Partial Identification of Local Average Treatment Effects with an Invalid Instrument

Carlos A. Flores
Department of Economics, University of Miami
caflores@miami.edu

Alfonso Flores-Lagunes
Department of Economics, State University of New York at Binghamton, and IZA
aflores@binghamton.edu

October, 2012

Abstract

We derive nonparametric bounds for local average treatment effects ($LATE$) without imposing the exclusion restriction assumption or requiring an outcome with bounded support. Instead, we employ assumptions requiring weak monotonicity of mean potential and counterfactual outcomes within or across subpopulations defined by the values of the potential treatment status under each value of the instrument. The key element in our derivation is a result relating $LATE$ to a causal mediation effect, which allows us to exploit partial identification results from the causal mediation analysis literature. The bounds are employed to analyze the effect of attaining a GED, high school, or vocational degree on future labor market outcomes using randomization into a training program as an invalid instrument. The resulting bounds are informative, indicating that the local effect when assigned to training for those whose degree attainment is affected by the instrument is at most 12.7 percentage points on employment and $64.4 on weekly earnings.

Key words: causal inference, instrumental variables, mediation analysis, nonparametric bounds, principal stratification

1 Introduction

Instrumental variable (IV) methods are widely used in economics and other fields to estimate causal treatment effects by exploiting exogenous treatment variation induced by exogenous variation in the instrument. Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) (hereafter IA and AIR, respectively) show that in the presence of heterogeneous effects and under some assumptions, IV estimators point identify the local average treatment effect ($LATE$) for individuals whose treatment status is changed because of the instrument. A critical assumption of IV methods is the exclusion restriction, which requires that the instrument affects the outcome only through its effect on the treatment. Unfortunately, in many applications, it is debatable whether the instrument employed is valid (i.e., satisfies the exclusion restriction), or it is difficult to find valid instruments. In this paper, we derive nonparametric bounds for $LATE$ without imposing the exclusion restriction assumption or requiring an outcome with bounded support. Instead, we employ weak monotonicity assumptions on the mean potential and counterfactual outcomes of strata defined by the values of the potential treatment status under each value of the instrument. The key element in the derivation of our bounds is an important result relating $LATE$ to a causal mediation effect. This result generalizes the one in IA and AIR to invalid instruments, and allows us to exploit partial identification results from the causal mediation analysis literature (e.g., Sjölander, 2009; Flores and Flores-Lagunes, 2010).

Our paper is related to the studies by Hirano et al. (2000) and Mealli and Pacini (2012), who extend the IV framework in AIR to identify effects of the instrument on the outcome for different latent subpopulations (i.e., local intention-to-treat effects) while allowing for violations of the exclusion restriction. The first study uses a parametric Bayesian approach to point identify those effects, while the second derives nonparametric bounds by exploiting restrictions implied by the randomization of the instrument on the joint distribution of the outcome and an auxiliary variable (e.g., a covariate). In contrast, our main focus is on local treatment effects (as opposed to instrument effects).

There is a growing literature on partial identification of treatment effects in IV models. A strand of this literature constructs bounds on treatment effects assuming the validity of the instrument

(Manski, 1990, 1994; Balke and Pearl, 1997; Heckman and Vytlacil, 1999, 2000; Bhattacharya et al., 2008; Kitagawa, 2009; Huber and Mellace, 2010; Shaikh and Vytlacil, 2011; Chen et al., 2012), while another strand considers invalid instruments. Conley et al. (2012) use information on a parameter summarizing the extent of violation of the exclusion restriction along with distributional assumptions in the form of deterministic or probabilistic priors. Nevo and Rosen (2010) derive analytic bounds on average treatment effects by employing assumptions on the sign and extent of correlation between the instrument and the error term in a linear model. Our approach is different in that it is nonparametric and does not require modeling the extent of invalidity of the instrument nor its correlation with an error term; however, it is currently limited to the case of a binary and randomly assigned instrument and a binary endogenous regressor.

Similar to Manski and Pepper (2000), we study nonparametric partial identification of treatment effects without a valid instrument employing weak monotonicity assumptions, except we focus on $LATE$ (while they focus on the population average treatment effect) and our bounds do not require an outcome with bounded support (while, in general, theirs are uninformative without a bounded outcome). Our approach is also different in that it contains elements from the principal stratification framework (Frangakis and Rubin, 2002), which is rooted in AIR and Hirano et al. (2000). This framework is useful to analyze causal effects when allowing the IV to causally affect the outcome through channels other than the treatment, hence allowing the exclusion restriction to be violated.

Deriving bounds for $LATE$ is important for several reasons. $LATE$ is a widely used parameter in applied work and its bounds can be employed as a robustness check when estimating it, as illustrated in our empirical application. In addition, the bounds on $LATE$ can be used as a building block to construct bounds for the population average treatment effect, as we later discuss, and they can be more informative than the bounds on other parameters (e.g., the population average treatment effect) in some applications. Finally, in some settings $LATE$ is a relevant parameter even if the instrument is invalid, such as in experiments or quasi-experiments with imperfect compliance. For instance, consider the example in AIR of using the Vietnam era draft lottery as an IV for estimating the effect of military service on civilian earnings. This IV would be invalid if the draft lottery has a separate

effect on civilian earnings through channels other than military service (e.g., by affecting schooling decisions to postpone conscription). Even in this case, $LATE$ is a policy-relevant parameter as it measures the effect of military service on earnings for those induced to the military by the draft.

The bounds derived herein are likely to be relevant to applied researchers. First, in many experiments and quasi-experiments the randomized variable may not satisfy the exclusion restriction. For instance, in AIR's example the draft status could have affected civilian earnings through both veteran status and education. Similarly, the randomized variable may fail to satisfy the exclusion restriction even when an experiment is specifically designed for analyzing the effect of a particular treatment. For example, Hirano et al. (2000) analyze the effect of influenza vaccination on flu-related hospitalizations in a study where physicians were randomly encouraged to provide flu shots to high-risk patients. They find evidence that the encouragement could have affected the outcome through channels other than the receipt of the vaccination (e.g., through the receipt of other medical treatment), thus invalidating the use of the encouragement as an IV for vaccination. In the examples above, our bounds can be employed to learn about $LATE$ without imposing the exclusion restriction.

Second, our bounds can be used to exploit existing experiments to learn about treatment effects other than those for which the experiments were originally designed for. We illustrate the use of our bounds in this case by using random assignment into a training program as an IV to analyze the effect of attaining certain types of academic degrees on labor market outcomes. Finally, our results are also useful in designing experiments when it is not possible to randomize the treatment of interest because of financial or ethical reasons. In such cases, one could randomize a variable that affects the treatment instead, and employ the methods herein to bound the effect of interest.

2 Partial Identification of LATE

2.1 Relationship between $LATE$ and Causal Mediation Effects

Assume we have a random sample of size n from a large population. For each unit i in the sample, let $D_i \in \{0, 1\}$ indicate whether the unit received the treatment of interest ($D_i = 1$) or the control treatment ($D_i = 0$). We exploit exogenous variation in a binary instrument Z to learn

about the effect of D on an outcome Y, with $Z_i \in \{0,1\}$. Let $D_i(1)$ and $D_i(0)$ denote the potential treatment status; that is, the treatment status individual i would receive depending on the value of Z_i. This setting gives rise to four subpopulations (called "principal strata" in Frangakis and Rubin, 2002) within which individuals share the same values of the vector $\{D_i(0), D_i(1)\}$: $\{1,1\}$, $\{0,0\}$, $\{0,1\}$ and $\{1,0\}$, commonly referred to as always takers, never takers, compliers, and defiers, respectively (AIR, 1996). Also, let $Y_i(z,d)$ denote the potential or counterfactual outcome individual i would obtain if she received a value of the instrument and the treatment of z and d, respectively. For each unit i, we observe the vector (Z_i, D_i, Y_i), where $D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$ and $Y_i = D_i Y_i(Z_i, 1) + (1 - D_i) Y_i(Z_i, 0)$. This notation implicitly imposes the stable unit treatment value assumption, which implies that there is no interference between individuals and that there are no different versions of the treatments being analyzed (Rubin 1978, 1980, 1990). To simplify notation, in the rest of the paper we write the subscript i only when deemed necessary.

We begin by describing the key result in IA and AIR. They impose the following assumptions:

Assumption 1 (*Randomly Assigned Instrument*). $\{Y(1,1), Y(0,0), Y(0,1), Y(1,0), D(0), D(1)\}$ *is independent of* Z.

Assumption 2 *(Nonzero Average Effect of* Z *on* D*).* $E[D(1) - D(0)] \neq 0$.

