IDENTIFICATION OF TREATMENT EFFECTS WITH MISMEASURED IMPERFECT INSTRUMENTS

DÉSIRÉ KÉDAGNI

Iowa State University - Department of Economics

Abstract. In this article, I develop a novel identification result for estimating the effect of an endogenous treatment using a proxy of an unobserved imperfect instrument. I show that the potential outcomes distributions are partially identified for the compliers. Therefore, I derive sharp bounds on the local average treatment effect. I write the identified set in the form of conditional moments inequalities, which can be implemented using existing inferential methods. I illustrate my methodology on the National Longitudinal Survey of Youth 1979 to evaluate the returns to college attendance using tuition as a proxy of the true cost of going to college. I find that the average return to college attendance for people who attend college only because the cost is low is between 29% and 78%.

Keywords: Potential outcome, unobserved invalid instruments, LATE, proxy, mixture models.

JEL subject classification: C14, C21, C25, C26.

1. Introduction

Ability and cost are the two main drivers of the college education decision. However, these two variables are in general unobserved by the econometrician when estimating the returns to college. Ability influences both wages and schooling, which makes education endogenous. A potential instrument for college education would be its cost (which includes the financial cost, the opportunity cost, as well as the psychological cost). However, high ability individuals tend to go to high quality schools (signaling), which often have higher cost. For this reason, the cost of college education would not be a valid instrument even if it were observed by the econometrician. Because the actual costs of education are unknown,

Date: The present version is of June 10, 2019. I thank Otávio Bartalotti, Helle Bunzel, Alfonso Flores-Lagunes, Christian Hansen, Marc Henry, Brent Kreider, Ismael Mourifié, Alex Zhylyevskyy, participants at the Iowa State econometrics workshop for helpful discussions and comments. All errors are mine. Corresponding address: 260 Heady Hall, 518 Farm House Lane, Ames, IA, 50011, USA. Email address: dkedagni@iastate.edu.

researchers often resort to proxy variables such as tuition fees, distance to college, local unemployment rate, local average earnings, etc, to instrument for education.

This paper proposes an identification result for estimating the returns to college when an unobserved imperfect instrument like cost is replaced by its proxy (e.g., tuition). I show that the local average treatment effect defined based on the unobserved imperfect instrument is partially identified under some mild assumptions. Moreover, I partially identify the potential outcome distributions for the compliers (e.g., people who go to college only because the cost is low).

The intuition behind my identification result is the following. The treatment variable college attendance and the unobserved binary cost variable[1] partition the population into four unobserved groups, commonly referred to as types: the always-takers, individuals who attend college regardless of the cost level; the never-takers, people who will not attend college regardless of the cost level; the compliers, those who attend college only because the cost is low; and the defiers, individuals who attend college only because the cost is high (which I assume away in the population for convenience). Under my identifying assumptions, the identified distribution of the observed outcome variable conditional on the treatment and the proxy is a mixture of the potential outcome distributions for the compliers and the always-takers or never-takers, where mixture weights depend on the proxy variable while the mixture components do not. Therefore, I use variations in the proxy variable to derive sharp bounds on the mixture components, which are the distributions of interest.

I write the identified set in the form of conditional moment inequalities, which is more conducive to inference. I apply my methodology on the National Longitudinal Survey of Youth 1979 (NLSY79) data and find that the average return to college attendance for individuals who attend college only because the cost is low is between 29% and 78%.

Related literature. Chalak (2017) discusses the interpretation of various estimands (Wald, Instrumental variable: IV, local instrumental variable: LIV) when the true instrument is valid, but mismeasured. In the current paper, the true instrument is invalid and mismeasured. This means that the estimands above will not have any clear causal interpretations. Kédagni (2018) studies identification of treatment effects when the true instrument is observed, but is invalid. He relies on the existence of a proxy of the true instrument to do

[1]The true cost of going to college may be continuous. However, I assume that the individuals make the schooling decision based on their ability and whether the cost is low or high. Each agent has its own subjective threshold above which she judges the cost high.

identification. This use of a proxy variable for identification purposes is not new in the literature, see Lubotsky and Wittenberg (2006) and references therein. However, using a proxy of an invalid unobserved instrumental variable to identify heterogeneous treatment effects appears novel.

This paper relates to the work by Ura (2018), Lewbel (2007), Mahajan (2006), Kreider et al. (2012) among others, as it allows for measurement errors in a variable. However, here the measurement error is in the instrumental variable while the papers above allow for measurement errors in the treatment variable. Like in Ura (2018), the measurement error is nonclassical (i.e., it can depend on the true instrument) and differential (i.e., it can be dependent on the outcome conditional on the true instrument). The paper uses results on identification of finite mixture models by Henry, Kitamura, and Salanié (2014) to partially identify parameters of interest. Even though, I derive bounds on the potential outcome distributions for the compliers, I focus on the local average treatment effect introduced by Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996) in the empirical illustration. This article also relates to Card (2001) as it proposes a way to do identification with an invalid instrument like cost, an issue raised by the author.

The remainder of the paper is organized as follows. Section 2 presents the model and the identifying assumptions. In Section 3, we provide the identification results. Section 4 shows some empirical evidence and Section 5 concludes. The proofs are presented in the appendix.

