Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps

Xuan Chen* Carlos A. Flores†

August 2012

Abstract

Randomized and natural experiments are now common in economics and other social fields to estimate the effect of different programs or interventions. Even when employing experimental data, assessing the impact of the treatment on some outcomes of interest is often complicated by the presence of sample selection (outcomes are only observed for a selected group) and noncompliance with the assigned treatment (some members of the treatment group do not receive the treatment while some control individuals do). We address both identification problems simultaneously and derive nonparametric bounds for average treatment effects within a principal stratification framework. We employ these bounds to empirically assess the wage effects of the Job Corps (JC) program, which is the most comprehensive and largest federally-funded job training program for disadvantaged youth in the United States. Our results strongly suggest positive average effects of JC on wages for those individuals who comply with their treatment assignment and who would be employed whether or not they enrolled in JC (the "always-employed compliers"). In particular, under relatively weak monotonicity and mean dominance assumptions, we find that this average effect is between 5.7 and 13.9 percent four years after randomization, and between 7.7 and 17.5 percent for Non-Hispanics.

Key words and phrases: Nonparametric partial identification; Principal stratification; Instrumental variables

JEL classification: C13, C21, J30

*Department of Economics, University of Miami. Email: xchen11@miami.edu.

†Department of Economics, University of Miami. Email: caflores@miami.edu

1 Introduction

Randomized and natural experiments are now common in economics and other social fields to estimate the effect of different programs or interventions. Even when employing experimental data, assessing the impact of the treatment on some outcomes of interest is often made difficult by two critical identification problems: sample selection and noncompliance with the assigned treatment. Our leading example in the paper is the problem of evaluating the effect of the Job Corps training program on wages using data on individuals who were randomly assigned to participate or not in the program. In this case, the sample selection issue arises from the fact that wages are only observed for those who are employed, with the employment decision itself being potentially affected by the program. The noncompliance problem appears because some individuals who were assigned to participate in the program did not enroll, while some control individuals enrolled. In this paper, we derive nonparametric bounds for treatment effects in settings where both identification problems are present, and we employ these bounds to empirically assess the effect of the Job Corps training program on its participants' wages.

Job Corps (JC) is the most comprehensive and largest federally-funded job training program for disadvantaged youth in the United States. It provides academic, vocational and social skills training, among many other services, at over 120 centers throughout the country. Assessing the effect of this and similar programs on wages is of great importance to policy makers. Most of the econometric evaluations of training programs, however, focus on their impact on total earnings, which are the product of the hourly wage and the hours worked. As discussed by Lee (2009), focusing only on total earnings fails to answer the relevant question of whether the programs lead to an increase in the participants' wages (e.g., through human capital accumulation), or to an increase in the probability of being employed (e.g., through counseling and job search assistance services) without any increase in wages. We analyze the effect of JC on wages using data from the National Job Corps Study (NJCS), a randomized social experiment undertaken in the mid-to-late nineties to evaluate the effectiveness of JC.

Standard approaches for point identification of treatment effects in the presence of sample selection require strong parametric assumptions or the availability of a valid instrument (e.g., Heckman, 1979). In settings where an instrument is not available, an alternative strategy is to bound or partially identify the effects under relatively mild assumptions (Zhang and Rubin, 2003; Zhang, Rubin and Mealli, 2008; Imai, 2008; Lee, 2009; Lechner and Melly, 2010; Huber and Mellace, 2010). Part of this literature makes use of principal stratification (Frangakis and Rubin, 2002), which provides a framework for studying causal effects when controlling for a variable that has been affected by the treatment, in our setting, the employment decision. The basic idea behind principal stratification is to compare individuals within common principal strata, which are subpopulations consisting of individuals that share the same potential values of the employment variable under both treatment arms. Since membership to a particular principal stratum is not affected by treatment assignment, sample selection is not an issue within principal strata and the estimated effects are causal effects.

Zhang, Rubin and Mealli (2008) (hereafter ZRM) and Lee (2009) derive bounds for the average effect of a training program on wages for a particular stratum, the "always-employed", which consists of those individuals who would be employed whether or not they were assigned to enroll in the training program. They focus on this stratum because it is the only one for which the individuals' outcomes are observed under both treatment arms. Following Zhang and Rubin (2003), ZRM consider two assumptions and derive bounds for this effect under each of them and when both are imposed. The first assumption is a monotonicity assumption on the effect of the treatment (training program) on the selection (employment), and the second is a stochastic dominance assumption relating the potential outcomes of the always-employed to those of some of the other strata. Lee (2009) uses an alternative approach to that in ZRM to derive bounds under the monotonicity assumption. Importantly, the bounds derived in these papers do not impose the assumption that the support of the outcome is bounded. Lee (2009) uses his bounds to evaluate the wage effects of Job Corps. More recently, Blanco et al. (2012) employ the bounds in those two papers, along with their extension to quantile treatment effects by Imai (2008), to study the wage effects of Job Corps for different demographic groups without adjusting for noncompliance.

Huber and Mellace (2010) and Lechner and Melly (2010) derive bounds for subpopulations other than the always-employed. Huber and Mellace (2010) consider a setting similar to that in ZRM and Lee (2009) and, using a principal stratification approach, extend their results to construct bounds on the effects for two other strata (those who would be employed only if assigned to the treatment group; and those who would be employed only if assigned to the control group), as well as for the "selected" subpopulation, which consists of those individuals whose wages are observed and is a mixture of different strata. Their assumptions are similar to those in ZRM and Lee (2009); however, since bounds are constructed for strata and subpopulations for which the outcome is never observed under one of the treatment states, additional assumptions are required, such as bounded support of the outcome. Lechner and Melly (2010) derive bounds for mean and quantile treatment effects for the "treated and selected" subpopulation, which consists of employed individuals who received training and is also a mixture of different strata. Contrary to the previously described literature, they do not follow a principal stratification approach to derive their results. The assumptions they consider involve monotonicity assumptions on the effect of the training program on employment (conditional on covariates), as well as stochastic dominance assumptions involving observed subpopulations (e.g., employed versus unemployed) rather than involving different strata. Similar to Huber and Mellace (2010), they require an outcome with bounded support to partially identify mean effects.

The previously discussed literature, with the exception of Lechner and Melly (2010), focuses on the intention-to-treat (ITT) effect of being offered to participate in the training program. It compares potential outcomes according to the assigned treatment and ignores possible noncompliance behavior by individuals. The noncompliance issue is important in practice. For example, in the data set we employ in this paper, which is based on that previously used by Lee (2009), only 73.8 percent of the individuals who were assigned to the treatment group enrolled in JC, and 4.4 percent of the individuals who were assigned to the control group enrolled in JC in the four years after random assignment.

In this paper, we extend the partial identification results in ZRM and Lee (2009) to account for noncompliance. Thus, we bound the effect of actually *enrolling* in the program, rather than the effect of *being allowed* to enroll in the program. Our approach to account for noncompliance is based on the work by Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) (hereafter, AIR), who use an instrumental variable (IV) approach to address noncompliance behavior in the absence of sample selection. Their approach is a special case of principal stratification. Based on the individuals' potential compliance behavior under the two treatment assignments, they stratify the population into four strata, the so-called always-takers, never-takers, compliers and defiers. They show that in the presence of heterogeneous effects and under some assumptions (including a no-defiers assumption and an exclusion restriction), IV estimators point identify the average treatment effect for those individuals who comply with their treatment assignment (the compliers).

We employ principal stratification to address the sample selection and noncompliance problems simultaneously, and derive bounds for the average effect of participating in a training program on wages for the stratum of always-employed compliers. This stratum consists of those individuals who comply with their treatment assignment and who would be employed whether or not they enrolled in the training program. Analogous to the cases analyzed in Imbens and Angrist (1994), AIR, ZRM, Lee (2009), among others, this is the only stratum for which wages are observed for individuals who enrolled and did not enroll in the training program. In the context of analyzing the effects of Job Corps on wages, this is also the largest stratum (about 40 percent of the population).

Principal stratification has often been used to address a single post-treatment complication. To the best of our knowledge, this is the first paper deriving bounds for treatment effects within this framework accounting for more than one identification problem. There are, however, a few papers that employ principal stratification to point identify treatment effects in the presence of more than one complication. A particularly relevant paper in our setting is the one by Frumento et al. (2012), who analyze the effects of JC on employment and wages using data from the NJCS. They perform a likelihood-based analysis to simultaneously address three problems: sample selection, noncompliance and missing outcomes due to non-response. They stratify the population based on the potential values of the compliance behavior and employment status to address the noncompliance and sample selection issues, and they employ a "missing at random" assumption (Rubin, 1976) to address the missing-outcome problem.[1] Under some

[1] The missing at random assumption states that the probability that the outcome is missing for a given individual is random conditional on a set of observable characteristics.

parametric assumptions, Frumento et al. (2012) point identify the effect of JC on wages for the always-employed compliers.[2] Our paper complements the work by Frumento et al. (2012) by constructing nonparametric bounds for the effect of JC on wages based on an alternative set of assumptions. In the empirical section of the paper, we also present results that account for missing values due to non-response by using weights constructed by the NJCS using non-public use data that account for sample design and non-response.

The contribution of this paper is twofold. First, we add to the partial identification literature by deriving nonparametric bounds for average treatment effects in the presence of sample selection and noncompliance. More generally, the bounds we derive can be employed in settings where two identification problems are present and there is a valid instrument to address one of them. Some of these complications may include sample selection, noncompliance, endogeneity of the treatment variable, missing outcomes, among others. For example, when assessing the effect of military service on future health using the Vietnam-era draft lottery as an instrument to address the endogeneity issue (e.g., Angrist et al., 2009), our results could be used to bound the average effect on those individuals who enrolled in the military because of the draft lottery (compliers) and who would live at the time the outcome is measured regardless of their veteran status. Second, we contribute to the literature on the evaluation of the JC program (Schochet et al., 2001; Schochet et al., 2008; Lee, 2009; Flores-Lagunes et al., 2010; Flores et al., 2011; Frumento et al., 2012; Blanco et al., 2012) by empirically assessing the effect of JC on its participants' wages. Our results suggest greater positive average effects of JC on wages than those found without adjusting for noncompliance in Lee (2009) and Blanco et al. (2012). Under a monotonicity and a mean dominance assumptions, we find that the average effect of JC on wages for the always-employed compliers is between 5.7 percent and 13.9 percent four years after random assignment, and between 7.7 and 17.5 percent for Non-Hispanics.

The paper is organized as follows. Section 2 presents the econometric framework and the partial identification results, and discusses estimation and inference. Section 3 briefly describes JC and the data we use, and empirically analyzes the wage effects of JC. Section 4 concludes.

[2] Another assumption in Frumento et al. (2012) is that the individuals in the control group never enroll in JC, which rules out the existence of "always-takers". This assumption may not be plausible in our application, especially when looking at outcomes four years after random assignment.

2 Econometric Framework

2.1 Setup, Principal Strata, and Parameter of Interest

Assume we have a random sample of size n from a large population. For each unit i in the sample, let $Z_i = z \in \{0, 1\}$ indicate whether the unit was randomly assigned to the treatment group ($Z_i = 1$) or to the control group ($Z_i = 0$). In our application, Z_i specifies whether individual i was randomly assigned to participate in JC or not. We let $D_i = d \in \{0, 1\}$ indicate whether individual i actually received the treatment, i.e., enrolled in JC ($D_i = 1$), or not ($D_i = 0$). Similarly, we let $S_i = s \in \{0, 1\}$ be a post-treatment sample selection variable indicating whether the latent outcome variable Y_i^* is observed ($S_i = 1$) or not ($S_i = 0$). In our setting, S_i specifies whether individual i is employed or not, and Y_i^* is the offered market wage. The observed outcome variable is $Y_i = Y_i^*$ if $S_i = 1$, and Y_i is missing if $S_i = 0$.

Let $D(z)$ denote the potential compliance behavior as a function of the treatment assignment. In addition, let $S(z, d)$ and $Y^*(z, d)$ denote the potential values of the selection indicator and the potential latent outcome, respectively, as a function of the treatment assignment (z) and the treatment received (d). We observe $\{Z_i, D_i(Z_i), S_i(Z_i, D_i(Z_i))\}$ for all units, and $Y_i^*(Z_i, D_i(Z_i))$ for those with $S_i = 1$. Our notation implicitly imposes the stable unit treatment value assumption (SUTVA) (Rubin 1978, 1980, 1990), which is common in the literature and implies that the potential outcomes for each unit are unrelated to the treatment assignment and treatment receipt of the other individuals. To simplify the notation, in what follows we omit the subscript i unless deemed necessary for clarity.

We follow AIR and address the noncompliance issue by employing randomization into JC (Z) as an instrument for enrollment into JC (D). As in AIR, we impose the following assumptions.

Assumption 1 (*Randomly Assigned Instrument*). *$\{Y^*(z, d), S(z, d), D(z)\}$ is independent of Z, for all $z, d \in \{0, 1\}$.*

Assumption 2 *(Exclusion Restriction of Z). $Y^*(z, d) = Y^*(z', d) = Y^*(d)$ and $S(z, d) = S(z', d) = S(d)$, for all $z, d \in \{0, 1\}$.*

Assumption 3 *(Nonzero Average Effect of Z on D). $E[D(1) - D(0)] \neq 0$.*

Assumption 2 states that any effect of the instrument Z on the potential outcomes Y^* and on the potential sample selection indicator S must be via an effect of Z on the treatment D. In other words, this assumption prevents the instrument from having a direct effect on Y^* and S. In the context of our empirical application, Assumption 2 requires that randomization into JC affects potential wages and potential employment only through its effect on enrollment into JC. Assumption 2 allows us to write the potential variables $Y^*(z, d)$ and $S(z, d)$ as a function of the treatment d only.

Assumption 3 requires the instrument (e.g., randomization into JC) to have a non-zero average effect on the probability of actually receiving treatment (enrolling into JC). As in Imbens and Angrist (1994) and AIR, a valid instrument in our context should satisfy Assumptions 1, 2 and 3 simultaneously. An important difference with respect to the assumptions in those two papers, however, is that we require Z to be a valid instrument for both Y^* and S.

We derive our bounds for the effect of enrolling in JC on wages accounting for sample selection and noncompliance within a principal stratification framework (Frangakis and Rubin, 2002). This framework, which generalizes the approach in AIR, is useful for studying causal effects when controlling for post-treatment or intermediate variables, i.e., variables that have been affected by the treatment. The basic principal stratification with respect to a given post-treatment variable is a partition of the population into groups such that, within each group, all individuals share the same potential values of the post-treatment variable under each treatment arm. A principal effect is then defined as a comparison of potential outcomes within a given stratum. Since membership to a particular stratum is not affected by treatment assignment, individuals in that group are comparable and thus principal effects are causal effects.

The intermediate variables we want to control for are the compliance behavior (D) and the sample-selection (S) indicator. Thus, in our setting, the principal strata are defined by the joint potential values of $\{D(z = 0), D(z = 1)\} \times \{S(z = 0), S(z = 1)\}$. Following AIR, ZRM and Frumento et al. (2012), we define the following subpopulations: $a = \{i : D_i(0) = D_i(1) = 1\}$, the "always-takers"; $n = \{i : D_i(0) = D_i(1) = 0\}$, the "never-takers"; $c = \{i : D_i(0) = 0, D_i(1) = 1\}$, the "compliers"; $d = \{i : D_i(0) = 1, D_i(1) = 0\}$, the "defiers", as well as $EE = \{i : S_i(0) = S_i(1) = 1\}$, the "always-employed", those who would be employed regardless of treatment assignment; $NN = \{i : S_i(0) = S_i(1) = 0\}$, the "never-employed", those who would be unemployed regardless of treatment assignment for them; $NE = \{i : S_i(0) = 0, S_i(1) = 1\}$, those who would be employed only if assigned to the treatment group; $EN = \{i : S_i(0) = 1, S_i(1) = 0\}$, those who would be employed only if assigned to the control group.