Assumption 3 *(Individual-Level Monotonicity of* Z *on* D*).* $D_i(1) \geq D_i(0)$ *for all* i.

Assumption 3 rules out the existence of defiers. IA and AIR also impose the following assumption:

$$\textit{Exclusion Restriction Assumption: } Y_i(0,d) = Y_i(1,d) \text{ for all } i \text{ and } d \in \{0,1\}. \tag{1}$$

This assumption requires that any effect of the instrument on the potential outcomes is through the treatment status only. Vytlacil (2002) shows that the IV assumptions imposed by IA and AIR are equivalent to those imposed in nonparametric selection models.

IA and AIR show that if the exclusion restriction holds, along with Assumptions 1 to 3, the local average treatment effect of D on Y for the compliers (which they call $LATE$) is point identified as:

$$E[Y(z,1) - Y(z,0) \,|D(1) - D(0) = 1] = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}. \tag{2}$$

The exclusion restriction is crucial to the point identification result in (2). Intuitively, an average effect of D on Y is obtained by dividing the reduced-form effect of Z on Y by the effect of Z on D because the exclusion restriction guarantees that all of the effect of Z on Y works through D.

In what follows, we maintain Assumptions 1 to 3 and drop the exclusion restriction. Hence, instead of assuming that all the effect of the instrument on the outcome works through the treatment status, we now let the instrument have a causal effect on the outcome through other channels. As a result, in order to employ Z as an instrument to learn about the effect of D on Y, we need to disentangle the part of the effect of Z on Y that works through D from the part that works through the other channels. To this end, we relate this problem to the causal mediation analysis literature.

We introduce some additional notation. We use $Y_i^Z(1) = Y_i(1, D_i(1))$ and $Y_i^Z(0) = Y_i(0, D_i(0))$ as shorthand for the outcome individual i would obtain if exposed or not exposed to the instrument, respectively. Thus, the reduced-form average treatment effect (or intention-to-treat effect) of the instrument on the outcome is $ATE_{ZY} \equiv E[Y^Z(1) - Y^Z(0)]$. Also, let the counterfactual outcome $Y_i(1, D_i(0))$ represent the outcome individual i would obtain if exposed to the instrument, but kept the treatment status at the value if not exposed to the instrument. Intuitively, $Y_i(1, D_i(0))$ is the outcome from an alternative counterfactual experiment in which the instrument is the same as the original one, but we block the effect of Z on D by holding D_i fixed at $D_i(0)$. Note that the use of the counterfactual outcome $Y_i(1, D_i(0))$, which is an "entirely hypothetical"outcome (Rubin, 1990), makes our approach somewhat different from the original principal stratification framework of Frangakis and Rubin (2002) that employs only potential outcomes $Y_i^Z(1)$ and $Y_i^Z(0)$. A discussion clarifying our departure from principal stratification is provided in the Appendix.

Following Flores and Flores-Lagunes (2010), let the mediation (or mechanism) average treatment effect, or $MATE$, be given by

$$MATE = E[Y(1, D(1)) - Y(1, D(0))], \tag{3}$$

and let the net average treatment effect, or $NATE$, be given by

$$NATE = E[Y(1, D(0)) - Y(0, D(0)]. \tag{4}$$

$NATE$ and $MATE$ have received different names in other fields. They are also called, respectively, the (average) pure direct and indirect effects (Robins and Greenland, 1992; Robins, 2003), or the (average) natural direct and indirect effects (Pearl, 2001).

$MATE$ gives the average effect on the outcome from a change in the treatment status that is due to the instrument, holding the value of the instrument fixed at one. $NATE$ gives the average effect of the instrument on the outcome when the treatment status of every individual is held constant at $D_i(0)$. Since $ATE_{ZY} = MATE + NATE$, these two effects decompose the total average effect of the instrument on the outcome into the part that works through the treatment status ($MATE$) and the part that is net of the treatment-status channel ($NATE$). Note that $ATE_{ZY} = MATE$ if all the effect of Z on Y works through D (i.e., under the exclusion restriction), whereas $ATE_{ZY} = NATE$ if none of the effect of Z on Y works through D. Finally, note that an alternative decomposition of ATE_{ZY} can be made by defining $MATE$ and $NATE$ using the counterfactual outcome $Y(0, D(1))$ instead of $Y(1, D(0))$. The two decompositions can differ because of treatment effect heterogeneity. We work with the definitions of $MATE$ and $NATE$ in (3) and (4), and discuss below the implication of the way the decomposition is made on the definition of $LATE$.

We now relate $MATE$ in (3) to a relevant effect of D on Y by writing it as:

$$\begin{aligned} MATE &= E[Y(1, D(1)) - Y(1, D(0))] \\ &= E\{[D(1) - D(0)] \cdot [Y(1,1) - Y(1,0)]\} \\ &= \Pr(D(1) - D(0) = 1) \cdot E[Y(1,1) - Y(1,0) | D(1) - D(0) = 1] \\ &\quad - \Pr(D(1) - D(0) = -1) \cdot E[Y(1,1) - Y(1,0) | D(1) - D(0) = -1]. \end{aligned} \tag{5}$$

The second line in (5) writes $MATE$ as the expected value of the product of the individual effect of the instrument on the treatment status times the individual effect from a change in the treatment status on the outcome, holding the value of the instrument fixed at one. The third line uses iterated expectations, and is the basis for the following proposition.

Proposition 1 *Under Assumptions 2 and 3 we can write*

$$LATE \equiv E\left[Y\left(1,1\right)-Y(1,0)|D\left(1\right)-D\left(0\right)=1\right]=\frac{MATE}{E\left[D(1)-D(0)\right]}. \tag{6}$$

Proposition 1 follows directly from (5) by ruling out the existence of defiers. As in IA and AIR, we refer to the parameter $E\left[Y\left(1,1\right)-Y(1,0)|D\left(1\right)-D\left(0\right)=1\right]$ as the local average treatment effect ($LATE$). It gives the average treatment effect for compliers under exposure to the instrument. Proposition 1 writes $LATE$ as a function of $MATE$ and the average effect of the instrument on the treatment status, where now $MATE$ plays the role of the reduced-form effect of the instrument on the outcome when the exclusion restriction holds.

Proposition 1 is important because it generalizes the IA and AIR result in (2) to include the case when the exclusion restriction is violated, and since the denominator in (6) is point identified under Assumption 1, it allows us to employ identification results on $MATE$ to identify $LATE$. In principle, point identification results on MATE (e.g., Robins and Greenland, 1992; Imai et al., 2010; Flores and Flores-Lagunes, 2011) could be employed to point identify $LATE$. However, these results require selection into the mechanism (in our case the treatment) to be based on observables, which runs contrary to the logic for using an IV strategy to control for unobservable factors (see also Mealli and Mattei, 2012). Therefore, we focus on partial identification of $LATE$.

Before presenting our bounds on $LATE$, it is instructive to relate Proposition 1 to the IA and AIR result in (2). The exclusion restriction in (1) implies that $MATE = ATE_{ZY}$, which obtains the result in IA and AIR as a special case of Proposition 1. In addition, the exclusion restriction implies that $E\left[Y\left(1,1\right)-Y(1,0)|D\left(1\right)-D\left(0\right)=1\right]=E\left[Y\left(0,1\right)-Y(0,0)|D\left(1\right)-D\left(0\right)=1\right]$, so in this case specifying whether the effect of the treatment on the outcome is under exposure to the instrument is irrelevant. In our setting, however, this distinction is important because we allow the instrument to have a net or direct effect on the outcome, so average treatment effects can be different depending on whether or not the individuals are exposed to the instrument. As a result, the $LATE$ in (6) is not exactly the same as that in (2) without further assumptions (e.g., assuming that $LATE$ under exposure to the IV is the same as under no exposure).

Two additional remarks are in order. First, similar to the $LATE$ in (2), the specific instrument employed is crucial in interpreting the $LATE$ in (6), as we illustrate in our empirical application. Second, it is possible to obtain a result analogous to (6) for the average treatment effect for compliers under no exposure to the instrument, $E[Y(0,1) - Y(0,0)|D(1) - D(0) = 1]$, by using the counterfactual outcome $Y(0, D(1))$ instead of $Y(1, D(0))$ in the definition of $MATE$ and $NATE$. The choice of $LATE$ to consider (whether or not under exposure to the instrument) depends on the particular application at hand.

2.2 Bounds on LATE

In this section, we present bounds on $LATE$ in (6) based on Proposition 1 and the partial identification results on $MATE$ in Flores and Flores-Lagunes (2010). Given bounds on $MATE$, bounds on $LATE$ follow from Proposition 1 and point identification of $E[D(1) - D(0)]$.

