

*INTERNET APPENDIX*

*OF*

PARTIAL IDENTIFICATION OF WAGE EFFECTS OF

TRAINING PROGRAMS

Michael Lechner

*Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen*

Blaise Melly

*Department of Economics, Brown University*

First version: July, 2007

This version: April 2010

Date this version has been printed: 22 April 2010

This appendix is available on the website of the authors at:

[www.sew.unisg.ch/lechner/earnings](http://www.sew.unisg.ch/lechner/earnings)

**Abstract:** This appendix complements the paper “Partial identification of wage effects of training programs”. It contains (1) The proof of the lemmas, (2) The complete proof of the theorem, inclusively the sharpness of the bounds, (3) two additional theorems, and (4) the results of a Monte Carlo simulation.

## 1. Proof of the Lemmas

### 1.1 Proof of Lemma 1

Using iterated expectations, we get  $E[g \mid Y] = E[E[g \mid Y \mid X = x] \mid Y]$ . Since there is no cross restriction between  $E[g \mid Y \mid X = x]$  and  $E[g \mid Y \mid X = \tilde{x}]$  for  $x \neq \tilde{x}$ , we can consider separately the bounds evaluated at different values of  $x$ . This implies that

$$E[E[g \mid Y \mid X = x]] \geq E[\underline{b}_g \mid x] \text{ and } E[E[g \mid Y \mid X = x]] \leq E[\bar{b}_g \mid x].$$

The bounds on the conditional expectation are tight and can be attained. Moreover, in the absence of cross restrictions, the lower (or the upper) bound could be attained at each point of the support of  $X$ . Therefore, the bounds on the unconditional expected value can be attained and are tight. The same arguments hold in all subpopulations defined by  $P$  and  $S$ .

### 1.2 Proof of Lemma 2

Note the following useful property:  $a \geq b \Rightarrow \bar{b}_{1 \mid Y \leq a} \geq \bar{b}_{1 \mid Y \leq b}$  and  $\underline{b}_{1 \mid Y \leq a} \geq \underline{b}_{1 \mid Y \leq b}$ . By definition,

$$\underline{b}_{1 \mid Y \leq b} \leq E[1 \mid Y \leq b] \leq E[1 \mid Y \leq b] + E[1 \mid b < Y \leq a] = E[1 \mid Y \leq a] \leq \underline{b}_{1 \mid Y \leq a}, \text{ that is the tight}$$

lower bound on  $F_Y \mid a$  cannot be lower than the tight lower bound on  $F_Y \mid b$ . Similarly,

$$\bar{b}_{1 \mid Y \leq a} \geq E[1 \mid Y \leq a] \geq E[1 \mid Y \leq b] \geq \bar{b}_{1 \mid Y \leq b}, \text{ that is the tight upper bound on } F_Y \mid b \text{ cannot be}$$

above the tight upper bound on  $F_Y \mid a$ .

a)  $\underline{b}_{QY} \mid \theta$  is a tight lower bound.

First, note that if  $\lim_{\tilde{y} \rightarrow -\infty} \bar{b}_{1 \mid Y \leq \tilde{y}} > \theta$ , we cannot exclude that  $\lim_{a \rightarrow -\infty} P(Y \leq a) > \theta$  and, therefore,

$F_Y^{-1} \mid \theta$  is unbounded from below. Otherwise, suppose that the lower bound is given by

$q < \underline{b}_{QY}(\theta)$ . By the definition of  $\underline{b}_{QY}(\theta)$ ,  $\bar{b}_{1_{Y \leq q}} < \theta$  and it is, therefore, impossible that

$F_Y(q) = \theta$ . It follows that  $F_Y^{-1}(\theta) \geq \underline{r}_{QY}(\theta)$ .

To show that this bound is tight, suppose that  $F_Y(\tilde{y}) = \bar{b}_{1_{Y \leq \tilde{y}}}$  for  $\forall \tilde{y}$  in the support of  $Y$ ,

which is possible because the bounds are sharp. Then,  $F_Y^{-1}(\theta) = \underline{b}_{QY}(\theta)$  holds by the definition of a quantile.

b)  $\bar{b}_{QY}(\theta)$  is a tight upper bound

First, note that if  $\lim_{\tilde{y} \rightarrow \infty} \underline{b}_{1_{Y \leq \tilde{y}}} < \theta$ , then we cannot exclude that  $\lim_{a \rightarrow \infty} P(Y \geq a) > 1 - \theta$  and, there-

fore,  $F_Y^{-1}(\theta)$  is unbounded from above. Otherwise, suppose that the upper bound is given by

$q > \bar{b}_{QY}(\theta)$ . By the definition of  $\bar{b}_{QY}(\theta)$ ,  $\underline{b}_{1_{Y \leq q}} > \theta$  and it is therefore impossible that

$F_Y(q) = \theta$ . It follows that  $F_Y^{-1}(\theta) \leq \bar{b}_{QY}(\theta)$ .

To show that this bound is tight, suppose that  $F_Y(\tilde{y}) = \underline{b}_{1_{Y \leq \tilde{y}}}$  for  $\forall \tilde{y}$  in the support of  $Y$ ,

which is possible because the bounds are sharp. Then,  $F_Y^{-1}(\theta) = \bar{b}_{QY}(\theta)$  holds by the definition of a quantile.

### 1.3 Proof of Lemma 3

a)  $E[S_{1-t} - S_0 | T=t]$

Applying the law of iterated expectation:<sup>1</sup>

$$E[S_{1-t} - S_0 | T=t] \stackrel{IE}{=} \int E[S_{1-t} | X=x, T=t] - E[S_0 | X=x, T=t] f_{X|T}(x; t) dx$$

All elements in the last expression are identified by Assumption 1.

---

<sup>1</sup> Notation: IE means by the law of iterated expectation, A1 means by Assumption 1, OR means by the observation rule, TP means by the law of total probability.

b) To show the identification of the mean as well as the quantile effects of  $D$  on  $Y$ , it is sufficient to show that respective counterfactual expectations of the  $g \cdot$  functions that enter the mean as well as the distribution functions are identified.

