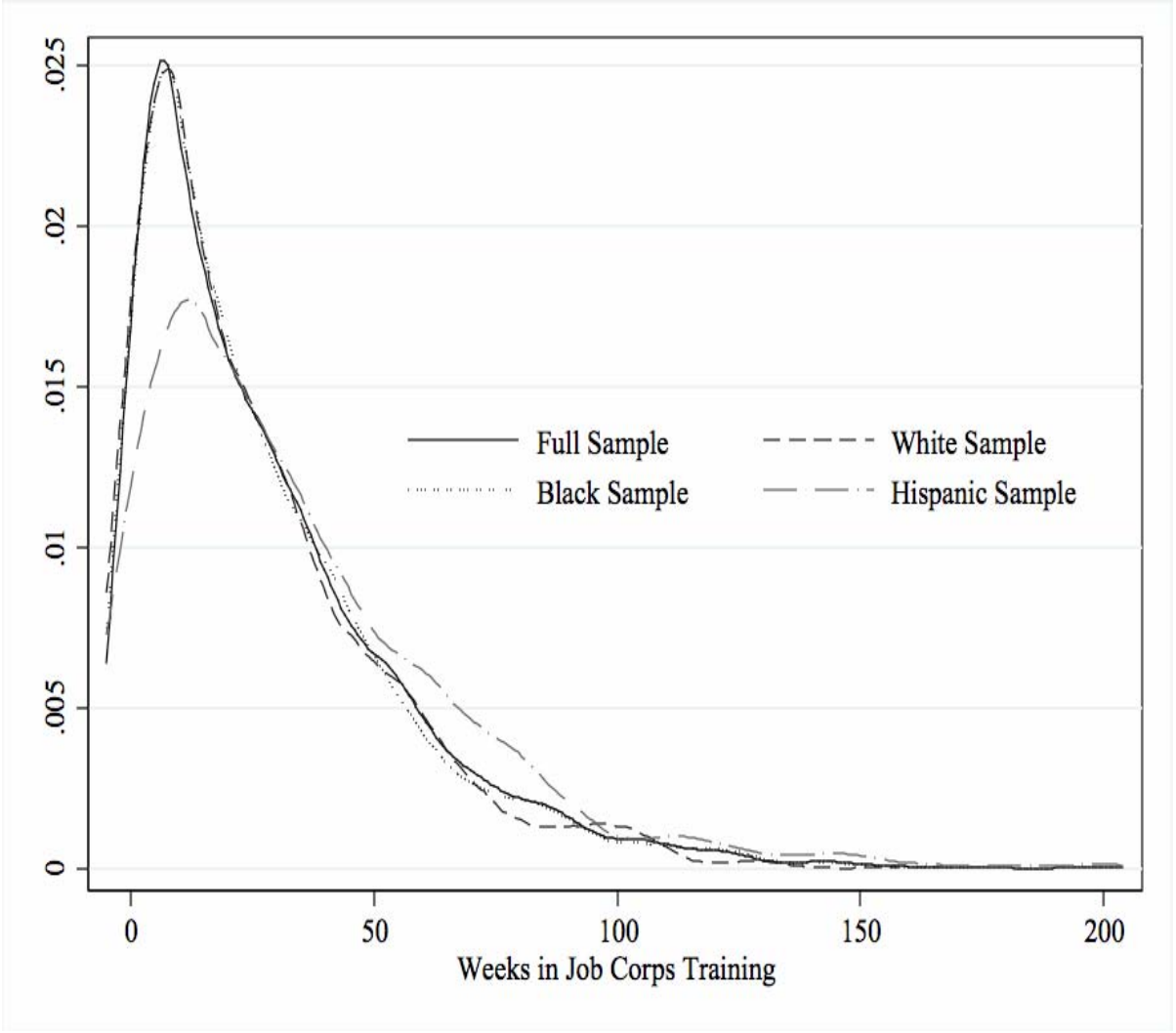


Table 1. Summary Statistics of Relevant Covariates and their Estimated Coefficients in the GPS Model