Partial identification of $MATE$ in Flores and Flores-Lagunes (2010) is attained from the level of the strata up. To simplify notation, we write at, nt, c and d to refer to the strata of always takers, never takers, compliers, and defiers, respectively. Define local versions of $MATE$ and $NATE$ as the corresponding average effects within strata:

$$LMATE_k = E[Y^Z(1)|k] - E[Y(1, D(0))|k], \text{ for } k = at, nt, c, d; \text{ and,} \tag{7}$$

$$LNATE_k = E[Y(1, D(0))|k] - E[Y^Z(0)|k], \text{ for } k = at, nt, c, d. \tag{8}$$

The fact that $D_i(0) = D_i(1)$ for the always and never takers implies that for these two strata $Y^Z(1) = Y(1, D(0))$, so $LMATE_k = 0$ and $LNATE_k = E[Y^Z(1) - Y^Z(0)|k]$ for $k = at, nt$. It also implies that the observed data contains information on $Y(1, D(0))$ only for those treated individuals in the nt and at strata (Flores and Flores-Lagunes, 2010; Mealli and Mattei, 2012). Note that $LNATE_{at}$ and $LNATE_{nt}$ are local intention-to-treat effects like those considered in Hirano et al. (2000) and Mealli and Pacini (2012). In addition, $LATE$ in (6) equals $LMATE_c$ since $LMATE_c = E[Y(1, D(1)) - Y(1, D(0))|c] = E\{[D(1) - D(0)] \cdot [Y(1,1) - Y(1,0)]|c\} = E[Y(1,1) - Y(1,0)|c]$.

To briefly motivate the bounds on $MATE$ in Flores and Flores-Lagunes (2010), consider Table 1 which summarizes the relationship between the compliance behavior of the individuals in the sample

and their observed treatment status (D_i) and instrument exposure (Z_i), under Assumption 3:

Table 1

		Z_i = 0	Z_i = 1
D_i	0	nt, c	nt
	1	at	at, c

Under Assumptions 1 and 3, it is possible to point identify the proportion of each of the strata in the population, and to point or partially identify the mean potential outcomes and local effects of certain strata. Let π_{nt}, π_{at}, and π_c be the population proportions of the principal strata nt, at, and c, respectively, and let $p_{d|z} \equiv \Pr(D_i = d|Z_i = z)$ for $d, z = 0, 1$. Then, $\pi_{nt} = p_{0|1}$, $\pi_{at} = p_{1|0}$, $\pi_c = p_{1|1} - p_{1|0} = p_{0|0} - p_{0|1}$, $E[Y^Z(0)|at] = E[Y|Z = 0, D = 1]$, and $E[Y^Z(1)|nt] = E[Y|Z = 1, D = 0]$. In addition, it is possible to construct bounds on $E[Y^Z(1)|at]$, $E[Y^Z(0)|nt]$, $E[Y^Z(0)|c]$, and $E[Y^Z(1)|c]$ by employing the trimming procedure in Lee (2009) and Zhang et al. (2008). For example, consider constructing an upper bound for $E[Y^Z(0)|nt]$. The average outcome for the individuals in the $(Z, D) = (0, 0)$ group can be written as:

$$E[Y|Z = 0, D = 0] = \frac{\pi_{nt}}{\pi_{nt} + \pi_c} \cdot E[Y^Z(0)|nt] + \frac{\pi_c}{\pi_{nt} + \pi_c} \cdot E[Y^Z(0)|c]. \tag{9}$$

Since $\pi_{nt}/(\pi_{nt} + \pi_c) = p_{0|1}/p_{0|0}$, $E[Y^Z(0)|nt]$ can be bounded from above by the expected value of Y for the $p_{0|1}/p_{0|0}$ fraction of the largest values of Y for those in the observed group $(Z, D) = (0, 0)$. A similar approach can be used to construct a lower bound on $E[Y^Z(0)|nt]$, as well as bounds on $E[Y^Z(0)|c]$, $E[Y^Z(1)|at]$, and $E[Y^Z(1)|c]$, where for the last two terms the observed group $(Z, D) = (1, 1)$ is employed. Moreover, note that the bounds on $E[Y^Z(0)|nt]$ and $E[Y^Z(1)|at]$ can be used to construct bounds on $LNATE_{nt}$ and $LNATE_{at}$, respectively, as the other term in the definition of each of these $LNATE$s is point identified (see equation (8)). As we further discuss in Section 4, the bounds on $LNATE_{nt}$ and $LNATE_{at}$ can shed light on the validity of the exclusion restriction assumption, as this assumption implies that $LNATE_{nt} = LNATE_{at} = 0$ (e.g., Hirano et al., 2000; Huber and Mellace, 2010, 2011; Mealli and Pacini, 2012).

Flores and Flores-Lagunes (2010) use the point identified quantities and trimming bounds above as building blocks to construct bounds on $MATE$ by writing it in different ways as a function of the

local effects and average potential and counterfactual outcomes of the three existing strata as:

$$
\begin{aligned}
& MATE \\
= \; & \pi_c LMATE_c && (10) \\
= \; & \pi_{nt} E\left[Y^Z(0)\,|nt\right] + \pi_{at} E\left[Y^Z(0)\,|at\right] + \pi_c E\left[Y^Z(1)\,|c\right] - \pi_c LNATE_c - E\left[Y^Z(0)\right] && (11) \\
= \; & E\left[Y^Z(1)\right] - \pi_{at} E\left[Y^Z(1)\,|at\right] - \pi_{nt} E\left[Y^Z(1)\,|nt\right] - \pi_c E\left[Y(1, D(0))\,|c\right] && (12) \\
= \; & E\left[Y^Z(1)\right] - E\left[Y^Z(0)\right] - \pi_{at} LNATE_{at} - \pi_{nt} LNATE_{nt} - \pi_c LNATE_c. && (13)
\end{aligned}
$$

Intuitively, bounds on $MATE$ are constructed by partially or point identifying quantities at the strata level and then employing equations (10) to (13) to bound $MATE$.

Note that to partially identify $MATE$, and hence $LATE$, further assumptions are needed because the data have no information on $Y(1, D(0))$ for compliers, so the terms $E[Y(1, D(0))\,|c]$, $LMATE_c$, and $LNATE_c$ appearing in equations (10) to (13) are not identified. As in Flores and Flores-Lagunes (2010), we consider two additional sets of assumptions. The first imposes weak monotonicity of mean potential and counterfactual outcomes within strata.

Assumption 4. (*Weak Monotonicity of Mean Potential and Counterfactual Outcomes within Strata).*

4.1. $E[Y^Z(1)\,|c] \geq E[Y(1, D(0))\,|c]$. *4.2.* $E[Y(1, D(0))\,|k] \geq E[Y^Z(0)\,|k]$, for $k = nt, at, c$.

Assumption 4.1 implies that $LMATE_c (= LATE) \geq 0$, so that the treatment has a non-negative average effect on the outcome for the compliers. When combined with Assumption 3, it also implies that $MATE = \pi_c LMATE_c \geq 0$. Assumption 4.2 states that, for each stratum, the instrument has a non-negative average effect on the outcome net of the effect that works through the treatment status. It requires that $LNATE \geq 0$ for all strata, which implies that $NATE \geq 0$. Hence, under Assumptions 3 and 4, we have $ATE_{ZY} \geq 0$, and the instrument is assumed to have a non-negative average effect on the outcome.

Assumptions similar to Assumption 4 have been considered for partial identification of average treatment effects in IV models (e.g., Manski and Pepper, 2000) and in other settings (Manski, 1997; Sjölander, 2009). For instance, Manski and Pepper (2000) consider the "monotone treatment

response" (MTR) assumption, which states that the individual-level effect of the treatment on the outcome is non-negative for every individual. In contrast to the MTR assumption, note that Assumption 4.1 allows some individual effects of the treatment on the outcome to be negative by imposing this monotone restriction on the average treatment effect for the compliers.

Let y_r^{zd} be the r-th quantile of Y conditional on $Z = z$ and $D = d$. For ease of exposition, suppose Y is continuous so that $y_r^{zd} = F^{-1}_{Y|Z=z,D=d}(r)$, with $F_{\cdot}(\cdot)$ the CDF of Y given $Z = z$ and $D = d$. Let $U^{z,k}$ and $L^{z,k}$ denote, respectively, the upper and lower bounds on the mean potential outcome $Y^Z(z)$ for stratum k derived using the trimming procedure described above, with $z \in \{0, 1\}$ and $k \in \{at, nt, c\}$. Proposition 2 presents bounds on $LATE$ under Assumptions 1 through 4.

