Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps[1]

Carlos A. Flores
Department of Economics
University of Miami
P.O. Box 248126, Coral Gables, FL 33124-6550
305-2842615
caflores@miami.edu

Alfonso Flores-Lagunes
Food and Resource Economics Department and Department of Economics
University of Florida and IZA, Bonn, Germany
P.O. Box 110240, Gainesville, FL 32611-0240
520-245-9182
alfonsofl@ufl.edu

Arturo Gonzalez
Office of the Comptroller of the Currency and IZA, Bonn, Germany
250 E Street S.W. (Mail Stop: 3-2), Washington, DC 20219
202-874-0978
Arturo.Gonzalez@occ.treas.gov

Todd C. Neumann
School of Social Sciences, Humanities & Arts
University of California, Merced
5200 North Lake Road, 349 Classroom Building, Merced, California 95343
209-228-4020
tneumann@ucmerced.edu

Revised version: February 2010

[1] Detailed comments from an anonymous referee and an editor greatly improved the paper and are gratefully acknowledged. We also thank the comments we received from Chung Choe, Oscar Mitnik, Ron Oaxaca, and William J. Smith. Kalena Cortes and David Green provided useful discussions on an earlier version of the paper at the ASSA Meetings in January 2007 and the Society of Labor Economists (SOLE) Meetings in May 2007, respectively. Useful comments were also provided at the Princeton Junior Faculty Presentation Series. Flores-Lagunes gratefully acknowledges financial support from the Industrial Relations Section at Princeton University. The views herein are those of the authors and do not necessarily represent the views of the Office of the Comptroller of the Currency.

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

Abstract

Length of exposure to instruction in a training program is important in determining the labor market outcomes of participants. Employing methods to estimate the causal effects from continuous treatments, we provide insights regarding the effects of different lengths of exposure to academic and vocational instruction in Job Corps (JC)—America's largest and most comprehensive job training program for disadvantaged youth. We semiparametrically estimate average causal effects (on the treated) of different lengths of exposure using the "generalized propensity score" under the assumption that selection into different lengths is based on a rich set of observed covariates and time-invariant factors. We find that the estimated effects on future earnings are increasing in the length of exposure—which is consistent with human capital accumulation—and that the marginal effects of additional instruction are decreasing with length of exposure. We also document differences in the estimated effects across demographic groups, which are particularly large between males and females. Finally, our results suggest an important "lock-in" effect in JC training.

JEL Classification: I38, C21, J24

Keywords: Training Programs, Continuous Treatments, Generalized Propensity Score, Dose-Response Function.

I. Introduction

An important feature of publicly sponsored job training programs is that participants enroll for different lengths of time, likely acquiring different levels of human capital. This is the case in the Job Corps (JC) program, America's largest and most comprehensive job training program for disadvantaged youth, where participants are exposed to different levels of academic and vocational (AV) instruction. If participants accumulate different amounts of human capital during their time in JC, the expectation is that their labor market outcomes will differ. For this reason, we consider it important to go beyond the estimation of the causal effects of JC employing a binary indicator of participation, and instead estimate the causal effects of receiving different "dosages" of AV instruction on future earnings. Estimating a "dose-response function" (DRF) provides more information regarding the effectiveness of the program by uncovering heterogeneities in the effects of AV instruction in JC along the different lengths of exposure. We accomplish this in the present paper by using semiparametric estimators of the DRF based on recent results for analyzing continuous treatments, and by highlighting the type of insights that can be learned about the effects of training when considering its continuous nature.

Our treatment variable is the number of *actual* hours of AV instruction received while enrolled in JC, which we interpret as a measure of human capital accumulation. This variable provides a nearly continuous measure of human capital accumulation since it only counts the actual time devoted to AV instruction while enrolled in JC. Consequently, we interpret the estimated derivatives of the DRF as the average marginal returns to additional human capital accumulation as measured by exposure to AV instruction in JC. These effects are important for policy since, for example, if the returns are positive, JC staff may encourage or provide incentives to participants to undertake additional AV instruction.

Our key identifying assumption in estimating the DRF is that selection into levels of the treatment is random conditional on a rich set of observable covariates (unconfoundedness). In addition, our preferred specification allows for time-invariant unobservable factors to influence selection by employing a difference-in-difference (DD) version of our estimators. A common approach for estimating causal effects under unconfoundedness in a binary-treatment setting is the use of the propensity score (Rosenbaum and Rubin, 1983). Hirano and Imbens (2004) introduced the concept of "generalized propensity score" (GPS) and extended the results in Rosenbaum and Rubin (1983) to the case of a continuous treatment. We use the GPS to estimate average treatment effects (on the treated) of different lengths of exposure to AV instruction in JC on earnings, thereby constructing a DRF. More specifically, we employ two-step semiparametric estimators of the DRF. The first step involves a parametric but flexible estimation of the GPS based on generalized linear models. The second step involves estimating the DRF using the estimated GPS by employing either a nonparametric partial mean or a nonparametric inverse-weighting estimator. In analogy to the advantages of the use of the propensity score in the binary-treatment case, we use the estimated GPS (i) to identify individuals for whom it is difficult to construct counterfactual outcomes by imposing an overlap condition, and (ii) to control for observed covariates in a more flexible way relative to OLS.

The data on JC we employ comes from the National Job Corps Study (NJCS), a randomized social experiment undertaken in the mid-to-late 1990s to evaluate the effectiveness of JC. This data set has several advantages. First, it contains a detailed measure of the hours of exposure to AV instruction. Second, it contains very detailed pre-treatment information about program participants, such as expectations and motivations for applying to JC, information about the specific training center attended, and the zip-code of residence that allows matching participants to measures of their local labor market unemployment rates (LURs). Our unconfoundedness assumption hinges on the richness of our pre-treatment covariates plus the LURs. Finally, these data provide an opportunity to estimate a DRF for different demographic groups (males, females, blacks, whites, and Hispanics) to learn if there are differences in the returns to length of exposure across them.[2]

A disadvantage of the NJCS data is that it does not contain detailed enough information on the participants' experiences while they are enrolled in Job Corps. Hence, despite the richness of our data, there is the possibility that our key identifying assumption (unconfoundedness) is not satisfied because of the existence of dynamic confounders, that is, variables whose realization occurs after starting training and that are related to both the individual's length of exposure to AV instruction and her outcome. There are at least two other general alternative approaches to our identification strategy that explicitly control for dynamic confounding (Abbring and Heckman, 2007). One is "dynamic matching" as in Robins (1997), which employs a dynamic version of unconfoundedness that allows treatment assignment at a given point in time to depend

[2] In our data, the self-reported race/ethnicity is formally non-Hispanic white, non-Hispanic black, and Hispanic. For simplicity, we refer to them as whites, blacks, and Hispanics.

on past observable outcomes and covariates.[3] The other approach (Heckman and Navarro, 2007), which is based on discrete-time dynamic discrete-choice models, allows dynamic selection on unobservables by explicitly modeling dynamics and information arrival to the decision-maker. Unfortunately, the lack of data on intermediate outcomes and covariates prevents us from pursuing these alternative approaches. However, the institutional characteristics of the JC program, the richness of our data, and indirect assessments of our identifying assumptions lead us to believe that dynamic confounding is unlikely to be driving our results.

Our results indicate that the estimation of a DRF is informative about the effectiveness of the JC training program over the different lengths of AV exposure. In particular, we find that the DRFs for two measures of average weekly earnings after leaving JC increase with length of exposure to AV instruction, and that the marginal effects of additional instruction decrease with length of exposure. These results imply average rates of return that are consistent with estimated returns to an additional year of formal schooling (e.g., Card, 1999), and also with previous estimates of the effect of participation in JC (e.g., Schochet, Burghardt, and Glazerman, 2001; Lee, 2009). We also document important differences in the estimated DRF across the demographic groups considered, especially between males and females. Additionally, our results suggest an important "lock-in" effect of JC, as the estimated effects increase notably when employing an outcome that holds constant the amount of time between training termination and when the earnings are measured.

This study contributes to different strands of literature. First, we advance the literature on the evaluation of the JC program (Schochet, Burghardt, and Glazerman, 2001; Schochet, Burghardt, and McConnell, 2008; Schochet and Burghardt, 2008; Flores-Lagunes, Gonzalez, and Neumann, 2009) and the evaluation of training programs in the United States in general (Heckman, LaLonde, and Smith, 1999; Angrist and Krueger, 1999). Second, we join a number of authors that have considered the estimation of effects from multi-valued or continuous labor market interventions or treatments under unconfoundedness. However, contrary to recent parametric applications of the GPS in a continuous treatment setting (Hirano and Imbens, 2004; Kluve et al., 2007; Mitnik, 2008; Bia and Mattei, 2008), our estimators are semiparametric in the sense that we control for the estimated GPS nonparametrically in the estimation of the DRF.

[3] Our unconfoundedness assumption can be thought of as a special case of this dynamic unconfoundedness assumption in which the conditioning is based on pre-treatment covariates and the local unemployment rates at time of exit from the program.

Finally, in analyzing different demographic groups, we contribute to the literature on how the effectiveness of active labor market programs varies across them (Heckman and Smith, 1999; Abadie, Angrist, and Imbens, 2002; Flores-Lagunes, Gonzalez, and Neumann, 2009).

The rest of this paper is organized as follows. Section II discusses the JC program, the NJCS data, and how the variability in the length of exposure to AV instruction in JC arises. Section III presents the identification results employed and our semiparametric estimators of the DRF. Section IV presents the main estimation results for the different samples and discusses their implications. In this section we also perform some exercises aimed at indirectly assessing our identifying assumptions, and discuss implications for our results in the event dynamic confounding is present. Finally, Section V concludes.

II. Job Corps, the National Job Corps Study, and the Data Employed

A. The Job Corps Program and The National Job Corps Study

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 to 24. JC provides academic, vocational, and social skills training at over 120 centers throughout the country, where most students reside during training.[4] In addition to academic and vocational training, JC also provides health services and a stipend during program enrollment (Schochet, Burghardt, and Glazerman, 2001). Individuals are eligible for JC 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, and who is a high school dropout (75%) reading at a seventh grade level (U.S. Department of Labor, 2005). The motivation for applying to JC varies with age. In particular, younger applicants are more interested in completing high school or obtaining a GED. Older applicants are more interested in vocational training. Nevertheless , the majority of JC participants (77%) take both academic and vocational

[4] Academic training consists mainly of reading and math classes at different levels, GED courses, courses leading to completion of high school, and other required courses such as “world of work” (general job-related skills such as how to look, find, and keep a job) and health education. Vocational training is offered in a wide variety of trades that vary by JC center, such as clerical work, automotive repair, building maintenance, and carpentry.

classes (Schochet, Burghardt, and Glazerman, 2001). Above all, they see JC training as a way of finding employment since the majority have never held a full-time job (Schochet, 1998).

The data used for this paper come from the National Job Corps Study (NJCS). The sampling frame for the NJCS consisted of first-time JC applicants in the 48 contiguous states and the District of Columbia. All pre-screened eligible applications (80,833) were randomly assigned to control, treatment, or program non-research groups between November 1994 and February 1996. Approximately 7% of the eligible applicants were assigned to the control group (N = 5,977) while 12% were 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. Control group members were barred from enrolling in JC for a period of three years. The control and treatment groups were tracked with a baseline interview immediately after randomization and then at 12, 30, and 48 months after randomization. Flores-Lagunes, Gonzalez, and Neumann (2009) discuss other features of the NJCS.[5]

The original NJCS estimates imply an overall (full sample) average gain of $22.1 weekly earnings at the 48-month after randomization, although it is not uniform across the 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.[6] Schochet, Burghardt, and Glazerman (2001) report that these differential impacts cannot be explained by individual and institutional variables. Flores-Lagunes, Gonzalez, and Neumann (2009) present evidence indicating that one explanation for these disparate effects lies with the different levels of local labor market unemployment rates (LUR) each of the groups faced during the time of the study, along with a differential LUR impact on their earnings. The analysis below sheds additional light on the differential returns to JC by analyzing differences in exposure to AV instruction and the corresponding DRFs across these groups.