We first note that

$$E[g(Y, d, S, d) | X = x, T = t] = p_{S=d | X, T}(x, t) E[g(Y, d, S, d) | X = x, T = t, S = d = 1] + [1 - p_{S=d | X, T}(x, t)] E[g(Y, d, S, d) | X = x, T = t, S = d = 0]$$

which is therefore identified by Assumption 1. Then, similarly to part a):

$$E[g(Y, d, S, d) | T = t] \stackrel{IE}{=} \int E[g(Y, d, S, d) | X = x, T = t] f_{X|T}(x; t) dx.$$

All elements in the last expression are identified by Assumption 1. Thus,  $ATE^{YS}(T = t)$  and  $QTE_{\theta}^{YS}(T = t)$  are identified.

#### 1.4 Proof of Lemma 4

$$E[S = 1 | X = x, D = 1] \stackrel{OR}{=} E[S | X = x, D = 1] \text{ is observed for } \forall x \in \mathcal{X}_{D=1}.$$

$$E[S = 0 | X = x, D = 1] \stackrel{A1'a}{=} E[S = 0 | X = x, D = 0] \stackrel{OR}{=} E[S | X = x, D = 0] \text{ is observed for } \forall x \in \mathcal{X}_{D=1}.$$

$$F_{Y=1}(\tilde{y} | X = x, D = 1, S = 1) \stackrel{OR}{=} F_Y(\tilde{y} | X = x, D = 1, S = 1) \text{ is observed for } \forall x \in \mathcal{X}_{D=1}.$$

$$F_{Y=0}(\tilde{y} | X = x, D = 1, S = 0) \stackrel{A1'a}{=} F_{Y=0}(\tilde{y} | X = x, D = 0, S = 0) \stackrel{OR}{=} F_Y(\tilde{y} | X = x, D = 0, S = 1)$$

is observed for  $\forall x \in \mathcal{X}_{D=1}$

$$F_{X|D}(x; 1) \text{ is observed for } \forall x \in \mathcal{X}_{D=1}.$$

#### 1.5 Proof of Lemma 5

We can apply Lemma 2.1 in Abadie (2002) conditional on  $X$  to obtain:

$$p_{S=0|X,C}(x,1) = \frac{E[S \cdot D - 1 | X=x, Z=1] - E[S \cdot D - 1 | X=x, Z=0]}{E[D | X=x, Z=1] - E[D | X=x, Z=0]} \text{ and}$$

$$p_{S=1|X,C}(x,1) = \frac{E[SD | X=x, Z=1] - E[SD | X=x, Z=0]}{E[D | X=x, Z=1] - E[D | X=x, Z=0]}.$$

By the definition of a conditional distribution:

$$\begin{aligned} F_{Y=1}(y|X=x, C=1, S=1) &= E[\mathbb{1}_{Y=1} \leq y | X=x, C=1, S=1] \\ &= \frac{E[\mathbb{1}_{Y=1} \leq y \cdot S | X=x, C=1]}{E[S | X=x, C=1]}. \end{aligned}$$

The denominator is already identified. For the numerator we apply again Lemma 2.1 in Abadie (2002):

$$\begin{aligned} &E[\mathbb{1}_{Y=1} \leq y \cdot S | X=x, C=1] \\ &= \frac{E[\mathbb{1}_{Y \leq y} \cdot SD | X=x, Z=1] - E[\mathbb{1}_{Y \leq y} \cdot SD | X=x, Z=0]}{E[D | X=x, Z=1] - E[D | X=x, Z=0]}. \end{aligned}$$

The same result holds for  $F_{Y=0}(y|X=x, C=1, S=0)$ . Finally, as noted by Frölich (2007),

$$\begin{aligned} f_{X|C}(x;1) &= f_X(x) \frac{\Pr(C=1 | X=x)}{\Pr(C=1)} \\ &= f_X(x) \frac{E[D | X=x, Z=1] - E[D | X=x, Z=0]}{\int (E[D | X=x, Z=1] - E[D | X=x, Z=0]) f_X(x) dx}. \end{aligned}$$

## 2. Proofs of the Theorems

### 2.1 Proof of Theorem 1

By the law of total probability, we obtain the following expression:

$$\begin{aligned}
& E[g \mid Y=0 \mid X=x, T=t, S_1=1] = \\
& E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=1] \frac{p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_1 \mid X, T}(x, t)} \\
& + E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=0] \frac{p_{S_1 \mid X, T}(x, t) - p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_1 \mid X, T}(x, t)}.
\end{aligned} \tag{1}$$

$E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=0]$  is unobserved and bounded only by  $\underline{K}_g$  and  $\bar{K}_g$

without further assumptions. Therefore, the worst case bounds are attained if  $p_{S_1, S_0 \mid X, T}(x, t)$

attains the smallest value compatible with the observed employment probabilities. The

following set of equations restricts this probability:

$$\begin{aligned}
& p_{S_1, S_0 \mid X, T}(x, t) + p_{S_1, 1-S_0 \mid X, T}(x, t) = p_{S_1 \mid X, T}(x, t) \\
& p_{S_1, S_0 \mid X, T}(x, t) + p_{1-S_1, S_0 \mid X, T}(x, t) = p_{S_0 \mid X, T}(x, t) \\
& p_{S_1, S_0 \mid X, T}(x, t) + p_{S_1, 1-S_0 \mid X, T}(x, t) + p_{1-S_1, S_0 \mid X, T}(x, t) + p_{1-S_1, 1-S_0 \mid X, T}(x, t) = 1.
\end{aligned}$$

We have three restrictions and four unknowns. Solving for the element of interest, we get

$$p_{S_1, S_0 \mid X, T}(x, t) = p_{S_0 \mid X, T}(x, t) + p_{S_1 \mid X, T}(x, t) - 1 + p_{1-S_1, 1-S_0 \mid X, T}(x, t).$$

We cannot exclude that  $p_{S_1, S_0 \mid X, T}(x, t) = 0$  if  $p_{S_0 \mid X, T}(x, t) + p_{S_1 \mid X, T}(x, t) \leq 1$ . In this case,

the observed values do not allow us to tighten the bounds on the support  $\underline{K}_g$  and  $\bar{K}_g$ .