Proposition 2 *If Assumptions 1 through 4 hold,*

$$0 \leq LATE \leq \frac{U}{E[D|Z=1] - E[D|Z=0]}, \text{ where}$$

$$\begin{aligned} U &= E[Y|Z=1] - E[Y|Z=0] - p_{1|0} \max\left\{0, L^{1,at} - E[Y|Z=0, D=1]\right\} \\ &\quad - p_{0|1} \max\left\{0, E[Y|Z=1, D=0] - U^{0,nt}\right\} \\ U^{0,nt} &= E[Y|Z=0, D=0, Y \geq y^{00}_{1-(p_{0|1}/p_{0|0})}] \\ L^{1,at} &= E[Y|Z=1, D=1, Y \leq y^{11}_{(p_{1|0}/p_{1|1})}]. \end{aligned}$$

Proof. See Internet Appendix.

Proposition 2 implies that under Assumptions 1 to 4 the upper bound on $LATE$ is at most equal to the usual IV estimand in (2), since $U \leq E[Y|Z=1] - E[Y|Z=0]$.

In contrast to Assumption 4, the next set of assumptions does not restrict the sign of $LATE$. It involves weak monotonicity of mean potential and counterfactual outcomes across strata.

Assumption 5. (*Weak Monotonicity of Mean Potential and Counterfactual Outcomes across Strata).*

5.1. $E[Y(1, D(0))|c] \geq E\left[Y^Z(1)|nt\right]$. *5.2.* $E\left[Y^Z(1)|at\right] \geq E[Y(1, D(0))|c]$. *5.3.* $E\left[Y^Z(0)|c\right] \geq E\left[Y^Z(0)|nt\right]$. *5.4.* $E\left[Y^Z(0)|at\right] \geq E\left[Y^Z(0)|c\right]$. *5.5.* $E\left[Y^Z(1)|c\right] \geq E\left[Y^Z(1)|nt\right]$. *5.6.* $E\left[Y^Z(1)|at\right] \geq E\left[Y^Z(1)|c\right]$.

This assumption states that the mean potential and counterfactual outcomes of the always takers are greater than or equal to those of the compliers, and that these in turn are greater than or equal to those of the never takers. Assumption 5 formalizes the notion that some strata are likely to have more favorable characteristics and thus better mean outcomes than others. For example, in our empirical application it requires that, on average, individuals who are more likely to attain an academic degree are also more likely to have favorable labor market outcomes.

Assumption 5 has two attractive features: it may be substantiated with economic theory, and the combination of Assumptions 1, 3, and 5 provide the testable implications $E[Y|Z = 0, D = 1] \geq E[Y|Z = 0, D = 0]$ and $E[Y|Z = 1, D = 1] \geq E[Y|Z = 1, D = 0]$ that can be used to falsify the assumptions. Moreover, it is possible to obtain indirect evidence about the plausibility of Assumption 5 by comparing relevant average baseline characteristics (e.g., pre-treatment outcomes) of the different strata (e.g., Flores and Flores-Lagunes, 2010; Frumento et al., 2012; Chen et al., 2012). These tools are illustrated and further discussed in our empirical analysis.

Assumption 5 is implicit in some selection and single-index models, and similar monotonicity assumptions across strata have been considered previously in other settings (e.g., Zhang and Rubin, 2003; Zhang et al., 2008, 2009). Assumption 5 is also related to, but different from, the monotone instrumental variable (MIV) assumption in Manski and Pepper (2000). The MIV assumption states that mean potential outcomes as a function of the treatment vary weakly monotonically across subpopulations defined by specific observed values of the instrument: $E[Y(d)|Z = 1] \geq E[Y(d)|Z = 0]$ for $d = \{0, 1\}$. Assumption 5 differs from the MIV assumption in at least two important ways. First, Assumption 5 refers to potential and counterfactual outcomes that explicitly allow the instrument to have a causal effect on the outcome (through D and other channels) by writing them as a function of the treatment and the instrument. Second, Assumption 5 imposes weak inequality restrictions across subpopulations defined by specific values of the potential treatment status (principal strata).

Proposition 3 presents bounds on $LATE$ employing Assumptions 1, 2, 3, and 5.

Proposition 3 *If Assumptions 1, 2, 3, and 5 hold,*

$$\frac{\overline{L}}{E[D|Z=1]-E[D|Z=0]} \leq LATE \leq \frac{\overline{U}}{E[D|Z=1]-E[D|Z=0]}, \text{ where}$$

$$\begin{aligned}
\overline{L} &= \left(p_{1|1}-p_{1|0}\right)\cdot\left(\max\left\{L^{1,c}, E\left[Y|Z=1,D=0\right]\right\}-U^{1,at}\right) \\
\overline{U} &= \left(p_{1|1}-p_{1|0}\right)\left(E\left[Y|Z=1,D=1\right]-E\left[Y|Z=1,D=0\right]\right) \\
U^{1,at} &= E[Y|Z=1,D=1,Y\geq y^{11}_{1-(p_{1|0}/p_{1|1})}] \\
L^{1,c} &= E[Y|Z=1,D=1,Y\leq y^{11}_{1-(p_{1|0}/p_{1|1})}].
\end{aligned}$$

Proof. See Internet Appendix.

The lower bound on $LATE$ in Proposition 3 is always less than or equal to zero because $p_{1|1} - p_{1|0} = \pi_c \geq 0$ and $U^{1,at}$ is always greater than or equal to $L^{1,c}$ and $E[Y|Z=1,D=0]$ (from the testable implications above). Therefore, the bounds in Proposition 3 cannot be used to rule out a negative $LATE$. Nevertheless, as illustrated in our empirical application, the upper bound on $LATE$ in this proposition can be informative.

Finally, Assumptions 1 through 5 can be combined to construct bounds on $LATE$. This yields an additional testable implication: $E[Y|Z=1,D=1] \geq E[Y|Z=0,D=0]$.

Proposition 4. *If Assumptions 1 through 5 hold,*

$$0 \leq LATE \leq \frac{\min\{\widetilde{U}_1, \widetilde{U}_2\}}{E[D|Z=1]-E[D|Z=0]}, \text{ where}$$

$$\begin{aligned}
\widetilde{U}_1 &= E\left[Y|Z=1\right]-p_{1|0}\max\{E\left[Y|Z=1,D=1\right], E\left[Y|Z=0,D=1\right]\} \\
&\quad -\left(p_{1|1}-p_{1|0}\right)\max\{E[Y|Z=1,D=0], E[Y|Z=0,D=0]\} \\
&\quad -p_{0|1}E\left[Y|Z=1,D=0\right] \\
\widetilde{U}_2 &= E\left[Y|Z=1\right]-E\left[Y|Z=0\right] \\
&\quad -p_{1|0}\max\{0, E\left[Y|Z=1,D=1\right]-E\left[Y|Z=0,D=1\right]\} \\
&\quad -p_{0|1}\max\{0, E\left[Y|Z=1,D=0\right]-E\left[Y|Z=0,D=0\right]\}-\left(p_{1|1}-p_{1|0}\right)\cdot \\
&\quad \max\{0, E\left[Y|Z=1,D=0\right]-U^{0,c}, E\left[Y|Z=1,D=0\right]-E\left[Y|Z=0,D=1\right]\} \\
U^{0,c} &= E[Y|Z=0,D=0,Y\geq y^{00}_{(p_{0|1}/p_{0|0})}].
\end{aligned}$$

Proof. See Internet Appendix.

Similar to Proposition 2, $\widetilde{U}_2$ implies that the upper bound on $LATE$ in Proposition 4 is at most equal to the IV estimand in (2), and Assumption 4.1 implies that the lower bound on $LATE$ is zero.

3 Estimation and Inference

The bounds derived in Section 2 involve minimum (min) and maximum (max) operators, which are problematic for standard estimation and inferences procedures. Sample analog estimates of this type of bounds tend to be narrower than the true bounds because of the concavity and convexity of the min and max functions, respectively, and the asymptotic distribution of the bound estimators is usually unavailable. Moreover, Hirano and Porter (2012) 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 references therein). We use the methodology proposed by Chernozhukov, Lee and Rosen (2011) (hereafter CLR) to obtain confidence regions for the true parameter value and half-median unbiased estimators for our lower and upper bounds. The half-median-unbiasedness property means that the upper bound estimator exceeds the true value of the upper bound with probability at least one half asymptotically, while the reverse holds for the lower bound.