2. Analytical Framework

Consider the following IV model

$$\begin{cases} Y &= \alpha D + U \\ D &= \delta Z + V \end{cases} \tag{2.1}$$

where the vector (Y, D, Z) is the observed data, $Y \in \mathcal{Y} \subset \mathbb{R}$, D and Z are binary, and (α, U, δ, V) is a vector of latent variables. Since D and Z are binary, for the model to be well-specified, we must have $V \in \{0,1\}$ and $\delta \in \{-1, 0, 1\}$. Indeed, the support of (δ, V) is {(0,1),(1,0),(0,0),(-1,1)}[2]. The standard LATE assumptions introduced by Imbens and Angrist (1994) are:

[2]To see this, suppose $Y = g(D, \tilde{U})$ and $D = h(Z, \tilde{V})$. Since D and Z are binary, we can rewrite $Y = \alpha D + U$ and $D = \delta Z + V$, where $\alpha = [g(1, \tilde{U}) - g(0, \tilde{U})]$, $U = g(0, \tilde{U})$, $\delta = [h(1, \tilde{V}) - h(0, \tilde{V})]$, $V = h(0, \tilde{V})$, and $h(z, \tilde{V}) \in \{0, 1\}$ for all $z \in \{0, 1\}$.

- $Z \perp\!\!\!\perp (\alpha, U, \delta, V)$ (full independence);
- $\delta \geq 0$, i.e., $\delta \in \{0, 1\}$ (monotonicity);
- $\mathbb{E}[\delta] \neq 0$ (non-zero effect of Z on D).

Under these assumptions, $LATE \equiv \mathbb{E}[\alpha|\delta = 1]$ is identified as

$$\alpha_{IV} \equiv \frac{Cov(Y, Z)}{Cov(D, Z)} = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]}.$$

First, I relax full independence $Z \perp\!\!\!\perp (\alpha, U, \delta, V)$ to conditional independence $Z \perp\!\!\!\perp (\alpha, U)|(\delta, V)$. The intuition is that if (δ, V) were independent of (α, U), there would be no endogeneity issue in the model as long as Z is exogenous. In such a case, $\mathbb{E}[\alpha]$ is identified by $\mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0]$. Therefore, controlling for (δ, V) would help deal with endogeneity in the model.

Suppose now that the econometrician does not observe Z, instead she observes a proxy W that satisfies: $(Z, W) \perp\!\!\!\perp (\alpha, U) \mid (\delta, V)$. As the instrument Z, its proxy W can be dependent on the unobserved heterogeneity (α, U, δ, V). This assumption helps partially identify the same $LATE = \mathbb{E}[\alpha|\delta = 1]$. All results derived in the paper hold conditionally on covariates. To ease the exposition, I drop exogenous covariates from the model.

As illustration, let D be an indicator for college education, Y be log wage, Z be an indicator for low college cost, and W be tuition fees (or an indicator for college proximity). High ability individuals tend to go to high quality schools (signaling), which often have higher cost (See Table 1 for the NLSY79 data). Therefore, Z would not be independent of (α, U, δ, V).

Table 1. Ability and tuition

tuition	OLS
ability	0.1071**
	(0.0538)
n	1230

Standard errors (in parentheses)
** stands for 5% significant.

We are interested in learning some characteristics of the latent variable α. Denote $Y_1 = \alpha + U$ and $Y_0 = U$. I now state my main identifying assumptions.

Assumption 1 (Selection on unobservables)**.** *The vector (Z, W) is independent of Y_d given the unobservables (δ, V), i.e., $(Z, W) \perp\!\!\!\perp Y_d \mid (\delta, V)$, for both $d = 0$ and $d = 1$.*

Assumption 2 (Monotonicity)**.** $\delta \geq 0$.

Assumption 1 states that the vector of the imperfect instruments (Z, W) is independent of the potential outcomes Y_d given the first stage unobserved heterogeneity (δ, V). It is weaker than the assumption $(Z, W) \perp\!\!\!\perp (\alpha, U) \mid (\delta, V)$ discussed above. Assumption 2 is the standard monotonicity assumption that rules out the the element $(-1, 1)$ from the support of (δ, V). It is used here for convenience. Relaxing this assumption will only increase the dimensionality of the parameters to be identified.[3]

In the treatment effect literature, the unobserved heterogeneity (δ, V) partition the population into four unobserved groups known as types or strata: the always-takers, the defiers, the compliers and the never-takers. For example, in my framework, the always-takers are

Table 2. Subpopulations

Types	δ	V
Always-takers	0	1
Compliers	1	0
Defiers	-1	1
Never-takers	0	0

people who would go to college regardless of the cost being low or high; the compliers are individuals who attend college only because the cost is low; the never-takers are people who will not go to college whether the cost is low or not; and the defiers are individuals who attend college only because the cost is high.

3. Identification results

In this section, I provide a heuristic derivation of the main results in the paper. Formal proofs are relegated to the appendix. I first write the distribution of the observed outcome Y conditional on the treatment $D = 1$ and the proxy variable $W = w$ as a mixture of distributions of the potential outcome Y_1 for the compliers group $(\delta = 1, V = 0)$ and the always-takers group $(\delta = 0, V = 1)$, where only the mixture weights depend on w. Therefore, I use variations in w to write the compliers and always-takers distributions of Y_1 as functions of two parameters that are partially identified. Furthermore, I show that point-identification can be achieved under some tail conditions.

[3]See Appendix B for more details.