---

[5] The NJCS experienced some non-response in the 48-month survey, for which the effective response rate was 79.9%. While Schochet (2001) reports that the pre-treatment characteristics of non-respondents are similar to those of respondents, Schochet, Burghardt, and McConnell (2008) find—using administrative data unavailable to us—that survey non-respondents have smaller impacts, suggesting that the original NJCS results may be biased upwards.

[6] These are difference-in-means estimators modified to account for non-compliance (Schochet, 2001). These estimators identify the average treatment effect for those individuals that comply with their treatment assignment (e.g., Imbens and Angrist, 1994).

B. Institutional Details of Job Corps and Sources of Variability in Exposure Lengths

In this subsection, we describe relevant institutional details of the JC program that allow an understanding of the sources of variability in the length of exposure to AV instruction and the selection mechanisms into those lengths. Our strategy to estimate a causal DRF relies on successfully controlling for all factors that simultaneously affect exposure lengths and future earnings (i.e., confounders). As further discussed in the following section, we assume that all possible confounders are accounted for by our observed pre-treatment covariates, the LURs, and time-invariant factors (when using the DD version of our estimators). In this setting, a potential threat to our identifying assumptions is the existence of dynamic confounders, such as those related to the participant's performance while in JC. However, as discussed in this subsection, we believe this type of dynamic confounding is not a serious threat in our application given the institutional characteristics of JC and the richness of our data. We close with a discussion of the sources of exogenous variation used to identify our parameters of interest.

An important characteristic of JC is that it is a self-paced program. For instance, all courses in JC are open-entry and open-exit. Each participant develops her own training plan with the help of a JC counselor based on her own needs (e.g., required academic training), preferences (e.g., desired vocational trade), and characteristics of the JC center attended (e.g., availability of trades and open slots).[7] As a result, the overall variability in hours of exposure to AV instruction in JC arises from a combination of individual decisions and institutional factors of the JC program. Fortunately, our data is rich enough to allow controlling for a large number of factors that are plausibly simultaneously related to the length of exposure to AV instruction and future earnings. These factors include individual-specific traits such as demographic characteristics, motivations and expectations upon enrolling, and two different local labor market unemployment rates, among others; as well as institutional features such as the specific center each individual attends.[8] We discuss in greater detail the set of covariates we control for in the next subsection.

The institutional features of JC weaken the potential role of dynamic confounders such as the individual's performance within JC. The only general and obligatory testing in JC is the test of adult basic education (TABE), which serves for tracking purposes. It takes place at the

[7] Upon arrival to the JC center participants are given some time, roughly between two to eight *calendar* weeks, to get acquainted with center life. It is during this time that the training plan is developed.

[8] We control for Job Corps center "fixed effects", which should also help account for differences in local labor market conditions across center locations in addition to the LURs.

beginning of enrollment in JC and is intended to help determine the extent of academic training needed by the individual.[9] Aside from this initial testing, there are only built-in diagnostic tests within the academic instruction that justify a student "moving" through the academic classes. These tests are not "formal" —in that they can be taken anytime and as many times as needed—and are mainly offered through the Computer Managed Instructional (CMI) system.[10] In terms of vocational training, "formal" testing is even less prevalent, with trades only having "proficiency levels" that allow determining the mastery level achieved by the student in a particular trade.

In addition to the lack of strong performance signals in AV classes, individuals take academic and vocational training jointly in the majority of centers. After determining the academic training needed, students choose their vocational training and begin taking "…a balanced schedule of one-half academic coursework and one-half vocational coursework" (Johnson et al., 1999) either as split days or split weeks. As students complete the minimum academic requirements, they move on towards taking mostly vocational training. Given this system in which students are simultaneously enrolled in both academic and vocational training with no comprehensive formal testing while in the program, it is unlikely that students will terminate or continue their training duration based on small signals such as the CMI diagnostic tests. Hence, we believe that the type of testing within JC is likely to play a small role as a dynamic confounder in our application, especially when compared to, for example, formal testing in the context of selecting a level of regular schooling (e.g., Abbring and Heckman, 2007).[11]

JC also monitors the progress of students through progress/performance evaluation panels (P/PEPs) that "…assess student performance in all major areas and guide the student in an ongoing self-assessment and goal-setting process" (Johnson et al., 1999). These evaluations take

---

[9] Unfortunately, we do not have information on individual TABE scores and hence cannot explicitly control for it. However, we do not think this is a serious confounder since it occurs at the beginning of enrollment in JC—likely before participants start accumulating AV instruction—and we do control for a rich set of pre-treatment covariates. Section IV.C below provides suggestive evidence to support this view.

[10] "The CMI assists teachers and students… by providing initial placement, lesson assignment, lesson and test scoring, and individual student tracking" (Johnson et al., 1999).

[11] More generally, we believe the issue of dynamic confounding is less severe in our application than in the context of estimating the effects of regular schooling on earnings. In addition to the JC program being much more flexible than regular schooling, the time frame in which JC occurs is much shorter, so more dynamic factors are likely to determine the level of regular schooling. Finally, the potential role of intermediate outcomes related to the labor market (e.g., forgone earnings) is reduced for JC participants as they receive a stipend and typically live at a JC center.

place for the first time after 45 days of enrollment and about every 60 days thereafter. The outcomes of these P/PEPs are recommendations "…regarding the student's training (course or schedule changes), social training performance, incentive awards, bonuses, and readiness to exit the program" (Johnson et al., 1999). These panels seldom recommend abrupt enrollment termination, though.[12] Furthermore, the JC philosophy states that students are encouraged to carry on training at their own pace, and counselors are not expected to discriminate between students by their length of stay in JC (U.S. Department of Labor, 2006). Nevertheless, while the main objective of the P/PEPs is to keep students focused and motivated and to establish realistic goals, they may also impact the length of the individual's enrollment in JC through suggestions of courses to take and recommendations about readiness to exit the program. If the recommendations are correlated with future earnings (for instance, if they are based on the individual's previous performance in JC), then they are potential confounders. Therefore, we implicitly assume that, after conditioning on covariates, any remaining correlation between the P/PEP outcomes and our treatment variable is due to factors not correlated to future earnings, such as the specific composition of the P/PEP.[13] Although we have no way to directly test this assumption, we provide some indirect evidence in section IV.C.

In addition to the institutional features of JC as previously discussed, the richness of our data helps reduce the influence that potential dynamic confounders may have on our results. For instance, consider the potential role of the individual's performance while in JC as a dynamic confounder. To the extent that this performance is correlated to the variables we control for, the potential bias arising for not directly controlling for it is reduced or eliminated. We provide some suggestive evidence about this possibility in section IV.C.

Finally, it is important to note that we need to have variables affecting the length of exposure to AV instruction in JC but not future earnings, conditional on our covariates—i.e., an "exogenous" source of variation.[14] Based on the structure of JC and the nature of the data, we believe that the variation in our treatment variable after controlling for covariates is largely due

[12] A recommendation of terminating JC enrollment by a panel is rare: "In the previous year, most centers had terminated only a few students through this process. Some centers had not terminated any students for months…" (Johnson et al., 1999, pp. 159).

[13] The P/PEPs are generally composed of a JC counselor and (ideally) instructors who are familiar with the participant's experience. There is a fair amount of variability in their composition arising from a combination of the specific center policies, the availability of staff to perform these duties, and the training path undertaken by the participant.

[14] These variables can be thought of as "unobserved instruments" (e.g., Busso et al., 2009).

to time-varying characteristics of the JC centers.[15] For instance, if the desired trade is filled up at the time an individual starts AV training, she will be placed on a waiting list and JC staff may suggest alternative classes/trades while slots in the desired trade become open (Johnson et al., 1999). Other examples are changes in the staff composition, the particular composition of the individual's P/PEP, and aggregate trade preferences among JC cohorts. These factors are plausibly not correlated to individual future earnings, ruling them out as confounders.[16,17]

C. Summary Statistics of the Data Employed

The variables of interest for this study are taken mainly from the public-use baseline and 48-month surveys of the NJCS. We concentrate on those individuals who enrolled in JC to study the effect of the length of exposure to AV instruction on two outcomes: (1) weekly earnings 48 months after randomization took place and (2) weekly earnings 12 months after their exit from JC. The first outcome embodies lock-in costs (e.g., lost labor market experience) to the participant from a longer enrollment spell, while the second outcome puts participants on an equal footing in terms of the time they have been in the labor market after completing their enrollment spell. Our treatment variable measures *actual* hours spent in AV training, as opposed to *calendar time* spent in JC, which is important given the self-paced philosophy of JC. For simplicity, we re-scale our treatment variable to weeks by assuming a 40-hour workweek.

Our sample consists of 3,715 individuals who completed at least one week of AV training; reported being white, black, or Hispanic; and had information on the outcomes and covariates of interest.[18] The complete list of covariates used in the estimation of the GPS model in Section IV is presented in the Data Appendix. We broadly classify them into demographic, education, health, past employment, arrest, household, and location characteristics; pre-treatment

---

[15] Time-invariant characteristics of the JC centers are controlled for through the inclusion of JC center attended fixed effects. However, those fixed effects do not control for time-varying characteristics of the JC center attended arising, for instance, from individuals in the sample enrolling in JC at different points in time.

[16] Additional sources of independent variation in the length of exposure to AV instruction could come, for instance, from the number of siblings or relatives also enrolled in the same JC center, proximity of the participant's place of residence to the JC center, etc.

[17] To provide some suggestive evidence that the variation in our treatment variable after controlling for covariates still reflects HC accumulation, we looked at its relationship with another HC measure: number of different AV classes taken while in JC. Both variables are strongly positively correlated even after controlling for covariates. This exercise is documented in the Internet Appendix.

[18] Relative to the original NJCS research sample individuals (15,386): 5,825 never enroll in JC; 1,096 are not of the race/ethnicities we consider; 3,872 do not answer the 48-month survey; 407 do not report quarter 16 earnings; 62 have less than one week of AV instruction; and 409 have missing values on other covariates of interest.

expectations about (and motivations for) enrolling in JC; LUR measures; NJCS predictions of the individual enrollment duration; and geographical variables such as state of residence and the JC center attended.

Table 1 presents summary statistics of selected pre-treatment variables for the full sample and five demographic groups: whites (27%), blacks (54%), Hispanics (19%), males (56%), and females (44%). According to our length of exposure variable, JC is a time-intensive program, with the average participant receiving 30.4 weeks of AV instruction. Whites, blacks, and males receive similar levels of instruction (about 29 weeks), but notably less than Hispanics and females, who receive 36.4 and 31.3 weeks, respectively. For comparison, a typical high school student receives the equivalent of 1,080 hours (27 weeks) of instruction during a school year (Schochet, Burghardt, and Glazerman, 2001, pp. 65). The table presents some percentiles of the distribution of the treatment variable for the different samples, which show that the longer exposure by Hispanics and females holds across all percentiles, with the Hispanics' differences increasing over the distribution.

Some of the covariates presented in Table 1 warrant further explanation. We employ two LUR variables to account for the local labor market conditions that may influence a participant's decision to terminate JC enrollment. Both LUR variables were constructed by matching county-level unemployment rates from the Local Area Unemployment Statistics (LAUS) for different years to the individual's zip-code of residence from a restricted-use NJCS data set obtained from Mathematica Policy Research, Inc. The first LUR variable employed further matches the individual's year of exit from JC to that same year's LUR. The second variable is a race/ethnicity-specific LUR for people ages 16 to 35 in the year 2000.[19] The GPS models estimated in Section IV indicate that the two variables often have separate explanatory power.

Another set of variables presented in Table 1 includes individual predicted probabilities for enrollment and training durations in JC of more than 30, 90, 180, and 270 *calendar* days (conditional on enrollment). They were originally computed by Johnson, Gritz, and Dugan (2000) in a NJCS report to analyze the programmatic experiences of JC participants, and are available in the NJCS public-use data. These probabilities were obtained employing multinomial logistic models using a myriad of covariates, including some not available in the public-use data

[19] The race/ethnicity-specific LUR is only available for the year 2000 since it is constructed from 2000 Census data. For details in the construction of all these LUR variables see the Internet Appendix to the paper.

such as variables reflecting the characteristics and particular practices of the Outreach and Admissions (OA) agency attended and the characteristics of the specific OA counselor who had contact with the individual.[20] These probabilities are included in the GPS models to account for potential factors related to the length of exposure to AV instruction and future earnings.