If  $p_{S_0 \mid X, T}(x, t) + p_{S_1 \mid X, T}(x, t) > 1$ , the smallest acceptable value for  $p_{S_1, S_0 \mid X, T}(x, t)$  is given

by  $p_{S_0 \mid X, T}(x, t) + p_{S_1 \mid X, T}(x, t) - 1$  and is strictly positive. In this case, we need to bound

$E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=1]$ . The distribution of

$g \mid Y=0 \mid X=x, T=t, S_0=1$  is identified by assumption 1. The population defined by

$X=x, T=t, S_0=1$  is a  $\frac{p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)}$  and  $\frac{p_{S_0 \mid X, T}(x, t) - p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)}$  mixture of

the population with  $X = x, T = t, S_0 = 1, S_1 = 1$  and  $X = x, T = t, S_0 = 1, S_1 = 0$ . Since

$p_{S_1, S_0 | X, T}(x, t) = p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) - 1$  the upper bound will be attained when the population with  $X = x, T = t, S_0 = 1, S_1 = 1$  represents the  $\frac{p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) - 1}{p_{S_0 | X, T}(x, t)}$

fraction of the population with  $X = x, T = t, S_0 = 1$  with the largest value of  $g(Y)$ .

Similarly, the lower bound will be attained when population with

$X = x, T = t, S_0 = 1, S_1 = 1$  represents the  $\frac{p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) - 1}{p_{S_0 | X, T}(x, t)}$  fraction of the

population with  $X = x, T = t, S_0 = 1$  with the smallest value of  $g(Y)$ . This is the result of Theorem 1.

We now show that the bounds can effectively be attained to prove the **sharpness** of the bounds.

We consider only the lower bound because of the symmetry of the arguments. We have to distinguish two cases.

1<sup>st</sup> case:  $p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) \leq 1$ . In this case, it is possible that  $p_{S_0, S_1 | X, T}(x, t) = 0$ ,

$p_{S_0, 1-S_1 | X, T}(x, t) = p_{S_0 | X, T}(x, t)$ ,  $p_{1-S_0, S_1 | X, T}(x, t) = p_{S_1 | X, T}(x, t)$ ,  $p_{1-S_0, 1-S_1 | X, T}(x, t) =$

$1 - p_{S_0 | X, T}(x, t) - p_{S_1 | X, T}(x, t)$ . In this case, it can be that  $g(Y_0) = \underline{b}_Y$  for the population

defined by  $X = x, T = t, S_1 = 1$  because  $Y_0$  is never observed for this population.

2<sup>nd</sup> case:  $p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) > 1$ . It is possible that  $p_{S_0, S_1 | X, T}(x, t) =$

$p_{S_0 | X, T}(x, t) + p_{S_1 | X, T}(x, t) - 1$ ,  $p_{S_0, 1-S_1 | X, T}(x, t) = 1 - p_{S_1 | X, T}(x, t)$ ,  $p_{1-S_0, S_1 | X, T}(x, t) =$

$1 - p_{S_0 | X, T}(x, t)$  and  $p_{1-S_0, 1-S_1 | X, T}(x, t) = 0$ . By the law of total probability and using these

values for the employment probabilities:

$$\begin{aligned}
& E[g(Y=0) | X=x, T=t, S_1=1] = \\
& E[g(Y=0) | X=x, T=t, S_1=1, S_0=0] \frac{1 - p_{S_0|X,T}(x,t)}{p_{S_1|X,T}(x,t)} \\
& + E[g(Y=0) | X=x, T=t, S_1=1, S_0=1] \frac{p_{S_0|X,T}(x,t) + p_{S_1|X,T}(x,t) - 1}{p_{S_1|X,T}(x,t)}.
\end{aligned}$$

It can be that  $g(Y=0) = \underline{b}_g$  for the population defined by  $X=x, T=t, S_1=1, S_0=0$

because  $Y=0$  is never observed for this population. We can construct a distribution of

$g(Y=0) | X=x, T=t, S_1=0, S_0=1$  equal to the distribution of

$g(Y=0) | X=x, T=t, S_0=1$  truncated below

$$F_{g(Y=0) | X,T,S_0}^{-1} \left( \frac{p_{S_0|X,T}(x,t) + p_{S_1|X,T}(x,t) - 1}{p_{S_0|X,T}(x,t)}; x, t, 1 \right). \text{ Similarly, we can construct a distribution}$$

of  $g(Y=0) | X=x, T=t, S_1=1, S_0=1$  equal to the distribution of

$g(Y=0) | X=x, T=t, S_0=1$  truncated above

$$F_{g(Y=0) | X,T,S_0}^{-1} \left( \frac{p_{S_0|X,T}(x,t) + p_{S_1|X,T}(x,t) - 1}{p_{S_0|X,T}(x,t)}; x, t, 1 \right). \text{ By construction, the mixture of these two}$$

distributions replicates the identified distribution of  $g(Y=0) | X=x, T=t, S_0=1$ .

## 2.2 Proof of Theorem 2

a) If  $g(Y)$  is a monotonic increasing function of  $Y$ , then Assumption 2 implies that the distribution of  $g(Y=0)$  given  $X=x, T=t, S_0=1$  stochastically dominates the distribution of  $g(Y=0)$  given  $X=x, T=t, S_0=0$ . The upper bound is attained when this assumption is just satisfied, that is when these two distributions are the same.

As shown in the proof of Theorem 1,



$$\begin{aligned}
& E[g(Y=0) \mid X=x, T=t, S_1=1] = \\
& E[g(Y=0) \mid X=x, T=t, S_1=1, S_0=1] \frac{p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_1 \mid X, T}(x, t)} \\
& + E[g(Y=0) \mid X=x, T=t, S_1=1, S_0=0] \frac{p_{S_1 \mid X, T}(x, t) - p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_1 \mid X, T}(x, t)}.
\end{aligned}$$

The populations defined by  $X=x, T=t, S_0=1$  and  $X=x, T=t, S_0=0$  are mixtures of the two sub-populations defined by  $S_1 \in \{0,1\}$ . Since, in the worst case,  $g(Y=0)$  given  $X=x, T=t, S_0=0$  has the same distribution as  $g(Y=0)$  given  $X=x, T=t, S_0=1$ , the upper bound is attained when the mixture proportions are the same for  $S_0=0$  and  $S_0=1$ :

$$\frac{p_{S_1, S_0 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)} = \frac{p_{S_1, 1-S_0 \mid X, T}(x, t)}{p_{1-S_0 \mid X, T}(x, t)} = p_{S_1 \mid X, T}(x, t)^2.$$