3.1. **Partial identification.** For $y \in \mathcal{Y}$, we have

$$\begin{aligned}
\mathbb{P}(Y \leq y|D = 1, W = w) &= \mathbb{P}(Y_1 \leq y|D = 1, W = w), \\
&= \mathbb{P}(Y_1 \leq y|D = 1, \delta = 1, V = 0, W = w)\mathbb{P}(\delta = 1, V = 0|D = 1, W = w) \\
&\quad + \mathbb{P}(Y_1 \leq y|D = 1, \delta = 0, V = 1, W = w)\mathbb{P}(\delta = 0, V = 1|D = 1, W = w), \\
&= \mathbb{P}(Y_1 \leq y|Z = 1, \delta = 1, V = 0, W = w)\mathbb{P}(\delta = 1, V = 0|D = 1, W = w) \\
&\quad + \mathbb{P}(Y_1 \leq y|D = 1, \delta = 0, V = 1, W = w)\mathbb{P}(\delta = 0, V = 1|D = 1, W = w),
\end{aligned}$$

where the second equality follows from the law of iterated expectations (LEI) and Assumption 2, and the third holds because the following equality holds: $\{D = 1, \delta = 1, V = 0\} = \{Z = 1, \delta = 1, V = 0\}$. Using Assumption 1 and this implication $\{\delta = 0, V = 1\} \Longrightarrow \{D = 1\}$ (the always-takers take the treatment), we have

$$\begin{aligned}
\mathbb{P}(Y \leq y|D = 1, W = w) \quad &= \quad \mathbb{P}(Y_1 \leq y|\delta = 1, V = 0)\mathbb{P}(\delta = 1, V = 0|D = 1, W = w) \\
&\quad + \mathbb{P}(Y_1 \leq y|\delta = 0, V = 1)\mathbb{P}(\delta = 0, V = 1|D = 1, W = w).
\end{aligned}$$

The identified distribution $\mathbb{P}(Y \leq y|D = 1, W = w)$ is a mixture of two distributions of interest $\mathbb{P}(Y_1 \leq y|\delta = 1, V = 0)$ and $\mathbb{P}(Y_1 \leq y|\delta = 0, V = 1)$ with unknown weights functions of w. Similarly, the identified distribution $\mathbb{P}(Y \leq y|D = 0, W = w)$ is a mixture of two distributions of interest $\mathbb{P}(Y_0 \leq y|\delta = 1, V = 0)$ and $\mathbb{P}(Y_0 \leq y|\delta = 0, V = 0)$ with unknown weights functions of w.

$$\begin{aligned}
\mathbb{P}(Y \leq y|D = 0, W = w) \quad &= \quad \mathbb{P}(Y_0 \leq y|\delta = 1, V = 0)\mathbb{P}(\delta = 1, V = 0|D = 0, W = w) \\
&\quad + \mathbb{P}(Y_0 \leq y|\delta = 0, V = 0)\mathbb{P}(\delta = 0, V = 0|D = 0, W = w).
\end{aligned}$$

For the sake of clarity of exposition, I use some additional notation.

Notation 1. *Denote* $\alpha^d(w) \equiv \mathbb{P}(\delta = 1, V = 0|D = d, W = w)$, $F(y|d, w) \equiv \mathbb{P}(Y \leq y|D = d, W = w)$, $\alpha^d(A) \equiv \mathbb{P}(\delta = 1, V = 0|D = d, W \in A)$, $F(y|d, A) \equiv \mathbb{P}(Y \leq y|D = d, W \in A)$, $F_d(y|1, 0) \equiv \mathbb{P}(Y_d \leq y|\delta = 1, V = 0)$, *and* $F_d(y|0, d) \equiv \mathbb{P}(Y_d \leq y|\delta = 0, V = d)$.

We have the following two-component mixture models:

$$F(y|d, w) \quad = \quad F_d(y|1, 0)\alpha^d(w) + F_d(y|0, d)(1 - \alpha^d(w)), \tag{3.1}$$

for each $d \in \{0, 1\}$. Therefore, I follow the results of Henry, Kitamura, and Salanié (2014) on identification of finite mixture models to show that the distributions $F_d(y|1, 0)$ and $F_d(y|0, d)$ are identified up to two scalar parameters that are partially identified. Indeed, I show that for some subsets A_0^d and A_1^d of $\mathcal{W}$, the distributions $F_d(y|0, d)$, $F_d(y|1, 0)$, and the probability weight $\alpha^d(w)$ can be written as functions of two parameters θ^d and η^d as follows:[4]

$$\begin{aligned}
F_d(y|0,d) &= F(y|d,A_0^d) - \eta^d\left[F(y|d,A_1^d) - F(y|d,A_0^d)\right], \\
F_d(y|1,0) &= F(y|d,A_0^d) + \left(\theta^d - \eta^d\right)\left[F(y|d,A_1^d) - F(y|d,A_0^d)\right], \qquad (3.2)\\
\alpha^d(w) &= \frac{1}{\theta^d}\left(\eta^d + \Lambda^d(w)\right),
\end{aligned}$$

where

$$\Lambda^d(w) = \frac{F(y^d|d,w) - F(y^d|d,A_0^d)}{F(y^d|d,A_1^d) - F(y^d|d,A_0^d)}$$

for some $y^d \in \mathcal{Y}$. An implicit assumption behind this result is that there exist y^d, A_0^d and A_1^d such that $F(y^d|d,A_1^d) \neq F(y^d|d,A_0^d)$.

Under a relevance assumption that I discuss below, if $F(y|d,A_1^d) = F(y|d,A_0^d)$ for all y, then the distributions $F_d(y|1,0)$ and $F_d(y|0,d)$ are equal and point-identified. Indeed, we have $F(y|d,A_1^d) - F(y|d,A_0^d) = [\alpha^d(A_1^d) - \alpha^d(A_0^d)][F_d(y|1,0) - F_d(0,d)]$. Suppose that $\alpha^d(A_1^d) \neq \alpha^d(A_0^d)$. Then $F(y|d,A_1^d) = F(y|d,A_0^d)$ implies $F_d(y|1,0) = F_d(y|0,d)$. By plugging the latter equality in Equation (3.1), it follows that the distributions $F_d(y|1,0)$ and $F_d(y|0,d)$ are identified and are both equal to the distribution of the outcome conditional on the treatment $F(y|d)$.