Variables pertaining to the expectation about and motivation for enrolling in JC (not reported in Table 1) are relevant since they help control for possible individual unobserved characteristics that may be related to both the outcome (earnings) and our treatment variable. The 29 variables in this group are listed in the Data Appendix. Finally, the state and JC center-attended indicators (not reported in Table 1) are intended to control for additional time-invariant local labor market and JC center characteristics.[21]

III. Econometric Methods

Let the units in our sample be indexed by $i = 1, \ldots, N$. Also, let $Y_i(t)$ be the potential outcome of unit i under treatment level $t \in \Im$, where in our case $\Im$ is an interval and t denotes the length of exposure to AV instruction in JC. We are interested in estimating the average dose-response function (DRF): $\mu(t) = \mathrm{E}[Y_i(t)]$ for all *t*. The observed variables for each unit *i* are a vector of covariates X_i, the level of the treatment received, T_i, and the observed outcome for the level of the treatment actually received $Y_i = Y_i(T_i)$.

The key identifying assumption used in this paper is that the selection into different treatment levels is unconfounded given the covariates, which include the rich set of pre-treatment characteristics and the LURs. In fact, for the theory presented below, we only require a weaker version of unconfoundedness as introduced in Hirano and Imbens (2004; hereafter HI) for the case of a continuous treatment:[22]

$$Y_i(t) \perp T_i \mid X_i \text{ for all } t \in \Im. \qquad (1)$$

This assumption—that the level of the treatment received (T_i) is independent of the potential

[20] OA agencies—which are in general not linked to a particular JC center—are involved in making *potential* participants aware of the opportunities JC offers, screen for eligibility, assign individuals to centers, and prepare candidates for enrollment (Johnson, Gritz, and Dugan, 2000, pp. 1-6).
[21] The number of centers represented by these indicators is 109 for the full sample, with no center having more than 5.2% of the individuals in the sample.
[22] HI refer to this assumption 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.

outcome $Y_i(t)$ conditional on observed covariates—is a natural extension of the common unconfoundedness assumption used in the binary-treatment literature (Heckman, LaLonde, and Smith, 1999; Imbens, 2004). Particularly, this assumption rules out any systematic "selection" into levels of the treatment based on unobservable characteristics not captured by our observed covariates. Although Assumption (1) rules out the existence of dynamic confounders, the characteristics of the JC program and the richness of the data considerably diminish their role, as explained in section II.B. In addition, we consider a difference-in-difference version of the estimators by employing the outcomes in differences relative to the individual's average weekly earnings before treatment, further controlling for time-invariant unobserved confounders (Heckman, et al., 1998; Abadie, 2005). In section IV.C, we provide some indirect assessments of Assumption (1) and its no-dynamic-confounding implication, and discuss the consequences for our results if dynamic confounders are important.

Under Assumption (1), the average DRF can be obtained by estimating average outcomes in subpopulations defined by covariates and different levels of the treatment. However, as the number of covariates increases, it becomes difficult to simultaneously adjust for all covariates in X. In the case of a binary treatment, the propensity score is commonly used to estimate average treatment effects under unconfoundedness (Imbens, 2004). Rosenbaum and Rubin (1983) show that adjusting for differences in the conditional probability of receiving treatment given pre-treatment covariates (the propensity score) eliminates selection bias between treated and untreated individuals, if selection into treatment is based on observable factors. This result implies that we only need to adjust for a scalar variable to control for imbalances in the covariates, leading to more flexible ways to estimate treatment effects. Another advantage of propensity score methods is that, by using this scalar measure, we are able to detect observations in the treatment and control groups for which it is not possible to find comparable units in the opposite group. Imbens (2000) extends the propensity score methodology to multi-valued treatments, while HI further extend the results to continuous treatments. Both papers employ the "generalized propensity score" (GPS) to reduce the conditioning set to one, just as in the binary case.

Following HI, the GPS is the conditional density of the treatment given the covariates:

$$r(t,x) = f_{\mathrm{T}|X}(t \mid X = x). \qquad (2)$$

For the discussion below, it is important to note that $r(t,x)$ represents different random variables. Let $R_i = r(T_i, X_i)$ denote the conditional density at the treatment actually received, and let $R_i^t = r(t, X_i)$ denote the family of random variables indexed by *t*. Clearly, for those units with $T_i = t$ we have $R_i = R_i^t$.

HI show that the GPS shares many of the attractive properties of the propensity score in the binary case, such as the "balancing property" (loosely speaking, $X \perp 1\{T = t\} \,|\, r(t, X)$) and the fact that weak unconfoundedness given the covariates implies weak unconfoundedness given the GPS (i.e., $f_T(t \,|\, R_i^t, Y_i(t)) = f_T(t \,|\, R_i^t)$). This last result allows the estimation of the average DRF by using the GPS to remove selection bias. In particular, HI show that under Assumption (1) we can identify the average DRF as (Theorem 3.1 in HI):

$$(i) \;\; \beta(t, r) = \mathrm{E}[Y_i(t) \,|\, R_i^t = r] = \mathrm{E}[Y_i \,|\, T_i = t, R_i = r]$$
$$(ii) \;\; \mu(t) = \mathrm{E}[\beta(t, R_i^t)]. \qquad (3)$$

The result in (3) suggests estimating the DRF at *t* using a partial mean, which is the average of a regression function over some of its regressors while holding others fixed (Newey, 1994). In our case, the regressor that is fixed in the second averaging is the treatment level *t*. Hence, the DRF function can be estimated using the GPS by following two steps. The first step involves estimating the conditional expectation of *Y* on *T* and *R*, $E[Y_i \,|\, T_i = t, R_i = r]$, while the second step involves averaging this conditional expectation over R_i^t to get the value of the DRF at *t*. HI implement this partial mean approach by assuming a (flexible) parametric form for the regression function of *Y* on *T* and *R*. Let $\hat{R}_i$ and $\hat{R}_i^t$ be estimators of R_i and R_i^t, respectively (discussed in the following section). HI estimate the regression

$$E[Y_i \,|\, T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 + \alpha_4 \hat{R}_i + \alpha_5 \hat{R}_i^2 + \alpha_6 \hat{R}_i^3 + \alpha_7 T_i \cdot \hat{R}_i, \qquad (4)$$

and estimate the DRF at *t* by averaging (4) over the distribution of $\hat{R}_i^t$:

$$\hat{\mu}(t)_{PPM} = \frac{1}{N} \sum_{i=1}^{N} [\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 t^3 + \hat{\alpha}_4 \hat{R}_i^t + \hat{\alpha}_5 \hat{R}_i^{t2} + \hat{\alpha}_6 \hat{R}_i^{t3} + \hat{\alpha}_7 t \cdot \hat{R}_i^t].$$

We refer to this estimator as the parametric partial mean (PPM) estimator.

Since there is no reason to commit ex-ante to any particular specification for the conditional expectation $\beta(t, r)$, and a misspecification could result in inappropriate bias removal, we also estimate the DRF employing partial means based on nonparametric kernel estimators as those previously used in Newey (1994) and more recently in Flores (2007). In particular, we use a local polynomial regression of order one (Fan and Gijbels, 1996) to estimate the regression function $\beta(t,r) = E[Y_i \mid T_i = t, \hat{R}_i = r]$. Let $K(u)$ be a kernel function with the usual properties; let h be a bandwidth satisfying $h \to 0$ and $Nh \to \infty$ as $N \to \infty$; and let $K_h(u) = h^{-1}K(u/h)$. Given that in this case we estimate a regression function with two regressors, we use a product kernel of the form $K_h(u,v) = K_h(u)K_h(v)$. Finally, let the nonparametric estimator of $\beta(t,r) = E[Y_i \mid T_i = t, \hat{R}_i = r]$ based on kernel $K_h(u,v)$ and bandwidth h be given by $\hat{\beta}(t,r;h,K_h)$. Then, our nonparametric partial mean (NPM) estimator of the DRF at t is given by

$$\hat{\mu}(t)_{NPM} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}(t,\hat{R}_i^t;h,K_h).$$

In addition to estimating the DRF within a partial mean framework, we also employ an approach based on weighting by the GPS. Weighting estimators in the binary-treatment case under an unconfoundedness assumption analogous to (1) are analyzed in Hirano, Imbens, and Ridder (2003), and have also been used in other settings such as multi-valued treatments (Cattaneo, 2010), difference-in-difference (Abadie, 2005) and instrumental variables models (Abadie, 2003). The approach we use here can be seen as a generalization of the weighting approach in Hirano Imbens, and Ridder (2003) to the continuous treatment case.[23] Let $\omega(T,X;t)$ be a function of the treatment and the covariates such that $E[\omega(T,X;t) \mid X]$ exists and is different from zero. Flores (2005) shows that, under Assumption (1), we can identify the DRF at t as

$$\mu(t) = E\left[\frac{\omega(T_i,X_i;t)Y_i}{E[\omega(T_i,X_i;t) \mid X_i]}\right]. \qquad (5)$$

In the case when T is binary so that $T = \{0,1\}$, we can let $\omega(T_i,X_i;1) = 1(T_i = 1)$, where $1(\cdot)$ is the usual indicator function. Hence, $E[Y(1)]$ can be estimated, for instance, by weighting the treated individuals by the estimated propensity score $\hat{p}(X) = \Pr(T = 1 \mid X)$: $N^{-1}\sum_{i=1}^{N}1(T_i = 1)Y_i[\hat{p}(X_i)]^{-1}$. The problem with estimating $\mu(t)$ in the continuous-treatment case by setting $\omega(T_i,X_i;t) = 1(T_i = t)$ is that the probability of having individuals with a particular

---

[23] An alternative approach to follow would be to use a continuous version of the approach developed in Abadie (2005) which, as discussed in that paper, can also be used within a "selection-on-observables" framework.

value of t is zero, and there will be an infinite number of values of T for which we will have no individuals with that treatment level. Hence, we use nonparametric methods to implement our weighting estimator of $\mu(t)$.

To motivate our estimator, let h be a sequence of positive numbers tending to zero as $N \to \infty$, $\Delta = [t-h, t+h]$, and $\omega(T_i, X_i; t) = 1(T_i \in \Delta)$. Then, for a small enough h, we can approximate the denominator inside the unconditional expectation in (5) as $\Pr(T_i \in \Delta \mid X_i) \approx 2hR_i^t$. Hence, we can define an estimator of $\mu(t)$ as $N^{-1}\sum_{i=1}^{N} 1(T_i \in \Delta)Y_i[2h\hat{R}_i^t]^{-1}$, where, as before, $\hat{R}_i^t$ is an estimator of the GPS at t. More generally, we can use a kernel function that assigns more weight to individuals closer to treatment level t, and let $\omega(T_i, X_i; t) = K_h(T_i - t)$.[24,25] In addition, just as in the binary-treatment case, we can normalize the weights of our estimator so that they add up to one (Imbens, 2004), and estimate the DRF at t as

$$\hat{\mu}(t) = \frac{\sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t) \cdot Y_i}{\sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)},$$

where $\tilde{K}_{h,X}(T_i - t) = K_h(T_i - t) / \hat{R}_i^t$. This estimator is simply the usual local constant regression (or Nadaraya-Watson) estimator, but now each individual's kernel weight is divided by her GPS at t. It is important to note that, in this weighting approach, we use $\hat{R}_i^t$, not $\hat{R}_i$, to weight the usual kernel in $\tilde{K}_h(T_i, X_i; t)$.

Rather than using this local-constant estimator, and in line with our NPM estimator, we implement the weighting approach to estimate $\mu(t)$ using a local linear regression of Y on T with weighted kernel function $\tilde{K}_{h,X}(T_i - t)$. We prefer the local linear estimator since it avoids boundary bias and it is also easier to work with for derivative estimation, which will be performed below. To explicitly write our estimator, let $S_j(t) = \sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)(T_i - t)^j$ and

---

[24] When using a second-order symmetric kernel, as the one employed here, the error in approximating $E[\omega(T_i, X_i; t) \mid X_i]$ by the GPS is of order $O(h^2)$.