Thus, the upper bound will be attained when the population with  $X=x, T=t$ ,

$S_0=1, S_1=1$  represents the  $p_{S_1 \mid X, T}(x, t)$  fraction of the population with  $X=x, T=t$ ,

$S_0=1$  with the largest value of  $g(Y=0)$ . Simultaneously, the population with

$X=x, T=t, S_0=0, S_1=1$  represents the  $p_{S_1 \mid X, T}(x, t)$  fraction of the population

$X=x, T=t, S_0=0$  with the largest value of  $g(Y=0)$ . The distribution of  $g(Y=0)$  is not

observed for this last population but, by the positive selection assumption, it is bounded by

the distribution of  $g(Y=0)$  for the population with  $X=x, T=t, S_0=1$ . Therefore,

$$\begin{aligned}
& E[g(Y=0) \mid X=x, T=t, S_1=1, S_0=1] \leq E[g(Y=0) \mid X=x, T=t, S_0=1] \\
& E[g(Y=0) \mid X=x, T=t, S_1=1, S_0=0] \leq E[g(Y=0) \mid X=x, T=t, S_0=1].
\end{aligned}$$

Inserting these two bounds gives the result of Theorem 2-a).

---

<sup>2</sup> If these mixture proportions were not the same, then it would be possible to get a higher upper bound by increasing slightly the smallest mixing proportion and decreasing slightly the highest mixing proportion.

We prove now the **sharpness** of this bound. We construct the joint employment statuses such

that they are independent:  $p_{S=1, S=0|X,T}(x,t) = p_{S=1|X,T}(x,t) p_{S=0|X,T}(x,t)$ ,  $p_{S=1, 1-S=0|X,T}(x,t)$

$= p_{S=1|X,T}(x,t) p_{1-S=0|X,T}(x,t)$ ,  $p_{1-S=1, 1-S=0|X,T}(x,t) = p_{1-S=1|X,T}(x,t) p_{1-S=0|X,T}(x,t)$ ,

$p_{1-S=1, 1-S=0|X,T}(x,t) = p_{1-S=1|X,T}(x,t) p_{1-S=0|X,T}(x,t)$ .

For each  $d \in [0,1]$ , we construct a distribution of  $g(Y=0 | X=x, T=t, S=1=1, S=0=d)$

equal to the distribution of  $g(Y=0 | X=x, T=t, S=0=1)$  truncated below

$F_{g(Y=0 | X,T,S=0)}^{-1}(p_{S=1|X,T}(x,t); x,t,1)$ . Similarly, we construct a distribution of

$g(Y=0 | X=x, T=t, S=1=0, S=0=d)$  equal to the distribution of

$g(Y=0 | X=x, T=t, S=0=1)$  truncated above  $F_{g(Y=0 | X,T,S=0)}^{-1}(p_{S=1|X,T}(x,t); x,t,1)$ .

The so-constructed distribution of  $g(Y=0 | X=x, T=t, S=0=1)$  replicates the identified

distribution and is equal to the distribution of  $g(Y=0 | X=x, T=t, S=0=0)$ , which satisfies the positive selection assumption.

b) If  $g(Y)$  is a monotonic decreasing function of  $Y$ , then Assumption 2 implies that the

distribution of  $g(Y=0)$  given  $X=x, T=t, S=0=1$  is stochastically dominated by the

distribution of  $g(Y=0)$  given  $X=x, T=t, S=0=0$ . This implies that the positive selection

assumption allows tightening the lower bound instead than the upper bound. The rest of the proof is along the lines of part a).

### 2.3 Proof of Theorem 3

Part a)  $P(S=1 \geq S=0 | X=x, T=t)=1$

Assumption 3a) excludes the existence of observations with  $S_0 = 1$  and  $S_1 = 0$ . There-

fore,  $p_{S_0, 1-S_1|X,T}(x, t) = 0$ ,  $p_{S_0, S_1|X,T}(x, t) = p_{S_0|X,T}(x, t)$ ,  $p_{1-S_0, S_1|X,T}(x, t) = p_{S_1|X,T}(x, t) - p_{S_0|X,T}(x, t)$  and  $p_{1-S_0, 1-S_1|X,T}(x, t) = 1 - p_{S_1|X,T}(x, t)$ .

By the law of total probability, Bayes' rule, and the implications of Assumption 3a) derived above:

$$\begin{aligned}
& E[g(Y=0) | X=x, T=t, S_1=1] = \\
& \stackrel{TP}{=} E[g(Y=0) | X=x, T=t, S_1=1, S_0=1] p_{S_0|X,T,S_1}(x, t, 1) + \\
& + E[g(Y=0) | X=x, T=t, S_1=1, S_0=0] [1 - p_{S_0|X,T,S_1}(x, t, 1)] \\
& \stackrel{BR}{=} E[g(Y=0) | X=x, T=t, S_1=1, S_0=1] \frac{p_{S_0, S_1|X,T}(x, t)}{p_{S_1|X,T}(x, t)} + \\
& + E[g(Y=0) | X=x, T=t, S_1=1, S_0=0] \frac{p_{1-S_0, S_1|X,T}(x, t)}{p_{S_1|X,T}(x, t)} \\
& \stackrel{A3a}{=} E[g(Y=0) | X=x, T=t, S_0=1] \frac{p_{S_0|X,T}(x, t)}{p_{S_1|X,T}(x, t)} + \\
& + E[g(Y=0) | X=x, T=t, S_1=1, S_0=0] \frac{p_{S_1|X,T}(x, t) - p_{S_0|X,T}(x, t)}{p_{S_1|X,T}(x, t)}. \tag{2}
\end{aligned}$$

The only unidentified element in (2) is  $E[g(Y=0) | X=x, T=t, S_1=1, S_0=0]$  bounded by  $\underline{b}_g$  and  $\bar{b}_g$ . Using those bounds, we obtain the result of Theorem 3a).