		FULL SAMPLE			WHITES*				BLACKS				HISPANICS			
	Variable	**Mean**	**Std. Dev.**	**GPS Coef.**	**Std. Err.**	**Mean**	**Std. Dev.**	**GPS Coef.**	**Std. Err.**	**Mean**	**Std. Dev.**	**GPS Coef.**	**Std. Err.**	**Mean**	**Std. Dev.**	**GPS Coef.**	**Std. Err.**
	Weeks in Training	28.15	27.32			25.99	25.11			27.26	26.53			34.09	31.71		
1	Black	0.54	0.50	0.26	0.10												
2	Hispanic	0.18	0.39	0.23	0.13												
3	Female	0.43	0.49	0.18	0.07	0.34	0.47	0.15	0.16	0.46	0.50	0.18	0.10	0.47	0.50	0.31	0.19
4	Has Child	0.17	0.38	-0.20	0.10	0.09	0.28	-0.23	0.29	0.21	0.41	-0.21	0.13	0.19	0.40	-0.18	0.27
5	Married	0.01	0.12	0.02	0.28	0.02	0.12	0.26	0.60	0.01	0.09	0.47	0.50	0.03	0.17	-0.22	0.50
6	Head of Household	0.10	0.30	-0.04	0.12	0.07	0.26	0.20	0.29	0.11	0.32	-0.33	0.16	0.09	0.28	0.36	0.33
7	Age	18.67	2.11	-2.87	3.41	18.67	2.04	-10.18	7.62	18.62	2.12	0.71	4.58	18.83	2.20	-5.35	8.95
8	Age-Squared/100	3.53	0.83	14.57	17.27	3.53	0.80	52.50	38.62	3.51	0.83	-3.11	23.21	3.59	0.87	23.93	45.18
9	Age-Cubed/1000	6.76	2.47	-2.42	2.89	6.74	2.37	-8.90	6.47	6.71	2.47	0.48	3.89	6.96	2.60	-3.49	7.53
10	Has High School Degree	0.18	0.38	-0.01	0.23	0.20	0.40	-0.24	0.41	0.17	0.37	0.18	0.33	0.18	0.38	-0.67	0.67
11	Has GED	0.05	0.21	0.06	0.25	0.08	0.27	0.26	0.40	0.02	0.15	-0.17	0.41	0.06	0.23	-0.05	0.68
12	Had Vocational Degree	0.02	0.13	0.12	0.25	0.02	0.14	0.34	0.50	0.02	0.12	0.46	0.37	0.03	0.16	-0.84	0.57
13	Attended Education or Training Program in Past Year	0.69	0.46	0.09	0.08	0.66	0.47	0.13	0.18	0.73	0.45	0.12	0.12	0.63	0.48	-0.03	0.19
14	Highest Grade Completed	10.03	1.51	0.02	0.16	10.04	1.51	-0.53	0.52	10.03	1.49	0.17	0.20	9.99	1.57	-0.28	0.36
15	Highest Grade Completed-Squared/100	1.03	0.30	-0.06	0.88	1.03	0.30	0.03	0.03	1.03	0.29	-1.12	1.14	1.02	0.31	1.90	1.96
16	Speaks English	0.89	0.31	-0.20	0.14	0.99	0.10	0.81	0.76	0.98	0.14	-0.46	0.33	0.48	0.50	-0.10	0.19
17	Good Health	0.40	0.49	0.00	0.07	0.44	0.50	-0.28	0.15	0.38	0.49	0.17	0.10	0.42	0.49	0.03	0.19
18	Fair Health	0.12	0.33	0.09	0.11	0.11	0.32	-0.11	0.23	0.12	0.33	0.18	0.14	0.13	0.34	-0.08	0.28
19	Poor Health	0.01	0.08	0.04	0.40	0.00	0.05	1.19	1.43	0.01	0.09	-0.23	0.52	0.01	0.11	0.33	0.80
20	Ever Smoked Cigarettes	0.51	0.50	-0.25	0.08	0.76	0.43	-0.34	0.18	0.40	0.49	-0.21	0.10	0.48	0.50	-0.29	0.22
21	Ever drank Alcohol	0.54	0.50	-0.15	0.08	0.69	0.46	0.00	0.17	0.46	0.50	-0.20	0.10	0.55	0.50	0.03	0.21
22	Ever Smoked Pot	0.31	0.46	0.04	0.08	0.36	0.48	0.08	0.16	0.29	0.45	0.07	0.11	0.29	0.45	-0.02	0.24
23	Weekly Earnings ($100)	1.15	4.26	0.01	0.01	1.32	1.23	0.00	0.00	0.97	1.13	0.01	0.05	1.04	1.14	0.12	0.09
24	Ever Worked	0.79	0.41	-0.06	0.09	0.88	0.33	-0.06	0.23	0.74	0.44	-0.07	0.12	0.78	0.42	0.04	0.25
25	Ever Arrested	0.23	0.42	-0.01	0.08	0.29	0.45	-0.18	0.16	0.21	0.41	0.11	0.11	0.20	0.40	-0.29	0.24
26	Lives in PMSA	0.32	0.47	0.27	0.13	0.17	0.37	0.11	0.26	0.36	0.48	0.32	0.20	0.43	0.50	0.93	0.39
27	Lives in MSA	0.47	0.50	-0.01	0.10	0.47	0.50	0.03	0.17	0.48	0.50	0.00	0.14	0.46	0.50	0.42	0.33
28	Lives with Parents	0.18	0.38	0.02	0.09	0.21	0.41	-0.09	0.17	0.13	0.34	0.16	0.13	0.27	0.45	0.24	0.21