[25] Note that, in the previous example, $K_h(u) = h^{-1}1(T_i \in \Delta)$ is the uniform kernel.

$D_j(t) = \sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)(T_i - t)^j Y_i$. Then, the weighted estimator used in this paper is

$$\hat{\mu}(t)_{IW} = \frac{D_0(t)S_2(t) - D_1(t)S_1(t)}{S_0(t)S_2(t) - S_1^2(t)}. \qquad (6)$$

We implement the NPM and IW estimators by using a product-normal and a normal kernel, respectively, and by choosing a global bandwidth based on the procedure proposed by Fan and Gijbels (1996). This procedure is based on estimating the unknown terms appearing in the optimal global bandwidth by employing a global polynomial of order *p* plus 3, with *p* being the order of the local polynomial fitted. This bandwidth selector has been previously used in economics (Ichimura and Todd, 2007), especially an adaptation of it to the regression discontinuity context (Lee and Lemieux, 2009).[26]

IV. Estimation Results

In this section, we estimate the average DRF of the length of exposure to AV instruction in JC and its derivative for the two measures of earnings described in section II.C. We first discuss the estimation of the GPS, followed by the presentation of our main estimation results. We conclude this section with some exercises aimed at indirectly assessing our identifying assumptions and with a discussion of the implications for our results if dynamic confounding is present.

A. Estimation of the GPS

The first step to implement the estimators from the previous section consists of modeling the conditional distribution of the treatment T_i (weeks of AV instruction) given the covariates, that is, the GPS. There are many choices available to do this. A commonly employed specification for a non-negative, continuously distributed variable is a lognormal distribution (Hirano and Imbens, 2004; Kluve et al., 2007). Although our treatment variable is also non-negative and continuous (Figure 1), we hesitate to commit ex-ante to any one specification and

[26] We analyze the sensitivity of the results in the next section to our specific choice of (i) nonparametric estimator, by considering the Nadaraya-Watson estimator at non-boundary points; (ii) kernel, by considering an Epanechnikov kernel; and (iii) bandwidth, by considering alternative bandwidths of the form $h=ah^*$, where h^* is the bandwidth selected and we let a vary. In general, our results are robust to these different specifications, especially in regions where we have enough data points.

instead estimate a number of flexible generalized linear models (McCullagh and Nelder, 1989) and choose the model that best fits our data.

In general, letting $g\{E(T)\} = X\gamma$ with γ denoting a vector of coefficients and *g* a "link function", various specifications can be obtained by choosing a distribution *F* for *T* and a functional form for *g*. For example, the log-normal specification is obtained as a special case with *F* as normal and *g* as the log function. We estimate several plausible specifications (outlined in the Internet Appendix) by maximum likelihood and choose the model that best fits the data according to the deviance measure of McCullagh and Nelder (1989), as well as the Akaike Information Criteria (AIC) and the value of the log-likelihood function.[27]

Recall that the estimated GPS model is the basis to controlling for selection bias into different lengths of exposure. The variables included in the generalized linear models are all those listed in the Data Appendix, along with a full set of interactions with indicator variables for race/ethnicity and gender and higher order polynomials of several continuous variables. Across all groups considered, a gamma model with a log link achieves the best model fit and is thus employed to model the GPS.[28] All estimated coefficients of the model for each sample, along with their respective robust standard errors, are relegated to the Internet Appendix.

Given that the GPS will be employed to make comparisons of individuals with different values of *T* but the same values of the GPS, it is important to verify that no values of the GPS are so extreme that individuals with comparable values of the GPS are impossible to find. For these extreme values, inference using the GPS will be poor, as no comparable individuals in terms of the GPS are available to undertake causal comparisons. Here, we follow common practice (Dehejia and Wahba, 2002; Imbens, 2004; Gerfin and Lechner, 2002; Lechner, 2002) and restrict the analysis to those individuals for which reliable inference can be obtained by concentrating the analysis to the subsample that satisfies the overlap or common support condition.

In the binary treatment literature, it is common to gauge the overlap in the covariate distributions between treated and non-treated groups by looking at the distribution of their

---

[27] The distributions considered are the log-normal, inverse Gaussian, and gamma. Within the inverse Gaussian and gamma distributions, we employ link functions corresponding to the inverse powers 0.5, 1, 1.5, 2, and a log link. We employ AIC to choose across the different distributions and the other two measures to choose among link functions (Hardin and Hilbe, 2007).

[28] Note that the gamma model with a log link function (and restricting the scale parameter to 1) is equivalent to the exponential regression model, which is commonly used in duration analysis. However, we do not restrict the scale parameter to equal one in our models.

estimated propensity scores, and restricting estimation to the common support region (Imbens, 2004). In the continuous-treatment setting, it is not straightforward to impose this condition. The reason is that there is a continuum of treatment levels, so we have an infinite number of "treatment groups" and generalized propensity scores (R_i^t) to compare.[29]

To gauge the extent of overlap in the support of different levels of the treatment, we utilize a straightforward extension of a method employed in the binary-treatment case (Dehejia and Wahba, 2002) by dividing the range of *T* into quintiles. Let the quintile each individual belongs to be denoted by $Q_i = \{1, 2, 3, 4, 5\}$. For each quintile *q*, we compute the value of the GPS at the median level of the treatment in that quintile for all individuals, call it $\hat{R}_i^q$. The common-support region with respect to quintile *q* is obtained by comparing the support of the distribution of $\hat{R}_i^q$ for those individuals with $Q_i = q$ to that of individuals with $Q_i \neq q$. Let CS_q denote the common-support subsample with respect to quintile *q*. Then, we define CS_q as

$$CS_q = \left\{ i : \hat{R}_i^q \in [\max\{\min_{\{j:Q_j=q\}} \hat{R}_j^q, \min_{\{j:Q_j\neq q\}} \hat{R}_j^q\}, \min\{\max_{\{j:Q_j=q\}} \hat{R}_j^q, \max_{\{j:Q_j\neq q\}} \hat{R}_j^q\}] \right\}.$$

We restrict our sample (for each demographic group analyzed) by keeping only those individuals that are comparable across all five quintiles simultaneously. Hence, our common-support subsample is given by $CS = \bigcap_{q=1}^{5} CS_q$.

The resulting common-support restricted samples are similar in size to the original samples in Table 1 for some groups, but unfortunately not for others. This reflects the difficulty of finding comparable individuals within those subpopulations. The restricted samples consist of (percentage of observations dropped in parentheses) 3,524 observations (5.1%) for the full sample, 1,830 (9.1%) for blacks, 1,825 (12.5%) for males, 1,407 (13.7%) for females, while the groups losing more observations due to the common support condition are whites with 726 restricted observations (28%) and Hispanics with 404 (41.7%). Given that using observations outside the common support of the GPS can result in misleading inference, we concentrate on these smaller samples for the rest of the analysis.[30]

[29] Note that, in the binary-treatment case, imposing the overlap condition on $p(x) = \Pr(T = 1 | X = x)$ is equivalent to imposing this condition on $p_0(x) = \Pr(T = 0 | X = x) = 1 - p(x)$. In the continuous case, we have an infinite number of scores.

[30] Histograms containing a graphical representation of the GPS support before and after imposing the overlap condition for each quintile in each of the groups analyzed are contained in the Internet Appendix to the paper.

An important characteristic of the estimated GPS that needs to be verified is its balancing property: the GPS "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. This balancing property can be employed to empirically assess the adequacy of the estimated GPS in a similar spirit in which it is done in the binary treatment case with the propensity score (Dehejia and Wahba, 2002; Smith and Todd, 2005). In the case of a continuous treatment, the approaches employed in the literature analyze the balancing of each covariate separately (Hirano and Imbens, 2004; Imai and van Dyk, 2004). The main idea is to compare the balancing of each covariate before and after accounting for the GPS.[31]

The approach we follow is closer in spirit to that in Imai and van Dyk (2004), which in our framework would be equivalent to employing regressions (or logit models for binary covariates) of each covariate on *T* with and without the GPS—$r(T, X)$, and comparing the significance of the coefficient for *T*. However, given the large number of covariates we employ, along with interactions between them and selected variables and higher-order polynomials, we instead estimate a gamma model with a log link for *T* (as in the GPS) that includes all covariates plus the GPS up to a cubic term (the unrestricted model).[32] Then, a likelihood ratio (LR) test is employed to compare that model with a restricted one that sets the coefficients of all covariates to zero. If the GPS sufficiently balances the covariates employed, then they should have little explanatory power conditional on the GPS. We find this to be the case in Table 2, as the LR tests strongly indicate that the restricted model is overwhelmingly preferred in all samples (top panel). Conversely, using the same unrestricted model, we test whether the GPS coefficients are all simultaneously equal to zero. This restriction is strongly rejected in the bottom panel of Table 2, speaking to the importance of the role played by the GPS. We regard this as evidence that the balancing property of the GPS is satisfied.[33]

B. Estimation of the Dose-Response Function and Its Derivative

[31] HI divide the levels of the treatment into three intervals. Then, within those intervals, they stratify individuals into five groups according to the values of the GPS evaluated at the median value of the treatment of the corresponding interval. Finally, they test whether the observed covariates are "balanced" within these GPS strata.

[32] We use a cubic specification of the GPS to make it consistent with the specification of the PPM estimator in (4). The same qualitative results are obtained with more flexible specifications of the GPS.

[33] Note that all these models are estimated only for the common-support restricted sample.

The DRF estimates are obtained at 99 different values of the length of exposure to AV instruction, corresponding to the percentiles of the empirical distribution for each of the samples analyzed. In addition to obtaining DRF estimates employing the estimators presented above (PPM, NPM, IW), we also obtain for comparison estimates of the DRF using OLS of weekly earnings on a cubic function of the length of exposure and the full set of covariates used in the estimation of the GPS. The difference between this OLS estimator and the GPS estimators is twofold. First, all three GPS methods provide a more flexible specification of the relationship between the covariates and the outcome, especially the semiparametric estimators NPM and IW. The second is that, by imposing the GPS-based common support condition, we are able to drop observations for which there are no comparable individuals across different treatment levels.

We generate a series of figures that plot the DRF of length of AV instruction (in weeks) on two different outcomes: average weekly earnings at quarter 16 after random assignment and average weekly earnings measured one year after the individual has exited JC. As previously mentioned, the first measure is taken at a fixed point in time and includes any potential lock-in costs to the participant, such as any lost labor market experience due to training. Conversely, the second measure fixes at one year the amount of time that elapses between the end of training and when earnings are measured.

Plots are produced for these two outcomes measured in levels and in differences relative to the individual's average weekly earnings before randomization. The latter corresponds to the DD specification. The estimates in levels and in differences are very similar, so we choose to present below results using the outcome specifications in differences and to relegate the results in levels to the Internet Appendix. Finally, for each outcome and specification, we also generate plots of the derivative of the DRF, which is informative about the "marginal" return to additional exposure to AV instruction in JC.[34] Our results are accompanied by 95% (point-wise and percentile-based) confidence bands obtained with 1,000 bootstrap replications that account for all estimation steps, including the estimation of the GPS and the imposition of the common support condition for the GPS-based estimates.

[34] For OLS and PPM estimators, the derivative at t is obtained as the "forward" change of one additional week of training: $\hat{\mu}(t+1)-\hat{\mu}(t)$. This is the usual approach when using the PPM estimator (Bia and Mattei, 2008). For the IW estimator, the derivative estimate at t equals the slope coefficient of the linear term from a local quadratic regression of Y on T using the re-weighted kernel defined in section III, $\tilde{K}_h(T_i, X_i; t)$. We choose the appropriate bandwidth by using the automatic procedure described in Fan and Gijbels (1996). For the reasons mentioned later in the text, we do not employ the NPM estimator for derivative estimation.