Notice that I turn the problem to a parametric model, which is easier to deal with as it reduces the dimensionality of the problem. The parameters η^d and θ^d are partially identified using the monotonicity condition of a cumulative distribution function and the condition on the probability weights as follows:

$$\begin{aligned}
&f(y|d,A_0^d) - \eta^d\left[f(y|d,A_1^d) - f(y|d,A_0^d)\right] \geq 0, \\
&f(y|d,A_0^d) + \left(\theta^d - \eta^d\right)\left[f(y|d,A_1^d) - f(y|d,A_0^d)\right] \geq 0, \qquad (3.3)\\
&0 \leq \frac{1}{\theta^d}\left(\eta^d + \Lambda^d(w)\right) \leq 1,
\end{aligned}$$

for all y and w, where $f(y|d,A)$ denotes the density (or probability mass) function of Y conditional on $(D = d, W \in A)$.

Before I summarize the above discussion, let me state clearly one relevance assumption that I use.

[4]See Appendix A.1 for more details.

Assumption 3 (Relevance). *For each $d \in \{0,1\}$, there exist two subsets A_0^d and A_1^d of $\mathcal{W}$ such that $\alpha^d(A_0^d) \neq \alpha^d(A_1^d)$.*

Proposition 1. *Under Assumptions 1, 2 and 3, the following holds.*

(1) *If $F(y^d|d, A_1^d) \neq F(y^d|d, A_0^d)$ for some $y^d \in \mathcal{Y}$, then the distributions $F_d(y|1,0)$ and $F_d(y|0,d)$ are partially identified as described by Equations (3.2) and (3.3). The identification region is sharp.*
(2) *If $F(y|d, A_1^d) = F(y|d, A_0^d)$ for all $y \in \mathcal{Y}$, then the distributions $F_d(y|1,0)$ and $F_d(y|0,d)$ are point-identified and are both equal to the conditional distribution $F(y|d)$.*

The proposition above shows that in general, the potential outcome distributions are not point identified, but partially identified. However, in some extreme cases, they are point-identified. As a corollary of this proposition, sharp bounds on the LATE can be obtained as the interval defined by the minimum and the maximum of the difference of the expectations of the potential outcomes distributions for compliers $F_1(y|1,0)$ and $F_0(y|1,0)$.

Denote by Ω^d the (sharp) identified set for (θ^d, η^d) and by $\mathbb{E}[F_d(y|1,0)]$ the expectation of the distribution $F_d(y|1,0)$, $d \in \{0,1\}$.

Corollary 1. *Under Assumptions 1, 2 and 3, the following holds.*

$$\begin{aligned} \inf_{(\theta^1,\eta^1)\in\Omega^1} \mathbb{E}[F_1(y|1,0)] - \sup_{(\theta^0,\eta^0)\in\Omega^0} \mathbb{E}[F_0(y|1,0)] \\ \leq \mathbb{E}[\alpha|\delta=1] \leq \\ \sup_{(\theta^1,\eta^1)\in\Omega^1} \mathbb{E}[F_1(y|1,0)] - \inf_{(\theta^0,\eta^0)\in\Omega^0} \mathbb{E}[F_0(y|1,0)]. \end{aligned}$$

These bounds are sharp.

The proof of Corollary 1 is straightforward since the distributions $F_d(y|1,0)$, $d \in \{0,1\}$, are linear in the parameters θ^d and η^d, and the identified set Ω^d is compact.

Similarly, we can obtain bounds for the local quantile treatment effect, the distributional treatment effect, and many other parameters that are functions of the potential outcomes distributions for compliers.

3.2. **Point-identification under tail restrictions.** In this subsection, I show that point-identification of the potential outcome distributions for the compliers can be obtained under the following assumption.

Assumption 4 (Tail restrictions (TR))**.** *$\lim_{y\downarrow y^\ell} \frac{F_0(y|1,0)}{F_0(y|0,0)} = 0$ and $\lim_{y\uparrow y^u} \frac{1-F_1(y|1,0)}{1-F_1(y|0,1)} = 0$, where y^ℓ and y^u are lower and upper bounds of the support $\mathcal{Y}$, repectively.*

The intuition behind this assumption is the following. Think of the always-takers as the high return individuals, the never-takers as the low return ones, and the compliers as the marginal individuals. The condition $\lim_{y\uparrow y^u} \frac{1-F_1(y|1,0)}{1-F_1(y|0,1)} = 0$ states that among people who attended college, the high earners among the always-takers earn an order of magnitude higher than the high earners among the compliers. Similarly, The assumption $\lim_{y\downarrow y^\ell} \frac{F_0(y|1,0)}{F_0(y|0,0)} = 0$ means that among people who did not attend college, the low earners among the never-takers earn an order of magnitude less than the low earners among the compliers. Jochmans, Henry, and Salanié (2017) also study nonparametric partial identification of finite mixtures with varying weights and fixed component distributions under some tail restrictions.