29	Worried about Attending JC	0.36	0.48	0.10	0.07	0.39	0.49	0.19	0.15	0.35	0.48	0.04	0.10	0.36	0.48	-0.01	0.19
30	Knew What job wanted to Train For	0.85	0.36	0.04	0.09	0.87	0.34	-0.12	0.21	0.84	0.36	0.09	0.12	0.83	0.37	0.07	0.24
31	Knew What Center wished to Attend	0.52	0.50	-0.08	0.07	0.53	0.50	0.00	0.14	0.54	0.50	-0.08	0.09	0.44	0.50	-0.24	0.18
32	Expected to imporve Math Skills	0.70	0.46	-0.10	0.08	0.12	0.12	0.92	0.99	0.73	0.44	0.06	0.11	0.77	0.42	-0.18	0.22
33	Expected to Imporve Reading Skills	0.54	0.50	0.07	0.07	0.08	0.11	2.39	1.04	0.56	0.50	-0.02	0.10	0.64	0.48	0.24	0.21
34	Expected to Imporve ability to get along	0.62	0.49	0.00	0.08	0.12	0.11	1.71	1.01	0.63	0.48	0.01	0.11	0.65	0.48	0.02	0.23
35	Expected to imporve Self Control	0.60	0.49	0.11	0.08	0.12	0.12	2.06	1.00	0.58	0.49	0.09	0.11	0.61	0.49	-0.08	0.22
36	Expected to imporce Self Esteem	0.59	0.49	0.10	0.08	0.11	0.11	2.32	1.03	0.59	0.49	0.03	0.11	0.59	0.49	0.37	0.23
37	Expected to get Training for Specific Job	0.96	0.19	0.22	0.17	0.26	0.18	1.94	0.89	0.95	0.21	0.14	0.21	0.97	0.18	0.83	0.48
38	Expected to Find Friends	0.71	0.45	-0.08	0.08	0.17	0.13	1.65	0.95	0.70	0.46	-0.07	0.10	0.73	0.45	-0.19	0.22
39	Hear about JC from Parents	0.11	0.31	0.09	0.11	0.11	0.31	-0.18	0.23	0.12	0.32	0.25	0.14	0.09	0.29	-0.31	0.33
40	Knew someone who Attended JC	0.69	0.46	0.24	0.07	0.55	0.50	0.13	0.15	0.78	0.41	0.27	0.11	0.61	0.49	0.27	0.19
41	Joined JC to Achive Career Goal	0.99	0.08	0.28	0.44	0.99	0.07	dropped		0.99	0.08	0.90	0.55	0.99	0.08	1.33	1.46
42	Joined JC to get Job Training	0.99	0.11	0.23	0.33	0.99	0.10	6.03	3.07	0.99	0.11	-0.14	0.42	0.99	0.10	-0.19	1.09
43	Joined JC to get HS or GED	0.75	0.44	0.07	0.20	0.67	0.47	3.18	2.40	0.78	0.41	0.05	0.30	0.74	0.44	-0.16	0.58
44	Joined JC to Find Work	0.91	0.28	-0.17	0.12	0.89	0.31	1.12	2.08	0.92	0.28	-0.11	0.16	0.93	0.26	-0.16	0.35
45	Joined JC to get Away from Community	0.64	0.48	-0.11	0.07	0.48	0.50	2.52	1.87	0.73	0.45	-0.12	0.11	0.61	0.49	-0.04	0.20
46	Joined JC to get away from Home	0.58	0.49	-0.03	0.07	0.51	0.50	3.21	1.84	0.62	0.49	-0.04	0.10	0.59	0.49	0.09	0.19
47	Joined JC for Other Reason	0.73	0.45	0.12	0.08	0.74	0.44	4.29	1.86	0.72	0.45	0.13	0.10	0.72	0.45	0.10	0.21
48	Designated for Non-Residential Program	0.15	0.35	-0.14	0.12	0.07	0.26	-0.19	0.30	0.17	0.38	0.07	0.16	0.18	0.39	-0.40	0.28
	Constant			19.45	22.24			63.75	49.75			-4.32	29.83			39.38	58.74
	Weekly Earnings at 48-month($100)	2.19	2.20			2.71	2.35			1.94	2.10			2.10	2.11		
	R^2 (GPS model)			9.41				25.62				11.51				24.76	
	F-stat p-value (demographic variables: 1-9)			0.07				0.74				0.07				0.41	
	F-state p-value (education variables: 10-16)			0.80				0.71				0.48				0.74	
	F-state p-value (health variables: 17-22)			0.00				0.18				0.03				0.86	
	F-state p-value (economic variables: 23-28)			0.25				0.91				0.35				0.10	
	F-stat p-value (expection & motivation variables: 29-48)			0.05				0.20				0.66				0.56	
	F-stat p-value (State variables)			0.72				0.48				0.25				0.95	
	F-stat p-value (JC center variables)			0.00				0.01				0.06				0.80	
	Observations			3406				947				1837				622	