For the **sharpness** of the lower bound, we construct  $p_{S_0, 1-S_1|X,T}(x, t) = 0$ ,

$p_{S_0, S_1|X,T}(x, t) = p_{S_0|X,T}(x, t)$ ,  $p_{1-S_0, S_1|X,T}(x, t) = p_{S_1|X,T}(x, t) - p_{S_0|X,T}(x, t)$  and

$p_{1-S_0, 1-S_1|X,T}(x, t) = 1 - p_{S_1|X,T}(x, t)$ . We construct  $g(Y=0) = \underline{b}_g$  for the population with

$S_0 = 0$ . We construct the distribution of  $g_{Y=0} | X=x, T=t, S_1=1, S_0=1$  equal to the distribution of  $g_{Y=0} | X=x, T=t, S_0=1$ .

Part b)  $P(S_1 \leq S_0 | X=x, T=t) = 1$ .

Assumption 3b) excludes the existence of observations with  $S_0 = 0$  and  $S_1 = 1$ . Thus, we

get  $p_{S_0, S_1 | X, T}(x, t) = p_{S_1 | X, T}(x, t)$ ,  $p_{S_0, 1-S_1 | X, T}(x, t) = p_{S_0 | X, T}(x, t) - p_{S_1 | X, T}(x, t)$ , and

$p_{1-S_0, 1-S_1 | X, T}(x, t) = 1 - p_{S_0 | X, T}(x, t)$ . These simplifications lead to the following equality:

$$E[g_{Y=0} | X=x, T=t, S_1=1] = E[g_{Y=0} | X=x, T=t, S_1=1, S_0=1].$$

The distribution of  $g_{Y=0} | X=x, T=t, S_0=1$  is identified by assumption 1. The

population defined by  $X=x, T=t, S_0=1$  is a mixture of the population

$X=x, T=t, S_0=1, S_1=1$  with probability  $\frac{p_{S_1 | X, T}(x, t)}{p_{S_0 | X, T}(x, t)}$  and of

$X=x, T=t, S_0=1, S_1=0$  with probability  $\frac{p_{S_0 | X, T}(x, t) - p_{S_1 | X, T}(x, t)}{p_{S_0 | X, T}(x, t)}$ . The upper

bound will be attained when the population  $X=x, T=t, S_0=1, S_1=1$  represents the

$\frac{p_{S_1 | X, T}(x, t)}{p_{S_0 | X, T}(x, t)}$  fraction of the population with  $X=x, T=t, S_0=1$  with the largest value of

$g_{Y=0}$ . Similarly, the lower bound will be attained when the population

$X=x, T=t, S_0=1, S_1=1$  represents the  $\frac{p_{S_1 | X, T}(x, t)}{p_{S_0 | X, T}(x, t)}$  fraction of the population with

$X=x, T=t, S_0=1$  with the smallest value of  $g_{Y=0}$ .

For the **sharpness** of the lower bound, we construct  $p_{S=0, S=1|X,T}(x,t) = p_{S=1|X,T}(x,t)$ ,

$$p_{S=0, 1-S=1|X,T}(x,t) = p_{S=0|X,T}(x,t) - p_{S=1|X,T}(x,t), \text{ and } p_{1-S=0, 1-S=1|X,T}(x,t) = 1 - p_{S=0|X,T}(x,t).$$

We construct a distribution of  $g(Y=0 | X=x, T=t, S=1=0, S=0=1)$  equal to the distribution

$$\text{of } g(Y=0 | X=x, T=t, S=0=1) \text{ truncated below } F_{g(Y=0 | X,T,S=0)}^{-1} \left( \frac{p_{S=1|X,T}(x,t)}{p_{S=0|X,T}(x,t)}; x, t, 1 \right).$$

Similarly, we can construct a distribution of  $g(Y=0 | X=x, T=t, S=1=1, S=0=1)$  equal to the

distribution of  $g(Y=0 | X=x, T=t, S=0=1)$  truncated above

$$F_{g(Y=0 | X,T,S=0)}^{-1} \left( \frac{p_{S=1|X,T}(x,t)}{p_{S=0|X,T}(x,t)}; x, t, 1 \right). \text{ By construction, the mixture of these two distributions}$$

replicates the identified distribution of  $g(Y=0 | X=x, T=t, S=0=1)$ .

## 2.4 Proof of Theorem 4

Part a): We have shown in equation (2) that Assumptions 1 and 3-a) imply the following:

$$\begin{aligned} E[g(Y=0 | X=x, T=t, S=1=1)] = \\ E[g(Y=0 | X=x, T=t, S=0=1)] \frac{p_{S=0|X,T}(x,t)}{p_{S=1|X,T}(x,t)} + \\ + E[g(Y=0 | X=x, T=t, S=1=1, S=0=0)] \frac{p_{S=1|X,T}(x,t) - p_{S=0|X,T}(x,t)}{p_{S=1|X,T}(x,t)}. \end{aligned}$$

The only unknown element is  $E[g(Y=0 | X=x, T=t, S=1=1, S=0=0)]$ . By Assumption 2

and because  $g(\cdot)$  is monotone increasing, the distribution of  $g(Y=0)$  given

$X=x, T=t, S=0=0$  is stochastically dominated by the (identified) distribution of  $g(Y=0)$

given  $X=x, T=t, S=0=1$ . The upper bound is attained when these two distributions are

identical. The distribution of  $g(Y=0)$  given  $X=x, T=t, S=0=0$  is mixture of the sub-

population  $S_1 = 0$  with probability  $\frac{p_{1-S_1|X,T}(x,t)}{p_{1-S_0|X,T}(x,t)}$  and of subpopulation  $S_1 = 1$  with

$\frac{p_{S_1|X,T}(x,t) - p_{S_0|X,T}(x,t)}{p_{1-S_0|X,T}(x,t)}$ . Therefore, the upper bound on

$E[g(Y=0) | X=x, T=t, S_1=1, S_0=0]$  is given by

$$\max \left| \frac{E[g(Y=0) | X=x, T=t, S_0=1]}{p_{1-S_0|X,T}(x,t)} \right|.$$

For the **sharpness** of the bound, we construct  $p_{S_0, 1-S_1|X,T}(x,t) = 0$ ,

$$p_{S_0, S_1|X,T}(x,t) = p_{S_0|X,T}(x,t), \quad p_{1-S_0, S_1|X,T}(x,t) = p_{S_1|X,T}(x,t) - p_{S_0|X,T}(x,t) \quad \text{and}$$

$$p_{1-S_0, 1-S_1|X,T}(x,t) = 1 - p_{S_1|X,T}(x,t). \quad \text{We construct a distribution of}$$