Proposition 2. *Under Assumptions 1, 2, 3 and 4, the distributions $F_1(y|1,0)$ and $F_0(y|1,0)$ are point-identified as follows:*

$$\begin{aligned} F_0(y|1,0) &= F(y|0,A_0^0) + \tfrac{1}{1-\zeta^0(A_1^0,A_0^1)}\left[F(y|0,A_1^0) - F(y|0,A_0^0)\right], \\ F_1(y|1,0) &= F(y|1,A_0^1) + \tfrac{1}{1-\pi^1(A_1^1,A_0^1)}\left[F(y|1,A_1^1) - F(y|1,A_0^1)\right], \end{aligned} \tag{3.4}$$

where

$$\zeta^0(A_1^0,A_0^0) = \lim_{y\downarrow y^\ell} \frac{F(y|0,A_1^0)}{F(y|0,A_0^0)}, \quad \text{and} \quad \pi^1(A_1^1,A_0^1) = \lim_{y\uparrow y^u} \frac{1-F(y|1,A_1^1)}{1-F(y|1,A_0^1)}.$$

The result of this proposition combined with that of Proposition 1 can serve as a test for Assumption 4. Indeed, under Assumptions 1, 2, 3 and 4, the bounds of Proposition 1 remain valid. Therefore, the point-identified distributions in Proposition 2 must lie within those bounds. Otherwise, Assumption 4 is not compatible with the other three. From Proposition 2, we obtain point-identification of the LATE.

In the next section, I illustrate how the bounds of Proposition 1 can be constructed in practice using the NLSY79 data.

4. Empirical results: Returns to college

In this section, I use the methodology developed in this paper to evaluate the returns to college attendance using tuition fees as a proxy for the cost of going to college. As discussed in the introduction, the cost of college education comprises the financial cost, the opportunity cost, the psychological cost (cost of effort), etc. High ability students usually tend to attend high quality schools (signaling), which often have higher attendance cost.[5] For this reason, the educational cost, even if it were observed, would not be a good instrument for schooling as it is not independent of ability. Card (2001) warns researchers against using variables that are related to educational institutions like tuition fees or distance to college as instruments for schooling.

For my analysis, I use the data of Heckman, Tobias, and Vytlacil (2001). The data is a sample of 1,230 white males taken from the National Longitudinal Survey of Youth 1979. The outcome variable is the log weekly wage labelled "lwage", the treatment variable "college" is the indicator of whether the individual completed at least 13 years of education or not. The proxy variable "tuit17" is the tuition fees at age 17. Table 3 shows the descriptive statistics. In the data, 43% of the individuals completed 13 years of education. The data also contains a measure of ability we use to show some evidence that tuition is correlated with ability. Notice that Kédagni and Mourifié (2015) reject the independence assumption between potential earnings and tuition fees.

Table 3. Summary Statistics

	Total
Observations	1,230
lwage	2.4138 (0.5937)
college	0.4325 (0.4956)
tuit17	8.5686 (4.1277)
abil	1.7966 (2.1844)

Average and standard deviation (in the parentheses)

I rewrite inequalities in (3.3) in the form of conditional moment inequalities, which allows me to use existing inference methods such as those of Chernozhukov, Lee, and Rosen (2013) and Chernozhukov et al. (2015). One could alternatively use Andrews and Shi (2013) and Andrews, Kim, and Shi (2016). See the appendix for the details of the implementation.

Table 4 displays the confidence regions for different parameters. The results show that the 95% confidence for the LATE is between is between 0.2545 and 0.5753 log points, that is, the average return to college for individuals who attend college only because the cost is low is between 29% and 78%. This parameter may be relevant for policies which aim at reducing the college attendance cost, as it shows that the return is positive and substantial

[5]Note however that high quality schools are more likely to provide scholarships to their students, which can compensate their high cost. But, still the attendance cost for these schools remains high.

for people who are sensitive to the cost. Covariates can be included in the analysis, but the

Table 4. Confidence sets for parameters

Parameters	Estimates	95% conf. LB	95% conf. UB
θ^1		0.1	9.1
η^1		0.5	50
θ^0		0.1	9.1
η^0		0.5	50
$\mathbb{E}[Y_1 \vert \delta = 1]$		2.5185	2.6822
$\mathbb{E}[Y_0 \vert \delta = 1]$		2.1069	2.2640
$LATE \equiv \mathbb{E}[\alpha \vert \delta = 1]$		0.2545	0.5753
OLS	0.4039***	0.3408	0.4671
	(0.0322)		

Standard errors (in parentheses); *** stands for 1% significant;
conf.: confidence; LB: lower bound; UB: upper bound.

size of the data is not big enough to allow me to do so.

5. Conclusion

This paper develops a new identification result for evaluating the effect of an endogenous treatment when the researcher observes a proxy of an imperfect instrumental variable. I derive sharp bounds on the potential outcomes distributions for compliers and hence on the local average treatment effect. I show that inference on the identified set can be done using existing results on intersection bounds. I apply my methodology on the National Longitudinal Survey of Youth 1979 and find that the average return to college attendance for people who attend college only because the cost is low is between 29% and 78%.

My approach requires the treatment variable and the imperfect instrument to be discrete. The next step of this research will be the extension of the methodology to continuous treatments and instruments. This is left for future research.