Note: The GPS specification for whites is slightly different from the others: the variable "Joined JC to Achieve Carrer Goal" is not included, and the "Expected to..." and "Joined JC..." variables are transformed into their percent of the total number of expectations and reasons to join Job Corps.

Figure 2.1 Standard Normal Quantile Plots for the Full Sample

(a) No GPS (b) With GPS

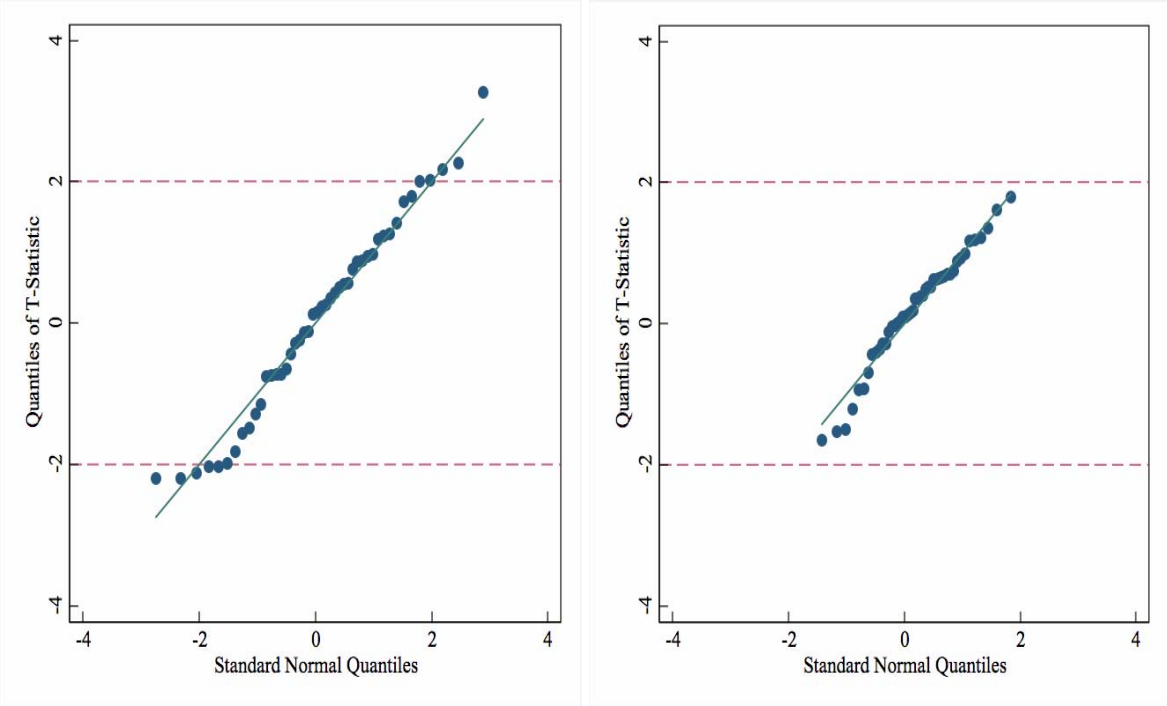