To briefly describe CLR's procedure as applied to our setting, let the bounds for a parameter θ_0 (e.g., $LATE$) be given by $[\theta_0^l, \theta_0^u]$, where $\theta_0^l = \max_{v \in \mathcal{V}^l = \{1,\ldots,m^l\}} \theta^l(v)$ and $\theta_0^u = \min_{v \in \mathcal{V}^u = \{1,\ldots,m^u\}} \theta^u(v)$. CLR refer to $\theta^l(v)$ and $\theta^u(v)$ as bounding functions. In our setting, v indexes the bounding functions, while m^l and m^u give, respectively, the number of terms inside the max and min operators. It is straightforward to write the bounds from the previous section in this form by manipulating the min and max operators. For example, the lower bound for $LATE$ in Proposition 3 can be written as $\theta_0^l = \max_{v \in \mathcal{V}^l = \{1,2\}} \theta^l(v) = \max\{\theta^l(1), \theta^l(2)\}$, with $\theta^l(1) \equiv L^{1,c} - U^{1,at}$ and $\theta^l(2) \equiv E[Y|Z=1, D=0] - U^{1,at}$. Similarly, the upper bound in Proposition 2 contains 4 bounding functions, while it contains 16 bounding functions in Proposition 4. In our case, sample analog

estimators of the bounding functions $\theta^l(v)$ and $\theta^u(v)$ are known to be consistent and asymptotically normally distributed, as they are simple functions of proportions, conditional means, and trimmed means (Lee, 2009; Newey and McFadden, 1994).

CLR address the issues related to estimation and inference for the bounds $[\theta_0^l, \theta_0^u]$ by employing precision-corrected estimates of the bounding functions before applying the min and max operators. The precision adjustment consists of adding to each estimated bounding function its pointwise standard error times an appropriate critical value, $\kappa(p)$, so that estimates with higher standard errors receive larger adjustments. Depending on the choice of $\kappa(p)$, it is possible to obtain confidence regions for either the identified set or the true parameter value, and half-median unbiased estimators for the lower and upper bounds.

More specifically, the precision-corrected estimator of the upper bound θ_0^u is given by

$$\widehat{\theta}^u(p) = \min_{v \in \mathcal{V}^u} [\widehat{\theta}^u(v) + \kappa^u_{\widehat{V}^u_n}(p) s^u(v)], \tag{14}$$

where $\widehat{\theta}^u(v)$ is the sample analog estimator of $\theta^u(v)$ and $s^u(v)$ is its standard error. CLR compute the critical value $\kappa^u_{\widehat{V}^u_n}(p)$ based on simulation methods and a preliminary estimator $\widehat{V}^u_n$ of $V^u = \arg\min_{v \in \mathcal{V}^u} \theta^u(v)$. Intuitively, $\widehat{V}^u_n$ selects those bounding functions that are close enough to binding to affect the asymptotic distribution of the estimator of the upper bound. A precision-corrected estimator of the lower bound θ_0^l is obtained in a similar way. Further details on the CLR methodology and our specific implementation steps are provided in the Internet Appendix.

4 Bounds on the Labor Market Effects of Attaining a Degree

In this section we illustrate the use of our bounds by analyzing the effect of attaining a general educational development (GED), high school, or vocational degree on labor market outcomes using randomization into the Job Corps (JC) training program as an instrument. We employ data from the National Job Corps Study (NJCS), a randomized experiment performed in the mid-1990s to evaluate the effectiveness of JC. This application is relevant given the large empirical literature analyzing the effects on earnings of education (e.g., Card, 1999) and degree attainment (e.g., Hungerford and Solon, 1987; Cameron and Heckman, 1993; Jaeger and Page, 1996; Flores-Lagunes and Light, 2010).

Recall that there are two leading situations where our bounds can be used. The first is when it is debatable whether the exclusion restriction holds. The second is when the randomized variable in an existing experiment is used as an IV to learn about treatment effects other than those for which the experiment was designed for, in which case the exclusion restriction is unlikely to hold. Our application illustrates the latter situation, which is subtler. In particular, we do not argue that random assignment in the NJCS is not a valid IV for analyzing the effect of JC participation adjusting for non-compliance with the assigned treatment. Instead, we employ random assignment as an IV to bound the effect of a different treatment (attaining a degree) from that for which the experiment was designed for (JC participation).

4.1 The Job Corps Program, Data, and Preliminary Analysis

JC is U.S.'s largest and most comprehensive job training program for economically disadvantaged youth aged 16 to 24 years old. In addition to academic and vocational training, JC provides its participants a variety of services such as health services, counseling, job search assistance, social skills training, and a stipend during program enrollment, as well as room and board for those residing at the JC centers during program enrollment. In the NJCS, a random sample of all pre-screened eligible applicants in the 48 contiguous states and the District of Columbia was randomly assigned into treatment and control groups, with the second group being denied access to JC for three years. Both groups were tracked with a baseline interview immediately after randomization and thereafter at 12, 30 and 48 months. For further details on JC and the NJCS see Schochet et al. (2001).

Since the original NJCS reports, which found statistically positive effects of JC participation on employment and weekly earnings (Schochet et al., 2001), several papers have analyzed different aspects of JC using the NJCS data. For instance, Flores-Lagunes et al. (2010) analyze why positive effects of JC are found for whites and blacks but not for Hispanics, while Flores et al. (2012) study the effects of different lengths of exposure to academic and vocational instruction in JC under a selection-on-observables assumption. The wage effects of JC adjusting for selection into employment have also been considered. Zhang et al. (2009) and Frumento et al. (2012) point identify wage effects

for certain principal strata, with the latter paper also addressing the issues of non-compliance and missing outcomes, and analyzing the average characteristics and outcomes of the different strata. Lee (2009) and Blanco et al. (2012) use trimming bounds analogous to those for $LNATE_{at}$ discussed in Section 2.2 to bound, respectively, the average and quantile wage effects of JC for individuals who would be employed regardless of their random assignment. Flores and Flores-Lagunes (2010) use their bounds on $NATE$ and $MATE$ to analyze the mechanisms through which JC affects labor market outcomes, focusing on the attainment of a high school, GED, or vocational degree as a potential mechanism. Importantly, the present application is different from theirs in that the focus here is on analyzing the returns to attaining a degree (as opposed to analyzing how JC works), employing random assignment as a potentially invalid instrument. In other words, disentangling the part of the effect of JC on labor market outcomes that is due to attaining a degree is not the purpose of the current analysis, but rather an intermediate step for bounding the effect of interest.

Our sample consists of all individuals with non-missing values on the randomized treatment status, the variables regarding the attainment of a GED, high school, or vocational degree, and the outcome variables considered. We concentrate on estimating the returns to attaining any combination of those three degrees because many JC participants attain at least two of them (a GED or high school degree plus a vocational degree), and thus breaking up the effects of the different degrees would require additional assumptions. We focus on the outcomes measured 36 months after randomization, which corresponds to the time the embargo from JC ended for the control group. We use as an instrument the randomized indicator for whether or not the individual was assigned to participate in JC, hereafter referred to as the "random assignment status". We note that there existed non-compliance with the assigned treatment in the NJCS (73 percent of those in the treatment group enrolled in JC, while 1.4 percent of those in the control group managed to enroll in JC). We discuss below the implications of this non-compliance for the interpretation of our results.

Table 2 presents point estimates for some relevant quantities. The population average effect, or intention-to-treat effect (ITT), of the random assignment status on the probability of being employed 36 months after randomization is 4 percentage points, while the ITT on weekly earnings is \$18.1. The

Table 2. Point Estimates of Interest

Parameters	*Estimate*		*Standard Error*	
Average Treatment Effects				
ITT of instrument on employment	0.041		*(0.011)*	
ITT of instrument on earnings	18.11		*(4.759)*	
Average effect of instrument on degree attainment	0.209		*(0.011)*	
LATE of degree attainment on employment	0.194		*(0.055)*	
LATE of degree attainment on earnings	86.63		*(22.769)*	
Strata Proportions				
π_{nt}	0.34		*(0.007)*	
π_{at}	0.45		*(0.009)*	
π_{c}	0.21		*(0.011)*	
	Employment		**Weekly Earnings**	
Conditional Means	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
E[Y|Z=0]	0.61	*(0.009)*	170.85	*(3.703)*
E[Y|Z=1]	0.65	*(0.007)*	188.96	*(2.933)*
Testable Implications				
E[Y|Z=0, D=1]-E[Y|Z=0, D=0]	0.09	*(0.018)*	48.87	*(7.365)*
E[Y|Z=1, D=1]-E[Y|Z=1, D=0]	0.15	*(0.014)*	70.50	*(5.894)*
E[Y|Z=1, D=1]-E[Y|Z=0, D=0]	0.13	*(0.015)*	64.23	*(5.708)*

Notes: The labor market outcomes (*Y*) are either weekly earnings or employment status in month 36 after randomization. The treatment (*D*) is the attainment of a high school, GED, or vocational degree. The instrumental variable (*Z*) is an indicator for whether the individual was randomly assigned to participate in JC. Sample size is 8,020 individuals: 2,975 with *Z*=0 and 5,045 with *Z*=1. Standard errors are based on 5,000 bootstrap replications.

average (first-stage) effect of the random assignment status on the probability of attaining a degree is 21 percentage points. All three effects are highly statistically significant. The IV point estimate for the effect of attaining a degree on employment and weekly earnings using random assignment status as an instrument is 19.4 percentage points and $86.63, respectively, and both are highly statistically significant. These are point estimates of the IA and AIR $LATE$ in (2) under Assumptions 1 to 3 plus the exclusion restriction assumption in (1), which requires that all the effect of the random assignment status on employment and earnings works through the attainment of a degree. This assumption is likely violated because JC may have an effect on the outcomes through other services, such as job search services or social skills training. In fact, below we present evidence that JC has an effect on both outcomes net of the effect that works through the attainment of a degree, which under the assumptions considered implies that the exclusion restriction is violated.