In total, we have sixteen strata: $\{a, n, c, d\} \times \{EE, NN, NE, EN\}$. These strata are the same to those in Frumento et al. (2012), and they result from combining the strata employed in AIR to account for noncompliance with those used in ZRM to address sample selection.

An important characteristic of principal strata is that they are latent subpopulations, meaning that, in general, we cannot observe to which stratum each individual belongs to. Thus, additional assumptions are usually imposed to point or partially identify effects of interest by reducing the number of strata that exist in the population. Note that Assumption 2 implies that the following four strata do not exist: aNE, aEN, nNE and nEN. The reason is that for the individuals in these four strata there exists an effect of the treatment assignment (Z) on employment (S) that is not through their enrollment-in-JC status (since $D_i(1) = D_i(0)$), which contradicts the exclusion restriction of Z.

We also impose the following assumption, which was also employed by AIR and further reduces the number of existing strata.

Assumption 4 *(Individual-Level Monotonicity of D in Z). $D_i(1) \geq D_i(0)$ for all i.*

Assumption 4 rules out the existence of defiers, thus eliminating the strata dEE, dNN, dEN and dNE. In the context of our application, it eliminates the existence of individuals who would enroll in JC only if assigned to the control group. A necessary condition for this assumption to hold is that Z has a non-negative average effect on D, which can be falsified by the data. As we further discuss in Section 2.4, it is possible to relax Assumption 4 by letting the direction of the monotonicity be unknown.

Imbens and Angrist (1994) and AIR show that under Assumptions 1 to 4 instrumental variable estimators point identify the average treatment effect of D on Y^* (and S) for the compliers in the absence of sample selection. If sample selection is present, however, those assumptions are not enough to point identify the average effect of D on Y^*.

In this paper, our parameter of interest is the average treatment effect of D on wages for the "always-employed compliers" (i.e., the cEE stratum):[3]

$$\Delta = E[Y^*(1) - Y^*(0)|cEE] = E[Y(1) - Y(0)|cEE]. \quad (1)$$

As can be seen from the definition of the different subpopulations above, this stratum is the only one for which wages are observed for individuals who enrolled and did not enroll into JC after imposing Assumption 4. The parameter in (1) is also considered in Frumento et al. (2012). It is the average effect for the intersection of the subpopulation Imbens and Angrist (1994) and AIR focus on when accounting for noncompliance with that Lee (2009) and ZRM focus on when addressing sample selection. In our application, this stratum is the largest one in the population, accounting for about 40 percent. In the following sections we construct bounds for (1) by considering two more assumptions.

2.2 Bounds under an Additional Monotonicity Assumption

We derive bounds for (1) by extending the trimming procedure bounds in Zhang and Rubin (2003), ZRM and Lee (2009) to allow for noncompliance. We impose the following assumption.

Assumption 5 *(Individual-Level Monotonicity of S in D). $S_i(1) \geq S_i(0)$ for all i.*[4]

Assumption 5 states that there is a non-negative effect of D on S for every unit. In our application, it assumes that there is a non-negative effect of JC on employment for every individual. This assumption is similar to the monotonicity assumption employed in ZRM and Lee (2009); however, it differs in that we impose monotonicity of S in the actual treatment received (D) rather than in the treatment assigned (Z). This type of monotonicity assumptions has been employed in the partial identification literature to address problems other than sample selection (e.g., AIR, Manski and Pepper, 2000; Flores and Flores-Lagunes, 2010a). A testable implication of Assumption 5 is that the average effect of D on S for compliers, which is point identified under Assumptions 1 to 4, is non-negative. Similar to Assumption 4, and as further

[3] Note that since for compliers we have that $Z = D$, we can also interpret the stratum cEE as those compliers who would be always employed regardless of treatment *receipt*.

[4] Under Assumptions 1 to 4, Assumption 5 can be relaxed as "$S_i(1) \geq S_i(0)$ for all compliers" in deriving our bounds.

discussed in Section 2.4, it is possible to relax Assumption 5 by letting the direction of the monotonicity be unknown.

Assumption 5 rules out strata where the selection indicator S is negatively affected by D. From the strata remaining after imposing Assumptions 1 to 4, Assumption 5 rules out the existence of the cEN stratum. Therefore, under Assumptions 1 to 5 there are seven strata in the population: aEE, aNN, nEE, nNN, cEE, cNN and cNE. The relation between these seven strata and the observed groups defined by the values of $\{Z, D(Z), S(Z, D(Z))\}$ is given by the following table:

Table 1: Observed Groups and Principal Strata

		$Z=0$				$Z=1$	
		D				D	
		0	1			0	1
S	0	cNE, cNN, nNN	aNN	S	0	nNN	cNN, aNN
	1	cEE, nEE	aEE		1	nEE	cNE, cEE, aEE

Thus, while some observed groups are composed of only one stratum, some of them are mixtures of two or more strata. Under Assumptions 1 to 5 we can identify the proportion of each stratum in the population. Let π_k denote the proportion of stratum k in the population, and let $p_{ds|z} \equiv \Pr(D = d, S = s|Z = z)$ and $q_{s|z} \equiv \Pr(S = s|Z = z)$. Then, we have:

$$\pi_{aNN} = p_{10|0}; \pi_{aEE} = p_{11|0}; \pi_{nNN} = p_{00|1}; \pi_{nEE} = p_{01|1} \tag{2}$$

$$\pi_{cEE} = p_{01|0} - p_{01|1}; \pi_{cNN} = p_{10|1} - p_{10|0}; \pi_{cNE} = q_{1|1} - q_{1|0}.$$

In addition, we can write the mean outcomes for those observed cells with $S = 1$ as a function of mean potential outcomes for different strata. Letting $\overline{Y}^{zds} \equiv E[Y|Z = z, D = d, S = s]$, we have:

$$\overline{Y}^{011} = E[Y(1)|aEE] \tag{3}$$

$$\overline{Y}^{101} = E[Y(0)|nEE] \tag{4}$$

$$\overline{Y}^{001} = E[Y(0)|cEE]\frac{\pi_{cEE}}{\pi_{cEE} + \pi_{nEE}} + E[Y(0)|nEE]\frac{\pi_{nEE}}{\pi_{cEE} + \pi_{nEE}} \tag{5}$$

$$\overline{Y}^{111} = \frac{\pi_{cEE} \cdot E[Y(1)|cEE]}{\pi_{cEE} + \pi_{cNE} + \pi_{aEE}} + \frac{\pi_{cNE} \cdot E[Y(1)|cNE]}{\pi_{cEE} + \pi_{cNE} + \pi_{aEE}} + \frac{\pi_{aEE} \cdot E[Y(1)|aEE]}{\pi_{cEE} + \pi_{cNE} + \pi_{aEE}} \tag{6}$$

The average potential outcomes under treatment and control are point identified for the nEE and aEE strata, respectively. Moreover, it is possible to point identify $E[Y(0)|cEE]$ by combining equations (2), (4) and (5) to obtain:

$$E[Y(0)|cEE] = \frac{p_{01|0}}{p_{01|0} - p_{01|1}} \overline{Y}^{001} - \frac{p_{01|1}}{p_{01|0} - p_{01|1}} \overline{Y}^{101}. \tag{7}$$

Thus, one of the terms of Δ in (1) is point identified. However, the term $E[Y(1)|cEE]$ is not point identified because two of the conditional means in (6) are not point identified. Next, we construct bounds for $E[Y(1)|cEE]$ and Δ.

In a setting without noncompliance, Zhang and Rubin (2003), ZRM and Lee (2009) construct bounds for the non-point identified expectation of the potential outcome in the definition of their average effect based on a cell containing only two strata. To illustrate the main idea behind their "worst-case" bounds, suppose that there were no individuals in the aEE stratum, so that $\pi_{aEE} = 0$ and the cell $\{Z = 1, D = 1, S = 1\}$ contained only two strata, cEE and cNE. Then, $E[Y(1)|cEE]$ would be bounded from above (below) by the mean of Y for the fraction $\pi_{cEE}/(\pi_{cEE} + \pi_{cNE})$ of the largest (smallest) values of Y for those individuals in that cell. A key difference between the bounds derived in those papers and ours is that in our setting the bounds for $E[Y(1)|cEE]$ are derived from a cell containing three strata.

From equations (3) and (6) we can see that, although the observed mean $\overline{Y}^{111}$ is a weighted average of the mean potential outcome of $Y(1)$ for three strata, the mean $E[Y(1)|aEE]$ is point identified. Thus, our bounds are constructed by considering "worst-case" scenarios that exploit the information that $\overline{Y}^{011} = E[Y(1)|aEE]$. To motivate the way we construct our bounds, we can think of the problem as finding "worst-case" scenarios for $E[Y(1)|cEE]$ subject to the constraint that $\overline{Y}^{011} = E[Y(1)|aEE]$. Our strategy to derive bounds for $E[Y(1)|cEE]$ is to solve the unconstrained problem first, and then check whether the value of $E[Y(1)|aEE]$ implied by this solution can satisfy the constraint that $\overline{Y}^{011} = E[Y(1)|aEE]$. If the constraint can be satisfied, then the unconstrained solution is just the solution to the constrained problem. Otherwise, we impose the constraint first and then obtain the solution to the constrained problem.

We introduce additional notations to describe our bounds. Let y_r^{111} be the r-th quantile of Y in the cell $\{Z = 1, D = 1, S = 1\}$, and let

$$\overline{Y}(y_{r'}^{111} \leq Y \leq y_r^{111}) \equiv E[Y|Z = 1, D = 1, S = 1, y_{r'}^{111} \leq Y \leq y_r^{111}]. \tag{8}$$

Hence, $\overline{Y}(y_{r'}^{111} \leq Y \leq y_r^{111})$ gives the mean outcome in the cell $\{Z = 1, D = 1, S = 1\}$ for those outcomes between the r'-th and r-th quantiles of Y in that cell. Suppose that we want to derive the lower bound for $E[Y(1)|cEE]$. First, we consider the problem without the constraint and ignore the information about aEE. In this case, we can directly apply the existing trimming procedure in ZRM and Lee (2009) and bound $E[Y(1)|cEE]$ from below by the expected value of Y for the $\pi_{cEE}/p_{11|1}$ fraction of the *smallest* values of Y in the cell $\{Z = 1, D = 1, S = 1\}$, or, $\overline{Y}(Y \leq y_{\pi_{cEE}/p_{11|1}}^{111})$, where $p_{11|1} = \pi_{cEE} + \pi_{cNE} + \pi_{aEE}$. Next, we check whether this solution is consistent with the constraint that $\overline{Y}^{011} = E[Y(1)|aEE]$. To do this, we construct the "worst-case" scenario lower bound for $E[Y(1)|aEE]$, call it $LY_{1,aEE}$, implied by the unconstrained solution by assuming that all the observations that belong to the aEE stratum are at the bottom of the remaining observations in the cell $\{Z = 1, D = 1, S = 1\}$. This yields $LY_{1,aEE} = \overline{Y}(y_{\pi_{cEE}/p_{11|1}}^{111} \leq Y \leq y_{1-(\pi_{cNE}/p_{11|1})}^{111})$. If $LY_{1,aEE} \leq \overline{Y}^{011}$, the unconstrained solution is consistent with the constraint and the lower bound for $E[Y(1)|cEE]$ is $\overline{Y}(Y \leq y_{\pi_{cEE}/p_{11|1}}^{111})$. If $LY_{1,aEE} > \overline{Y}^{011}$, then the unconstrained solution is not consistent with $\overline{Y}^{011} = E[Y(1)|aEE]$. Intuitively, having $LY_{1,aEE} > \overline{Y}^{011}$ implies that some observations from the aEE stratum must be at the bottom $\pi_{cEE}/p_{11|1}$ fraction of the smallest values of Y in the cell $\{Z = 1, D = 1, S = 1\}$ and, thus, $\overline{Y}(Y \leq y_{\pi_{cEE}/p_{11|1}}^{111})$ is not a sharp lower bound for $E[Y(1)|cEE]$. In this case, we construct the "worst-case" scenario lower bound for $E[Y(1)|cEE]$ by placing all the observations in the aEE and cEE strata at the bottom of the distribution of Y in the cell $\{Z = 1, D = 1, S = 1\}$. Thus, if $LY_{1,aEE} > \overline{Y}^{011}$, the lower bound for $E[Y(1)|cEE]$, call it $LY_{1,cEE}$, is derived from the equation:

$$\overline{Y}(Y \leq y_{1-(\pi_{cNE}/p_{11|1})}^{111}) = \frac{\pi_{cEE}}{\pi_{cEE} + \pi_{aEE}} LY_{1,cEE} + \frac{\pi_{aEE}}{\pi_{cEE} + \pi_{aEE}} \overline{Y}^{011}, \tag{9}$$

where $\overline{Y}(Y \leq y^{111}_{1-(\pi_{cNE}/p_{11|1})})$ is the mean of Y for the $1-(\pi_{cNE}/p_{11|1})$ fraction of the smallest values of Y in the cell $\{Z=1, D=1, S=1\}$.

Note that the lower bound $LY_{1,cEE}$ derived from equation (9) does not yield a sharp lower bound for $E[Y(1)|cEE]$ if $LY_{1,aEE} \leq \overline{Y}^{011}$. For example, if $\overline{Y}^{011} = E[Y(1)|aEE]$ is large, so that it is impossible that all individuals from the aEE stratum are at the bottom $1 - (\pi_{cNE}/p_{11|1})$ fraction of the smallest values of Y in the cell $\{Z=1, D=1, S=1\}$, then the lower bound derived from (9) will be much lower than $\overline{Y}(Y \leq y^{111}_{\pi_{cEE}/p_{11|1}})$, the lower bound derived without using the information on $E[Y(1)|aEE]$. Intuitively, in this case the value of $\overline{Y}^{011} = E[Y(1)|aEE]$ is so large that it provides little information about the "worst-case" lower bound scenario for $E[Y(1)|cEE]$ (but it will provide valuable information for the upper bound of $E[Y(1)|cEE]$).

The upper bound for $E[Y(1)|cEE]$ is derived in a similar way to the lower bound, but now placing the observations in the corresponding strata in the upper part of the distribution of Y in the cell $\{Z=1, D=1, S=1\}$. Once we construct bounds for $E[Y(1)|cEE]$, they can be combined with the point identification of $E[Y(0)|cEE]$ in (7) to construct bounds for the average effect of the always-selected compliers, Δ in (1). The following proposition presents bounds for Δ under Assumptions 1 to 5.

Proposition 1 *If Assumptions 1 to 5 hold, then $L_{cEE} \leq \Delta \leq U_{cEE}$. L_{cEE} and U_{cEE} are lower and upper bounds for Δ given by:*

$$L_{cEE} = LY_{1,cEE} - \overline{Y}^{001}\frac{p_{01|0}}{p_{01|0}-p_{01|1}} + \overline{Y}^{101}\frac{p_{01|1}}{p_{01|0}-p_{01|1}}$$

$$U_{cEE} = UY_{1,cEE} - \overline{Y}^{001}\frac{p_{01|0}}{p_{01|0}-p_{01|1}} + \overline{Y}^{101}\frac{p_{01|1}}{p_{01|0}-p_{01|1}},$$

where

$$LY_{1,cEE} = \begin{cases} \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}), \ \textit{if} \ \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) \leq \overline{Y}^{011} \\ \overline{Y}(Y \leq y^{111}_{1-\alpha_{cNE}})\frac{q_{1|0}-p_{01|1}}{p_{01|0}-p_{01|1}} - \overline{Y}^{011}\frac{p_{11|0}}{p_{01|0}-p_{01|1}}, \ \textit{otherwise} \end{cases}$$

$$
\begin{aligned}
UY_{1,cEE} &= \begin{cases} \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}}), \ \textit{if} \ \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}}) \geq \overline{Y}^{011} \\ \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})\frac{q_{1|0}-p_{01|1}}{p_{01|0}-p_{01|1}} - \overline{Y}^{011}\frac{p_{11|0}}{p_{01|0}-p_{01|1}}, \quad \textit{otherwise} \end{cases} \\
\alpha_{cEE} &= \frac{\pi_{cEE}}{p_{11|1}} = \frac{p_{01|0}-p_{01|1}}{p_{11|1}}, \ \textit{and} \\
\alpha_{cNE} &= \frac{\pi_{cNE}}{p_{11|1}} = \frac{q_{1|1}-q_{1|0}}{p_{11|1}}.
\end{aligned}
$$

Proof. See Appendix.