Appendix A. Proofs

A.1. **Proof of Proposition 1. Validity of the bounds**: Consider the mixture model given by Equation (3.1)

$$F(y|d,w) \quad = \quad F_d(y|1,0)\alpha^d(w) + F_d(y|0,d)(1-\alpha^d(w)). \tag{A.1}$$

We have

$$F(y|d,A_1^d) - F(y|d,A_0^d) = \Big[\alpha^d(A_1^d) - \alpha^d(A_0^d)\Big]\Big[F_d(y|1,0) - F_d(y|0,d)\Big]. \tag{A.2}$$

Therefore, under Assumption 3, we can write

$$F(y|1,0) = F_d(y|0,d) + \frac{1}{\alpha^d(A_1^d) - \alpha^d(A_0^d)}\Big[F(y|d,A_1^d) - F(y|d,A_0^d)\Big],$$

which together with (A.1) imply

$$F(y|d,A_0^d) = F_d(y|0,d) + \frac{\alpha^d(A_0^d)}{\alpha^d(A_1^d) - \alpha^d(A_0^d)}\Big[F(y|d,A_1^d) - F(y|d,A_0^d)\Big].$$

Hence,

$$F_d(y|0,d) \quad = \quad F(y|d,A_0^d) - \eta^d\Big[F(y|d,A_1^d) - F(y|d,A_0^d)\Big], \tag{A.3}$$

$$F_d(y|1,0) \quad = \quad F(y|d,A_0^d) + (\theta^d - \eta^d)\Big[F(y|d,A_1^d) - F(y|d,A_0^d)\Big], \tag{A.4}$$

where $\theta^d = \frac{1}{\alpha^d(A_1^d)-\alpha^d(A_0^d)}$ and $\eta^d = \frac{\alpha^d(A_0^d)}{\alpha^d(A_1^d)-\alpha^d(A_0^d)}$. From there, it is straightforward that (2) holds, i.e., if $F(y|d,A_1^d) - F(y|d,A_0^d)$ for all $y \in \mathcal{Y}$, then $F_d(y|0,d) = F_d(y|1,0) = F(y|d,A_0^d)$.

Suppose now that there exists y^d such that $F(y^d|d,A_1^d) - F(y^d|d,A_0^d) \neq 0$. Then

$$\frac{F(y^d|d,w) - F(y^d|d,A_0^d)}{F(y^d|d,A_1^d) - F(y^d|d,A_0^d)} = \frac{\alpha^d(A_1^d) - \alpha^d(A_0^d)}{\alpha^d(w) - \alpha^d(A_0^d} = \theta^d\alpha^d(w) - \eta^d \equiv \Lambda^d(w).$$

Therefore,

$$\alpha^d(w) = \frac{1}{\theta^d}\Big(\Lambda^d(w) + \eta^d\Big), \tag{A.5}$$

where

$$\Lambda^d(w) = \frac{F(y^d|d,w) - F(y^d|d,A_0^d)}{F(y^d|d,A_1^d) - F(y^d|d,A_0^d)}.$$

For Equations (A.3) and (A.4) to represent cumulative distribution functions, θ^d and η^d must satisfy the monotonicity condition of a distribution function, i.e.,

$$f(y|d,A_0^d) - \eta^d\Big[f(y|d,A_1^d) - f(y|d,A_0^d)\Big] \geq 0, \tag{A.6}$$

$$f(y|d,A_0^d) + \Big(\theta^d - \eta^d\Big)\Big[f(y|d,A_1^d) - f(y|d,A_0^d)\Big] \geq 0. \tag{A.7}$$

Also, for Equation (A.5) to represent a probability weight, θ^d and η^d must satisfy the non-negativity condition and the condition that a probability is no greater than 1, i.e.,

$$0 \le \frac{1}{\theta^d}\Big(\eta^d + \Lambda^d(w)\Big) \le 1. \tag{A.8}$$

Thus, the bounds are valid. It remains to show that they are sharp.

Sharpness: Given the above constraints A.6, A.7 and A.8, we need to find for each $(\theta^0, \theta^1, \eta^0, \eta^1)$ a joint distribution on $(Y_0, Y_1, \delta, V, Z, W)$ that generates a joint distribution on the data (Y, D, W) through the model (2.1) and satisfies Assumptions 1, 2 and 3. Define $F_d(y|1,0)$, $F_d(y|0,d))$ and $\alpha^d(w)$ as in (A.3), (A.4) and (A.5), respectively. I propose the following conditional distribution for (δ, V, Z) given $W = w$

$$\begin{aligned}
\mathbb{P}(\delta = 1, V = 0, Z = 1|w) &\equiv \alpha^1(w)\mathbb{P}(D = 1|W = w),\\
\mathbb{P}(\delta = 1, V = 0, Z = 0|w) &\equiv \alpha^0(w)\mathbb{P}(D = 0|W = w),\\
\mathbb{P}(\delta = 0, V = 1, Z = 0|w) &\equiv 0.25(1 - \alpha^1(w))\mathbb{P}(D = 1|W = w),\\
\mathbb{P}(\delta = 0, V = 1, Z = 1|w) &\equiv 0.75(1 - \alpha^1(w))\mathbb{P}(D = 1|W = w),\\
\mathbb{P}(\delta = 0, V = 0, Z = 1|w) &\equiv 0.25(1 - \alpha^0(w))\mathbb{P}(D = 0|W = w),\\
\mathbb{P}(\delta = 0, V = 0, Z = 0|w) &\equiv 0.75(1 - \alpha^0(w))\mathbb{P}(D = 0|W = w),
\end{aligned}$$

and the following joint distribution for (Y_0, Y_1, δ, V, Z) conditional on $W = w$

$$\begin{aligned}
\mathbb{P}(Y_0 \le y_0, Y_1 \le y_1, \delta = 1, V = 0, Z = z|w) &\equiv F_0(y_0|1,0)F_1(y_1|1,0)\mathbb{P}(\delta = 1, V = 0, Z = z|w),\\
\mathbb{P}(Y_0 \le y_0, Y_1 \le y_1, \delta = 0, V = 1, Z = z|w) &\equiv F_0(y_0|0,1)F_1(y_1|0,1)\mathbb{P}(\delta = 0, V = 1, Z = z|w),\\
\mathbb{P}(Y_0 \le y_0, Y_1 \le y_1, \delta = 0, V = 0, Z = z|w) &\equiv F_0(y_0|0,0)F_1(y_1|0,0)\mathbb{P}(\delta = 0, V = 0, Z = z|w).
\end{aligned}$$