To conserve space, we present here a selection of plots that provide the main insights of our analysis (the full set of plots can be found in the Internet Appendix to the paper). We start with a plot of the DRF for each of the outcomes in differences (Figure 2) using all estimators considered for the full sample. We do not include confidence bands for readability. In this and the following figures, the vertical lines represent the length of exposure corresponding to the 25th, 50th, and 75th percentiles of the within-sample distribution. Figure 2 gives a general idea of the shape of the DRF for each of the outcomes in differences and allows us to point out that (a) the two semiparametric GPS estimators are very similar to each other, especially in regions with several observations and (b) there are important differences between the semiparametric GPS estimators and both the PPM estimator and OLS.[35] For instance, while in the second panel the semiparametric estimators suggest an increasing relation between the outcome and length of exposure to AV instruction, the OLS estimator suggests an inverted-U relationship.

Given the greater flexibility allowed by the semiparametric estimators, we will concentrate our discussion below on one of them: the IW estimator. We focus on the IW estimator over the NPM for several reasons. First, it is considerably faster to compute since to estimate the DRF at t we only need to compute a local linear regression with one regressor (T), while to calculate the NPM estimator we need to compute a local linear regression with two regressors (T and $\hat{R}_i$) N times in the first step (see Section III). This is particularly relevant when obtaining bootstrap confidence intervals.[36] Second, even though nonparametric partial means and nonparametric regressions of the same dimension converge at the same speed (Newey, 1994), the fact that in the NPM estimator we need to estimate a higher dimensional regression in the first step is likely to make it more variable than the IW estimator in finite samples, especially in regions where data is sparse. Finally, in our application, we observe that, in regions with a relatively large number of observations, the NPM and IW estimators are very close to each other.

We start our analysis of the results by focusing on the outcome average weekly earnings at quarter 16 in differences. Figure 2 illustrates that, for the full sample, the difference between average weekly earnings in quarter 16 and average weekly earnings at baseline is an increasing function of the length of exposure to AV instruction, but that past a particular level of exposure this difference begins to decrease, although this occurs in regions where the data is sparse. This

[35] These features are even more evident in the other samples analyzed.
[36] Therefore, we do not report confidence bands or derivative estimates for the NPM estimator.

general pattern holds for all samples analyzed, with differences arising on the magnitude of the DRF's derivative and the width of the corresponding confidence bands. Figure 3 presents, for each of the samples, plots of the derivative of the DRF calculated using the IW estimator on weekly earnings in quarter 16 (in differences), along with 95% confidence bands.[37]

Several features are worth pointing out. First, the point estimates of the derivatives are mostly positive across samples—consistent with a positive relationship between the outcome in differences and weeks of exposure to AV instruction. The second important feature to observe in Figure 3 is that, despite the positive point estimates of the derivative of the DRF documented above, the 95% confidence bands indicate that very few estimates are statistically different from zero. In fact, only two samples show a range of statistically significant point estimates: the full sample from 14.5 (34^{th} percentile) to 28.4 weeks (57^{th} percentile) and males from 22 (47^{th} percentile) to 36 weeks (67^{th} percentile).[38] Nevertheless, none of the estimates are statistically negative, so non-negative effects cannot be ruled out throughout.[39]

A way to summarize the derivative estimates of the DRF in Figure 3 is to consider the average derivative over different treatment levels. We present some of these average derivatives (along with their statistical significance) in Table 3 for selected ranges of *T* that correspond to quantiles 1-99, 1-25, 1-50, 1-75, and 25-75. Hence, for instance, the average derivative for the full sample between the 25^{th} and 75^{th} quantiles—where data is not too sparse and thus our inference is more reliable—is \$0.8 per week (top panel) and is statistically different from zero at the 99 percent level. This means that the change in the difference between average weekly earnings at quarter 16 and at baseline in this range increases by \$0.8 per week for each marginal increase in weeks of AV instruction. Note that the estimates indicate that between the 25^{th} and

[37] Note that, as expected, the confidence intervals get very wide as we move to the right of the graph to places where the data are sparse, and when we analyze smaller samples (e.g., whites and Hispanics).

[38] The OLS estimates show a larger proportion of statistically significant estimates in the full and male samples, but at the (potential) cost of restricting the flexibility of the dependence between the outcome and the covariates. Interestingly, the PPM estimates are less precise than the IW estimates.

[39] The fact that only few estimates are statistically significant may be due to a myriad of factors. For instance, it may be that the relatively small sample sizes (in particular of whites and Hispanics) do not allow enough precision of the IW estimator. In this regard, it is important to remember that the speed of convergence of nonparametric estimators of derivatives are slower than that of nonparametric estimators of the function itself. On the other hand, there is precedent that the positive and statistically significant estimated impacts of participation in JC in the original NJCS (Schochet, Burghardt, and Glazerman, 2001) that use the same outcome as we do are not robust to some perturbations such as considering only the group of Hispanics (Schochet, Burghardt, and Glazerman, 2001), alternative administrative data on the outcome (Schochet, McConnell, and Burghardt, 2008), or controlling for labor market experience gained by individuals during the study or the LURs they face (Flores-Lagunes, Gonzalez, and Neumann, 2009).

75th quantiles the average derivative for males is almost twice relative to the full sample and highly significant, while it is -0.1 and statistically insignificant for females. Finally, note that both the average derivatives in Table 3 and the graphs in Figure3 suggest the presence of diminishing marginal returns to time spent receiving AV instruction.

In order to explore a potential role played by the fact that individuals with shorter spells of training have more time in the labor market before the 48-month survey—that is, the "lock-in effect" (van Ours, 2004)—we consider the alternative outcome that fixes the time between the end of training and when the individual's earnings are measured at one year.[40] The second panel in Figure 2 suggests an increasing relationship between the gain in average weekly earnings from baseline to one year after exiting JC and the length of exposure to AV instruction in JC. Figure 4 shows plots of the derivative of the DRF similar to those of Figure 3 for this alternative outcome (in differences), and the second panel of Table 3 shows the corresponding average derivatives. Now, both the magnitude of the estimates and their statistical significance are higher. These results suggest that lock-in effects in JC are important. As with the previous outcome, no derivative estimate is statistically negative. Furthermore, in every sample—except Hispanics—there are ranges of statistically significant estimates at the 95%. The full sample has significant point estimates from 1.5 (1st percentile) to 31.1 weeks (61st percentile), whites from 15.8 (29th percentile) to 21.75 weeks (39th percentile), blacks from 3.5 (7th percentile) to 23.9 weeks (52th percentile), females from 10.1 (21th percentile) to 17 weeks (33th percentile) and males from 1.8 (2nd percentile) to 35 weeks (66th percentile).

Similar to the previous outcome, the differences between the marginal returns to AV instruction for males and females are striking. For instance, the estimated average derivatives for males shown in Table 3 are usually more than twice those of females. Moreover, the difference between the average derivatives from the 25th to the 75th quantiles for these two groups is statistically different at a 95 percent level (not shown in Table 3). The differences in the estimated average derivatives that are statistically significant for both black and white are not statistically different from each other. Regarding Hispanics, the derivatives and average derivatives are again estimated very imprecisely, and hence it is more difficult to draw

[40] Note that another difference between these two outcomes is that "weekly earnings at quarter 16" represents a longer-term outcome for some individuals, especially for those who enrolled in JC earlier and exited sooner. On the other hand, "weekly earnings one year after exiting JC" represents the same relatively shorter-term outcome for all individuals. Thus, to the extent that JC effects increase or decrease over time, this will also be a source of difference between the two measures.

conclusions for this particular group. Finally, the results for this outcome also suggest the existence of diminishing returns to the time spent in AV instruction.

C. Indirect Assessment of the Identifying Assumptions

We now present some exercises aimed at indirectly assessing the unconfoundedness assumption in (1), and its implied assumption of no dynamic confounding conditional on observed covariates. Since both assumptions are not directly testable, the evidence presented below is only indirect and suggestive. Hence, although finding favorable evidence does not imply that the assumptions hold, unfavorable evidence does weaken their plausibility considerably. We end this subsection with a discussion of the consequences for our results if dynamic confounding is present.

First, we perform placebo tests in the spirit of Heckman and Hotz (1989). This is a common type of exercise employed in the static binary treatment literature to indirectly assess the unconfoundedness assumption (Imbens and Wooldridge, 2009). It is based on estimating the average effect of the treatment on a variable known to be unaffected by it (a placebo outcome) so the effect is known to be zero. If the estimated effect is not zero, this is interpreted as evidence against unconfoundedness. For these tests, the common practice is to use lagged outcomes as placebo outcomes.

It is important to clarify the scope of this exercise in a dynamic setting. In the static-framework, placebo tests are informative only about potential unobserved confounders that are related to both the outcome of interest and the placebo outcome. Since dynamic confounders are variables revealed after starting treatment, the placebo tests can only provide indirect information on the potential importance of dynamic confounders to the extent that they are correlated to both pre-treatment uncontrolled factors and the placebo outcome. Hence, these tests are less informative when applied to a dynamic setting. For example, consider test scores while enrolled in JC as a potential dynamic confounder, which are likely correlated to variables that are realized before AV instruction starts (e.g., the initial TABE score, race, gender). Our placebo exercise can only shed light on the potential importance of test scores while enrolled in JC as a dynamic confounder to the extent that they are correlated to both, pre-treatment uncontrolled factors (e.g., the TABE score) and to the placebo outcomes (e.g., pre-training earnings). Thus, if we were to find a strong and significant correlation between our treatment variable and a placebo outcome conditional on covariates, we would strongly believe that an important pre-treatment confounder is not being controlled for (e.g., the TABE score), *and* that dynamic confounders correlated to that pre-treatment confounder (e.g., test scores while in JC) are likely important. Clearly, although failing to reject the placebo test does not imply that dynamic confounders are not present or important, it is a lot more difficult to argue that dynamic confounding may not be present if it is not rejected.

We consider three variables measured before randomization as placebo outcomes. The first is individual earnings during the full year prior to randomization. Since this earnings measure can be affected by a dip in earnings right before seeking the training program (Ashenfelter, 1978), we also construct a measure of earnings in the first three quarters of the year prior to randomization.[41] Finally, we consider a variable that measures the number of weeks that the individual was employed during the first three quarters of the year prior to randomization. Figure 5 presents plots similar to those of Figures 3 and 4 employing the earnings (in levels) in the first three quarters of the year prior to randomization as an outcome, and the last panel of Table 3 shows the corresponding average derivatives. Using any of the other two placebo variables as well as employing the other GPS estimators yield essentially the same conclusions and are available in the Internet Appendix to the paper. None of the derivative estimates in Figure 5 are statistically distinguishable from zero, except in a small range of treatment values (45-55 weeks) in the sample of whites, where the estimates are statistically less than zero. Similarly, none of the average derivatives in Table 3 is statistically different from zero.[42]

Another suggestive exercise aimed at indirectly assessing the potential importance of dynamic confounders related to the individual's performance in JC relies on three skill test scores obtained when most participants had already exited JC (and available only for a

[41] The NJCS data contain information on whether the individuals were employed in each week of the previous year, their total earnings over the previous year, and the average weekly earnings at their most recent job in the previous year. We estimate total earnings in quarter 4 of the previous year as the average weekly earnings at the most recent job times the number of weeks worked in quarter 4. We then construct a measure of earnings in the first three quarters of the previous year by subtracting the estimated quarter 4 earnings from the total earnings in the previous year.

[42] For the computation of the results employing the placebo outcomes we re-estimate the GPS excluding the variables that are closely related to the placebo outcomes—such as average weekly earnings at baseline—to avoid mechanically controlling for the placebo outcome in the GPS. We also note that the placebo outcomes are not included in the GPS used in the previous section since closely related variables are already included.

subsample).[43] These scores are likely highly correlated with that type of potential dynamic confounders, and thus can be used as proxy variables for them to test whether they are uncorrelated to our treatment variable conditional on covariates. We run log-gamma models (as in the GPS) of our treatment variable on each one of the skill test scores with and without controlling for the GPS. In each of the models, the test score coefficients are not statistically significant when controlling for the GPS, and although only the prose literacy score was unconditionally statistically significant, the magnitude of all point estimates decreases considerably when controlling for the GPS. For brevity, the results are shown in the Internet Appendix.[44] While not ruling out the existence of dynamic confounders, this exercise provides additional suggestive evidence that they do not play a significant role in our application after conditioning on covariates.