2.3 Bounds under Mean Dominance

In this section we consider an assumption that narrows the bounds presented in Proposition 1. This assumption states that the mean of $Y(1)$ for the cEE stratum is greater or equal than the mean of $Y(1)$ for the cNE stratum.

Assumption 6 *(Mean Dominance).* $E[Y(1)|cEE] \geq E[Y(1)|cNE]$.

The intuition behind Assumption 6 is that we expect some strata to have more favorable characteristics and thus better potential outcomes than others. In our application, this assumption states that the mean potential outcome under treatment of the always-employed compliers is greater than or equal to that of those who would be employed only if they enrolled in JC. Assumption 6 implies a positive correlation between employment and wages, which is supported by standard economic models of labor supply. Zhang and Rubin (2003), ZRM (2008) and Huber and Mellace (2010) consider stochastic-dominance versions of Assumption 6. For example, in our setting, such an assumption would state that the potential outcome under treatment of the cEE stratum at any rank of the outcome distribution is weakly less than that of the cNE stratum. For our purposes, stochastic dominance is much stronger than needed.

Even though Assumption 6 is not directly testable, it is possible to get indirect evidence about its plausibility by comparing average baseline characteristics of the cEE and cNE strata that are closely related to the outcome of interest (e.g., values of the outcome prior to randomization). Assumption 6 is less likely to hold if these comparisons suggest that the cNE stratum has better characteristics at baseline than the cEE stratum. Under Assumptions 1 to 5 it is possible to point identify the average characteristics for all seven strata at baseline. This can be seen by noting that the observed average characteristics at baseline for each of the observed groups $\{Z, D, S\}$ in Table 1 is a weighted average of the average characteristics for the different strata (see, for reference, equations (5) and (6)), with the weights being point identified from (2). Based on these equations, we can estimate the average characteristics at baseline for all seven strata by solving an overidentified GMM problem. We implement this tool in Section 3.2 and provide further details in the Appendix.

The mean dominance assumption above tightens the bounds in Proposition 1 by increasing the lower bound on $E[Y(1)|cEE]$. To get the new lower bound note that, similar to equation (6), we can write

$$\overline{Y}^{111} = \frac{(\pi_{cEE} + \pi_{cNE}) \cdot E[Y(1)|cEE, cNE]}{\pi_{cEE} + \pi_{cNE} + \pi_{aEE}} + \frac{\pi_{aEE} \cdot E[Y(1)|aEE]}{\pi_{cEE} + \pi_{cNE} + \pi_{aEE}}, \tag{10}$$

where the stratum proportions and $E[Y(1)|aEE]$ are point identified. By Assumption 6 we have that $E[Y(1)|cEE] \geq E[Y(1)|cEE, cNE]$, which provides a lower bound for $E[Y(1)|cEE]$ that is greater than or equal to the one obtained in Proposition 1. The following proposition presents bounds for Δ under Assumptions 1 to 6.

Proposition 2 *If Assumptions 1 to 6 hold, then $L_{cEE} \leq \Delta \leq U_{cEE}$. L_{cEE} and U_{cEE} are lower and upper bounds for Δ, where U_{cEE} is equal to the upper bound for Δ given in Proposition 1 and L_{cEE} equals:*

$$L_{cEE} = LY_{1,cEE} - \overline{Y}^{001} \frac{p_{01|0}}{p_{01|0} - p_{01|1}} + \overline{Y}^{101} \frac{p_{01|1}}{p_{01|0} - p_{01|1}},$$

with

$$LY_{1,cEE} = \frac{p_{11|1}\overline{Y}^{111} - p_{11|0}\overline{Y}^{011}}{p_{11|1} - p_{11|0}}.$$

Proof. See Appendix.

2.4 Remarks

Remark 1. It is possible to relax the individual-level monotonicity assumptions we employ (Assumptions 4 and 5) by not requiring prior knowledge about their direction. This is closely related to the work by Shaikh and Vytlacil (2011), who derive bounds for average treatment effects on binary outcomes with a valid instrument by imposing monotonicity (or threshold crossing models) assumptions similar to those in Assumptions 4 and 5 without specifying the direction of the monotonicity (see also Bhattacharya el al., 2008 and Chen, Flores and Flores-Lagunes, 2012). In our setting, we can replace Assumptions 4 and 5 with the following assumptions.

Assumption 4' *(Individual-Level Monotonicity of D in Z, unknown direction). Either $D_i(1) \geq D_i(0)$ for all i or $D_i(1) \leq D_i(0)$ for all i.*

Assumption 5' *(Individual-Level Monotonicity of S in D, unknown direction). Either $S_i(1) \geq S_i(0)$ for all i or $S_i(1) \leq S_i(0)$ for all i.*

Assumption 7 $E[S|Z = 1] - E[S|Z = 0] \neq 0$.

Under Assumptions 1, 2, 3, 4', 5' and 7, it is possible to infer the directions of the monotonicity assumptions above from the observed data, and hence, derive bounds for the average effect of D on Y for either the cEE or dEE stratum by the procedure described in Section 2.2.

Remark 2. It is possible to construct bounds on Δ without Assumption 5, in which case the stratum cEN is not ruled out and appears in the observed cells $\{Z = 0, D = 0, S = 1\}$ and $\{Z = 1, D = 1, S = 0\}$ in Table 1. Although we can still point identify the proportions of the strata aEE, aNN, nEE and nNN, neither the proportions of the strata cEE, cNN, cNE and cEN nor the term $E[Y(0)|cEE]$ is now point identified. To construct bounds for Δ in this case, we can combine our approach in Section 2.3 with that followed by Zhang and Rubin (2003), ZRM, Imai (2008) and Huber and Mellace (2010) in a setting with sample selection but without the noncompliance issue. However, the bounds in this case will be wider than those presented in Proposition 1, and may result in uninformative bounds (see e.g., Blanco et al., 2012).

Remark 3. In the absence of Assumptions 5 and 6, the lower bound for Δ in Proposition 2 provides information for another parameter of interest, $ATE_{cEE,cNE}$, which is defined as the weighted average of Δ and the ATE for cNE. L_{cEE} in Proposition 2 can be viewed as the lower bound for $ATE_{cEE,cNE}$, under Assumptions 1 to 4 and the following assumption.

Assumption 5'' $E[Y(0)|cEE, cEN] \geqslant E[Y(0)|cEE, cNE]$, where $E[Y(0)|k_1, k_2]$ is the weighted average of $Y(0)$ between two strata k_1 and k_2.

This assumption states that the mean value of $Y(0)$ (i.e., the potential wage if individual did not attend JC) for compliers who would be employed if they did not attend JC (cEE & cEN) is greater than or equal to that for compliers who would be employed if they attended JC (cEE & cNE). This assumption exploits the positive correlation between employment and wages implied by standard models of labor supply, as the cEN are employed under the control treatment but the cNE are not. Indirect evidence regarding the plausibility of Assumption 5" can be obtained by comparing the weighted average baseline characteristics, $E[X|cEE, cEN]$ and $E[X|cEE, cNE]$, derived from the cells $\{Z = 0, D = 0, S = 1\}$ and $\{Z = 1, D = 1, S = 1\}$, respectively. In applications where Assumption 5 or 6 are difficult to justify, Assumption 5" may become attractive. Furthermore, the mixture of the strata cEE and cNE seems to be an interesting target group, since those are individuals who would comply with treatment assignment and who would be employed if attended JC.

Remark 4. In this paper we focus on the average treatment effect of the cEE stratum. It is possible to combine the methods in the previous sections with those in Huber and Mellace (2010) to construct bounds for the average effects of other subpopulations. For instance, we could consider the average effect of the treated and selected individuals (those with $D = 1$ and $S = 1$), or of the treated and selected compliers. As can be seen from Table 1, these other subpopulations are mixtures of different strata for which, with the exception of cEE, wages are unobserved under one treatment arm. Thus, additional assumptions (e.g., a bounded outcome) are needed to partially identify the effects for other strata or subpopulations.

2.5 Estimation and Inference

With the exception of the lower bounds in Proposition 2, our bounds involve minimum (min) and maximum (max) operators, which create complications for estimation and inference. As an illustration, the upper bound $UY_{1,cEE}$ for $E[Y(1)|cEE]$ in Proposition 1 can be written as $UY_{1,cEE} = \min\{\overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}}), \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})\frac{q_{1|0}-p_{01|1}}{p_{01|0}-p_{01|1}} - \overline{Y}^{011}\frac{p_{11|0}}{p_{01|0}-p_{01|1}}\}$. The first complication is that because of the concavity (convexity) of the min (max) function, sample analog estimators of the bounds can be severely biased in small samples. Another issue is that closed-form characterization of the asymptotic distribution of estimators for parameters involving min or max functions are very difficult to derive and, thus, usually unavailable. Furthermore, Hirano and Porter (2011) show that there exist no locally asymptotically unbiased estimators and no regular estimators for parameters that are nonsmooth functionals of the underlying data distribution, such as those involving min or max operators. These issues have generated a growing literature on inference methods for partially identified models of this type (see Tamer, 2010, and the references therein).

To address those issues, we employ the methodology proposed by Chernozhukov, Lee and Rosen (2011) that are applicable to bounds of the form $[\theta_0^l, \theta_0^u]$, where $\theta_0^l = \sup_{v \in \mathcal{V}} \theta^l(v)$, $\theta_0^u = \inf_{v \in \mathcal{V}} \theta^u(v)$, $\theta^l(v)$ and $\theta^u(v)$ are bounding functions, and $\mathcal{V}$ is the set over which the infimum and supremum are taken. They employ precision-corrected estimates of the bounding functions before applying the infimum and supremum operators. The precision adjustment consists of adding to each estimated bounding function its pointwise standard error times an appropriate critical value. Hence, estimates with higher standard errors require larger adjustments. Depending on the choice of the critical value, it is possible to obtain confidence regions for either the identified set or the true parameter value, as well as half-median unbiased estimators for the lower and upper bounds. The half-median-unbiasedness property means that the upper (lower) bound estimator exceeds (falls below) the true value of the upper (lower) bound with probability at least one half asymptotically. This property is important because achieving local asymptotic unbiasedness is impossible (Hirano and Porter, 2011). In the application, we report the half-median-unbiased estimators for our bounds and confidence regions for the true parameter values. The details on implementing this methodology are provided in the Appendix.

3 The Wage Effects of Job Corps

3.1 Job Corps and the National Job Corps Study

Job Corps (JC) is the largest and most comprehensive education and job training program in the United States. Since 1964 it has been a central part of the federal government efforts to provide job training and employment assistance to disadvantaged youth people. It offers academic education, vocational training, residential living, health care and health education, counseling and job placement assistance. These services are delivered at more than 120 centers nationwide. An eligible individual must be a legal resident of the United States, be between 16 and 24 years old and come from a low-income household. According to Department of Labor (2005), a typical JC student lives at a local JC center for eight months and receives about 1100 hours of academic and vocational instruction, which is equivalent to approximately one year in high school.

We employ data from the National Job Corps Study (NJCS), a randomized experiment undertaken in the mid-to-late nineties and funded by Department of Labor to evaluate the effectiveness of JC. Eligible individuals who applied for JC for the first time between November 1994 and December 1995 (80,833 individuals) were randomly assigned to a program, control, or program non-research group. Individuals in the control group (5,977) were embargoed from the program for a period of three years, while those in the program or treatment group (9,409) were allowed to enrolled in JC. The research sample was interviewed at random assignment and at 12, 30, and 48 months after random assignment.

The specific sample we use is based on the same sample from the NJCS employed by Lee (2009). This sample involves only individuals with non-missing values for weekly earnings and weekly hours worked for every week after random assignment (9,145 individuals). We construct our data set by adding enrollment information at week 208 (i.e., 48 months) after random assignment. This is a binary variable indicating whether or not the individual was ever enrolled in JC during the 208 weeks after random assignment. We drop 55 observations from Lee's sample due to missingness of the enrollment variable, resulting in a final sample size of 9,090 individuals (3,599 and 5,491 individuals in the control and treatment groups, respectively). Wages at week 208 are obtained by dividing weekly earnings by weekly hours worked at that week. We regard the individual as unemployed when the wage is missing, and regard the individual as employed otherwise. Finally, due to both design and programmatic reasons, some subpopulations were randomized in the NJCS with different (but known) probabilities (Schochet et al., 2001). Hence, we employ the NJCS design weights through our analysis.

The first four columns of Table 2 report the average baseline characteristics of our entire sample by treatment assignment status, along with the percentage of missing values for each of those variables. The pre-treatment variables include demographic characteristics, education and background variables, employment and earnings information at baseline, as well as labor market outcomes in the year prior to randomization. As one would expect, the average pre-treatment characteristics of the treatment and control groups are similar, with the difference in means being statistically different from zero at the five percent level for only one variable (weekly hours worked at baseline). Thus, our sample maintains the balance of baseline variables between the control and treatment groups.

The first three columns of Table 3 show the averages of some relevant post-treatment variables by treatment assignment status, along with their differences, at week 208 after randomization. The first row shows information on the enrollment-into-JC variable. By week 208, 73.8 percent of those assigned to the treatment group and 4.4 percent of the control individuals had ever enrolled in JC.[5] The difference of these two numbers, which equals the proportion of compliers in the population, is 69.4 percent. The rest of the rows in Table 3 present the intention-to-treat (ITT) effect and the (local) average treatment effect for compliers ($LATE$) of JC on various labor market outcomes at week 208. The second type of effects adjust for noncompliance by employing the randomly assigned treatment as an instrument for enrollment into JC. All the ITT effects of JC on weekly hours worked, weekly earnings and employment are positive and statistically significant. The $LATE$ estimates for those three variables are also positive and statistically significant, and they are larger than the ITT estimates by about 44, 44 and 50 percent, respectively. The estimated average effects of JC on earnings and employment for compliers is 39.9 dollars and 6 percentage points, respectively. These results are consistent with the findings in the NJCS (Burghardt et al., 2001).

For reference, Table 3 also shows the estimated ITT and $LATE$ effects of JC on ln(wage) for employed individuals. The $LATE$ estimate implies an average effect of JC on ln(wage) of about 5.4 percent for compliers. However, these estimates are biased because of sample selection.

3.2 Assessment of Assumptions

Assumptions 1 to 4 are commonly used in the literature to address noncompliance in experimental settings, and they have been previously used in the NJCS to estimate the effect of

[5] From these controls, 3.2 percent enrolled after the end of the embargo period, while 1.2 percent of them enrolled in the program despite the three-year embargo imposed on them.

JC on labor market outcomes that are not affected by sample selection (Burghardt et al., 2001; Schochet, 2001; Schochet et al., 2001). Thus, we concentrate our discussion on the plausibility of Assumptions 5 and 6.

Assumption 5, which is also employed in Lee (2009), states that there is a non-negative effect of JC on employment for every individual at week 208. This assumption seems plausible in our application given the objectives and services provided by JC (e.g., academic and vocational training, job search assistance). A testable implication of this assumption is that the $LATE$ for compliers of JC on employment is non-negative. As discussed above, this effect is positive and highly statistically significant.

There are two potential threats to the validity of Assumption 5. First, individuals who enroll in JC may be less likely to be employed while undergoing training than those who do not enroll, which is usually referred to as the "lock-in" effect (van Ours, 2004). Second, trained individuals may raise their reservation wages because of the JC training, which may lead them to reject some job offers that they would otherwise accept if they had not received training. Both potential threats, however, are likely to become less relevant in the long run, as trained individuals are no longer "locked-in" away from employment, and individuals who chose to remain unemployed in the short run because of raising their reservation wages find jobs in the long run. Consistent with this view, Schochet et al. (2001) and Lee (2009) find negative effects of JC on employment in the short run, and positive effects in the long run. Thus, we focus our analysis on wages at week 208 after random assignment, which is the latest wage measure available in the NJCS.