$g(Y=0) | X=x, T=t, S_1=0, S_0=0$  equal to the distribution of

$g(Y=0) | X=x, T=t, S_0=1$  truncated above

$$F_{g(Y=0) | X,T,S_0}^{-1} \left( \frac{p_{S_1|X,T}(x,t) - p_{S_0|X,T}(x,t)}{p_{1-S_0|X,T}(x,t)}; x, t, 1 \right). \quad \text{Similarly, we construct a distribution of}$$

$g(Y=0) | X=x, T=t, S_1=1, S_0=0$  equal to the distribution of

$g(Y=0) | X=x, T=t, S_0=1$  truncated above

$$F_{g(Y=0) | X,T,S_0}^{-1} \left( \frac{p_{S_1|X,T}(x,t) - p_{S_0|X,T}(x,t)}{p_{1-S_0|X,T}(x,t)}; x, t, 1 \right). \quad \text{By construction, the mixture of these two}$$

distributions replicates  $g(Y=0) | X=x, T=t, S_0=1$  and  $g(Y=0) | X=x, T=t, S_0=0$ ,

which is allowed by the positive selection assumption.

Part b) The proof is similar to part a), but with the stochastic dominance inverted.

## 2.5 Proof of Theorem 5

a) We have shown in equation (2) that Assumptions 1 and 3-a) imply the following:

$$\begin{aligned} E[g \mid Y=0 \mid X=x, T=t, S_1=1] = \\ E[g \mid Y=0 \mid X=x, T=t, S_0=1] \frac{p_{S_0=1 \mid X, T}(x, t)}{p_{S_1=1 \mid X, T}(x, t)} + \\ + E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=0] \frac{p_{S_1=1 \mid X, T}(x, t) - p_{S_0=1 \mid X, T}(x, t)}{p_{S_1=1 \mid X, T}(x, t)}. \end{aligned}$$

By Assumptions 3 and 4 and because  $g \cdot$  is monotonic increasing,

$$\begin{aligned} E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=0] \\ \stackrel{A4}{\leq} E[g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=1] \\ \stackrel{A3a}{=} E[g \mid Y=0 \mid X=x, T=t, S_0=1]. \end{aligned} \tag{3}$$

Inserting (3) in (2) we get the result of Theorem 5-a).

For the **sharpness** of the bound, we construct  $p_{S_0=1, 1-S_1 \mid X, T}(x, t) = 0$ ,

$$p_{S_0=0, S_1 \mid X, T}(x, t) = p_{S_0 \mid X, T}(x, t), \quad p_{1-S_0, S_1 \mid X, T}(x, t) = p_{S_1 \mid X, T}(x, t) - p_{S_0 \mid X, T}(x, t) \quad \text{and}$$

$$p_{1-S_0=0, 1-S_1 \mid X, T}(x, t) = 1 - p_{S_1 \mid X, T}(x, t). \quad \text{We construct a distribution of}$$

$$g \mid Y=0 \mid X=x, T=t, S_1=0, S_0=0 \text{ as well as a distribution of}$$

$$g \mid Y=0 \mid X=x, T=t, S_1=0, S_0=1 \text{ equal to the distribution of}$$

$$g \mid Y=0 \mid X=x, T=t, S_0=1. \text{ This construction replicates the identified quantities and satisfy}$$

the positive selection assumption.

b) Similar to part a).

### 3. Additional Theorems

The following 2 theorems do not appear in the main paper since they are not useful for the application. However, they may be of interest in many other applications. These results are therefore stated and proved here.

#### 3.1 Exclusion restriction

*Assumption A1 (exclusion restriction)*

a) There is a random variable  $Z$  with support  $\mathcal{Z}$  such that:

$$Y = 0 \perp Z \mid X = x, T = t, S = 1, \forall x \in \mathcal{X}_{T=t}.$$

b) Assumption 1 holds with  $Z$  included in the list of control variables  $X$ .

*Theorem A1 (exclusion restriction)*

Assumptions 1 and A1 hold. For the case  $p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) > 1$ , define:

$$\begin{aligned} \underline{b}_g(z) &= \frac{E[g(Y=0) \mid X=x, Z=z, T=t, S=0] \times \min \left| \frac{p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) - 1}{p_{S=0|X,Z,T}(x, z, t)} \right|}{\times \frac{p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) - 1}{p_{S=1|X,Z,T}(x, z, t)} + \underline{b}_g \frac{1 - p_{S=0|X,Z,T}(x, z, t)}{p_{S=1|X,Z,T}(x, z, t)}} \\ \bar{b}_g(z) &= \frac{E[g(Y=0) \mid X=x, Z=z, T=t, S=0] \times \max \left| \frac{p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) - 1}{p_{S=0|X,Z,T}(x, z, t)} \right|}{\times \frac{p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) - 1}{p_{S=1|X,Z,T}(x, z, t)} + \bar{b}_g \frac{1 - p_{S=0|X,Z,T}(x, z, t)}{p_{S=1|X,Z,T}(x, z, t)}} \end{aligned}$$

If  $p_{S=0|X,Z,T}(x, z, t) + p_{S=1|X,Z,T}(x, z, t) \leq 1$ , we get  $\underline{b}_g(z) = \underline{b}_g$  and  $\bar{b}_g(z) = \bar{b}_g$ . The lower bound

on  $E[g(Y=0) \mid X=x, T=t, S=1] = 1$  is given by  $\sup_{z \in \mathcal{Z}} \underline{b}_g(z)$  and the upper bound by

$$\inf_{z \in \mathcal{Z}} \bar{b}_g(z).$$



### *Proof of Theorem A1*

All the results of Theorem 1 hold also conditionally on  $Z = z$ . By Assumption A1-a), for

$\forall x \in \mathcal{X}_{T=t}$  and  $\forall z \in Z$ , we get

$$E[g \mid Y=0 \mid X=x, Z=z, T=t, S=1] = E[g \mid Y=0 \mid X=x, T=t, S=1].$$

Therefore,  $E[g \mid Y=0 \mid X=x, T=t, S=1]$  lies in the intersection of the bounds evaluated at different  $z$ .