We acknowledge that despite the institutional characteristics of JC, the richness of our data, and the suggestive evidence presented above, the possibility of dynamic confounders biasing our estimated effects cannot be completely ruled out. If those dynamic factors result in individuals with higher future-earning potential systematically exiting JC early and receiving less AV instruction, this would attenuate our estimated effects, rendering them as a lower bound of the true effects. This could be the case, for instance, if even after controlling for covariates the evaluation panels systematically encourage students that perform well in JC to exit JC earlier by judging them ready. On the other hand, if those dynamic factors result in individuals with higher future-earning potential systematically acquiring more AV instruction, our estimated effects would overestimate the true effects. This could occur, for instance, if individuals that have

[43] During the 30-month follow-up survey of the NJCS, there was a round of literacy and numeracy skills measurement. Due to budget constraints, only a random sample of 3,750 individuals from the original study sample participated in the "literacy study" (Glazerman, Schochet, and Burghardt, 2000). As a result, our subsample of individuals with skills test scores contains 520 individuals only. In comparing our full sample with this subsample, a Kolmogorov-Smirnov test easily fails to reject the equality in the distribution of length of exposure to AV training (p-value=0.34). Similarly, equality-of-mean tests between the two samples are not rejected for most of the covariates in Table 1. Information is available on three literacy measures: prose literacy (skills to understand and use information from texts), document literacy (skills necessary to use information in graphs, tables, maps, etc.), and quantitative literacy (skills necessary to perform arithmetic operations). Glazerman, Schochet, and Burghardt (2000) found that the impact of participation in JC on these scores were all positive, and significant for prose and quantitative literacy. They also report a positive association between the literacy scores and earnings.

[44] The unconditional (conditional) estimated coefficients for the prose, document and quantitative literacy measures are -0.0021 (-0.0006), -0.0010 (-0.0005), and -0.0008 (-0.0005), respectively. We use a cubic specification for the GPS. Very similar results are obtained if a linear regression is employed or the square of the skill test scores is included in the estimation.

trouble getting through the diagnostic tests in JC classes systematically exit JC sooner, even after controlling for covariates.

V. Conclusion

This study estimates the average causal impact of the length of exposure to academic and vocational (AV) instruction in Job Corps (JC) on weekly earnings for different demographic groups. We employ semiparametric estimators of the dose-response function (DRF) based on the generalized propensity score (GPS) under the assumption that, conditional on observable characteristics, the length of exposure is independent of future weekly earnings. In our preferred specification, we further control for time-invariant unobservables by using a difference-in-difference version of our estimators. The richness of our data, the institutional characteristics of JC, and a number of exercises performed increase the credibility of our identifying assumption.

Our estimates reveal important heterogeneous effects along the different lengths of exposure to AV training and across the different demographic groups considered. In general, there are positive returns to spending additional time in AV instruction in JC, although these returns diminish with length of exposure. This important information is uncovered by considering the continuous nature of the length of exposure to AV instruction in JC. The differences between the estimated marginal returns for males and females are worthy of note. For instance, the overall average marginal returns of AV instruction in JC for males is about two times that of females for both outcomes analyzed in this paper. The marginal returns for blacks are similar to those of the full sample, albeit of smaller magnitude. The smaller sample sizes for whites and Hispanics make it difficult to draw conclusions for these groups, although some statistically-significant positive effects are found for whites. Finally, our results suggest that lock-in effects are important in JC, since the outcome measure that holds constant the amount of time between JC enrollment termination and when earnings are measured clearly results in higher marginal returns to the length of exposure to AV instruction for all groups.

Our results are consistent with estimates of the average rate of return to an additional year of regular schooling in the literature. Using the fact that a typical high school student receives the equivalent of 27 weeks of instruction during a schooling year (Schochet, Burghardt, and Glazerman 2001), the estimates focusing on weekly earnings at quarter 16 after randomization imply an average rate of return of 14.6 percent to a school-year-equivalent of AV instruction in

JC.[45] The magnitude of this estimate is similar to those of instrumental variable (IV) estimates of the return to schooling based on characteristics of the school system (e.g., compulsory schooling laws), which are "most likely to affect the schooling choices of individuals who would otherwise have relatively low schooling" (Card, 1999).[46] This finding is also consistent with previous studies documenting that the average (binary) treatment effect of JC is roughly equivalent to that of one regular school year (Schochet, Burghardt and Glazerman, 2001; Lee, 2009).

Some policy implications can be derived from our analysis. While retention efforts may be beneficial for most of the groups under analysis, the relevance of the lock-in effect cannot be understated given the documented differences in returns between our two alternative outcomes. This poses a challenge to designers of JC as they try to balance higher individual human capital investments with the corresponding lock-in costs. A more efficient transmission of the skills provided by JC to participants may be the key to reconcile these seemingly opposite goals.

[45] This rate is computed by taking the number of weeks at the 55th percentile of the lengths of exposure to AV instruction (27 weeks) and multiplying it by the average derivative from the 1st to the 55th percentile (1.13); the result is then divided by the average value of the DRF from the 1st to the 55th percentile (207.6).

[46] In Table 4 of the survey of the literature by Card (1999), these IV estimates range from 6 to 15.3 percent, with most of them above 9 percent.

References

Abadie, A. (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113, 231-263.

Abadie, A. (2005), "Semiparametric Differences-in-Differences Estimators." *Review of Economic Studies*, 72, 1-19.

Abadie, A., Angrist, J., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica*, 70, 91-117.

Abbring, J. and Heckman, J. (2007), "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation." In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 6B. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 5145-5303.

Abbring, J. and van den Berg, G. (2003), "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica*, 71, 1491-1517.

Angrist, J., and Krueger, A. (1999), "Empirical Strategies in Labor Economics." In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, 3A. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 1277-1366.

Ashenfelter, O. (1978), "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*, 6(1), 47-57.

Bia, M. and Mattei, A. (2008). "A STATA Package for the Estimation of the Dose-Response Function through Adjustment for the Generalized Propensity Score." *The Stata Journal*, 8, 354-373.

Card, D. (1999), "The Causal Effect of Education on Earnings." In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, 3A. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 1801-1863.

Cattaneo, M. (2010). "Effcient Semiparametric Estimation of Multi-valued Treatment Effects Under Ignorability." *Journal of Econometrics*, 155, 138-154.

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

Fan. J. and Gibels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall: London.

Flores, C. (2005), "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment." Doctoral Dissertation, University of California at Berkeley.

Flores, C. (2007), "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment." Working Paper, University of Miami.

Flores, C., Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (2010), "Internet Appendix to 'Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps'." Available at http://www.fred.ifas.ufl.edu/~alfonsofl/IA.pdf.

Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (2009), "Learning but Not Earning? The Value of Job Corps Training for Hispanic Youth." Forthcoming, *Economic Inquiry*.

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

Glazerman, S., Schochet, P., and Burghardt, J. (2000), *National Job Corps Study: The Impacts of Job Corps on Participants' Literacy Skills*. 8140-930. Mathematica Policy Research, Inc., Princeton, NJ.

Hardin, J, and Hilbe, J. (2007), *Generalized Linear Models and Extensions*, 2nd Ed. College Station: Stata Press.
Heckman, J. and Hotz, J. (1989), “Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training.” *Journal of the American Statistical Association*, 84(408), 862-874.
Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998), “Characterizing Selection Bias Using Experimental Data.” *Econometrica*, 66(5), 1017-1098.
Heckman, J., LaLonde, R., and Smith, J. (1999), “The Economics and Econometrics of Active Labor Market Programs.” In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, 3A. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 1865-2097.
Heckman, J. and Navarro, S. (2007), “Dynamic Discrete Choice and Dynamic Treatment Effects.” *Journal of Econometrics*, 136, 341-396.
Heckman, J. and Smith, J. (1999), “The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies.” *Economic Journal*, 109, 313-348.
Heckman, J., Smith, J., and Clements, N. (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, K. and Imbens, G. (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.
Hirano, K., Imbens, G., and Ridder, G. (2003), “Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score.” *Econometrica*, 71, 1161-1189.
Ichimura, H. and Todd, P. (2007), “Implementing Nonparametric and Semiparametric Estimators.” In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 6B. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 5369-5468.
Imai, K. and van Dyk, D. (2004), “Causal Inference with General Treatment Regimes: Generalizing the Propensity Score.” *Journal of the American Statistical Association*, 99, 854-866.
Imbens, G. (2000), “The Role of the Propensity Score in Estimating Dose-Response Functions.” *Biometrika*, 87, 706-710.
------------- (2004), “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review.” *Review of Economics and Statistics*, 86(1), 4-29.
Imbens, G. and Angrist, J. (1994), “Identification and Estimation of Local Average Treatment Effects.” *Econometrica*, 62(2), 467-475.
Imbens, G. and Wooldridge, J. (2009), “Recent Developments in the Econometrics of Program Evaluation.” *Journal of Economic Literature*, 47(1), 5-86.
Johnson, T., Gritz, M., and Dugan, M. (2000), *National Job Corps Study: Job Corps Applicants’ Programmatic Experiences*. Mathematica Policy Research, Inc., Princeton, NJ.
Johnson, T., Gritz, M., Jackson, R., Burghardt, J., Boussy, C., Leonard, J., and Orians, C. (1999), *National Job Corps Study: Report on the Process Analysis*. 8140-510. Mathematica Policy Research, Inc., Princeton, NJ.
Kluve, J., Schneider, H., Uhlendorff, A. and Zhao, Z. (2007), “Evaluating Continuous Training Programs Using the Generalized Propensity Score.” IZA Discussion Paper No. 3255.

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.
Lee, D. (2009), “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *Review of Economic Studies*, 76, 1071-1102.
Lee, D. and Lemieux, T. (2009), “Regression Discontinuity Designs in Economics.” NBER Working Paper No. w14723.
McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Second Edition. Chapman and Hall/CRC.
Mitnik, O. (2008), “Intergenerational Transmission of Welfare Dependency: The Effects of Length of Exposure.”, Working Paper, University of Miami.
Newey, W. (1994), “Kernel Estimation of Partial Means and a General Variance Estimator.” *Econometric Theory*, 10, 233-253.
Robins, J. (1997), “Causal Inference from Complex Longitudinal Data.” In M. Berkane (ed.), *Latent Variable Modeling and Applications to Causality, Lecture Notes in Statistics*, New York: Springer, 69-117.
Rosenbaum, P. and Rubin, D. (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 70(1), 41-55.
Schochet, P. (1998), *National Job Corps Study: Eligible Applicants' Perspectives on Job Corps Outreach and Admissions*. Mathematica Policy Research, Inc., Princeton, NJ.
------------ (2001), *National Job Coprs Study: Methodological Appendixes on the Impact Analysis.* 8140-530. Mathematica Policy Research, Inc., Princeton, NJ.
Schochet, P. and Burghardt, J. (2008), “Do Job Corps Performance Measures Track Program Impacts?” *Journal of Policy Analysis and Management*, 27, 556–576.
Schochet, P., Burghardt, J., and Glazerman, S. (2001), *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*. 8140-530. Mathematica Policy Research, Inc., Princeton, NJ.
Schochet, P., Burghardt, J., and McConnell S. (2008), “Does Job Corps Work? Impact Findings from the National Job Corps Study.” *American Economic Review*, 98, 1864-1886.
Smith, J. and Todd, P. (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).
------------- (2006), *Policy and Requirements Handbook*. Washington, DC.
van Ours, J. (2004), “The Locking-in Effect of Subsidized Jobs.” *Journal of Comparative Economics*, 32, 37-55.