4.2 Discussion of Assumptions

Assumption 1 holds by design, and the results in Table 2 suggest that Assumption 2 is satisfied. Assumption 3 states that there are no defiers: individuals who would obtain a GED, high school, or

vocational degree only if they were assigned to not participate in JC. This assumption is plausible given that JC facilitates the obtainment of such a degree. In our application, the never takers are those individuals who would never obtain a degree regardless of whether or not they are assigned to participate in JC, while the always takers are those who would always obtain a degree regardless of their assignment. The compliers are those who would obtain a degree only if assigned to participate in JC. Table 2 shows the estimated proportions of each of these strata (under Assumptions 1 and 3).

Assumption 4.1 states that the attainment of a degree has a non-negative average effect on employment and earnings for the compliers, which is consistent with conventional human capital models. Assumption 4.2 states that the combination of the rest of the channels through which the random assignment status affects the outcome has a non-negative average effect on labor market outcomes for all strata. This assumption seems plausible because the other components of JC (e.g., job search assistance, social skills training) also aim to improve the participants' future labor market outcomes. Nevertheless, a potential threat to the validity of this assumption is that individuals are "locked-in" away from employment while undergoing training, which could negatively affect their labor market outcomes. Note, however, that the fact that our outcomes are measured 36 months after randomization (when most individuals have finished training) decreases the relevance of this issue. In addition, this assumption is imposed on the $LNATE$s (as opposed to the individual net effects), which increases its plausibility as it allows some individuals to experience a negative net effect. We present empirical evidence below that $LNATE_{at} > 0$ (under Assumptions 1, 3, and 5).

Finally, Assumption 5 states that the average potential and counterfactual outcomes of the compliers are no less than the corresponding average potential outcomes of the never takers, and no greater than those of the always takers. This assumption seems plausible in our application given the characteristics of the individuals expected to belong to each strata. For instance, we would expect individuals with more favorable traits to succeed in the labor market (e.g., discipline) to belong to the always-taker stratum rather than to the complier stratum, with the reverse holding when comparing the compliers to the never takers.

Although Assumption 5 is not directly testable, indirect evidence on its plausibility can be gained

from comparing mean baseline characteristics that are closely related to the outcomes for the different strata (e.g., Flores and Flores-Lagunes, 2010; Frumento et al., 2012; Chen et al., 2012). If these comparisons suggest that compliers have better average baseline characteristics than always takers, or worse characteristics than never takers, then Assumption 5 is less likely to hold. In this application, the probability of being employed in the year prior to randomization for never takers, compliers, and always takers are, respectively (standard errors in parenthesis), 0.153 (.009); 0.205 (.010); and 0.216 (.011). The corresponding numbers for the mean weekly earnings in the year prior to randomization are \$86.26 (2.57); \$117.73 (12.95); and \$109.74 (3.09). For both always takers and compliers, the two mean characteristics are statistically greater than those of the never takers, while the differences of those two means between always takers and compliers are not statistically significant. Thus, the data do not provide indirect evidence against Assumption 5. In addition, the last three rows of Table 2 verify that the testable implications of Assumptions 4 and 5 discussed in Section 2.2 hold. In sum, although not innocuous, Assumptions 3 to 5 appear acceptable in our application.

4.3 Results

Table 3 shows half-median unbiased estimates of the bounds on $LATE$ in Propositions 2 through 4 for employment and weekly earnings, along with 95 percent confidence intervals for the true parameter value based on the methodology discussed in Section 3. Table 3 also presents results for $MATE$, $LNATE_{nt}$, and $LNATE_{at}$. We begin by focusing on $LNATE_{nt}$ and $LNATE_{at}$. Their bounds can provide evidence on the validity of the exclusion restriction, as this assumption implies that $LNATE_{nt} = LNATE_{at} = 0$. The idea of analyzing the direct or net effect of the instrument on the outcome for particular strata to assess the plausibility of the exclusion restriction has been considered previously by Hirano et al. (2000), and similar ideas appear in Huber and Mellace (2010, 2011) and Mealli and Pacini (2012) (see also Heckman and Vytlacil, 2005, and Kitagawa, 2008, 2009, for more on testing implications of the $LATE$ assumptions). The specific formulas for the bounds on $LNATE_{nt}$ and $LNATE_{at}$ are provided in the Internet Appendix. The first two columns of Table 3 show bounds on these two parameters under Assumptions 1 and 3, which correspond to the trimming

Table 3. Estimated Bounds on the Effect of Degree Attainment on Labor Market Outcomes Using Random Assignment Status as an Invalid Instrument

			Proposition 2		Proposition 3		Proposition 4	
Assumptions:	1 and 3		1, 2, 3, and 4		1, 2, 3, and 5		1, 2, 3, 4 and 5	
Parameters	LB	UB	LB	UB	LB	UB	LB	UB
			Employment Outcome					
1. $LNATE_{nt}$	-0.365	0.241	0.000	0.241	-0.019	0.241	0.000	0.241
	[-0.419, 0.289]		[0.000, 0.289]		[-0.048, 0.289]		[0.000, 0.289]	
2. $LNATE_{at}$	-0.095	0.341	0.000	0.341	0.044	0.341	0.044	0.341
	[-0.129, 0.364]		[0.000, 0.364]		[0.019, 0.364]		[0.019, 0.364]	
3. MATE			0.000	0.044	-0.094	0.032	0.000	0.027
			[0.000, 0.064]		[-0.104, 0.037]		[0.000, 0.036]	
4. LATE			0.000	0.211	-0.448	0.151	0.000	0.127
			[0.000, 0.308]		[-0.470, 0.175]		[0.000, 0.166]	
			Weekly Earnings Outcome					
1. $LNATE_{nt}$	-96.41	111.09	0.00	111.09	-6.28	111.09	0.00	111.09
	[-113.51, 122.47]		[0.00, 122.47]		[-16.59, 122.47]		[0.00, 122.47]	
2. $LNATE_{at}$	-98.76	114.74	0.00	114.74	15.35	114.74	15.35	114.74
	[-112.19, 131.75]		[0.00, 131.75]		[3.49, 131.75]		[3.49, 131.75]	
3. MATE			0.00	19.48	-35.94	14.74	0.00	13.60
			[0.00, 27.93]		[-41.83, 17.15]		[0.00, 16.71]	
4. LATE			0.00	86.70	-169.92	70.50	0.00	64.43
			[0.00, 129.19]		[-185.45, 80.23]		[0.00, 77.53]	

Notes: The Table shows half-median unbiased estimates of the bounds and 95 percent confidence intervals (in square brackets) for the true value of the parameter based on the method proposed by Chernozhukov, Lee and Rosen (2011), as discussed in Section 3. For the bounds with no minimum or maximum operators, the bound estimates and confidence intervals closely correspond to the analog estimates and the Imbens and Manski (2004) confidence intervals, respectively. The variance-covariance matrix for all the estimated bounding functions is based on 5,000 bootstrap replications, and the number of draws from a normal distribution used to get the estimated bounds and confidence intervals (see Internet Appendix for details) is 100,000. Sample size is 8,020 individuals: 2,975 with Z=0 and 5,045 with Z=1. The labor market outcomes (Y) are either employment status or weekly earnings in month 36 after randomization. The treatment (D) is the attainment of a high school, GED, or vocational degree. The instrumental variable (Z) is an indicator for whether the individual was randomly assigned to participate in JC ("random assignment status"). $LNATE_k$, the local net average treatment effect for stratum k (never or always takers), gives the part of the effect of the random assignment status (Z) on labor market outcomes (Y) that works through channels other than the attainment of a degree (D) for stratum k. MATE, the mechanism average treatment effect, is the part of the intention-to-treat effect of Z on Y that works through D. LATE is the local average treatment effect of degree attainment on labor market outcomes for those individuals assigned to the program and whose degree attainment is affected by their random assignment status.

bounds discussed in Section 2.2. These estimated bounds are not able to reject that those two net effects are zero. Bounds on $LNATE_{nt}$ and $LNATE_{at}$ under Assumption 4 are presented for completeness, since adding Assumption 4.2 ($LNATE_k \geq 0$) is not attractive for assessing the validity of the exclusion restriction. Employing Assumption 5 substantially increases the lower bounds shown in the first column. The estimated lower bounds on $LNATE_{at}$ for employment and earnings equal 0.044 and \$15.35, respectively, and the 95 percent confidence intervals for $LNATE_{at}$ exclude zero for both outcomes. Thus, under Assumptions 1, 3, and 5, the data reject the validity of the instrument because the random assignment status has a strictly positive average direct effect on the outcome for the *at* stratum, suggesting the unreliability of the conventional IV point estimates in Table 2.