The sharpness of the bounds follows from the sharpness of the bounds of Theorem 1.

### **3.2 Alternative definition of positive selection**

Note that neither Theorem 3 nor Theorem 4 allows to tighten the bounds if

$P(S=1 \leq S=0 \mid X=x, T=t=1)$ . The intuition for this result is that in this case, all observations with  $S=1=1$  also have  $S=0=1$ . Thus, the problem for identifying the counterfactual mean is not that we do not know the value for the population with  $S=0=0$  (this is irrelevant for the estimation of the effects on our target population), but that we do not know which of the observations with  $S=0=1$  have  $S=1=1$  as well. To tighten the bounds in this particular case, we suggest Assumption A2:

*Assumption A2 (positive selection into employment for  $Y=0$  with respect to  $S=1$ )*

$$F_{Y=0 \mid X, T, S=0, S=1}(\tilde{y}; x, t, 1, 0) \geq F_{Y=0 \mid X, T, S=0, S=1}(\tilde{y}; x, t, 1, 1).$$

Note that this assumption is conceptually different from Assumptions 2 and 4 because it relates the control outcome to the treated employment status and is therefore more restrictive. Similar assumptions have been made by Angrist, Bettinger, Bloom, King, and Kremer (2002, especially footnote 20), Zhang and Rubin (2003, Assumption 2) and Angrist, Bettinger, and

Kremer (2006, especially proposition 1). To motivate this assumption, suppose that

$Y_1 + \alpha = Y_0$ , with  $\alpha \geq 0$  and suppose further that unemployed individuals accept a job if their potential earnings exceeds the reservation wage  $Y^R$ :  $S_d = 1 \mid Y_d \geq Y^R$ , for  $d \in \{0,1\}$

. This implies the following inequalities:

$$\begin{aligned} E[Y_0 \mid S_0 = 1, S_1 = 1] &= E[Y_0 \mid Y_1 \geq Y^R] = E[Y_0 \mid Y_0 \geq Y^R + \alpha] \\ &\geq E[Y_0 \mid Y_0 \geq Y^R] = E[Y_0 \mid S_0 = 1]. \end{aligned}$$

Since  $E[Y_0 \mid S_0 = 1]$  is a weighted average of  $E[Y_0 \mid S_0 = 1, S_1 = 1]$  and

$E[Y_0 \mid S_0 = 1, S_1 = 0]$ , the inequality implies that Assumption A2 is satisfied:

$$E[Y_0 \mid S_0 = 1, S_1 = 1] \geq E[Y_0 \mid S_0 = 1, S_1 = 0].$$

*Theorem A2 (positive selection into employment with respect to  $S_1$  and monotonicity)*

a) Assumptions 1, 3-b), and A2 hold. If  $g(\cdot)$  is a monotone increasing function, then:

$$E[g(Y_0) \mid X = x, T = t, S_0 = 1] \leq E[g(Y_0) \mid X = x, T = t, S_1 = 1].$$

b) Assumptions 1, 3-b), and A2 hold. If  $g(\cdot)$  is a monotone decreasing function, then:

$$E[g(Y_0) \mid X = x, T = t, S_0 = 1] \geq E[g(Y_0) \mid X = x, T = t, S_1 = 1].$$

*Proof of Theorem A2*

Part a): By the law of total probability, and Assumption 3-b):

$$\begin{aligned}
& E[g \mid Y=0 \mid X=x, T=t, S_1=1] \\
& \stackrel{TP}{=} E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=1] \frac{p_{S_0, S_1 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)} \\
& \quad + E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=0] \frac{p_{S_0 \mid X, T}(x, t) - p_{S_0, S_1 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)} \quad (4) \\
& \stackrel{A3b}{=} E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=1] \frac{p_{S_1 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)} \\
& \quad + E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=0] \frac{p_{S_0 \mid X, T}(x, t) - p_{S_0, S_1 \mid X, T}(x, t)}{p_{S_0 \mid X, T}(x, t)}.
\end{aligned}$$

By Assumptions A2 and the fact that  $g \cdot$  is monotonic increasing,

$$\begin{aligned}
& E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=0] \\
& \stackrel{AA2}{\leq} E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=1].
\end{aligned} \quad (5)$$

Inserting (5) into (4), we obtain:

$$\begin{aligned}
& E[g \mid Y=0 \mid X=x, T=t, S_1=1] \leq \\
& E[g \mid Y=0 \mid X=x, T=t, S_0=1, S_1=1] \\
& \stackrel{A3b}{=} E[g \mid Y=0 \mid X=x, T=t, S_0=1]
\end{aligned}$$

which is the result of Theorem A2-a).

For the **sharpness** of the lower bound, we construct  $p_{S_0, S_1 \mid X, T}(x, t) = p_{S_1 \mid X, T}(x, t)$ ,

$$p_{S_0, 1-S_1 \mid X, T}(x, t) = p_{S_0 \mid X, T}(x, t) - p_{S_1 \mid X, T}(x, t), \text{ and } p_{1-S_0, 1-S_1 \mid X, T}(x, t) = 1 - p_{S_0 \mid X, T}(x, t).$$

We construct a distribution of  $g \mid Y=0 \mid X=x, T=t, S_1=1, S_0=1$  equal to the distribution

of  $g \mid Y=0 \mid X=x, T=t, S_0=1$ . This replicates the identified quantities, satisfies monotonicity

and positive selection and attains the bound.

Part b) Proof is similar to the proof of part a).

## 4. Monte Carlo simulations

In this section, we evaluate the estimation and inference procedures proposed in section 6.2 of the paper. The codes in *R* available at [www.sew.unisg.ch/lechner/earnings](http://www.sew.unisg.ch/lechner/earnings) replicate all the results in this section.

### 4.1 Data generating processes

We consider only average treatment effects because of time and space constraints. We consider only the most informative bounds, which are the bounds resulting from our Theorem 5. Therefore, our data generating processes will respect the positive selection and conditional monotonicity assumptions.