DATA APPENDIX.
VARIABLES EMPLOYED IN THE ESTIMATION OF THE GPS

Treatment and outcomes
Weeks of av instruction
Weekly earnings at month 48
Weekly earnings one year after exiting JC

Demographics
Female
White
Black
Hispanic
Non-residential slot
Age
Age squared
Age cubed

Local unemployment rates (lur)
Lur at time of JC exit
Lur of 16 to 35 year olds of individual's race

NJCS predictions
Months between rand. And JC enrollment
Estimated probability will not enroll in JC
Estimated probability will stay 30+ days
Estimated probability will stay 90+ days
Estimated probability will stay 180+ days
Estimated probability will stay 270+ days

Education
Had HS diploma
Had GED
Had voc. Degree
Attended edu. or train. Program in last yr.
Highest grade
Highest grade squared
English native language

Past employment, arrest and location characteristics
Avg. Weekly earnings at baseline
Ever worked
Lived in PMSA
Lived in MSA
Ever arrested

Household characteristics
Has child
Married
Head of household
Live with 2 parents

Health characteristics
Good health
Fair health
Poor health
Smoke
Drink
Smoke marijuana

Expectations about JC
Worried about JC
Heard about JC from parents
Knew what center wished to attend
Wanted clerical training
Wanted health training
Wanted auto training
Wanted carpentry training
Wanted welding training
Wanted electrical training
Wanted construction training
Wanted food svc. training
Wanted cosmetology training
Wanted electronics training
Wanted other training
Expect to improve math
Expect to Improve reading
Expect to improve social skills
Expect to improve self control
Expect to improve self esteem
Expect to help find specific job
Expect to make friends
Knew desired job training
Joined to achieve career goal
Joined to get job training
Joined to get GED
Joined to find work
Joined to get away from community problems
Joined to get away from home
Joined for other reason

Fixed effects
JC Center Attended (109 centers)
State of Residence (48 states)

Interactions
Models are fully Interacted by Female,
White, Black, and Hispanic indicators

Table 1. Summary Statistics of Selected Covariates for Different Demographic Groups

		Full Sample		White Sample		Black Sample		Hispanic Sample		Male Sample		Female Sample	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Treatment & Outcomes	Weeks of AV Instruction	30.38	27.27	28.70	24.94	29.16	26.57	36.35	31.46	29.64	26.99	31.32	27.61
	Weeks of AV Instruction--10th Percentile	4.38		4.00		4.00		6.00		4.00		5.00	
	Weeks of AV Instruction--25th Percentile	9.63		9.00		9.09		12.50		8.94		10.38	
	Weeks of AV Instruction--Median	22.75		22.57		21.39		28.56		22.00		23.99	
	Weeks of AV Instruction--75th Percentile	42.50		40.13		41.00		52.57		42.38		42.75	
	Weeks of AV Instruction--90th Percentile	67.31		63.80		65.83		77.00		66.14		68.50	
	Weekly Earnings at Month 48	217.26	218.26	271.42	236.27	193.47	209.21	207.64	202.77	247.95	237.22	178.02	184.05
	Weekly Earnings 12 Months after JC exit	153.90	170.90	202.37	197.30	132.60	155.18	146.18	159.58	174.30	183.78	128.15	149.22
Demographics	Female	0.44	0.50	0.33	0.47	0.47	0.50	0.49	0.50				
	Age	18.73	2.16	18.70	2.07	18.66	2.16	18.94	2.26	18.60	2.09	18.89	2.23
LUR	LUR at time of JC Exit	6.15	2.83	5.43	2.22	5.98	2.17	7.69	4.34	6.15	2.83	6.15	2.83
	LUR of 16 to 35 Year olds of Same Race in 2000	11.19	5.46	5.38	2.11	14.48	4.31	10.07	4.21	10.68	5.40	11.84	5.47
JC Study Predictions	Prob. will not Enroll in JC	0.24	0.14	0.23	0.14	0.24	0.15	0.26	0.15	0.22	0.13	0.27	0.15
	Prob. will stay 30+ days	0.83	0.10	0.83	0.11	0.82	0.10	0.85	0.10	0.82	0.11	0.84	0.10
	Prob. will stay 90+ days	0.75	0.14	0.75	0.13	0.74	0.14	0.80	0.14	0.73	0.13	0.77	0.14
	Prob. will stay 180+ days	0.69	0.16	0.69	0.16	0.67	0.16	0.73	0.15	0.68	0.16	0.69	0.16
	Prob. will stay 270+ days	0.62	0.20	0.60	0.21	0.61	0.19	0.67	0.19	0.61	0.20	0.63	0.19
Education	Had HS Diploma	0.18	0.38	0.20	0.40	0.16	0.37	0.17	0.38	0.14	0.35	0.22	0.41
	Had GED	0.04	0.20	0.08	0.27	0.02	0.15	0.05	0.22	0.05	0.22	0.04	0.19
	Had Voc. Degree	0.02	0.14	0.02	0.14	0.02	0.12	0.03	0.16	0.02	0.13	0.02	0.14
	Highest Grade	10.03	1.53	10.05	1.54	10.03	1.48	10.00	1.63	9.92	1.49	10.18	1.56
	English Native Lang.	0.88	0.32	0.99	0.12	0.97	0.16	0.47	0.50	0.89	0.31	0.87	0.34
Past Employment	Avg. Weekly Earnings	113.30	408.94	130.78	121.01	96.02	110.16	138.08	915.98	119.22	120.98	105.72	602.04
	Ever Worked	0.79	0.41	0.88	0.33	0.75	0.43	0.78	0.41	0.81	0.39	0.77	0.42
Household Characteristics	Has Child	0.19	0.39	0.09	0.28	0.23	0.42	0.22	0.41	0.09	0.29	0.31	0.46
	Married	0.02	0.13	0.02	0.13	0.01	0.09	0.04	0.19	0.01	0.10	0.02	0.15
	n	3715		1008		2014		693		2085		1630	

Note: A complete list of summary statistics for all the covariates employed is available in the Internet Appendix to the paper, as well as the estimated coefficients of the GPS. A list of other types of variables included in the analysis is included in the Data Appendix.

Figure 1. Histograms of the Length of Exposure to AV Instruction in JC for Each Group

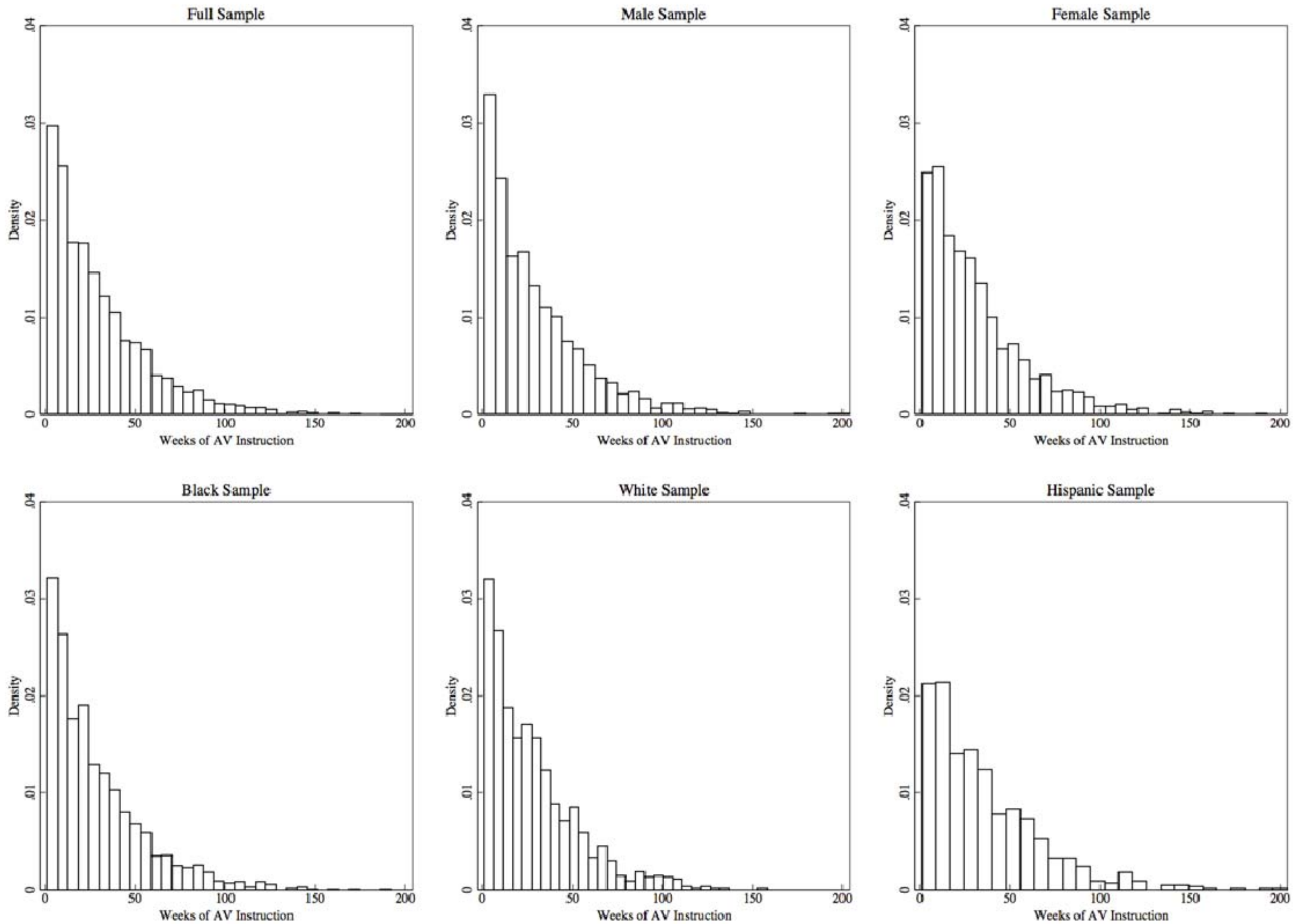

Table 2. Balancing Tests

	Sample					
	Full	**White**	**Black**	**Hispanic**	**Male**	**Female**
Unrestricted model: T on GPS, GPS^2, GPS^3, and X's						
Test restriction that X's can be exluded from the unrestricted model						
LL Restricted	-14554	-3088	-7505	-1846	-7526	-5946
LL Unrestricted	-14520	-3069	-7483	-1828	-7496	-5924
Test Statistic	69	38	44	35	61	43
p-value	1.00	1.00	1.00	1.00	1.00	1.00
Number of Restrictions	405	253	243	209	333	322
Test restriction that GPS coefficients can be exluded from the unrestricted model						
LL Restricted	-15392	-3186	-7940	-1871	-7932	-6229
LL Unrestricted	-14520	-3069	-7483	-1828	-7496	-5924
Test Statistic	1745	235	915	87	873	610
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Number of Restrictions	3	3	3	3	3	3
N	3524	726	1830	404	1825	1407

Figure 2. Dose response function (DRF) on two outcomes (in differences) using alternative estimators in the full sample