Based on a likelihood-based analysis, Frumento et al. (2012) provide evidence that there may be a positive proportion of compliers in the population for whom JC has a negative effect on employment at week 208, even though this proportion falls over time after randomization. Therefore, to further increase the plausibility of Assumption 5 in our application, we also consider a sample that excludes Hispanics. Hispanics were the only demographic group in the NJCS for which negative (although statistically insignificant) effects of JC on both employment and earnings were found (Schochet et al., 2001; Flores-Lagunes et al., 2010), and therefore, Assumption 5 may not be appropriate for them. The last set of columns in Tables 2 and 3 present summary statistics of pre- and post-treatment variables for the Non-Hispanics sample (7,529 individuals). As expected, the estimated ITT and $LATE$ effects of JC on labor outcomes for Non-Hispanics are stronger than those for the entire sample, with a statistically significant average effect on employment for compliers of 7.2 percentage points.

Assumption 6 states that the mean potential outcome under treatment of the always-employed compliers (cEE stratum) is greater than or equal to that of individuals who would be employed only if they enrolled in JC (cNE stratum). As discussed in Section 2.3, it is possible to indirectly shed some light on the plausibility of this assumption by comparing average baseline characteristics of the cEE and cNE strata that are likely to be highly correlated to wages at week 208. In Appendix Table 1 we present average characteristics of those two strata for the entire and Non-Hispanics samples, obtained by estimating the overidentified GMM procedure described in the Appendix. Focusing on the Non-Hispanics sample we find that, relative to individuals in the cNE stratum, individuals in the cEE stratum are more likely to be male and white, to have never been arrested at baseline, and to have better labor market outcomes the year before randomization. These differences, however, are not statistically different from zero. We conclude from these results that the data does not provide indirect evidence against Assumption 6, and that the point estimates of the differences suggest that the assumption is plausible.

3.3 Empirical Results

We start our analysis by bounding the average effect of being allowed to enroll in JC on wages (ITT effect) for the individuals who would be always employed regardless of treatment *assignment*, and then we bound the average effect of enrolling in JC on wages for those *compliers* who would be always employed regardless of treatment *receipt* (Δ in (1)). The first parameter is the one considered in Lee (2009) and ZRM, and it ignores the noncompliance issue. In their setting, the principal strata are EE, NN, NE and EN, where the last stratum is ruled out by assuming monotonicity of S in Z.

Table 4 presents bounds on the average ITT effect of JC on ln(wage) for always-employed individuals (EE stratum). The first column of Table 4 presents results for our entire sample.

The proportion of the always-employed individuals (EE) in the population equals 56.6 percent, and the proportion of those in the NE stratum (which equals the ITT effect of JC on employment) equals 4.1 percent. Under the monotonicity of S in Z assumption, the estimated lower and upper bounds for the ITT effect of JC on ln(wage) for the EE stratum are -0.022 and 0.100, respectively. These results are very similar to those obtained by Lee (2009).[6] As noted by Lee (2009), although the bounds include zero, they rule out plausible negative effects. Moreover, as also noted by Lee (2009), these particular lower bounds are based on the extreme and unintuitive assumption that wages are perfectly negatively correlated with the probability of being employed.[7] This is contradicted by standard models of labor supply, in which those individuals with higher wages are also more likely to be employed.

The last set of rows in Table 4 present the bounds on the ITT effect of JC on ln(wage) for the EE stratum under the mean dominance assumption that the average potential wage under $z = 1$ of the EE stratum is greater than that of the NE stratum. This assumption can be seen as a way to rule out the implausible extreme case mentioned above by implying a positive correlation between wages and employment. In this case, the estimated lower bound for the ITT effect for the EE stratum is now 0.038. Thus, under this additional assumption, the bounds are able to rule out a negative ITT effect of JC on wages, which illustrates the identifying power of this assumption. Table 4 also presents 95 percent confidence intervals, which are calculated based on the results from Imbens and Manski (2004) and asymptotically cover the true value of the parameter with 0.95 probability.[8] As above, while the 95 percent confidence intervals do not rule out a zero effect under the monotonicity assumption, they rule out negative effects once the mean dominance assumption is employed.[9]

[6] Our results are not numerically the same as those in Lee (2009) because he uses a transformed wage variable to calculate his bounds in order to minimize the effect of outliers. Specifically, the entire observed wage distribution was split into 20 percentile groups ($5^{th}, 10^{th}, \ldots, 95^{th}$), and then the mean wage within each of the 20 groups was assigned to each individuals. Our paper, however, uses the original wage variable. In addition, we drop 55 observations from Lee's sample because of missing enrollment information. For reference, the corresponding lower and upper bounds in Lee (2009) are $-.019$ and .093, respectively.

[7] Remember that to obtain the lower bound all the EE individuals are placed at the bottom of the wage distribution in the cell $\{Z = 1, S = 1\}$, which is a mixture of the EE and NE strata.

[8] Imbens and Manski (2004) confidence intervals are valid for the ITT effect of JC on wages, because the bounding functions don't involve min or max operators.

[9] Specifically, the Imbens and Manski (2004) confidence interval at 95% level is calculated from $(\widehat{\Delta}_{LB} - \overline{C}_n * \widehat{\sigma}_{LB}/\sqrt{N}, \widehat{\Delta}_{UB} + \overline{C}_n * \widehat{\sigma}_{UB}/\sqrt{N})$, where $\widehat{\sigma}_{LB} = \sqrt{\widehat{V(\widehat{\Delta}_{LB})}}$, $\widehat{\sigma}_{UB} = \sqrt{\widehat{V(\widehat{\Delta}_{UB})}}$, and $\overline{C}_n$ satisfies $\Phi(\overline{C}_n +$

The second column of Table 4 presents results for Non-Hispanics, for whom the monotonicity assumption of S in Z is more plausible. In general, the lower and upper bounds under the two sets of assumptions are larger for Non-Hispanics than for the entire population. Under the monotonicity and mean dominance assumptions, the estimated lower and upper bounds on the ITT effect of JC on ln(wage) for the EE stratum are 0.050 and 0.119, respectively, and the 95 percent confidence interval is $[0.029, 0.144]$.

Table 5 shows the estimation results for our parameter of interest, the average effect of enrolling in JC on ln(wage) for always-employed compliers. This table shows the estimated proportions, some relevant quantities used in the estimation of the bounds, and the unbiased half-median estimators for our bounds and CLR confidence intervals for true parameters under Assumptions 1 to 5 as well as under Assumptions 1 to 6. As above, the first column presents the results for the entire sample, and the second shows the results for Non-Hispanics. For both samples, the largest stratum is the cEE stratum, with an estimated proportion of almost 40 percent, while the stratum of always-employed always-takers (aEE) is the smallest stratum, with an estimated proportion of about 1.6 and 1.8 percent for the entire and the Non-Hispanics samples, respectively. The estimated proportion of always-takers ($\pi_{aEE} + \pi_{aNN}$) is 4.4 percent for the entire sample, while the proportion of never-takers is 26.2 percent. These proportions are slightly higher for the Non-Hispanics sample. All the estimated stratum proportions in Table 5 are statistically different from zero.

For the entire population, the estimated lower and upper bounds on the average effect of JC on ln(wage) for the cEE stratum under Assumptions 1 to 5 are -0.022 and 0.130, respectively, while the corresponding numbers for Non-Hispanics are -0.014 and 0.161. Given the weak effects of JC on labor market outcomes for Hispanics, it is not surprising that the bounds for Non-Hispanics cover a larger positive region than those for the overall population. For both samples, the estimated lower and upper bounds are larger than the corresponding bounds for the ITT effect presented in Table 4, especially the upper bound (e.g., for Non-Hispanics, 0.119 versus 0.161, or a 35.3 percent increase). From Tables 4 and 5, the positive region covered by the bounds on the effect of enrolling in JC on wages for the cEE stratum is larger than the

$\sqrt{N}(\widehat{\Delta}_{UB} - \widehat{\Delta}_{LB})/\max(\widehat{\sigma}_{LB}, \widehat{\sigma}_{UB})) - \Phi(-\overline{C}_n) = .95.$

positive region covered by the bounds on the ITT effect of JC on wages for the EE stratum. This suggests that the effects of JC on wages obtained by Lee (2009) were conservative, as the effect was weakened by noncompliance to the assigned treatment.

As in Lee (2009), we are not able to rule out zero effects of JC on wages at week 208 employing only the monotonicity assumption on the effect of JC on employment. However, as before, our lower bounds are constructed under the implausible "worst-case" scenario of a perfect negative correlation between employment and wages, which is contradicted by standard economic models. The mean dominance assumption we employ rules out this implausible extreme case and helps to increase the lower bound and narrow the bounds. The last set of rows in Table 5 show results when the monotonicity and mean dominance assumptions are both used. Under Assumptions 1 to 6, the estimated lower bound on the average effect of JC on ln(wage) for the cEE stratum is 0.055 for the entire population, and it is 0.074 for Non-Hispanics. Therefore, under all six assumptions, our results imply positive average effects of JC on wages for the cEE stratum in the entire and Non-Hispanics samples. These results also reinforce the notion that the ITT effects of JC on wages are likely to be lower than the effect of actually enrolling in JC on wages. Also note that, as already mentioned in Remark 3, the lower bound for Δ in Proposition 2 can be interpreted as the lower bound for $ATE_{cEE,cNE}$ under Assumptions 1 to 4 and 5".

We draw the following conclusions from our empirical analysis of the effects of JC on wages. First, our results strongly suggest a positive average effect of participating in JC on wages four years after random assignment for the always employed compliers. For Non-Hispanics, for whom the monotonicity assumption on the effect of JC on employment is more likely to hold, our estimated bounds under Assumptions 1 to 6 imply that, on average, JC increases the average wage (as oppose to ln(wage)) of the always-employed compliers who participate in JC by between 7.7 and 17.5 percent. Therefore, this evidence suggests that JC has an effect on the earnings of its participants not only by increasing their probability of being employed but also by increasing their wages, which is most likely a consequence of their human capital accumulation during enrollment in JC.

Second, our analysis implies that the results from the study of the ITT effects of JC on wages (e.g., Lee, 2009) are conservative, in the sense that the noncompliance issue is likely to dilute the effect of actually enrolling in JC on wages. In particular, we find that for the two samples we consider, and regardless of whether or not we employ Assumption 6, the positive region covered by the bounds on the effect of enrolling in JC on wages for the cEE stratum is larger than the positive region covered by the bounds on the ITT effect of JC on wages for the EE stratum. This is consistent with the results presented in Section 3.1 regarding the effect of JC on other labor market outcomes not suffering from sample selection, which show that the $LATE$ estimates of the effects are larger than the ITT estimates. This conclusion is also consistent with the literature on point estimation of the wage effects of JC.[10]

Finally, to analyze the sensitivity of our results to the fact that some observations are dropped because of missing values of relevant variables, and to compare our results to those in Frumento et al. (2012), we estimate bounds of our effect of interest that account for this issue. We do this by employing another weight constructed by the NJCS using non-public data that accounts for both sample design and non-response.[11] Thus, the key assumption is that the probability that the information is missing for a given individual is random conditional on the set of variables used to construct the weight. Frumento et al. (2012) employ the same assumption but condition on variables available in the public version of the NJCS data. To construct the data used in this exercise, we take all those individuals who responded to the 48-month interview and we drop those with missing values of weekly working hours, weekly earnings or enrollment information.[12]

The estimation results are presented in Table 6. the estimated lower and upper bounds for the effect of JC on ln(wages) for the cEE stratum are below those estimated in Table 5, which ignore the non-response issue. Focusing on Non-Hispanics, the positive region covered by the bounds under both sets of assumptions (monotonicity and mean dominance) is slightly less

[10] Frumento et al. (2012) find that their point estimate of the effect of enrolling in JC on wages at week 208 for the always-employed compliers is larger than the point estimate of the ITT effect in Zhang et al. (2009), who estimate the effect of JC on wages for the always-employed individuals without adjusting for non-compliance.

[11] More specifically, the weights we employ address sample design, 48-month interview design and 48-month interview non-response.

[12] As discussed in Section 3.1, the sample used in the previous tables includes only individuals with non-missing values for weekly earnings and weekly hours worked for every week after random assignment. This is done to make the results comparable to those in Lee (2009).

than the corresponding region covered by the bounds in Table 5. The estimated upper bound in Table 6 equals 0.153, which is close to the one presented in Table 5 (0.161). The estimate of the lower bound in Table 6 is also lower than, but still relatively close to, that presented in Table 5. In general, we conclude that, although adjusting for non-response slightly weakens our previous findings, the results still strongly suggest a positive average effect of enrolling in JC on wages four years after random assignment for the always-employed compliers.

As we previously mentioned, Frumento et al. (2012) point identify the average effect of enrolling in JC on wages adjusting for sample selection, non-compliance and missing outcomes by imposing a different set of assumptions from the two sets we consider. Employing a different sample from the one we use, they estimate this effect to be about 3.8 percent for the always-employed compliers in the population. This point estimate is consistent with the estimated bounds for this effect presented in Table 6 for the entire population under our Assumptions 1 to 5 and, although 3.8 is below the estimated lower bound under Assumptions 1 to 6 (4.4 percent), it falls inside the 95 percent confidence interval constructed under our six assumptions. Thus, the point estimate of the effect of JC on wages for the cEE stratum in Frumento et al. (2012) is consistent with our bounds adjusting for non-response.

4 Conclusion

This paper derives nonparametric bounds for average treatment effects in the presence of both sample selection and noncompliance under relatively weak assumptions, and it employs these bounds to empirically assess the effects of participating in Job Corps on wages. The first contribution of the paper is to extend the partial identification results in Zhang and Rubin (2003), Zhang et al. (2008) and Lee (2009), which address the sample selection problem, to settings where both sample selection and noncompliance problems are present. More generally, our results can be used in settings where there are two identification problems (e.g., endogeneity, missing outcomes) and there is a valid instrument to address one of them. Thus, our paper provides an important extension of the work previously mentioned, as it is common in empirical work to face more than one identification problem. Moreover, the bounds developed in those papers are becoming increasingly used in the partial identification literature to address other problems. For instance, Huber and Mellace (2011) employ those bounds to test implications of the exclusion restriction assumption in just-identified models. Similarly, Flores and Flores-Lagunes (2010a, 2010b) use those bounds as a building block to construct bounds on local average treatment effects ($LATE$) in instrumental variable models when the exclusion restriction is violated, and to derive bounds on the part of the effect of a treatment on an outcome that works through a given mechanism (i.e., direct or net effects). The bounds derived in the present paper can be employed to extend the results in those papers to address more than one complication.

The second contribution of the paper is to analyze the wage effects of participating in Job Corps, which is one of the largest federally-funded job training programs in the United States. Our results strongly suggest a positive average effect of participating in Job Corps on wages four years after random assignment for those individuals who comply with their treatment assignment and who would be employed whether or not they enrolled in Job Corps (the always-employed compliers). Our results also suggest larger positive effects of Job Corps on wages than those found without adjusting for noncompliance in Lee (2009) and Blanco et al. (2012).

The two key assumptions we consider are a monotonicity assumption on the effect of Job Corps on employment, and a mean dominance assumption stating that the average potential wage under treatment of the always-employed compliers is greater than that of those compliers who would be employed only if they participated in Job Corps. We argue in the paper that both assumptions are likely to hold, especially for Non-Hispanics. Under both assumptions, we find that for Non-Hispanics always-employed compliers the estimated bounds on the average effect of participating in Job Corps on wages is between 7.7 and 17.5 percent. This evidence suggests that Job Corps has positive effects not only on the employability of its participants but also on their wages, implying that Job Corps is likely to have positive effects in their human capital. Therefore, it is very important to consider the potential benefits of Job Corps and other training programs on wages when evaluating their effects.

References

[1] Angrist, J., Chen, S. and Frandsen, B. (2009), "Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health", NBER Working Paper No. 14781.