Table 3 shows that under Proposition 2 the estimated upper bound on $MATE$ equals 4 percentage points for employment and \$19.48 for earnings, which are very close to the ITT of the instrument on the outcomes in Table 2. Replacing Assumption 4 with Assumption 5 (Proposition 3) yields an estimated upper bound of 3 percentage points for employment and \$14.7 for earnings. This implies, for example, that at most 75 percent of the ITT effect of the instrument on employment is due to the attainment of a degree. Lastly, when both Assumptions 4 and 5 are used (Proposition 4), the bounds on $MATE$ imply that the part of the average effect of random assignment status on employment that is due to the attainment of a degree is at most 66 percent, while the corresponding quantity is 75 percent for earnings. The fact that the estimated upper bounds on $MATE$ under Propositions 3 and 4 are considerably below the ITT of random assignment status on the outcome is consistent with our previous finding regarding the existence of a non-zero net average effect of the instrument on the outcome and the invalidity of the instrument.

We now focus on the $LATE$ results. Here, $LATE$ in (6) is interpreted as the average effect for compliers of attaining a GED, high school, or vocational degree on the outcome when assigned to participate in JC. Note that, given the imperfect compliance with the random assignment present in the NJCS, $LATE$ cannot be interpreted as the average effect for compliers when enrolled in JC.

Under Proposition 2, the lower bound on $LATE$ for both outcomes is zero, and the estimated upper bounds equal 0.211 for employment and \$86.70 for weekly earnings. Although the upper bound

on $LATE$ in Proposition 2 is at most equal to the usual IV estimand in (2), both estimated upper bounds are slightly above the corresponding IV point estimates in Table 2 because of the half-median unbiased correction we employ. Note that under Proposition 2 the data do not provide any additional information to narrow the bounds in our application, as they come directly from Assumption 4.

The estimated upper bounds on $LATE$ obtained by employing Assumption 5 instead of Assumption 4 (Proposition 3) are more informative. They equal 15.1 percentage points for employment and \$70.5 for earnings. When all five assumptions are combined (Proposition 4), the lower bound on $LATE$ is zero for both outcomes, which comes directly from Assumption 4, and the estimated upper bounds imply that $LATE$ is at most 12.7 percentage points for employment and \$64.43 for weekly earnings. The 95 percent confidence intervals for the true value of $LATE$ under Proposition 4 for employment and earnings are, respectively, $[0, 0.166]$ and $[0, 77.53]$. Thus, in this case we can statistically reject effects larger than 0.166 for employment and \$77.53 for earnings, while based on the IV point estimates in Table 2 we cannot reject effects as large as 0.30 and \$131, respectively.

For each outcome, the IV point estimate in Table 2 falls outside the 95 percent confidence interval for $LATE$ under Proposition 4. In a setting where the $LATE$ in (6) and (2) are equal, this implies that the invalidity of the instrument results in IV point estimates that are severely upward biased (under Assumptions 4 and 5). To test this more formally, let β_{IV} denote the IV estimand in (2). The CLR-based 95 percent confidence interval on the partially identified parameter $(LATE - \beta_{IV})$ for employment is $[-0.2987, -0.0184]$, while the corresponding 90 percent confidence interval for earnings is $[-122.97, -1.25]$. These CLR-based confidence intervals take into account that the correlation between the estimates of β_{IV} and those of many of the upper bounding functions in Proposition 4 is very high and positive, and that some estimated bounding functions are well below the IV estimates in Table 2. When using the bounds under Proposition 3, although the point estimates of β_{IV} in Table 2 fall outside the 95 percent confidence interval for both outcomes, the corresponding CLR-based 95 percent confidence interval on $(LATE - \beta_{IV})$ for employment and the 90 percent one for earnings are given by $[-0.734, 0.057]$ and $[-293.94, 16.38]$, respectively, both of which include zero.

Our results also suggest that sample analog estimates of bounds involving min and max operators

can be severely biased in practice and, thus, that appropriate estimation and inference methods should be used. For instance, the sample analog estimates of the upper bounds on $LATE$ under Proposition 4 for employment and earnings equal 0.1 and \$53.9, respectively, both of which are well below the corresponding half-median unbiased estimates in Table 3.

Next, we compare our estimated bounds on $LATE$ to two sets of bounds on the population average treatment effect (ATE) derived by Manski and Pepper (2000). The first set is under their monotone instrumental variable and monotone treatment response assumptions (MIV-MTR), while the other is under their monotone treatment selection and MTR assumptions (MTS-MTR). The MTS assumption specializes the MIV assumption to the case when the IV is the realized treatment, in which case their bounds do not require a bounded outcome. In these comparisons, it is important to keep in mind that ATE may differ from $LATE$ in the presence of heterogeneous effects. For employment, the estimated MIV-MTR bounds on ATE are $[0, 0.49]$, while those for earnings are $[0, \$870.6]$ (these bounds require a bounded outcome; for earnings, we use the in-sample maximum, $\$2,358.4$, as the upper bound). The upper bounds for ATE are above all those presented in Table 3 for $LATE$. The estimated MTS-MTR bounds on ATE, $[0, 0.13]$ for employment and $[0, \$64.14]$ for earnings, are close to our bounds on $LATE$ under Proposition 4. Note that, with a binary treatment, the upper bound on ATE under the MTS-MTR assumptions is $E[Y|D = 1] - E[Y|D = 0]$.

Finally, we relate our results to the empirical literature on the returns to years of schooling. The average number of actual hours of academic and vocational instruction received while enrolled in JC for those individuals who participated and obtained a degree is 1,448. Considering that a typical high school student receives the equivalent of 1,080 hours of instruction during a school year (Schochet et al., 2001), obtaining a degree is equivalent to about 1.34 years of schooling. Hence, our results above suggest an upper bound on the average return to a year of schooling of about 28 percent (38%/1.34). Card (1999) surveyed IV estimates of the return to one year of schooling that exploit institutional features of school systems, which estimate the effect for individuals who would otherwise have relatively low educational attainment (as in our application). These IV estimates range from 6 to 15.3 percent, and thus fall within our estimated bounds on $LATE$.

5 Conclusion

This paper derived nonparametric bounds for local average treatment effects employing an invalid instrument and allowing the outcome to have an unbounded support. We substitute the exclusion restriction assumption in Imbens and Angrist (1994) and Angrist et al. (1996) with assumptions requiring weak monotonicity of potential and counterfactual outcomes within or across principal strata. Our bounds on the effect of attaining a degree on labor market outcomes illustrate their identifying power: they indicate that the local effect when assigned to training for those whose degree attainment is affected by the instrument (random assignment to a training program) is at most 12.7 percentage points on employment and \$64.43 on weekly earnings.

The analysis herein can be extended to derive bounds on the population ATE with invalid instruments. Under the monotonicity assumption of the instrument on the treatment, the ATE equals the weighted average of the (local) average treatment effects for the always takers, never takers, and compliers. Our general approach can be used to construct bounds on ATE with invalid instruments by first deriving bounds on each of those local average effects. A related approach has been employed in the literature to derive bounds on ATE when the exclusion restriction holds (e.g., Angrist et al., 1996; Huber and Mellace, 2010; Chen et al., 2012). Another possible extension of our results is to settings where the instrument is not randomly assigned. One could combine the ideas herein with work allowing estimation of $LATE$ when the instrument is assumed to be random conditional on a set of covariates (Hirano et al., 2000; Abadie, 2003; Frölich, 2007).