Assumption 2 holds by construction. We can check that Assumption 1 holds. For instance,

$$\begin{aligned}
\mathbb{P}(Y_1 \le y_1|\delta = 1, V = 0, Z = z, W = w) &= \frac{\mathbb{P}(Y_1 \le y_1, \delta = 1, V = 0, Z = z|W = w)}{\mathbb{P}(\delta = 1, V = 0, Z = z|W = w)},\\
&= \frac{\lim_{y_0\uparrow\infty} \mathbb{P}(Y_0 \le y_0, Y_1 \le y_1, \delta = 1, V = 0, Z = z|W = w)}{\mathbb{P}(\delta = 1, V = 0, Z = z|W = w)},\\
&= \frac{F_1(y_1|\delta = 1, V = 0)\mathbb{P}(\delta = 1, V = 0, Z = z|w)}{\mathbb{P}(\delta = 1, V = 0, Z = z|w)},\\
&= F_1(y_1|\delta = 1, V = 0).
\end{aligned}$$

The reasoning is similar for all other cases. This completes the proof.

A.2. **Inference.** Consider the following conditions that θ^d and η^d need to satisfy:

$$f(y|d, A_0^d) - \eta^d \Big[f(y|d, A_1^d) - f(y|d, A_0^d)\Big] \geq 0, \tag{A.9}$$

$$f(y|d, A_0^d) + \Big(\theta^d - \eta^d\Big) \Big[f(y|d, A_1^d) - f(y|d, A_0^d)\Big] \geq 0, \tag{A.10}$$

$$0 \leq \frac{1}{\theta^d} \Big(\eta^d + \Lambda^d(w)\Big) \leq 1. \tag{A.11}$$

By Bayes' rule, we have

$$f(y|d, A_\ell^d) = \frac{\mathbb{P}(D = d, W \in A_\ell^d | Y = y) f(y)}{\mathbb{P}(D = d, W \in A_\ell^d)}, \qquad \ell \in \{0, 1\},$$

where $f(y)$ is the probability density function of Y. Thus, for all y such that $f(y) > 0$, we can rewrite the first two inequalities (A.9) and (A.10) as

$$\frac{\mathbb{P}(D = d, W \in A_0^d | Y = y)}{\mathbb{P}(D = d, W \in A_0^d)} - \eta^d \Big[\frac{\mathbb{P}(D = d, W \in A_1^d | Y = y)}{\mathbb{P}(D = d, W \in A_1^d)} - \frac{\mathbb{P}(D = d, W \in A_0^d | Y = y)}{\mathbb{P}(D = d, W \in A_0^d)}\Big] \geq 0,$$

$$\frac{\mathbb{P}(D = d, W \in A_0^d | Y = y)}{\mathbb{P}(D = d, W \in A_0^d)} + (\theta^d - \eta^d) \Big[\frac{\mathbb{P}(D = d, W \in A_1^d | Y = y)}{\mathbb{P}(D = d, W \in A_1^d)} - \frac{\mathbb{P}(D = d, W \in A_0^d | Y = y)}{\mathbb{P}(D = d, W \in A_0^d)}\Big] \geq 0,$$

which are respectively equivalent to

$$\mathbb{E}\Big[1\{D = d, W \in A_0^d\} c_1 - \eta^d (1\{D = d, W \in A_1^d\} c_0 - 1\{D = d, W \in A_0^d\} c_1) | Y = y\Big] \geq 0,$$

$$\mathbb{E}\Big[1\{D = d, W \in A_0^d\} c_1 + (\theta^d - \eta^d)(1\{D = d, W \in A_1^d\} c_0 - 1\{D = d, W \in A_0^d\} c_1) | Y = y\Big] \geq 0,$$

where $c_0 = \mathbb{P}(D = d, W \in A_0^d)$ and $c_1 = \mathbb{P}(D = d, W \in A_1^d)$.

By multiplying the last condition (A.11) by $\theta^d sign(\theta^d)$, it can be written as

$$\begin{cases} sign(\theta^d) \left(\eta^d + \Lambda^d(w)\right) & \geq & 0 \\ sign(\theta^d) \left(\theta^d - \eta^d - \Lambda^d(w)\right) & \geq & 0 \end{cases}$$

where $sign(\theta^d) = 1\{\theta^d > 0\} - 1\{\theta^d < 0\}$, which can equivalently be rewritten as

$$\begin{cases} \mathbb{E}\Big[sign(\theta^d) \left(\eta^d + 1\{Y \leq y^d\} k_1 - k_2 | D = d, W = w\right)\Big] & \geq & 0 \\ \mathbb{E}\Big[sign(\theta^d) \left(\theta^d - \eta^d - 1\{Y \leq y^d\} k_1 + k_2 | D = d, W = w\right)\Big] & \geq & 0 \end{cases}$$

where $k_1 = \frac{1}{F(y^d|d, A_1^d) - F(y^d|d, A_0^d)}$, and $k_2 = \frac{F(y^d|d, A_0^d)}{F(y^d|d, A_1^d) - F(y^d|d, A_0^d)}$.