Figure 2.2 Standard Normal Quantile Plots for Whites

(a) No GPS (b) With GPS

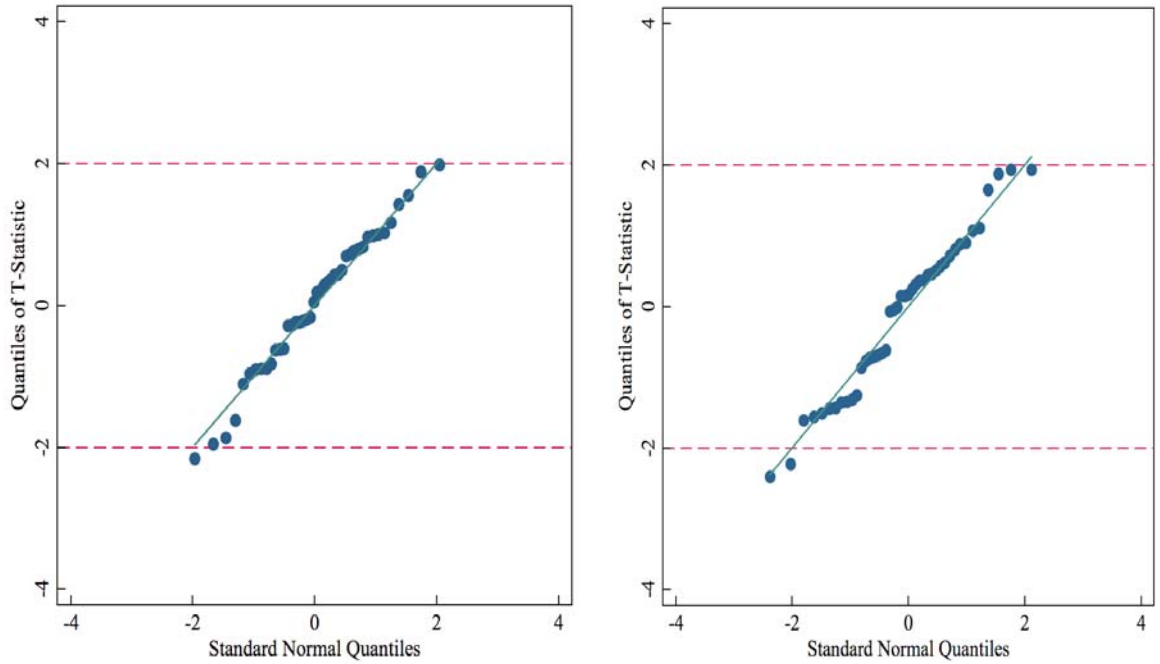


Figure 2.3 Standard Normal Quantile Plots for Blacks

(a) No GPS (b) With GPS

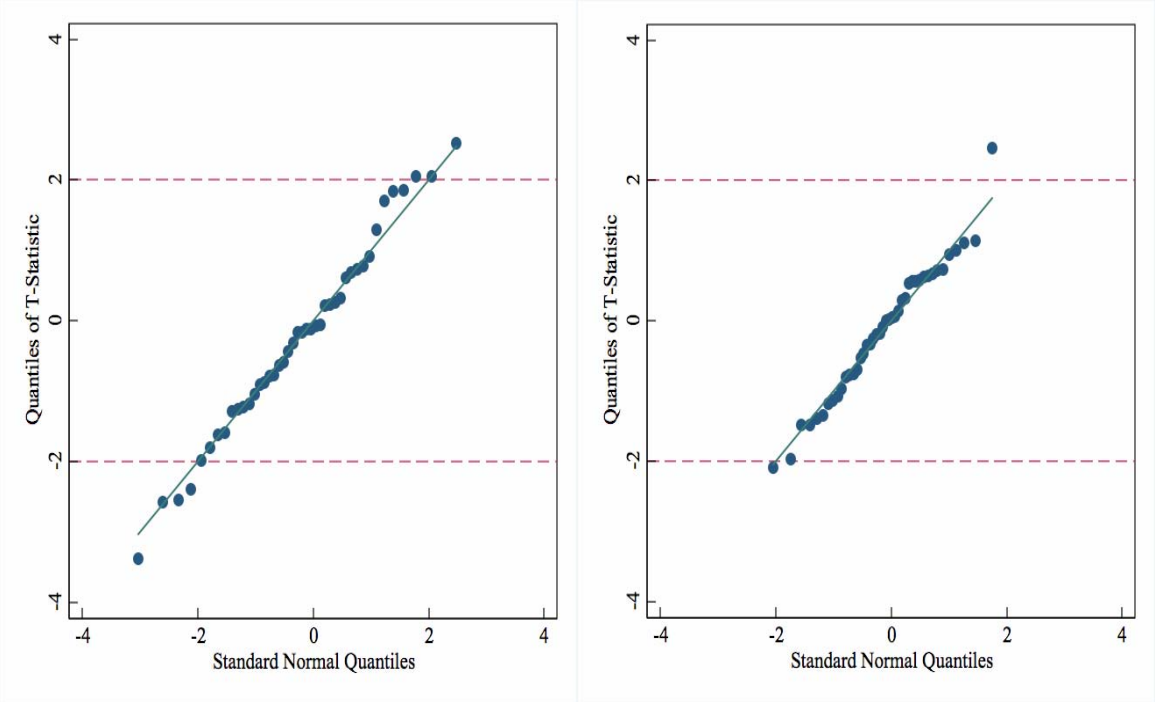


Figure 2.4 Standard Normal Quantile Plots for Hispanics

(a) No GPS (b) With GPS

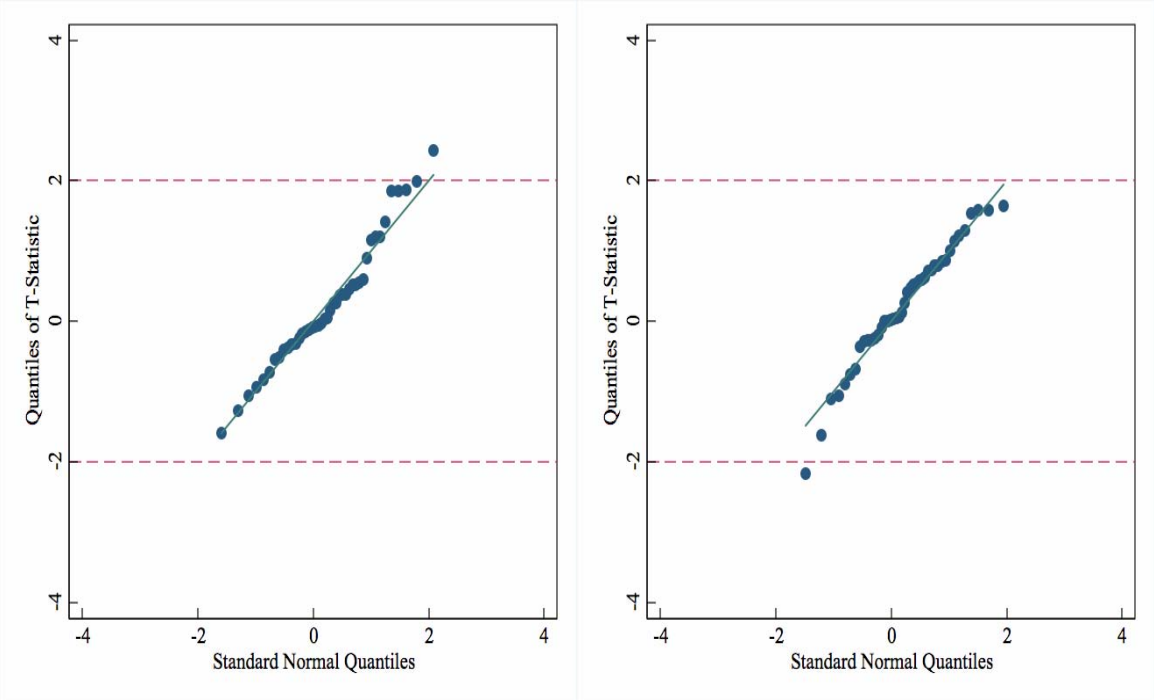
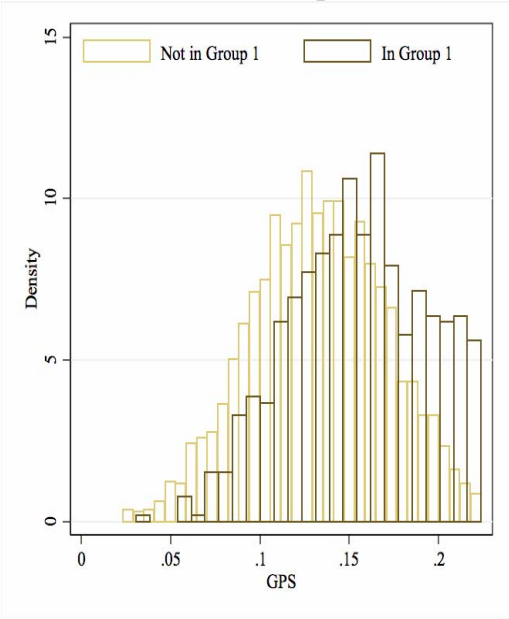
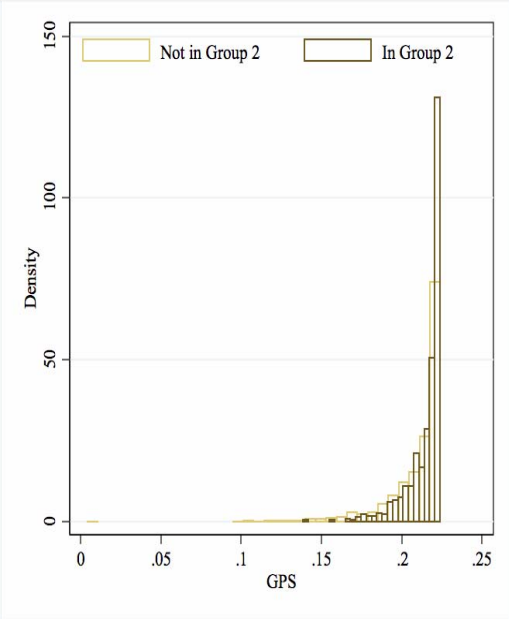