6 Appendix: Relationship to Principal Stratification

We discuss how our approach differs from the original Principal Stratification (PS) framework of Frangakis and Rubin (2002). In contrast to the original PS framework that employs only potential outcomes ($Y_i^Z(1)$ and $Y_i^Z(0)$), we also work with a priori counterfactual outcomes ($Y_i(1, D_i(0))$). The difference is that, while potential outcomes can be observed in the data depending on the instrument assignment (e.g., $Y_i^Z(1)$ is observed for those with $Z_i = 1$), a priori counterfactual outcomes are never observed regardless of the instrument assignment (e.g., the outcome compliers would have if $Z_i = 1$, but we could "force" them to have $D_i = 0$). As a result, we also consider partial identification of effects that are not "principal strata effects", defined as comparisons of potential outcomes within a given principal stratum. For example, the object of interest when estimating direct or net effects within a PS framework is the so-called principal strata average direct effect, which is defined as $E[Y^Z(1) - Y^Z(0) | D(1) = D(0)]$; that is, the average effect for those individuals whose D is unaffected by Z. As compared to $NATE$ and $MATE$ (which are not principal strata effects), this effect has the attractive feature that it does not require additional assumptions for employing the information available in the data to learn about counterfactual outcomes that are never observed. However, in a setting with heterogeneous effects, it does not decompose the ATE_{ZY} into a net and mechanism effect the way $NATE$ and $MATE$ do (e.g., VanderWeele, 2008; Flores and Flores-Lagunes, 2010, 2011; Mealli and Mattei, 2012). For further discussion on these topics see, for example, Frangakis and Rubin (2002), Rubin (2004), Pearl (2011), Mattei and Mealli (2011), and Mealli and Mattei (2012). Mealli and Mattei (2012) provide an interesting discussion of principal stratification and its connections and differences with IV methods, along with an alternative presentation of the exclusion restriction assumption.

7 Acknowledgments

Detailed comments from the Editor, Associated Editor and two anonymous referees greatly improved the paper and are gratefully acknowledged. Useful comments were also provided by Alberto Abadie, Joshua Angrist, Xuan Chen, Guido Imbens, Chris Parmeter, and participants at the 7th IZA Confer-

ence on Labor Market Policy Evaluation at Harvard University, the 2012 Society of Labor Economists Meeting, the 2011 North American Summer Meeting of the Econometric Society, and the 2010 NY Camp Econometrics. This work has been partially supported by the NSF under grants SES-0852211 and SES-0852139. Flores also acknowledges funding from the McLamore summer research award. Competent research assistance was provided by Maria Bampasidou. All errors are our own.

References

[1] Abadie, A. (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231-263.

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

[3] Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

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

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

[6] Cameron, S., and Heckman, J. (1993), "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11, 1–47.

[7] Card, D. (1999), "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, Vol. 3A, O.Ashenfelter and D. Card (Eds.), Amsterdam: North Holland, pp. 1801–1863.

[8] Chen, X., Flores, C., and Flores-Lagunes, A. (2012), "Bounds on Population Average Treatment Effects with an Instrumental Variable,"mimeo, University of Miami, Dept. of Economics.

[9] Chernozhukov, V., Lee, S., and Rosen, A. (2011), "Intersection Bounds: Estimation and Inference,"mimeo, Massachusetts Institute of Technology, Dept. of Economics.

[10] Conley, T., Hansen, C., and Rossi, P. (2012), "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260-272.

[11] Flores, C., and Flores-Lagunes, A. (2010), "Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects,"mimeo, University of Miami, Dept. of Economics.

[12] Flores, C., and Flores-Lagunes, A. (2011), "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness," mimeo, University of Miami, Dept. of Economics.

[13] Flores, C., Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (2012), "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps," *Review of Economics and Statistics*, 94, 153-171.

[14] Flores-Lagunes, A., and Light, A. (2010), "Interpreting Degree Effects in the Returns to Education," *Journal of Human Resources*, 45, 439-467.

[15] 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, 651-667.

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

[17] Frölich, M. (2007), "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35-75.

[18] Frumento P., Mealli F., Pacini B., and Rubin D. (2012) "Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data," *Journal of the American Statistical Association*, 107, 458-466.

[19] Heckman, J., and Vytlacil, E. (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730-4734.

[20] Heckman, J., and Vytlacil, E. (2000), "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," NBER Technical Working Paper 259.

[21] Heckman, J., and Vytlacil, E. (2005), "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

[22] Hirano, K., Imbens, G., Rubin, D., and Zhou, X. (2000), "Assessing the Effect of an Influenza Vaccine in an Encouragement Design with Covariates," *Biostatistics*, 1, 69-88.

[23] Hirano, K., and Porter, J. (2012), "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769-1790.

[24] Huber, M., and Mellace, G. (2010), "Sharp IV Bounds on Average Treatment Effects under Endogeneity and Noncompliance," Discussion Paper no. 2010-31, University of St. Gallen, Dept. of Economics.

[25] Huber, M., and Mellace, G. (2011), "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints," mimeo, University of St. Gallen, Dept. of Economics.

[26] Hungerford, T., and Solon, G. (1987), "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, 69, 175-177.

[27] Imai, K., Keele, L., and Yamamoto, T. (2010), "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25(1), 51-71.

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

[29] Jaeger, D., and Page, M. (1996), "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, 78, 733-740.

[30] Kitagawa, T. (2008), "A Bootstrap Test for Instrument Validity in the HeterogeneousTreatment Effect Model,"mimeo, Brown University, Dept. of Economics.

[31] Kitagawa, T. (2009), "Identification Region of the Potential Outcome Distributions under Instrument Independence,"Cemmap Working Paper CWP30/09, London, UK.

[32] Lee, D. (2009), "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071-102.

[33] Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review: Papers and Proceedings*, 80, 319-323.

[34] Manski, C. (1994), "The Selection Problem,"in C. Sims et al. (Eds), *Advances in Econometrics: Sixth World Congress*, Cambridge, MA: Cambridge University Press, pp. 143-170.

[35] Manski, C. (1997), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

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

[37] Mattei, A., and Mealli, F. (2011), "Augmented Designs to Assess Principal Strata Direct Effects," *Journal of the Royal Statistical Society, Series B*, 73, 729-752.

[38] Mealli, F. and Mattei, A. (2012), "A Refreshing Account of Principal Stratification," *International Journal of Biostatistics*, 8(1), 1-37.

[39] Mealli, F, and Pacini, B. (2012), "Using Secondary Outcomes and Covariates to Sharpen Inference in Randomized Experiments with Noncompliance,"Working Paper 2012/04, Dipartimento di Statistica, Universita di Firenze.

[40] Nevo, A., and Rosen, A. (2010), "Identification with Imperfect Instruments," *Review of Economics and Statistics* (forthcoming).

[41] Newey, W., and McFadden, D. (1994) "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics,* vol. 4, R. Engle and D. McFadden (Eds.), Amsterdam: North Holland, pp. 2111-2245.

[42] Pearl, J. (2001), "Direct and Indirect Effects," *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, pp. 411-20.

[43] Pearl, J. (2011), "Principal Stratification – a Goal or a Tool?" *International Journal of Biostatistics*, 7(1), 1-13.

[44] Robins, J. M. (2003), "Semantics of Causal DAG Models and the Identification of Direct and Indirect Effects,"in *Highly Structured Stochastic Systems*, P.J. Green, N.L. Hjort, and S. Richardson (Eds.), NY: Oxford University Press, pp. 70-81.

[45] Robins, J., and Greenland, S. (1992), "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143-155.

[46] Rubin, D. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34-58.

[47] Rubin, D. (1980), "Comment on: Randomization Analysis of Experimental Data: The Fisher Randomization Test by D. Basu," *Journal of the American Statistical Association*, 75, 591-593.

[48] Rubin, D. (1990), "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, 5, 472-480.

[49] Rubin, D. (2004), "Direct and Indirect Causal Effects via Potential Outcomes"(with Discussion), *Scandinavian Journal of Statistics*, 31, 161-198.

[50] Schochet, P., Burghardt, J., and Glazerman, S. (2001), *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes,* Princeton, NJ: Mathematica Policy Research, Inc.

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

[52] Sjölander, A. (2009), "Bounds on Natural Direct Effects in the Presence of Confounded Intermediate Variables," *Statistics in Medicine*, 28, 558-71.

[53] Tamer, E. (2010), "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167-195.

[54] VanderWeele, T.J. (2008), "Simple Relations Between Principal Stratification and Direct and Indirect Effects," *Statistics and Probability Letters*, 78, 2957-2962.

[55] Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

[56] 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, 353-368.

[57] Zhang, J.L., Rubin, D., and Mealli, F. (2008), "Evaluating the Effects of Job Training Programs on Wages Through Principal Stratification," in *Advances in Econometrics vol XXI*, D. Millimet et al. (Eds.), Amsterdam: North Holland, pp. 117-145.

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