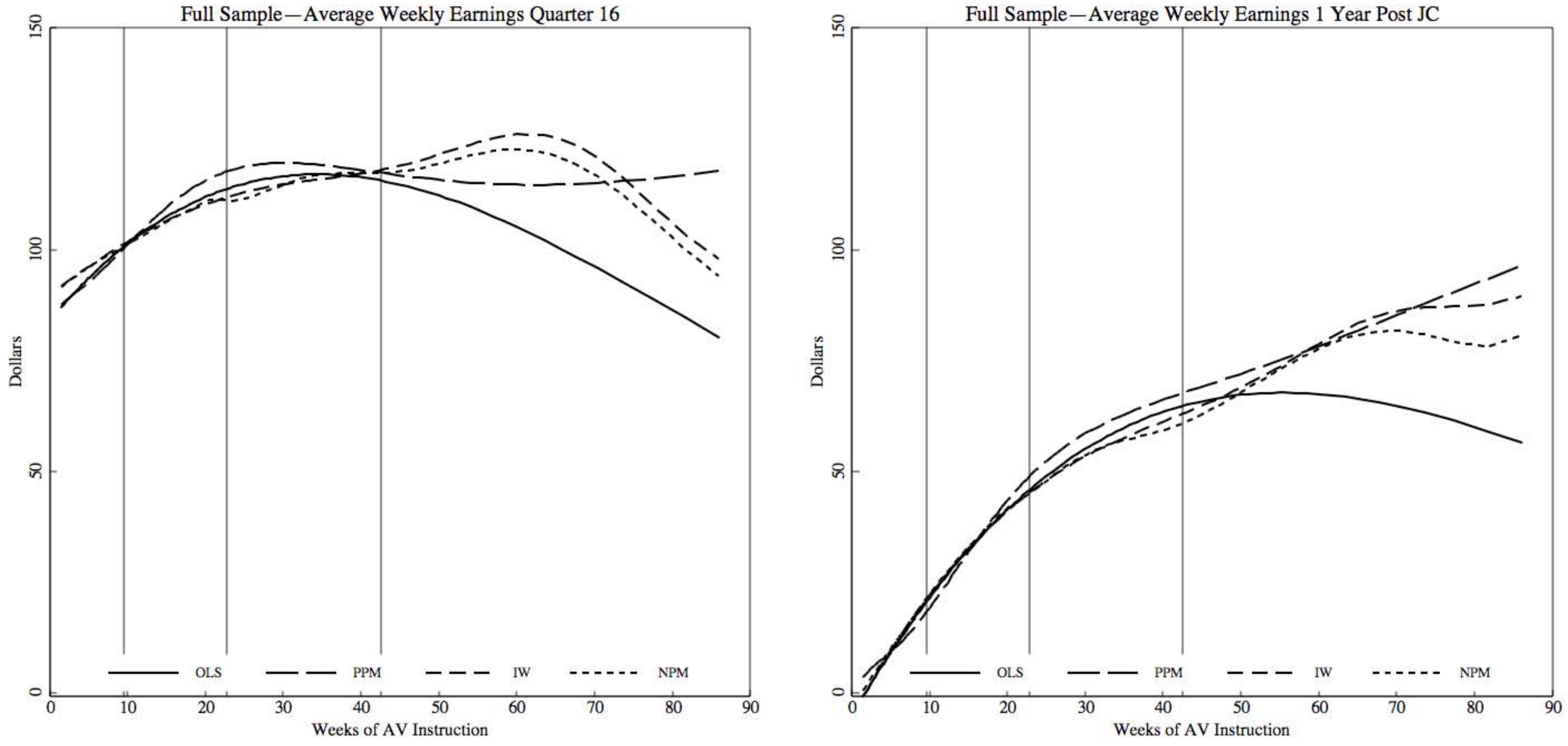

Figure 3. Estimated derivative of the DRF on weekly earnings in quarter 16 (in differences) using the IW estimator

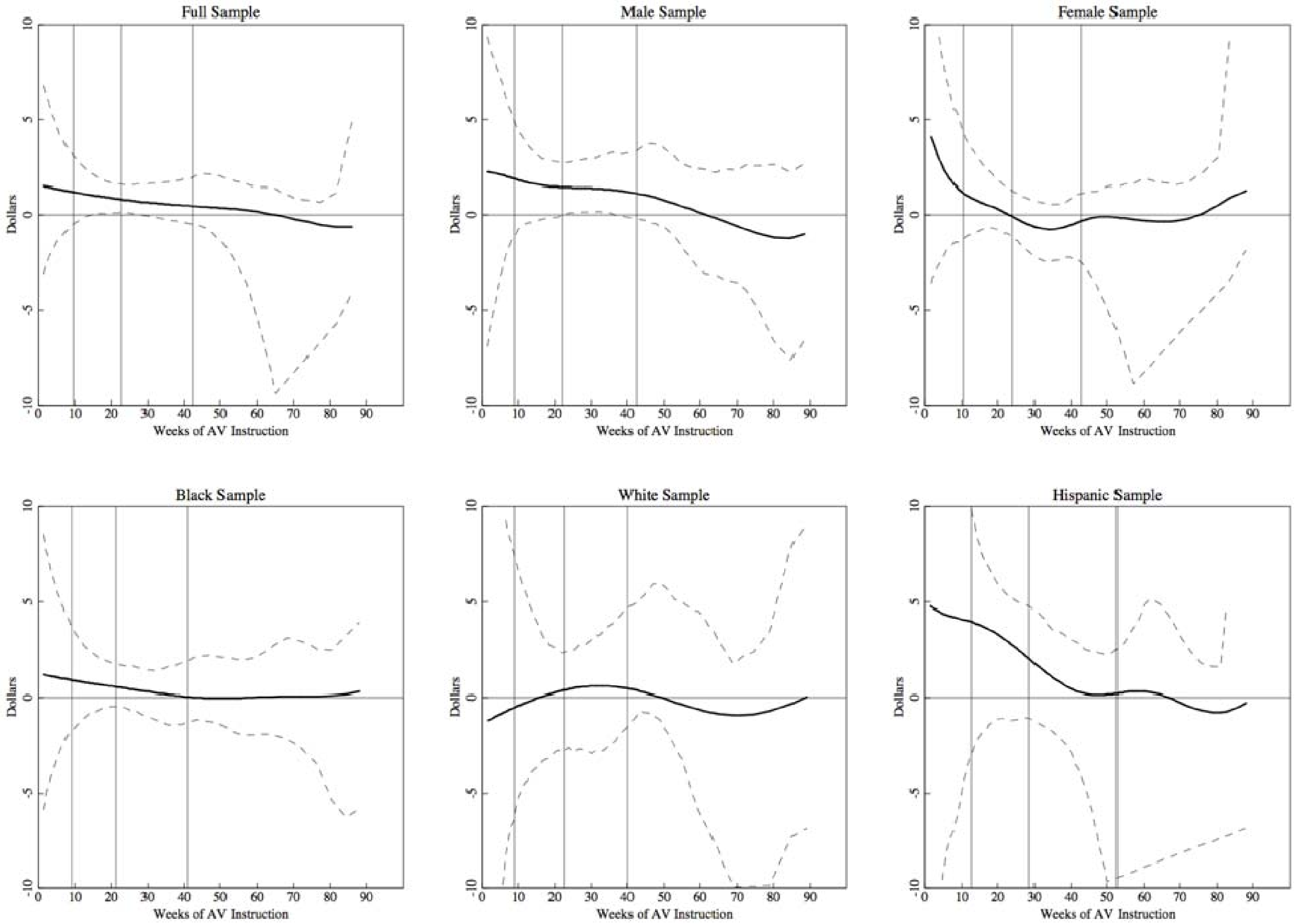

Table 3. Average Derivatives for Selected Quantiles Employing the IW Estimator

	Quantiles				
	1-99	**1-25**	**1-50**	**1-75**	**25-75**
Outcome: Weekly Earnings in Quarter 16 in differences					
Full	0.8 *	1.4	1.2	1.0 **	0.8 ***
Male	1.3 *	2.1	1.8	1.7 *	1.5 ***
Female	0.6	2.1	1.2	0.6	-0.1
Black	0.6	1.0	0.9	0.7	0.5
White	0.0	-0.7	-0.2	0.0	0.4
Hispanic	1.5	4.1	3.0	2.1	1.2
Outcome: Weekly Earnings One Year Post JC in differences					
Full	2.1 ***	4.2 ***	3.2 ***	2.5 ***	1.7 ***
Male	2.5 ***	4.9 **	4.0 ***	3.1 ***	2.3 ***
Female	1.3	2.3 *	1.7 **	1.3 **	0.8 **
Black	2.0 ***	4.1 **	3.0 ***	2.3 ***	1.4 ***
White	1.8 *	2.8	2.5	2.2 *	1.9 ***
Hispanic	1.3	3.4	2.1	1.6	0.8
Placebo Outcome: Total Earnings in the Three Quarters of the Year Prior to Randomization					
Full	4.4	24.8	14.2	9.1	1.6
Male	-3.3	5.2	6.4	0.4	-1.7
Female	0.2	-0.8	0.0	0.7	1.4
Black	5.4	23.9	15.8	9.4	2.5
White	-7.9	34.1	2.0	-2.3	-20.7
Hispanic	15.4	38.3	27.6	20.7	12.1

NOTE: *, **, and *** indicate statistically significant estimates at the 90, 95 or 99 percent level. These tests are computed using the bootstrap estimates employing the percentile

Figure 4. Estimated derivative of the DRF on weekly earnings one year post JC (in differences) using the IW estimator

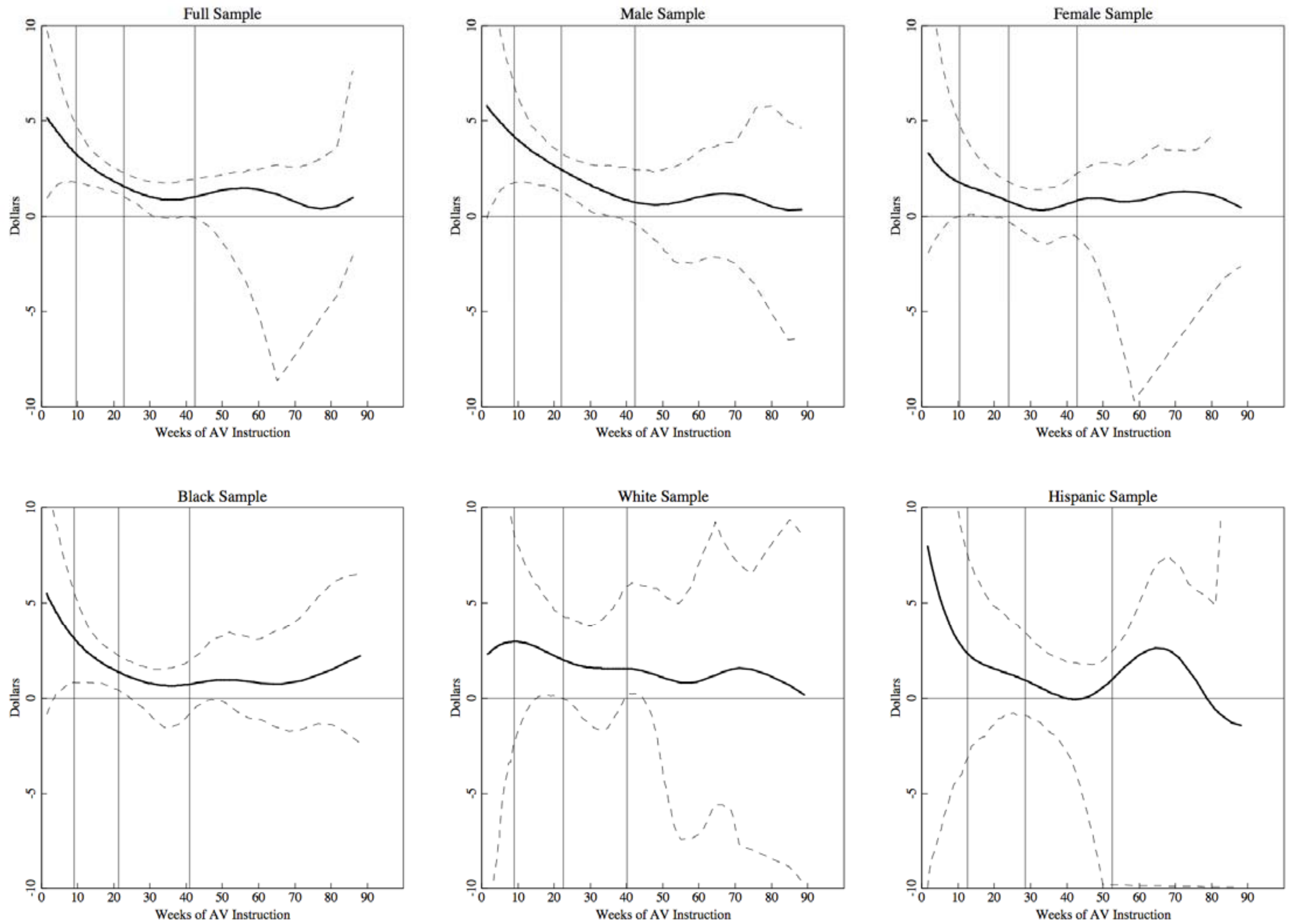

Figure 5. Estimated derivative of the DRF on a placebo outcome (total earnings in the first three quarters of the year prior to randomization) using the IW estimator