Figure 3. GPS Support Condition for Full Sample

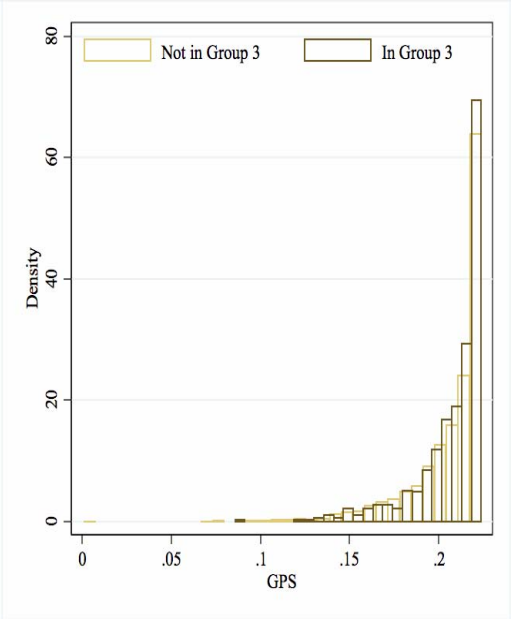
(a) Group 1



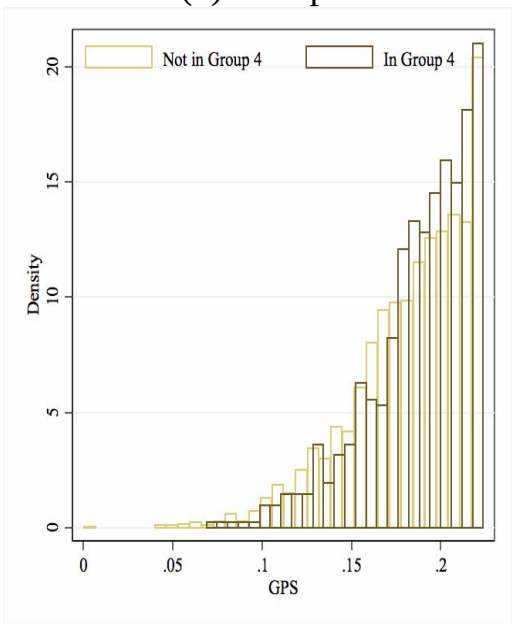
(b) Group 2



(c) Group 3



(d) Group 4



(e) Group 5

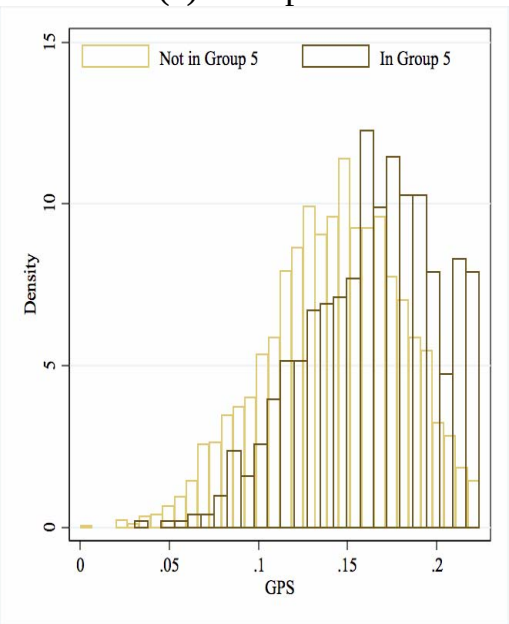


Table 2. Estimated Parameters of the Conditional Distribution of Earnings given Weeks Spent in Training and the GPS

Variable	FULL **Estimates (SE)**	WHITE **Estimates (SE)**	BLACK **Estimates (SE)**	HISPANIC **Estimates (SE)**
Weeks in Training	0.57	1.82	0.33	0.14
	(0.74)	(1.28)	(0.93)	(1.21)
(Weeks in Training^2)/100	-0.65 *	-1.38	-0.31	-0.98 *
	(0.35)	(0.85)	(0.51)	(0.50)
GPS	-7.19	229.32	111.31	-68.87
	(286.38)	(513.59)	(356.59)	(570.78)
GPS^2	121.83	-378.92	-273.10	-13.26
	(1152.42)	(1996.24)	(1338.15)	(2079.97)
GPS x Hours Spent in training	2.32	-2.82	1.19	8.18 *
	(3.28)	(5.23)	(3.88)	(4.87)
Female	-66.39 ***	-98.45 ***	-40.95 ***	-95.99 ***
	(7.52)	15.93	(9.85)	16.57
Black	-69.50 ***			
	(8.66)			
Hispanic	-55.50 ***			
	(11.27)			
Constant	274.08 ***	261.25 ***	192.40 ***	228.90 ***
	16.37	(30.38)	(21.66)	(35.83)
F-statistic of GPS & covariates coefficients	26.75	9.97	4.47	7.50
(p-value)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	3406	947	1837	622