[2] Angrist, J., Imbens, G. and Rubin, D. (1996), "Identification of Causal Effects Using Instrumental Variables", Journal of the American Statistical Association, 91, 444-472.

[3] Bhattacharya, J., Shaikh, A. and Vytlacil, E. (2008), "Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization", American Economic Review: Papers & Proceedings, 98 (2), 351-356.

[4] Blanco, G., Flores, C. and Flores-Lagunes, A. (2012), "Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages", forthcoming, Journal of Human Resources.

[5] Burghardt, J., Schochet, P.Z., McConnell, S., Johnson, T., Gritz, R.M., Glazerman, S., Homrighausen, J. and Jackson, R. (2001), Does Job Corps Work? Summary of the National Job Corps Study, 8140-530, Mathematica Policy Research, Inc., Princeton, NJ.

[6] Chen, X., Flores, C. and Flores-Lagunes, A. (2012), "Bounds on Population Average Treatment Effects with an Instrumental Variable".

[7] Chernozhukov, V., Lee, S. and Rosen, A. (2011), "Intersection Bounds: Estimation and Inference", last revised November 2011.

[8] Flores, C. and Flores-Lagunes, A.(2010a), "Partial Identification of Local Average Treatment Effects with an Invalid Instrument", Working Paper.

[9] Flores, C. and Flores-Lagunes, A. (2010b), "Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects", Working Paper.

[10] Flores, C., Flores-Lagunes, A., Gonzalez, A. and Neumann, T. (2011), "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps", forthcoming, The Review of Economics and Statistics.

[11] Flores-Lagunes, A., Gonzalez, A. and Neumann, T. (2010), "Learning but not Earning? The Impact of Job Corps Training on Hispanic Youth", Economic Inquiry, 48 (3), 651-667.

[12] Frangakis, C.E. and Rubin D. (2002), "Principal Stratification in Causal Inference", Biometrics, 58, 21-29.

[13] Frumento, F., Mealli, F., Pacini, B. and Rubin, D. (2012), "Evaluating the Effect of Training on Wages in the Presence of Noncompliance and Missing Outcome Data", forthcoming, Journal of the American Statistical Association.

[14] Heckman, J. (1979), "Sample Selection Bias as a Specification Error", Econometrica, 47, 153-162.

[15] Hirano, K. and Porter, J. (2011), "Impossibility Results for Nondifferentiable Functionals", forthcoming, Econometrica.

[16] Horowitz, J.L. and Manski, C.F. (1995), "Identification and Robustness with Contaminated and Corrupted Data", Econometrica, 63 (2), 281-302.

[17] Huber, M. and Mellace G. (2010), "Sharp Bounds on Average Treatment Effects under Sample Selection", Working Paper.

[18] Imai, K. (2008), "Sharp Bounds on the Causal Effects in Randomized Experiments with 'Truncation-by-Death'", Statistics & Probability Letters, 78, 141-149.

[19] Imbens, G. and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects", Econometrica, 62 (2), 467-475.

[20] Imbens, G. and Manski, C.F. (2004), "Confidence Intervals for Partially Identified Parameters", Econometrica, 72 (6), 1845-1857.

[21] Lechner, M. and B. Melly (2010), "Partial Identification of Wage Effects of Training Programs", Working Paper.

[22] Lee, D. (2009), "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", Review of Economic Studies, 76, 1071-102.

[23] Manski, C. and Pepper, J. (2000), "Monotone Instrumental Variables: with an Application to the Returns to Schooling", Econometrica 68 (4), 997-1010.

[24] Newey, W. and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing", Handbook of Econometrics, vol. IV, Chapter 36, 2111-2245.

[25] Rubin, D. (2005), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions", Journal of the American Statistical Association, 100, 322-331.

[26] Schochet, P. (2001), National Job Corps Study: Methodological Appedixes on the Impact Analysis, 8140-530, Mathematica Policy Research, Inc., Princeton, NJ.

[27] 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.

[28] Schochet, P., Burghardt, J. and McConnell, S. (2008), “Does Job Corps Work? Impact Findings from the National Job Corps Study”, American Economic Review, 98 (5), 1864-1886.

[29] Shaikh, A. and Vytlacil, E. (2005), "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis", Technical Working Paper 307, National Bureau of Economic Research.

[30] Shaikh, A. and Vytlacil, E. (2011), "Partial Identification in Triangular Systems of Equations with Binary Dependent Variables", Econometrica, 79 (3), 949-955.

[31] Tamer, E., 2010. Partial identification in econometrics. Annual Review of Economics 2, 167-195.

[32] van Ours, J. (2004), "The Locking-in Effect of Subsidized Jobs", Journal of Comparative Economics, 32, 37-55.

[33] Zhang, J.L. and Rubin, D. (2003), "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by ‘Death’", Journal of Educational and Behavioral Statistics, 28 (4), 353-368.

[34] Zhang, J.L., Rubin, D. and Mealli, F. (2008), "Evaluation the Effects of Job Training Programs on Wages Through Principal Stratification, in D. Millimet et al. (eds) Advances in Econometrics, XXI, Elsevier.

[35] Zhang, J.L., Rubin, D. and Mealli, F. (2009), "Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification.", Journal of the American Statistical Association, 104 (485), 166-176.

Table 2: Summary Statistics of Baseline Variables

	Entire Sample				Non-Hispanics			
	Missing Prop.	$Z=1$	$Z=0$	Diff.(Std.Err.)	Missing Prop.	$Z=1$	$Z=0$	Diff.(Std.Err.)
Female	0	.454	.458	-.004 (.011)	0	.454	.454	-.010 (.012)
Age at Baseline	0	18.44	18.35	.087 (.046**)	0	18.44	18.34	.096 (.050**)
White, Non-hispanic	0	.265	.263	.002 (.009)	0	.319	.318	.001 (.011)
Black, Non-Hispanic	0	.494	.491	.003 (.011)	0	.595	.593	.002 (.012)
Hispanic	0	.169	.172	-.003 (.008)	–	–	–	–
Other Race/Ethnicity	0	.072	.074	-.002 (.006)	0	.087	.089	-.003 (.007)
Never married	.017	.917	.916	.002 (.006)	.018	.926	.924	.002 (.006)
married	.017	.020	.023	-.003 (.003)	.018	.015	.018	-.003 (.003)
Living together	.017	.039	.040	-.002 (.004)	.018	.035	.037	-.002 (.004)
Separated	.017	.024	.021	.003 (.003)	.018	.023	.020	.003 (.003)
Has Child	.007	.189	.193	-.004 (.008)	.006	.187	.190	-.004 (.009)
Number of children	.010	.270	.268	.002 (.014)	.180	.269	.271	-.003 (.015)
Personal Education	.018	10.12	10.11	.013 (.033)	.018	10.14	10.12	.022 (.036)
Mother's Education	.188	11.49	11.46	.030 (.061)	.182	11.81	11.83	-.021 (.055)
Father's Education	.383	11.40	11.54	-.145 (.077**)	.379	11.72	11.86	-.147 (.072*)
Ever Arrested	.017	.248	.249	-.001 (.009)	.017	.255	.257	-.002 (.010)
Household Inc.: <3000	.358	.253	.251	.001 (.012)	.357	.248	.244	.004 (.013)
3000-6000	.358	.205	.208	-.003 (.011)	.357	.202	.213	-.012 (.012)
6000-9000	.358	.117	.114	.003 (.009)	.357	.119	.105	.015 (.009)
9000-18000	.358	.246	.245	.001 (.011)	.357	.244	.248	-.003 (.013)
>18000	.358	.180	.182	-.002 (.010)	.357	.187	.191	-.003 (.012)
Personal Inc.: <3000	.079	.788	.789	-.001 (.009)	.077	.787	.788	-.001 (.010)
3000-6000	.079	.128	.131	-.003 (.008)	.077	.129	.136	-.007 (.008)
6000-9000	.079	.053	.046	.007 (.005)	.077	.052	.043	.009 (.005**)
>9000	.079	.031	.034	-.003 (.004)	.077	.031	.033	-.001 (.004)
At Baseline:								
Have job	.021	.198	.192	.007 (.009)	.021	.204	.187	.017 (.009**)
Weekly hours worked	0	21.83	20.91	.922 (.447*)	0	21.97	20.75	1.216 (.491*)
Weekly earnings	0	111.08	102.89	8.183 (5.134)	0	107.79	102.28	5.516 (2.804*)
Had job, Prev. Yr.	.017	.635	.627	.008 (.010)	.017	.642	.627	.015 (.011)
Months Employed,Prev.Yr.	0	3.603	3.530	.074 (.091)	0	3.654	3.512	.143 (.100)
Earnings, Prev.Yr.	.062	2911.0	2810.5	100.56 (117.58)	.064	2900.3	2794.7	105.57 (106.34)
Numbers of observations	9090	5491	3599		7529	4551	2978	

Note: Z denotes whether the individual was randomly assigned to participate ($Z = 1$) or not ($Z = 0$) in the program. Numbers in parentheses are standard errors. * and ** denote that difference is statistically different from 0 at 5% and 10% level, respectively. Computations use design weights.

Table 3: Summary Statistics of Post-treatment Variables at Week 208 by Random Assignment

	Entire Sample			Non-Hispanics		
	$Z = 1$	$Z = 0$	Difference (Std.Err.)	$Z = 1$	$Z = 0$	Difference (Std.Err.)
Enrollment Variable						
Ever enrolled in JC	73.82%	4.42%	69.41% (.007*)	73.66%	4.76%	68.90% (.008*)
Intention-to-Treat (ITT) Effects						
Hours per week	27.80	25.83	1.967 (.559*)	28.05	25.53	2.523 (.617*)
Earnings per week	228.19	200.50	27.69 (5.121*)	230.22	194.66	35.57 (5.555*)
Employed	.607	.566	.041 (.011*)	.609	.559	.050 (.012*)
ln(wage)	2.029	1.991	.038 (.011*)	2.028	1.977	.050 (.013*)
Local ATE for Compliers, LATE (IV estimates)						
Hours per week			2.834 (.782*)			3.661 (.867*)
Earnings per week			39.90 (6.457*)			51.62 (6.960*)
Employed			.060 (.015*)			.072 (.017*)
ln(wage)			.054 (.016*)			.073 (.017*)

Note: Z denotes whether the individual was randomly assigned to participate ($Z = 1$) or not ($Z = 0$) in the program. Numbers in parentheses are standard errors. * and ** denote that difference is statistically different from 0 at 5% and 10% level, respectively. Computations use design weights. The standard error of the effect for employed compliers is calculated by a ML estimator, where the endogenous dummy variable is the treatment receipt indicator. The treatment assignment indicator is used as the exclusion restriction and all baseline characteristics (where mean values were imputed for missing values) are included in both the selection and outcome equations.

Table 4: Bounds for ITT Effects on ln(wage) at Week 208

	Entire Sample	Non-Hispanics
Proportions of strata:		
Always-employed(EE)	.566 (.009*)	.559 (.009*)
Never-employed(NN)	.393 (.007*)	.391 (.007*)
Employed only if assigned to program(NE)	.041 (.011*)	.050 (.012*)
$E[Y(Z=0)\|EE]$	1.991 (.009*)	1.977 (.010*)
Proportion of EE in cell$\{Z=1, S=1\}$	.932 (.017*)	.918 (.019*)
Bounds with Monotonicity		
Lower bound for the ITT effect for EE stratum	-.022 (.016)	-.018 (.017)
Upper bound for the ITT effect for EE stratum	.100 (.014*)	.119 (.015*)
Imbens and Manski 95% Confidence Interval	[-.048, .123]	[-.047, .144]
Bounds with Monotonicity and Mean Dominance		
Lower bound for the effect for EE stratum	.038 (.012*)	.050 (.013*)
Upper bound for the ITT effect for EE stratum	.100 (.014*)	.119 (.015*)
Imbens and Manski 95% Confidence Interval	[.019, .123]	[.029, .144]

Note: Numbers in parentheses are standard errors. * denotes that difference is statistically different from 0 at 5% level. Computations use design weights. The standard error is calculated by 5,000-repetition bootstrap. Imbens and Manski 95% confidence interval is calculated as $(\widehat{\Delta}_{LB}-1.645*\widetilde{\sigma}_{LB}, \widehat{\Delta}_{UB}+1.645\widetilde{\sigma}_{UB})$, where $\overline{C}_n = 1.645$ and $\widetilde{\sigma}_{LB}$ and $\widetilde{\sigma}_{UB}$ are calculated by bootstrap.

Table 5: Bounds for the Effects of Job Corps on ln(wage) for *cEE* Stratum at Week 208

	Entire Sample	Non-Hispanics
π_{aEE}	.016 (.002*)	.018 (.002*)
π_{nEE}	.158 (.005*)	.160 (.005*)
π_{cEE}	.391 (.010*)	.381 (.011*)
π_{cNE}	.041 (.011*)	.050 (.012*)
π_{aNN}	.028 (.003*)	.030 (.003*)
π_{nNN}	.104 (.004*)	.104 (.005*)
π_{cNN}	.261 (.007*)	.258 (.008*)
α_{cEE}	.872 (.023*)	.849 (.025*)
$E[Y(1)\|aEE]$	2.033 (.059*)	2.016 (.061*)
$E[Y(0)\|nEE]$	2.033 (.016*)	2.033 (.017*)
$E[Y(0)\|cEE]$	1.972 (.015*)	1.952 (.016*)
$\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$	2.429 (.066*)	2.376 (.057*)
$\overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$	1.676 (.034*)	1.703 (.035*)
Bounds with Monotonicity (Proposition 1)		
$[LY_{1,cEE}, UY_{1,cEE}]$	[1.951, 2.102]	[1.938, 2.113]
CLR 95% level confidence interval	(1.921, 2.128)	(1.907, 2.140)
$[L_{cEE}, U_{cEE}]$	[-.022, .130]	[-.014, .161]
CLR 95% level confidence interval	(-.061, .168)	(-.057, .201)
Bounds with Monotonicity and Mean Dominance (Proposition 2)		
$[LY_{1,cEE}, UY_{1,cEE}]$	[2.027, 2.102]	[2.026, 2.113]
CLR 95% level confidence interval	(2.011, 2.129)	(2.008, 2.141)
$[L_{cEE}, U_{cEE}]$	[.055, .130]	[.074, .161]
CLR 95% level confidence interval	(.023, .170)	(.039, .202)

Note: Numbers in parentheses from 2nd to 14th row are standard errors and * denotes estimate is statistically different from 0 at 5% level. Computations use design weights. The standard error is calculated by 5,000-repetition bootstrap. Numbers in parentheses in the bottom eight rows are CLR 95% level confidence intervals, while numbers in square brackets are identified sets determined by the half-median unbiased estimators.

Table 6: Bounds for the Effects of Job Corps on ln(wage) for *cEE* Stratum at Week 208 Adjusting for Non-Response

	Entire Sample	Non-Hispanics
Number of observations	10520	8701
π_{aEE}	.017 (.002*)	.018 (.002*)
π_{nEE}	.162 (.005*)	.162 (.005*)
π_{cEE}	.391 (.009*)	.381 (.010*)
π_{cNE}	.038 (.010*)	.049 (.011*)
π_{aNN}	.026 (.003*)	.028 (.003*)
π_{nNN}	.108 (.004*)	.108 (.004*)
π_{cNN}	.258 (.006*)	.254 (.007*)
α_{cEE}	.877 (.022*)	.851 (.024*)
$E[Y(1)\|aEE]$	2.010 (.050*)	2.001 (.052*)
$E[Y(0)\|nEE]$	2.033 (.015*)	2.032 (.016*)
$E[Y(0)\|cEE]$	1.985 (.013*)	1.961 (.015*)
$\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$	2.449 (.065*)	2.383 (.056*)
$\overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$	1.666 (.036*)	1.702 (.034*)
Bounds with Monotonicity (Proposition 1)		
$[LY_{1,cEE}, UY_{1,cEE}]$	[1.957, 2.102]	[1.940, 2.114]
CLR 95% level confidence interval	(1.928, 2.127)	(1.911, 2.140)
$[L_{cEE}, U_{cEE}]$	[-.028, .117]	[-.021, .153]
CLR 95% level confidence interval	(-.065, .153)	(-.060, .191)
Bounds with Monotonicity and Mean Dominance (Proposition 2)		
$[LY_{1,cEE}, UY_{1,cEE}]$	[2.029, 2.102]	[2.028, 2.114]
CLR 95% level confidence interval	(2.013, 2.128)	(2.011, 2.142)
$[L_{cEE}, U_{cEE}]$	[.044, .117]	[.067, .153]
CLR 95% level confidence interval	(.015, .155)	[.035, .193]

Note: Numbers in parentheses from 2nd to 14th row are standard errors and * denotes estimate is statistically different from 0 at 5% level. Computations use weights accounting for sample design, interview design and interview non-response. The standard error is calculated by 5,000-repetition bootstrap. Numbers in parentheses in the bottom eight rows are CLR 95% level confidence intervals, while numbers in square brackets are identified sets determined by the half-median unbiased estimators.