We calibrate the data generating processes to match many characteristics of the data in our application. Their values in the dataset determine the number of observations, the distribution of the control variables  $X$ , the conditional treatment and employment probabilities and the conditional distribution of the outcome variable. However, for computational reasons, we have to limit the number of covariates to 3 (instead of 89).

In detail, the joint distribution of  $X_1$  (gender),  $X_2$  (age) and  $X_3$  (education) is given by the empirical distribution in our sample. 306 different combinations of value for  $X_1$ ,  $X_2$  and  $X_3$  are observed in the data. The treatment probability,  $Pr(D=1|X_1, X_2, X_3)$ , is also given by the empirical distribution in the sample.

The employment probabilities  $Pr(S=0|X_1, X_2, X_3)$  and  $Pr(S=1|X_1, X_2, X_3)$  cannot be taken from the nonparametric empirical distribution because there are not enough treated observations. Therefore, we estimated these probabilities by linear probit regressions of  $S$  on  $X_1$ ,  $X_2$  and  $X_3$  in the treated and non-treated samples. The values of these coefficients along with randomly generated normal errors are used to generate  $S=0$  and  $S=1$ .

The conditional distribution of  $Y_0$  and  $Y_1$  given  $X$  were estimated by 100 quantile regressions of  $Y$  on  $X_1$ ,  $X_2$ ,  $X_2^2$ ,  $X_3$ ,  $X_3^2$ , and  $X_1 \cdot X_3$ . Then, for each observation, we randomly draw one percentile and impute the fitted value given  $X$  at this quantile.

In these Monte Carlo simulations, we consider three scenarios. In all of them the true treatment effect is zero because we simulate both conditional distributions using the coefficients obtained in the non-treated sample. In the first scenario, the true treatment effect corresponds to the lower bound, because the positive sample selection assumption is just satisfied. This means that sample selection is random. In the second case, the true effect is about half way between both bounds because we have positive selection. In the third scenario, the true effect corresponds to the upper bound. We implement that by setting the unobserved control wages to 0.

## 4.2 Inference procedure

We apply the estimators of the lower and upper bounds described in the paper to each simulated sample and obtain  $\hat{\theta}_l$  and  $\hat{\theta}_u$ . We bootstrap the bounds 50 times to estimate the standard errors of the bounds,  $\hat{\sigma}_l$  and  $\hat{\sigma}_u$ .<sup>3</sup> We construct the confidence interval as

$$CI_\alpha = \left[ \hat{\theta}_l - \frac{c_\alpha \hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \frac{c_\alpha \hat{\sigma}_u}{\sqrt{n}} \right]$$

where  $1 - \alpha$  is the confidence level and  $c_\alpha$  solves

$$\Phi \left( c_\alpha + \frac{\sqrt{n} (\hat{\theta}_u - \hat{\theta}_l)}{\max(\hat{\sigma}_l, \hat{\sigma}_u)} \right) - \Phi(-c_\alpha) = 1 - \alpha$$

## 4.3 Results

The number of Monte Carlo simulations was set to 500. Table A1 reports statistics that are useful to evaluate the quality of the point estimates of the bounds. Especially of interest are

---

<sup>3</sup> In the application, the number of bootstrap draws is 200.

the lower bound in scenario 1 and the upper bound in scenario 2 because the true effect, which is equal to zero, corresponds to these bounds. The mean and median biases are small, always less than 10% of the standard deviation.

Another interesting result of Table A1 concerns the quality of the normal approximation. In all cases, the skewness is very close to 0 and the kurtosis slightly below 3. The  $p$ -values of the Jarque-Bera test confirm this impression.

**Table A1: Point estimates**

	<i>Scenario 1</i>		<i>Scenario 2</i>		<i>Scenario 3</i>	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Mean	-23.66	4703.26	-3537.55	3437.93	-7074.67	-99.19
Median	-30.97	4764.57	-3534.64	3457.98	-7046.25	-91.74
Standard deviation	726.06	825.53	532.35	884.03	792.84	1036.25
Skewness	-0.02	-0.13	-0.004	-0.11	-0.11	0.001
Kurtosis	2.57	2.72	2.66	2.77	2.79	2.83
Jarque-Bera p-value	0.1436	0.2132	0.3059	0.3600	0.3896	0.7353

Table A2 evaluates the quality of the inference procedure. We first compare the standard errors observed in the Monte Carlo simulations with the mean bootstrap based estimated standard errors. The differences are small with a maximum difference of 5% and with one exception the estimated standard errors are slightly larger than the observed ones, which should lead to conservative inference.

Table A2 also reports the observed rejection probabilities of the correct null hypothesis that the treatment effect is equal to zero. Since we bound the treatment effect, we reject this null hypothesis either if the lower bound of the confidence interval is above zero or if the upper bound of this interval is below zero. However, since the bounds are wide, it never happened that the upper bound was negative in scenario 1 or that the lower bound was positive in

scenario 3. In the second scenario, 0 was always between both bounds and, therefore, we never rejected the null hypothesis.

We report the results for three levels (1, 5 and 10%). In the two cases where the true treatment effect is equal to one bound, the observed rejection probabilities are very close to the theoretical ones and the differences are never statistically significant. This shows that the inference procedure used in our application has a correct size when the true treatment effect is equal to a bound and is conservative when the true effect is between the bounds.

**Tables A2: Inference**

	<i>Scenario 1</i>		<i>Scenario 2</i>		<i>Scenario 3</i>	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Observed s.e.	726.06	825.53	532.35	884.03	792.84	1036.25
Mean estimated s.e.	710.46	845.64	575.73	893.44	838.06	1053.25
Rejection probabilities						
$\alpha = 0.01$	0.01	0	0	0	0	0.008
$\alpha = 0.05$	0.036	0	0	0	0	0.05
$\alpha = 0.10$	0.094	0	0	0	0	0.122

## 5. Additional references

Alberto Abadie (2002): “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models”, *Journal of the American Statistical Association*, 97, 284-292.

Markus Frölich (2007): “Nonparametric IV estimation of local average treatment effects with covariates”, *Journal of Econometrics*, 139, 35-75.

Joshua Angrist, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer (2002): “Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment”, *American Economic Review*, 92, 1535-1558.