Figure 4. Dose Response Function
Full Sample

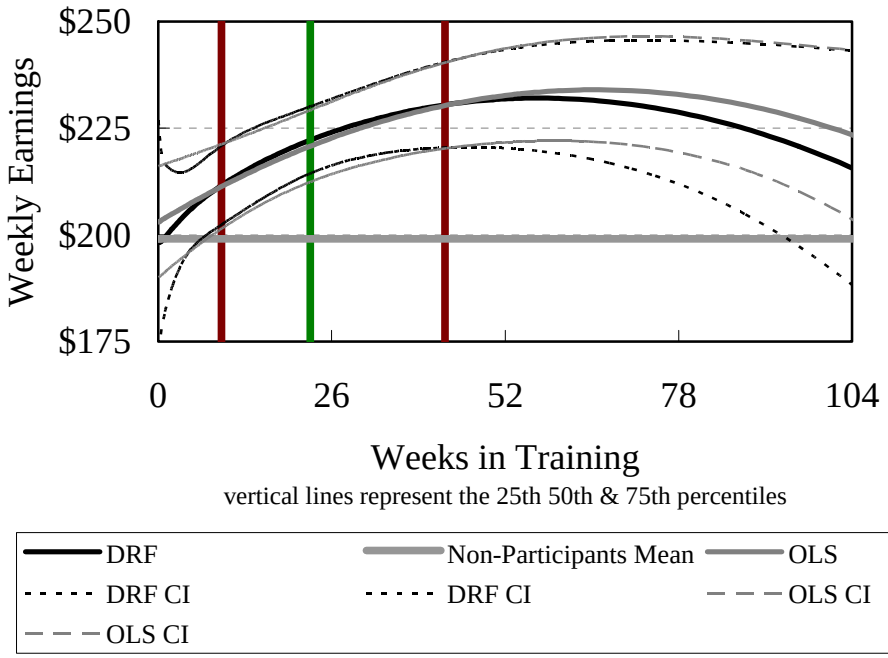


Figure 5. Dose Response Function
White Sample

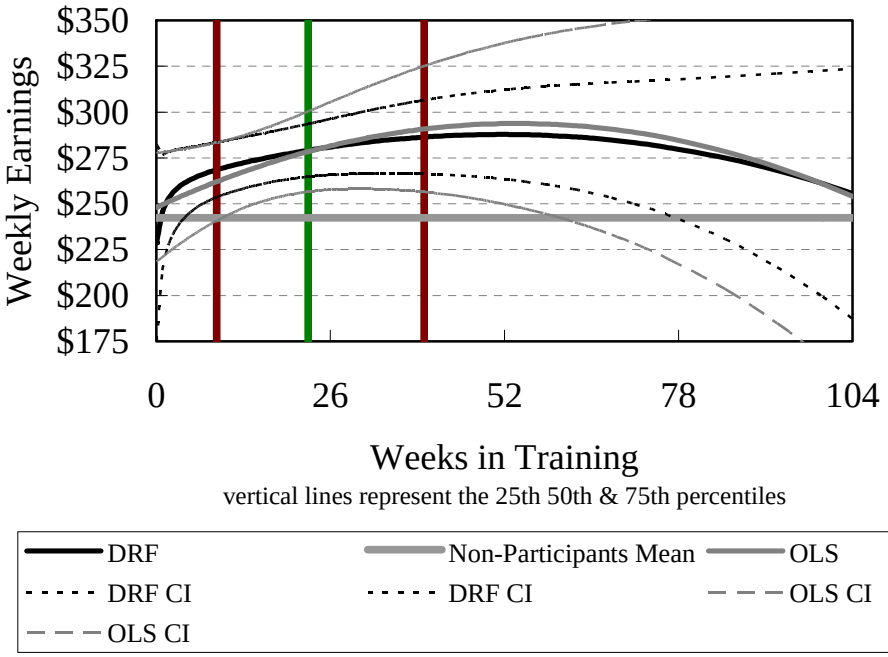


Figure 6. Dose Response Function
Black Sample

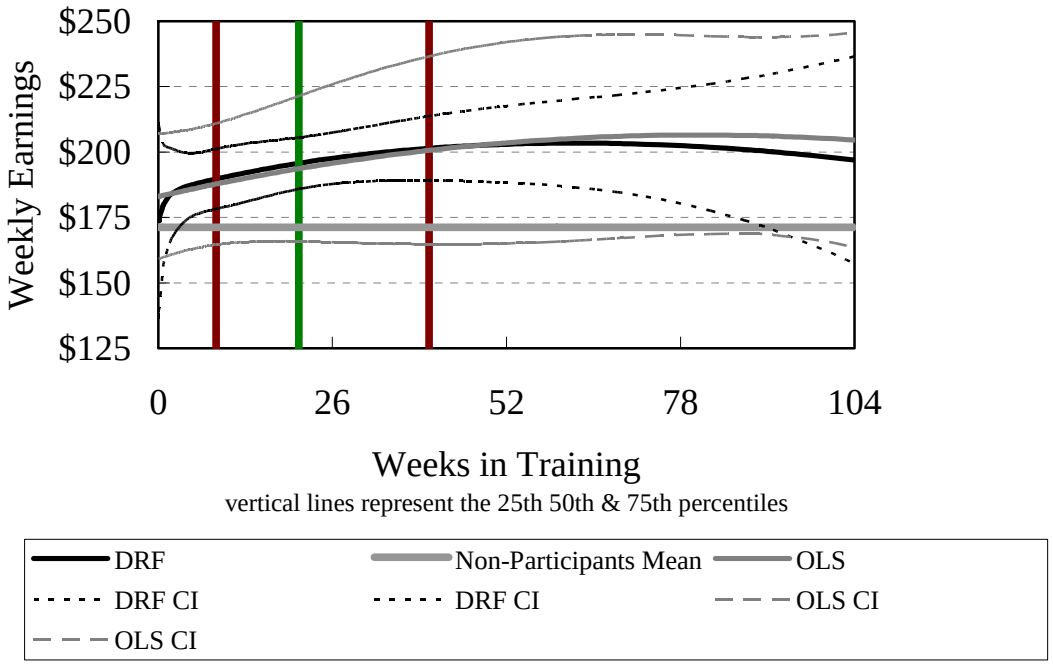


Figure 7. Dose Response Function
Hispanic Sample

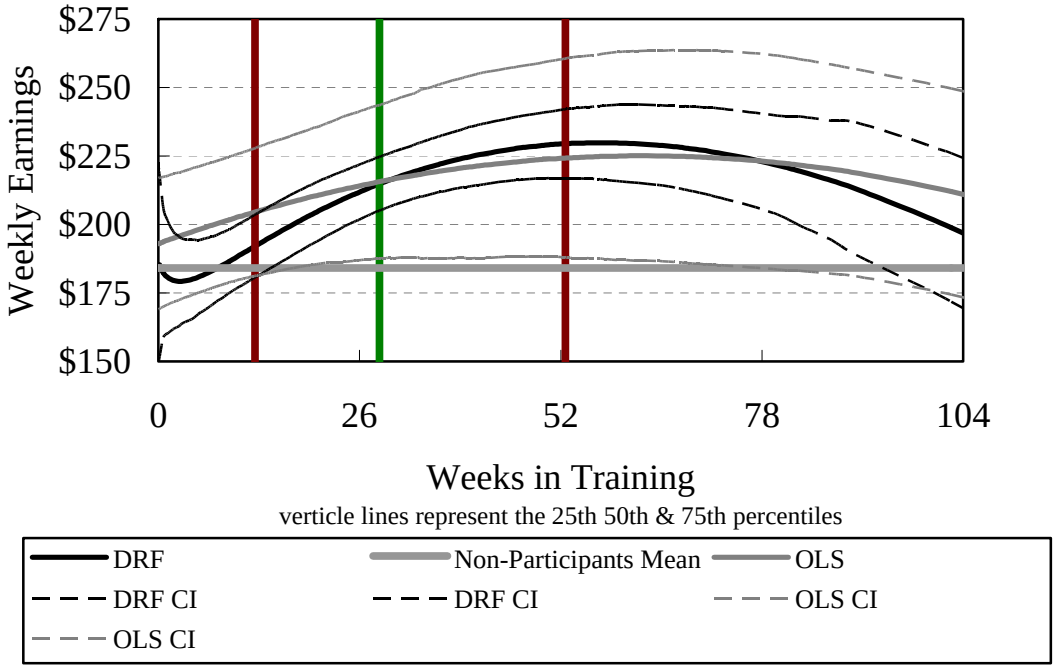


Table 3. Estimated DRF and DRF-DIFF for select percentiles of the Distribution of Treatment

	Full	White	Black	Hispanic
Covariate-Adjusted Non-Participants Mean	$199.11	$242.41	$171.23	$184.15
		25th Percentile		
DRF	$209.42	$265.96	$188.82	$187.89
DRF-DIFF	$0.94	$1.06	$0.53	$1.62
Weeks of Training	7.5	6.9	7.1	10.0
		50th Percentile		
DRF	$220.84	$276.48	$195.19	$212.13
DRF-DIFF	$0.77	$0.73	$0.46	$1.32
Weeks of Training	20.6	18.5	19.8	26.2
		75th Percentile		
DRF	$229.74	$285.70	$201.24	$228.93
DRF-DIFF	$0.45	$0.29	$0.30	$0.63
Weeks of Training	40.1	37.5	39.5	50.0

Full sample empirical distribution of treatment levels

	Full	White	Black	Hispanic
		25th Percentile		
DRF	-	$266.79	$189.22	$183.90
DRF-DIFF	-	$1.04	$0.53	$1.57
Percentile of own distribution	-	27th	26th	19th
		50th Percentile		
DRF	-	$277.92	$195.71	$205.32
DRF-DIFF	-	$0.68	$0.45	$1.46
Percentile of own distribution	-	53rd	52nd	42nd
		75th Percentile		
DRF	-	$286.61	$201.35	$224.58
DRF-DIFF	-	$0.21	$0.29	$0.92
Percentile of own distribution	-	78th	76th	68th