A Appendix (not Intended for Publication)

A.1 Proof of Proposition 1

First, we show that under Assumptions 1−5 L_{cEE} and U_{cEE} are the smallest and largest possible values, respectively, for the average treatment effect for the stratum cEE. Next, we prove that for $\forall\Delta \in [L_{cEE}, U_{cEE}]$, there exist distributions for cEE, aEE and cNE consistent with the observed data of Y in $\{Z = 1, D = 1, S = 1\}$ and the contraint that $E[Y(1)|aEE] = \overline{Y}^{011}$. In other words, the interval $[L_{cEE}, U_{cEE}]$ contains any other bounds that are consistent with Assumptions 1 to 5. The first-step proof is similar to that in Horowitz and Manski (1995), except that a binding constraint should be satisfied under the lower and upper bounds. Since both L_{cEE} and U_{cEE} depend on the range of $\overline{Y}^{011}$, we need to discuss multiple cases in the proof.

Proof. First by Assumptions 1−5, the proportions of each stratum are uniquely determined by the observed data. Thus, the proof is completed given the proportions of the strata. Second since $E[Y(0)|cEE]$ is point identified by Assumptions 1−5, the proof can be completed with respect to $E[Y(1)|cEE]$ instead of the average treatment effect for cEE. Let $\theta = E[Y(1)|cEE]$, and then $\Delta = \theta - E[Y(0)|cEE]$. Third, since both $LY_{1,cEE}$ and $UY_{1,cEE}$ depend on the range of $\overline{Y}^{011}$, we have to discuss multiple cases.

Let Qy be the observed distribution of Y in the cell $[Z = 1, D = 1, S = 1]$. $\overline{y}_{aEE} = \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$, $\widetilde{y}_{aEE} = \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$. Denote the expression of $LY_{1,cEE}$ as $\widetilde{LY}_{1,cEE}$ when $\overline{y}_{aEE} > \overline{Y}^{011}$ and the expression of $UY_{1,cEE}$ as $\widetilde{UY}_{1,cEE}$ when $\widetilde{y}_{aEE} < \overline{Y}^{011}$, The probability density functions for each stratum are fy_{cEE}, fy_{aEE} and fy_{cNE}, and their corresponding distributions are Fy_{cEE}, Fy_{aEE} and Fy_{cNE}.

For the first-step proof, we discuss the two cases to show that $LY_{1,cEE}$ is the smallest possible value for $E[Y(1)|cEE]$. The other two cases for the $UY_{1,cEE}$ can be shown in the same way.

1. $LY_{1,cEE} = \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})$, when $\overline{Y}^{011} \geq \overline{y}_{aEE}$.

Let $Gy = \begin{cases} \frac{Qy}{\alpha_{cEE}}, & \text{if } y \leq y^{111}_{\alpha_{cEE}} \\ 1, & \text{if } y > y^{111}_{\alpha_{cEE}} \end{cases}$.

$\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})$ is the smallest value is shown by $Gy \geq Fy_{cEE}$, for all $Fy_{cEE} \in \{\alpha_{cEE}Fy_{cEE} + \alpha_{aEE}Fy_{aEE} + \alpha_{cNE}Fy_{cNE} = Qy,\ E[Y(1)|aEE] = \overline{Y}^{011}\}$ and all $y \in \mathcal{R}$.

If $y \leq y^{111}_{\alpha_{cEE}}$, $Gy < Fy_{cEE} \Rightarrow Qy < \alpha_{cEE}Fy_{cEE} \Rightarrow Qy < \alpha_{cEE}Fy_{cEE} + \alpha_{aEE}Fy_{aEE} + \alpha_{cNE}Fy_{cNE}$. This contradicts the feasible set of Fy_{cEE}.

If $y > y^{111}_{\alpha_{cEE}}$, $Gy - Fy_{cEE} = 1 - Fy_{cEE} \geq 0$.

Next, we show that the distributions exist, when $E[Y(1)|cEE] = \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})$ and $E[Y(1)|aEE] = \overline{Y}^{011}$.

When $\overline{Y}^{011} \geq \overline{y}_{aEE}$ and $\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$, let ty be the observed density of $Y(Y \geq y^{111}_{1-\alpha_{aEE}})$ and hy the observed density of $Y(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$. Then $\exists\tau \in [0, 1]$, *s.t.* $\tau\overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}}) + (1-\tau)\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) = \overline{Y}^{011}$. Therefore, $gy = fy_{cEE}$,

$\tau ty + (1-\tau)hy = fy_{aEE}$ and $\frac{(1-\tau)\pi_{aEE}}{\pi_{cNE}}ty + (1 - \frac{(1-\tau)\pi_{aEE}}{\pi_{cNE}})hy = fy_{cNE}$.

When $\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}^{011} \leq \overline{y}_{aEE}$, let ty be the observed density of $Y(Y \geq y^{111}_{1-\alpha_{cNE}})$ and hy the observed density of $Y(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$. Then $\exists\tau \in [0,1]$, *s.t.* $\tau\overline{Y}(Y \geq y^{111}_{1-\alpha_{cNE}}) + (1-\tau)\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) = \overline{Y}^{011}$. Similarly, $gy = fy_{cEE}$, $\tau ty + (1-\tau)hy = fy_{aEE}$ and $(1 - \frac{\tau\pi_{aEE}}{\pi_{cNE}})ty + \frac{\tau\pi_{aEE}}{\pi_{cNE}}hy = fy_{cNE}$.

2. $LY_{1,cEE} = \widetilde{LY}_{1,cEE}$, when $\overline{Y}^{011} \leq \overline{y}_{aEE}$.

In this case, we first prove that$\overline{Y}(Y \leq y^{111}_{p1-1,cNE})$ is the smallest feasible value for the quantity $\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}E[Y(1)|cEE] + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}E[Y(1)|aEE]$, and then show $\widetilde{LY}_{1,cEE}$ is the smallest feasible value for $E[Y(1)|cEE]$.

Let $Gy = \begin{cases} \frac{Qy}{1-\alpha_{cNE}}, & \text{if } y \leq y^{111}_{1-\alpha_{cNE}} \\ 1, & \text{if } y > y^{111}_{1-\alpha_{cNE}} \end{cases}$.

$\overline{Y}(Y \leq y^{111}_{p1-1,cNE})$ is the smallest value for $\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}E[Y(1)|cEE] + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}E[Y(1)|aEE]$ is shown by $Gy \geq \frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{cEE} + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{aEE}$, for all $Fy_{cEE} \in \{\alpha_{cEE}Fy_{cEE} + \alpha_{aEE}Fy_{aEE} + \alpha_{cNE}Fy_{cNE} = Qy,\ E[Y(1)|aEE] = \overline{Y}^{011}\}$ and all $y \in \mathcal{R}$.

If $y \leq y^{111}_{1-\alpha_{cNE}}$, $Gy < \frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{cEE} + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{aEE} \Rightarrow Qy < \alpha_{cEE}Fy_{cEE} + \alpha_{aEE}Fy_{aEE} \Rightarrow Qy < \alpha_{cEE}Fy_{cEE} + \alpha_{aEE}Fy_{aEE} + \alpha_{cNE}Fy_{cNE}$. This contradicts the feasible set of Fy_{cEE}.

If $y > y^{111}_{1-\alpha_{cNE}}$, $Gy - (\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{cEE} + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{aEE}) = 1 - \frac{\pi_{cEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{cEE} - \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}Fy_{aEE} \geq 0$.

Since $E[Y(1)|aEE] = \overline{Y}^{011}$, $\widetilde{LY}_{1,cEE}$ is the smallest value for $E[Y(1)|cEE]$.

Next, we show the distributions exist, when $E[Y(1)|cEE] = \widetilde{LY}_{1,cEE}$ and $E[Y(1)|aEE] = \overline{Y}^{011}$.

When $\overline{Y}^{011} \leq \overline{y}_{aEE}$ and $\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$, let sy be the observed density of $Y(Y \geq y^{111}_{1-\alpha_{cNE}})$, ty the observed density of $Y(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$ and hy the observed density of $Y(Y \leq y^{111}_{\alpha_{aEE}})$. Then $\exists\tau \in [0,1]$, *s.t.* $\tau\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{\alpha_{aEE}}) = \overline{Y}^{011}$. Therefore, $sy = fy_{cNE}$, $\tau ty + (1-\tau)hy = fy_{aEE}$ and $(1 - \frac{\tau\pi_{aEE}}{\pi_{cEE}})ty + \frac{\tau\pi_{aEE}}{\pi_{cEE}}hy = fy_{cEE}$.

When $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) \leq \overline{Y}^{011} \leq \overline{y}_{aEE}$, let sy be the observed density of $Y(Y \geq y^{111}_{1-\alpha_{cNE}})$, ty the observed density of $Y(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$ and hy the observed density of $Y(Y \leq y^{111}_{\alpha_{cEE}})$. Then $\exists\tau \in [0,1]$, *s.t.* $\tau\overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}) = \overline{Y}^{011}$. Therefore, $sy = fy_{cNE}$, $\tau ty + (1-\tau)hy = fy_{aEE}$ and $\frac{(1-\tau)\pi_{aEE}}{\pi_{cEE}}ty + (1 - \frac{(1-\tau)\pi_{aEE}}{\pi_{cEE}})hy = fy_{cEE}$.

For the second-step proof, we have 4 cases to discuss taking into account the lower and upper bound simultaneously. Since they form different segmentation of Qy, in each case we use some cutoff values to discuss $\forall\theta \in [LY_{1,cEE}, cutoff]$ and $\forall\theta \in [cutoff, UY_{1,cEE}]$ separately. In either interval for θ, we have to discuss the range of $\overline{Y}^{011}$as we have done in the first-step proof. The 4 cases are listed as follows.

1. $LY_{1,cEE} = \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})$, $UY_{1,cEE} = \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})$

It happens when $\overline{y}_{aEE} \leq \overline{Y}^{011} \leq \widetilde{y}_{aEE}$, in other words, $\pi_{cNE} \geq \pi_{cEE}$. Since $\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}) \leq \overline{Y}^{111} \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cEE}})$ and $\overline{Y}(Y \leq y^{111}_{1-\alpha_{cEE}}) \leq \overline{Y}^{111} \leq \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})$, thus, $cutoff = \overline{Y}^{111}$.

For $\forall\theta \in [\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}), cutoff]$, it is necessary to discuss two cases when $\overline{Y}^{011} \geq \overline{y}_{aEE}$ $\&\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and $\overline{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

For $\forall\theta \in [cutoff, \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})]$, it is necessary to discuss two cases when $\overline{Y}^{011} \leq \widetilde{y}_{aEE}$ $\&\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$ and $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}}) \leq \overline{Y}^{011} \leq \widetilde{y}_{aEE}$.

2. $LY_{1,cEE} = \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})$, $UY_{1,cEE} = \widetilde{UY}_{1,cEE}$

In this case, we have to discuss the relationship between $\overline{y}_{aEE}$ and $\widetilde{y}_{aEE}$ first. We solve this problem with two different cutoffs. For either cutoff, we have:

For $\forall\theta \in [\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}), cutoff]$, two cases to discuss: $\overline{Y}^{011} \geq \overline{y}_{aEE}$ $\&\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and $\overline{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

For $\forall\theta \in [cutoff, \widetilde{UY}_{1,cEE}]$, two cases to discuss: $\overline{Y}^{011} \geq \widetilde{y}_{aEE}$ $\&\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

3. $LY_{1,cEE} = \widetilde{LY}_{1,cEE}$, $UY_{1,cEE} = \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})$

Like case 2, we have to discuss the relationship between $\overline{y}_{aEE}$ and $\widetilde{y}_{aEE}$ first. We solve this problem with two different cutoffs. For either cutoff, we have:

For $\forall\theta \in [\widetilde{LY}_{1,cEE}, cutoff]$, two cases to discuss: $\overline{Y}^{011} \leq \overline{y}_{aEE}$ $\&\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$ and $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) \leq \overline{Y}^{011} \leq \overline{y}_{aEE}$.

For $\forall\theta \in [cutoff, \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})]$, two cases to discuss: $\overline{Y}^{011} \leq \widetilde{y}_{aEE}$ $\&\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$ and $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}}) \leq \overline{Y}^{011} \leq \widetilde{y}_{aEE}$.

4. $LY_{1,cEE} = \widetilde{LY}_{1,cEE}$, $UY_{1,cEE} = \widetilde{UY}_{1,cEE}$

It happens when $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{y}_{aEE}$, in other words, $\pi_{cNE} \leq \pi_{cEE}$. In this case, it is difficult to get a uniform cutoff. We discuss multiple cases conditional on the proportions of the strata.

When $\pi_{cNE} \leq \pi_{cEE}$ and $\pi_{aEE} \leq \pi_{cEE}$, we discuss the intervals $\forall\theta \in [\widetilde{LY}_{1,cEE}, \overline{y}_{aEE}]$ and $\forall\theta \in [\widetilde{y}_{aEE}, \widetilde{UY}_{1,cEE}]$ to complete the proof in the entire range, i.e., $\forall\theta \in [\widetilde{LY}_{1,cEE}, \widetilde{UY}_{1,cEE}]$. For $\forall\theta \in [\widetilde{LY}_{1,cEE}, \overline{y}_{aEE}]$, two cases are $\overline{Y}^{011} \leq \overline{y}_{aEE}$ $\&\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$ and $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}}) \leq \overline{Y}^{011} \leq \overline{y}_{aEE}$. For $\forall\theta \in [\widetilde{y}_{aEE}, \widetilde{UY}_{1,cEE}]$, two cases are: $\overline{Y}^{011} \geq \widetilde{y}_{aEE}$ $\&\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

When $\pi_{cNE} \leq \pi_{cEE} \leq \pi_{aEE}$, we discuss the intervals $\forall\theta \in [\widetilde{LY}_{1,cEE}, \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})]$ and $\forall\theta \in [\overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}), \widetilde{UY}_{1,cEE}]$ to complete the proof in the entire range. For $\forall\theta \in [\widetilde{LY}_{1,cEE}, \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})]$, we have $\overline{Y}^{011} \leq \overline{y}_{aEE}$ $\&\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$. For $\forall\theta \in [\overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}), \widetilde{UY}_{1,cEE}]$, we have $\overline{Y}^{011} \geq \widetilde{y}_{aEE}$ $\&\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

From all the above, we can find that though cutoffs are different in the 4 cases, discussion of

$\overline{Y}^{011}$ is repeated. Case 2 and Case 3 compose a complete discussion of $\overline{Y}^{011}$. In the following, we only write the proof of Case 2. Case 3 can be shown in a similar way.

First, let us discuss the relationship between $\overline{y}_{aEE}$ and $\widetilde{y}_{aEE}$ in Case 2 and derive two different cutoffs.

When $\overline{y}_{aEE} \leq \widetilde{y}_{aEE} \leq \overline{Y}^{011}$(i.e., $\pi_{cNE} \geq \pi_{cEE}$), $LY_{1,cEE} \leq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cEE}})$ and $\widetilde{UY}_{1,cEE} = \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})\frac{\pi_{cEE}+\pi_{aEE}}{\pi_{cEE}} - \overline{Y}^{011}\frac{\pi_{aEE}}{\pi_{cEE}} \geq \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})\frac{\pi_{cEE}+\pi_{aEE}}{\pi_{cEE}} - \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}})\frac{\pi_{aEE}}{\pi_{cEE}} = \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$. The first inequality is derived from $\overline{Y}^{011} \leq \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}})$, and the last inequality is derived from $\pi_{cNE} \geq \pi_{cEE}$. Thus, $\widetilde{UY}_{1,cEE} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \geq \overline{Y}(Y \leq y^{111}_{\alpha_{cNE}})$. Therefore, the cutoff value $cutoff = \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

When $\widetilde{y}_{aEE} \leq \overline{y}_{aEE} \leq \overline{Y}^{011}(\pi_{cNE} \leq \pi_{cEE})$, $LY_{1,cEE} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}}) \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cEE}})$. The last inequality is derived from $\pi_{cNE} \leq \pi_{cEE}$. Thus, $LY_{1,cEE} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cEE}})$. And $\widetilde{UY}_{1,cEE} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) \geq \overline{Y}(Y \leq y^{111}_{\alpha_{cNE}})$. Its derivation is the same as that in the last paragraph. Therefore, $cutoff = \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

Second, $\forall \theta \in [cutoff, \widetilde{UY}_{1,cEE}]$, where $cutoff$ is the corresponding value in either case $\overline{y}_{aEE} \leq \widetilde{y}_{aEE}$ or $\widetilde{y}_{aEE} \leq \overline{y}_{aEE}$, $\exists \lambda \in (0,1]$, *s.t.* $\lambda \widetilde{UY}_{1,cEE} + (1-\lambda)\overline{Y}(Y \leq y^{111}_{\alpha_{cNE}}) = \theta$, since $\overline{Y}(Y \leq y^{111}_{\alpha_{cNE}}) \leq \theta \leq \widetilde{UY}_{1,cEE}$. ($\lambda = \frac{\theta - \overline{Y}(Y \leq y^{111}_{\alpha_{cNE}})}{\widetilde{UY}_{1,cEE} - \overline{Y}(Y \leq y^{111}_{\alpha_{cNE}})}$.)

To construct fy_{aEE}, it is necessary to discuss the value of $\overline{Y}^{011}$. One case is $\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$. It happens when either $\pi_{cEE} \leq \pi_{aEE}$, or $\pi_{cEE} \geq \pi_{aEE}$ but aEE take up very top quantiles of the observed distribution. The other case is $\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$. This is true only when $\pi_{cEE} \geq \pi_{aEE}$.

(1) $\overline{Y}^{011} \geq \widetilde{y}_{aEE}$ & $\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$

Let ty be the observed density of $Y(y \geq y^{111}_{1-\alpha_{aEE}})$, hy the density of $Y(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and gy the density of $Y(y \leq y^{111}_{\alpha_{cNE}})$. Since $\overline{Y}(Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}^{011} \leq \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}})$, $\exists \tau \in (0,1]$, *s.t.* $\tau \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{1-\alpha_{aEE}}) = \overline{Y}^{011}$. $\tau ty + (1-\tau)\frac{\pi_{cEE}}{\pi_{cNE}+\pi_{cEE}}hy + (1-\tau)\frac{\pi_{cNE}}{\pi_{cNE}+\pi_{cEE}}gy = fy_{aEE}$. Since $\widetilde{UY}_{1,cEE}$ is obtained by temporarily assuming aEE above $y^{111}_{\alpha_{cNE}}$, $\exists \phi \in [0,1]$, *s.t.* $\phi \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}}) + (1-\phi)\overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}}) = \overline{Y}^{011}$. Thus, $\lambda\{\frac{\pi_{aEE}}{\pi_{cEE}}ty + hy - [\phi ty + (1-\phi)hy]\frac{\pi_{aEE}}{\pi_{cEE}}\} + (1-\lambda)gy = fy_{cEE}$. Since $\alpha_{aEE}ty + \alpha_{cEE}hy + \alpha_{cNE}gy = \alpha_{aEE}fy_{aEE} + \alpha_{cEE}fy_{cEE} + \alpha_{cNE}fy_{cNE}$, the corresponding density for cNE is $\frac{\pi_{aEE}}{\pi_{cNE}}[1-\tau-\lambda(1-\phi)]ty + \frac{\pi_{cEE}}{\pi_{cNE}}[1-(1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}} - \lambda(1-(1-\phi)\frac{\pi_{aEE}}{\pi_{cEE}})]hy + [1-(1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}} - (1-\lambda)\frac{\pi_{cEE}}{\pi_{cNE}}]gy = fy_{cNE}$.

(2) $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$

Let ty be the observed density of $Y(y \geq y^{111}_{1-\alpha_{cEE}})$, hy the density of $Y(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$ and gy the density of $Y(y \leq y^{111}_{\alpha_{cNE}})$. Since $\overline{Y}(Y \leq y^{111}_{\alpha_{cNE}}) \leq \overline{Y}^{011} \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})$, $\exists \tau \in (0,1)$, *s.t.* $\tau \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{\alpha_{cNE}}) = \overline{Y}^{011}$. $\tau\frac{\pi_{cEE}}{\pi_{aEE}+\pi_{cEE}}ty + \tau\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cEE}}hy + (1-\tau)gy = fy_{aEE}$. As in (1), since $\widetilde{UY}_{1,cEE}$ is obtained by temporarily assuming aEE above $y^{111}_{\alpha_{cNE}}$,

$\exists\phi \in [0,1]$, *s.t.* $\phi\overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}}) + (1-\phi)\widetilde{y}_{aEE} = \overline{Y}^{011}$. Thus, $\lambda\{ty + \frac{\pi_{aEE}}{\pi_{cEE}}hy - [\phi ty + (1-\phi)hy]\frac{\pi_{aEE}}{\pi_{cEE}}\} + (1-\lambda)gy = fy_{cEE}$. Similarly as in (1), we finally get $\frac{\pi_{cEE}}{\pi_{cNE}}[1 - \tau\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}} - \lambda(1-\phi\frac{\pi_{aEE}}{\pi_{cEE}})]ty + \frac{\pi_{aEE}}{\pi_{cNE}}[1-\tau\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}} - \lambda\phi]hy + [1-(1-\tau)\frac{\pi_{aEE}}{\pi_{cNE}} - (1-\lambda)\frac{\pi_{cEE}}{\pi_{cNE}}]gy = fy_{cNE}$.

Third, $\forall\theta \in [\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}), cutoff]$, $\exists\lambda \in [0,1)$, *s.t.* $\lambda\overline{Y}(Y \geq y^{111}_{\alpha_{cEE}}) + (1-\lambda)\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}) = \theta$, since $\overline{Y}(Y \leq y^{111}_{\alpha_{cEE}}) \leq \theta \leq \overline{Y}(Y \geq y^{111}_{\alpha_{cEE}})$. ($\lambda = \frac{\theta - \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})}{\overline{Y}(Y \geq y^{111}_{\alpha_{cEE}}) - \overline{Y}(Y \leq y^{111}_{\alpha_{cEE}})}$.)

To to discuss the value of $\overline{Y}^{011}$, one case is $\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$. It happens when either $\pi_{cNE} \leq \pi_{aEE}$, or $\pi_{cNE} \geq \pi_{aEE}$ but aEE take up very top quantiles of the observed distribution. The other case is $\overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ when $\pi_{cNE} \geq \pi_{aEE}$.

(1) $\overline{Y}^{011} \geq \overline{y}_{aEE}$ & $\overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$

Let ty be the observed density of $Y(y \geq y^{111}_{1-\alpha_{aEE}})$, hy the density of $Y(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$ and gy the density of $Y(y \leq y^{111}_{\alpha_{cEE}})$. Since $\overline{Y}(Y \leq y^{111}_{1-\alpha_{aEE}}) \leq \overline{Y}^{011} \leq \overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}})$, $\exists\tau \in (0,1]$, *s.t.* $\tau\overline{Y}(Y \geq y^{111}_{1-\alpha_{aEE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{1-\alpha_{aEE}}) = \overline{Y}^{011}$. $\tau ty + (1-\tau)\frac{\pi_{cNE}}{\pi_{cNE}+\pi_{cEE}}hy + (1-\tau)\frac{\pi_{cEE}}{\pi_{cNE}+\pi_{cEE}}gy = fy_{aEE}$. For cEE, we have $\lambda\frac{\pi_{aEE}}{\pi_{cNE}+\pi_{aEE}}ty + \lambda\frac{\pi_{cNE}}{\pi_{cNE}+\pi_{aEE}}hy + (1-\lambda)gy = fy_{cEE}$. Finally, $\frac{\pi_{aEE}}{\pi_{cNE}}(1-\tau-\lambda\frac{\pi_{cEE}}{\pi_{aEE}+\pi_{cNE}})ty + [1-(1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}} - \lambda\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}}]hy + \frac{\pi_{cEE}}{\pi_{cNE}}[\lambda - (1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}})]gy = fy_{cNE}$.

(2) $\overline{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$

Let ty be the observed density of $Y(y \geq y^{111}_{1-\alpha_{cNE}})$, hy the density of $Y(y^{111}_{\alpha_{cEE}} \leq Y \leq y^{111}_{1-\alpha_{cNE}})$ and gy the density of $Y(y \leq y^{111}_{\alpha_{cEE}})$. Since $\overline{Y}(Y \leq y^{111}_{1-\alpha_{cNE}}) \leq \overline{Y}^{011} \leq \overline{Y}(Y \geq y^{111}_{1-\alpha_{cNE}})$, $\exists\tau \in (0,1]$, *s.t.* $\tau\overline{Y}(Y \geq y^{111}_{1-\alpha_{cNE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{1-\alpha_{cNE}}) = \overline{Y}^{011}$. $\tau ty + (1-\tau)\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cEE}}hy + (1-\tau)\frac{\pi_{cEE}}{\pi_{aEE}+\pi_{cEE}}gy = fy_{aEE}$. For cEE, we have $\lambda\frac{\pi_{cNE}}{\pi_{cNE}+\pi_{aEE}}ty + \lambda\frac{\pi_{aEE}}{\pi_{cNE}+\pi_{aEE}}hy + (1-\lambda)gy = fy_{cEE}$. Finally, $(1-\tau\frac{\pi_{aEE}}{\pi_{cNE}} - \lambda\frac{\pi_{cEE}}{\pi_{aEE}+\pi_{cNE}})ty + \frac{\pi_{aEE}}{\pi_{cNE}}[1-(1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}} - \lambda\frac{\pi_{cEE}}{\pi_{aEE}+\pi_{cNE}}]hy + \frac{\pi_{cEE}}{\pi_{cNE}}[\lambda - (1-\tau)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{aEE}}]gy = fy_{cNE}$. ■

A.2 Proof of Proposition 2

The proof of Proposition 2 is similar to that of Proposition 1, except two differences: first, multiple cases to discuss reduce to two due to L_{cEE}; second, the constructed distributions for cEE and cNE should also satisfy the mean dominance assumption.

Proof. As in the proof of Proposition 1, we first show that $LY_{1,cEE}$ is the smallest feasible value for $E[Y(1)|cEE]$, and then show $\forall\theta \in [\,LY_{1,cEE}, UY_{1,cEE}]$, there exist distributions Fy_{cEE}, Fy_{aEE} and Fy_{cNE} satisfying Assumptions 1−6.

For $\forall\theta \in [LY_{1,cEE}, UY_{1,cEE}]$, by equation (6), we have:

$$E[Y(1)|cEE] - E[Y(1)|cNE]$$
$$= \theta - (\overline{Y}^{111}\frac{\pi_{cEE}+\pi_{cNE}+\pi_{aEE}}{\pi_{cNE}} - \theta\frac{\pi_{cEE}}{\pi_{cNE}} - \overline{Y}^{011}\frac{\pi_{aEE}}{\pi_{cNE}})$$
$$\geq (1+\frac{\pi_{cEE}}{\pi_{cNE}})LY_{1,cEE} + \overline{Y}^{011}\frac{\pi_{aEE}}{\pi_{cNE}} - \overline{Y}^{111}\frac{\pi_{cEE}+\pi_{cNE}+\pi_{aEE}}{\pi_{cNE}} = 0.$$

If there was another lower bound smaller than $LY_{1,cEE}$, $E[Y(1)|cEE] - E[Y(1)|cNE]$ would be negative when $E[Y(1)|cEE]$ reached that lower bound. This contradicts to Assumption 6.

Thus, $LY_{1,cEE}$ is the smallest value for $E[Y(1)|cEE]$ under Assumptions 1−6.

As in Proposition 1, we have to show the distributions exist when $E[Y(1)|cEE] = LY_{1,cEE}$ and $E[Y(1)|aEE] = \overline{Y}^{011}$. According to the range of $\overline{Y}^{011}$, we have 4 cases to discuss:

(1) $\overline{Y}^{011} \leq \widetilde{y}_{aEE} \& \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$;

(2) $\overline{Y}(y^{111}_{\alpha_{aEE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}}) \leq \overline{Y}^{011} \leq \widetilde{y}_{aEE}$;

(3) $\overline{Y}^{011} \geq \widetilde{y}_{aEE} \& \overline{Y}^{011} \geq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$;

(4) $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$.

These 4 cases correspond with those in the second-step proof that for $\forall\theta \in [LY_{1,cEE}, UY_{1,cEE}]$, there exist distributions Fy_{cEE}, Fy_{aEE} and Fy_{cNE} satisfying Assumptions 1−6. Since for $\forall\theta \in [LY_{1,cEE}, UY_{1,cEE}]$, $\exists\lambda \in [0,1]$, $s.t.$ $\lambda UY_{1,cEE} + (1-\lambda)LY_{1,cEE} = \theta$, the proof that the distributions exist when $E[Y(1)|cEE] = LY_{1,cEE}$ and $E[Y(1)|aEE] = \overline{Y}^{011}$ is a special case of the second-step proof, i.e. $\lambda = 0$. In the following, we take Case (4) as an example to illustrate the second-step proof of Proposition 2.

(4) $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$

Let ty be the observed density of $Y(y \geq y^{111}_{1-\alpha_{cEE}})$, hy the density of $Y(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{cEE}})$ and gy the density of $Y(y \leq y^{111}_{\alpha_{cNE}})$.

$\forall\theta \in [LY_{1,cEE}, \widetilde{UY}_{1,cEE}], \exists\lambda \in [0,1]$, $s.t.$ $\lambda\widetilde{UY}_{1,cEE} + (1-\lambda)LY_{1,cEE} = \theta$.

As in the proof of Proposition 1, since $\widetilde{y}_{aEE} \leq \overline{Y}^{011} \leq \overline{Y}(y^{111}_{\alpha_{cNE}} \leq Y \leq y^{111}_{1-\alpha_{aEE}})$, $\exists\tau \in (0,1]$, $s.t.$ $\tau\overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}}) + (1-\tau)\overline{Y}(Y \leq y^{111}_{1-\alpha_{cEE}}) = \overline{Y}^{011}$. $\tau ty + (1-\tau)\frac{\pi_{aEE}}{\pi_{cNE}+\pi_{aEE}}hy + (1-\tau)\frac{\pi_{cNE}}{\pi_{cNE}+\pi_{aEE}}gy = fy_{aEE}$. Since $\widetilde{UY}_{1,cEE}$ is obtained by temporarily assuming aEE above $y^{111}_{\alpha_{cNE}}$, $\exists\phi \in [0,1]$, $s.t.$ $\phi\overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}}) + (1-\phi)\widetilde{y}_{aEE} = \overline{Y}^{011}$. Thus, $\lambda\{ty + \frac{\pi_{aEE}}{\pi_{cEE}}hy - [\phi ty + (1-\phi)hy]\frac{\pi_{aEE}}{\pi_{cEE}}\} + (1-\lambda)(\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}}ty + \frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}}hy + \frac{\pi_{cNE}}{\pi_{cEE}+\pi_{cNE}}gy - fy_{aEE}\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}}) = fy_{cEE}$. After some algebra, we have $[\lambda(1-\phi\frac{\pi_{aEE}}{\pi_{cEE}}) + (1-\lambda)(\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}} - \tau\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}})]ty + \{\lambda\phi\frac{\pi_{aEE}}{\pi_{cEE}} + (1-\lambda)\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}}[1-(1-\tau)\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cNE}}]\}hy + (1-\lambda)\frac{\pi_{cNE}}{\pi_{cEE}+\pi_{cNE}}[1-(1-\tau)\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cNE}}]gy = fy_{cEE}$. Then, the corresponding density for cNE is $\{-\frac{\pi_{aEE}}{\pi_{cNE}}\tau + \frac{\pi_{cEE}}{\pi_{cNE}}[1-\lambda(1-\phi\frac{\pi_{aEE}}{\pi_{cEE}}) - (1-\lambda)(\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}} - \tau\frac{\pi_{aEE}}{\pi_{cEE}+\pi_{cNE}})]\}ty + \frac{\pi_{aEE}}{\pi_{cNE}}\{-\lambda\phi - [1-(1-\lambda)\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}}][1-(1-\tau)\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cNE}}]\}hy + [1-(1-\lambda)\frac{\pi_{cEE}}{\pi_{cEE}+\pi_{cNE}}][1-(1-\tau)\frac{\pi_{aEE}}{\pi_{aEE}+\pi_{cNE}}]gy = fy_{cNE}$.

The inequality in the second paragraph of this proof shows that Assumption 6 holds, as long as $E[Y(1)|cEE] \geq LY_{1,cEE}$. Since $\theta \in [LY_{1,cEE}, UY_{1,cEE}]$ holds by construction, the constructed densities satisfy Assumption 6. ■

A.3 Procedure for Estimation and Inference

Following the methodology proposed by Chernozhukov, Lee and Rosen (2011) (hereafter CLR), as an illustration, the upper bound for $\theta_0 = E[Y(1)|cEE]$ in Proposition 1 is given by $\theta^u_0 = \min_{v\in\mathcal{V}=\{1,2\}} \theta^u(v)$, with $\theta^u(1) = \overline{Y}(Y \geq y^{111}_{1-\alpha_{cEE}})$ and $\theta^u(2) = \overline{Y}(Y \geq y^{111}_{\alpha_{cNE}})\frac{q_{1|0}-p_{01|1}}{p_{01|0}-p_{01|1}} - \overline{Y}^{011}\frac{p_{11|0}}{p_{01|0}-p_{01|1}}$. Then the precision-corrected estimate of θ^u_0 is given by $\widehat{\theta}^u(p) = \min_{v\in\{1,2\}}[\widehat{\theta}^u(v)+$

$k(p)s(v)]$, where $\widehat{\theta}^u(v)$ is a consistent estimate of $\theta^u(v)$, $s(v)$ is its standard error and $k(p)$ is a critical value.

The selection of $k(p)$ relies on standardized process $Z_n(v) = \{\theta^u(v) - \widehat{\theta}^u(v)\}/\sigma(v)$, where $\sigma(v)/s(v) \to 1$ uniformly in v. CLR approximate this process by a standardized Gaussian process $Z_n^*(v)$. Specifically, for any compact set V, CLR approximate by simulation the p-th quantile of $\sup_{v\in V} Z_n^*(v)$, denoted by $k_{n,V}(p)$, and use it in place of $k(p)$. Since setting $V = \mathcal{V}$ leads to asymptotically valid but conservative inference, CLR propose a preliminary set estimator $\widehat{V}_n$ of $V_0 = \arg\min_{v\in\mathcal{V}} \theta^u(v)$ for the upper bound (where arg min is replaced by arg max for the lower bound), which they call an adaptive inequality selector. Intuitively, $\widehat{V}_n$ selects the bounding functions that are close enough to binding to affect the asymptotic distribution of the estimators of the upper and lower bounds.

We now describe the precise procedure we employ to obtain half-median-unbiased estimators for the upper bounds of Δ and $E[Y(1)|cEE]$ in Proposition 1.[13] These upper bounds can be written as $\theta^u_{n0} = \min_{v\in\mathcal{V}=\{1,2\}} \theta^u_n(v)$, where the bounding functions $\theta^u_n(v)$ are given in Proposition 1, as illustrated in the example above.[14] Let $\gamma_n = [\theta^u_n(1)\ \theta^u_n(2)]'$ be the vector containing the two bounding functions and let $\widehat{\gamma}_n = [\widehat{\theta}^u_n(1)\ \widehat{\theta}^u_n(2)]'$ denote its sample analog estimator, which can be shown to be consistent and asymptotically normally distributed using standard results (e.g., Newey and McFadden, 1994; Lee, 2009). The specific steps are as follows.

1. Let Ω_n denote the asymptotic variance of $\sqrt{n}(\widehat{\gamma}_n - \gamma_n)$. We obtain a consistent estimate of Ω_n, $\widehat{\Omega}_n$, by a 5000-repetition bootstrap. Let $\widehat{g}_n(v)'$ denote the v^{th} row of $\widehat{\Omega}_n^{1/2}$, $s_n(v) = \|\widehat{g}_n(v)\|/\sqrt{n}$ and, following CLR, set $c_n = 1 - (.1/\log n)$.

2. We simulate $R = 1,000,000$ draws from $\mathcal{N}(0, I_2)$, denoted $Z_1, ..., Z_R$, where I_2 is a 2×2 identity matrix, and let $Z_r^*(v) = \widehat{g}_n(v)'Z_r/\|\widehat{g}_n(v)\|$ for $r = 1, ..., R$.

3. Let $Q_p(X)$ denote the p-th quantile of a random variable X. We compute $k_{n,\mathcal{V}}(c_n) = Q_{c_n}(\max_{v\in\mathcal{V}} Z_r^*(v), r = 1, ..., R)$; that is, for each replication r we calculate the maximum of $Z_r^*(1)$ and $Z_r^*(2)$, and we take the c-th quantile of those R values. We then use this critical value to compute the set estimator $\widehat{V}_n = \{v \in \mathcal{V} : \widehat{\theta}^u_n(v) \le \min_{\widetilde{v}\in\mathcal{V}}\{[\widehat{\theta}^u_n(\widetilde{v}) + k_{n,\mathcal{V}}(c_n)s_n(\widetilde{v})] + 2k_{n,\mathcal{V}}(c_n)s_n(v)\}$.

4. We compute $k_{n,\widehat{V}_n}(p) = Q_p(\max_{v\in\widehat{V}_n} Z_r^*(v), r = 1, ..., R)$, so that the critical value is based on $\widehat{V}_n$ instead of $\mathcal{V}$.

5. To get the half-median-unbiased estimator of θ^u_{n0}, $\widehat{\theta}^u_n(1/2)$, we set $p = 1/2$ and compute $\widehat{\theta}^u_n(1/2) = \min_{v\in\mathcal{V}}[\widehat{\theta}^u_n(v) + k_{n,\widehat{V}_n}(1/2)s_n(v)]$.

[13] For further details on the procedure see CLR, especially Appendix A and Section 4.1

[14] The subscript "n" indicates local parameters.

To obtain half-median-unbiased estimators for the lower bounds in Proposition 1, which have the form $\theta^l_{n0} = \max_{v \in \mathcal{V}=\{1,2\}} \theta^l_n(v)$, $\widehat{V}_n$ in step 3 above is replaced by $\widehat{V}_n = \{v \in \mathcal{V} : \widehat{\theta}^l_n(v) \geq \max_{\widetilde{v} \in \mathcal{V}}[\widehat{\theta}^l_n(\widetilde{v}) - k_{n,\mathcal{V}}(c_n)s_n(\widetilde{v})] - 2k_{n,\mathcal{V}}(c_n)s_n(v)\}$, and in step 5 we set $\widehat{\theta}^l_n(1/2) = \max_{v \in \mathcal{V}}[\widehat{\theta}^l_n(v) - k_{n,\widehat{V}_n}(1/2)s_n(v)]$.[15]

To describe the construction of confidence intervals for the true parameter θ_0, let $\widehat{\theta}^u_n(p) = \min_{v \in \mathcal{V}}[\widehat{\theta}^u_n(v) + k_{n,\widehat{V}_n}(p)s_n(v)]$ and $\widehat{\theta}^l_n(p) = \max_{v \in \mathcal{V}}[\widehat{\theta}^l_n(v) - k_{n,\widehat{V}_n}(p)s_n(v)]$, where the critical values are obtained as described above. Following CLR, let $\widehat{\Gamma}_n = \widehat{\theta}^u_n(1/2) - \widehat{\theta}^l_n(1/2)$, $\widehat{\Gamma}^+_n = \max(0, \widehat{\Gamma}_n)$, $\rho_n = \max\{\widehat{\theta}^u_n(3/4) - \widehat{\theta}^u_n(1/4), \widehat{\theta}^l_n(1/4) - \widehat{\theta}^l_n(3/4)\}$, $\tau_n = 1/(\rho_n \log n)$ and $\widehat{p}_n = 1 - \Phi(\tau_n \widehat{\Gamma}^+_n)\alpha$, where $\Phi(\cdot)$ is the standard normal CDF. Note that $\widehat{p}_n \in [1-\alpha, 1-\alpha/2]$, with $\widehat{p}_n$ approaching $1-\alpha/2$ when the length of the identified set tends to zero. Then, an asymptotically valid $1-\alpha$ confidence interval for θ_0 is given by $[\widehat{\theta}^l_n(\widehat{p}_n), \widehat{\theta}^u_n(\widehat{p}_n)]$, i.e., $\inf_{\theta_0 \in [\theta^l_{n0}, \theta^u_{n0}]} P(\theta_0 \in [\widehat{\theta}^l_n(\widehat{p}_n), \widehat{\theta}^u_n(\widehat{p}_n)]) \geq 1 - \alpha + o(1)$.

Finally, note that the lower bounds in Proposition 2 do not involve the max operator. However, for consistency with the bounds in Proposition 1, in our application we employ the same methodology described above (where $\mathcal{V}$ is a singleton for the lower bound) to construct confidence intervals and we report the half-median-unbiased estimates of the lower bounds (which, not surprisingly, are practically the same as the sample analog estimates).

A.4 Estimation of Average Baseline Characteristics of the Strata

We write the moment functions for average baseline characteristics of all the strata based on the conditional expectation in each cell defined by $\{Z, D, S\}$. Let $\overline{x}_k$ denote the expectation of a scalar baseline variable for a certain stratum k. The moment function for this variable is defined as:

$$g(\{\overline{x}_k\}) = \begin{bmatrix} (x - \overline{x}_{aNN})(1-Z)D(1-S) \\ (x - \overline{x}_{aEE})(1-Z)DS \\ (x - \overline{x}_{nNN})Z(1-D)(1-S) \\ (x - \overline{x}_{nEE})Z(1-D)S \\ (x - \overline{x}_{cEE}\frac{\pi_{cEE}}{p_{01|0}} - \overline{x}_{nEE}\frac{\pi_{nEE}}{p_{01|0}})(1-Z)(1-D)S \\ (x - \overline{x}_{cNN}\frac{\pi_{cNN}}{p_{10|1}} - \overline{x}_{aNN}\frac{\pi_{aNN}}{p_{10|1}})ZD(1-S) \\ (x - \overline{x}_{cNE}\frac{\pi_{cNE}}{p_{00|0}} - \overline{x}_{cNN}\frac{\pi_{cNN}}{p_{00|0}} - \overline{x}_{nNN}\frac{\pi_{nNN}}{p_{00|0}})(1-Z)(1-D)(1-S) \\ (x - \overline{x}_{cNE}\frac{\pi_{cNE}}{p_{11|1}} - \overline{x}_{cEE}\frac{\pi_{cEE}}{p_{11|1}} - \overline{x}_{aEE}\frac{\pi_{aEE}}{p_{11|1}})ZDS \\ x - \sum_k \pi_k \overline{x}_k \end{bmatrix}$$

where $\{\overline{x}_k\} = \{\overline{x}_{aNN}, \overline{x}_{aEE}, \overline{x}_{nNN}, \overline{x}_{nEE}, \overline{x}_{cNN}, \overline{x}_{cEE}, \overline{x}_{cNE}\}$. By Law of Iterated Expectation, $E[g(\{\overline{x}_k\})] = 0$ when evaluated at the true value of $\{\overline{x}_k\}$.

[15] Note that, because of the symmetry of the normal distribution, no changes are needed when computing the quantiles in steps 3 and 4.

Alternatively, we could also write the moment function for the proportions of all the strata and then estimate the model together with the average baseline characteristics simultaneously by GMM. However, such GMM estimators do not behave well in our data. Thus, in our application, we first identify the proportions of all the strata by equation (2) in the paper, and then estimate all the average baseline characteristics given the identified proportions. As seen in $g(\{\overline{x}_k\})$, for each variable, we have 9 equations (8 derived from the conditional expectations defined by $\{Z, D, S\}$ plus one from the expectation for the entire sample) to identify 7 means, i.e., $\{\overline{x}_k\}$. Since the standard errors obtained from this GMM model do not take into account the fact that the proportions of the strata are also estimated, we employ a 500-repetition bootstrap to calculate the standard errors of the estimated average baseline characteristics.

Appendix Table A1: Average Baseline Characteristics for the cEE and cNE Strata

	Entire Sample			Non-Hispanics		
	cEE	cNE	$cEE - cNE$	cEE	cNE	$cEE - cNE$
Female	.396 (.015*)	.630 (.165*)	-.234 (.174)	.390 (.016*)	.544 (.134*)	-.154 (.145)
Age at Baseline	18.44 (.056*)	19.19 (.699*)	-.749 (.735)	18.39 (.068*)	19.18 (.592*)	-.786 (.632)
White, Non-hispanic	.299 (.012*)	.260 (.126*)	.039 (.133)	.369 (.015*)	.259 (.126*)	.110 (.135)
Black, Non-Hispanic	.445 (.013*)	.622 (.158*)	-.177 (.166)	.550 (.015*)	.624 (.136*)	-.074 (.147)
Has Child	.161 (.011*)	.229 (.112*)	-.068 (.119)	.151 (.012*)	.210 (.111**)	-.059 (.119)
Number of children	.215 (.018*)	.356 (.187**)	-.141 (.200)	.209 (.019*)	.280 (.179)	-.071 (.192)
Personal Education	10.22 (.040*)	10.34 (.504*)	-.123 (.529)	10.24 (.048*)	10.27 (.402*)	-.036 (.434)
Ever Arrested	.230 (.012*)	.223 (.128**)	.007 (.136)	.228 (.013*)	.292 (.112*)	-.064 (.121)
At Baseline						
Have job	.241 (.011*)	.174 (.108)	.068 (.115)	.244 (.012*)	.159 (.102)	.084 (.110)
Weekly hours worked	24.07 (.583*)	25.27 (6.365*)	-1.196 (6.766)	24.05 (.613*)	25.23 (5.768*)	-1.187 (6.160)
Weekly earnings	113.86 (3.987*)	120.08 (39.90*)	-6.219 (40.90)	115.48 (3.500*)	142.57 (34.31*)	-27.09 (36.51)
Had job, Prev. Yr.	.714 (.013*)	.585 (.141*)	.129 (.151)	.718 (.014*)	.588 (.126*)	.130 (.136)
Months Employed,Prev.Yr.	4.346 (.122*)	3.286 (1.201*)	1.060 (1.280)	4.435 (.137*)	2.935 (1.105*)	1.500 (1.201)
Earnings, Prev.Yr.	3396.2 (128.63*)	3136.2 (1185.0*)	260.02 (1250.6)	3377.6 (128.55*)	2879.7 (1009.5*)	497.88 (1095.9)

Note: Numbers in parentheses are standard errors. * and ** denote that difference is statistically different from 0 at 5% and 10% level, respectively. Computations use design weights. Missing values for each of the baseline variables were imputed with the mean of the variable.