To summarize, we have the following conditional moment inequalities:

$$\mathbb{E}\Big[1\{D=d,W\in A_0^d\}c_1-\eta^d(1\{D=d,W\in A_1^d\}c_0-1\{D=d,W\in A_0^d\}c_1)|Y=y\Big]\geq 0,$$
$$\mathbb{E}\Big[1\{D=d,W\in A_0^d\}c_1+(\theta^d-\eta^d)(1\{D=d,W\in A_1^d\}c_0-1\{D=d,W\in A_0^d\}c_1)|Y=y\Big]\geq 0,$$
$$\mathbb{E}\Big[sign(\theta^d)\left(\eta^d+1\{Y\leq y^d\}k_1-k_2|D=d,W=w\right)\Big]\geq 0$$
$$\mathbb{E}\Big[sign(\theta^d)\left(\theta^d-\eta^d-1\{Y\leq y^d\}k_1+k_2|D=d,W=w\right)\Big]\geq 0.$$

Implementation. In my empirical illustration, Y=lwage, D=college, W=tuit17, and Z=indicator of low cost (unobserved). I use the clrbound command of Chernozhukov et al. (2015) with the local linear method. I set $A_1^d=\mathbb{1}\{W\leq F_W^{-1}(0.5)\}$ and $A_0^d=\mathbb{1}\{W>F_W^{-1}(0.5)\}$ for each $d=0$ and $d=1$, where $F_W^{-1}(\alpha)$ is the α-th quantile of W. I replace c_0, c_1, k_1, k_2 by their sample analogs. Mourifié and Wan (2017) showed the validity of this plug-in approach in the intersection bounds context. Let

$$m(D,W;\theta^d,\eta^d|1,0)=1\{D=d,W\in A_0^d\}c_1-\eta^d(1\{D=d,W\in A_1^d\}c_0-1\{D=d,W\in A_0^d\}c_1.$$

I do a grid search of (θ^d,η^d) over $[-M,M]\times[-L,L]$, where M and L are arbitrarily large. After applying the clrbound command on the inequalities above, I keep the values of (θ^d,η^d) for which the test is not rejected. For each such (θ^d,η^d), I obtain the estimate $\hat{m}(y;\theta^d,\eta^d|1,0)$ of $\mathbb{E}[m(D,W;\theta^d,\eta^d|1,0)|Y=y]$, its standard error $\hat{s}(y;\theta^d,\eta^d|1,0)$ and the critical value $k_{0.95}$. From there, I get the estimate of the density $f_d(y;\theta^d,\eta^d|1,0)$:

$$\hat{f}_d^{0.95}(y;\theta^d,\eta^d|1,0)=\left[\hat{m}(y;\theta^d,\eta^d|1,0)+k_{0.95}\hat{s}(y;\theta^d,\eta^d|1,0)\right]\hat{f}(y),$$

where $\hat{f}(y)=\frac{1}{nh}\sum_{i=1}^n K(\frac{y-Y_i}{h})$, $K(u)=\frac{3}{4\sqrt{5}}(1-\frac{1}{5}u^2)\mathbb{1}\left\{|u|\leq\sqrt{5}\right\}$, $h=n^{-1/5}\left[0.9\min(\sigma_Y,\frac{Q_3-Q_1}{1.349})\right]$, σ_Y, Q_1, Q_3 are empirical standard deviation, first and third quartiles of Y, respectively.

A.3. **Proof of Proposition 2.** I adapt the proof in Kédagni (2018) to the current setting.

Proof. Under Assumption 1, Equation (A.1) holds and we have for $d=1$,

$$1-F(y|1,w)=\alpha^1(w)\left[1-F_1(y|1,0)\right]+(1-\alpha^1(w))\left[1-F_1(y|0,1)\right].$$

Under Assumption 3, at least one of the weights $\alpha^1(A_1^1)$ and $\alpha^1(A_0^1)$ is different from 1. Assume without loss of generality that $\alpha^1(A_0^1)\neq 1$. Then

$$\lim_{y\uparrow y^u}\frac{1-F(y|1,A_1^1)}{1-F(y|1,A_0^1)}=\lim_{y\uparrow y^u}\frac{\alpha^1(A_1^1)\frac{1-F_1(y|1,0)}{1-F_1(y|0,1)}+1-\alpha^1(A_1^1)}{\alpha^1(A_0^1)\frac{1-F_1(y|1,0)}{1-F_1(y|0,1)}+1-\alpha^1(A_0^1)}=\frac{1-\alpha^1(A_1^1)}{1-\alpha^1(A_0^1)}\equiv\pi^1(A_1^1,A_0^1),$$

where the second equality holds under Assumption 4.

Under Assumption 3, we have $\frac{1}{1-\pi^1(A_1^1,A_0^1)} = \frac{1-\alpha^1(A_0^1)}{\alpha^1(A_1^1)-\alpha^1(A_0^1)}$. Then

$$\begin{aligned}
\frac{1}{1-\pi^1(A_1^1,A_0^1)}\left[F(y|1,A_1^1)-F(y|1,A_0^1)\right] &= \frac{\alpha^1(A_1^1)-\alpha^1(A_0^1)}{1-\pi^1(A_1^1,A_0^1)}\left[F_1(y|1,0)-F_1(y|0,1)\right], \\
&= (1-\alpha^1(A_0^1))\left[F_1(y|1,0)-F_1(y|0,1)\right], \\
&= F_1(y|1,0)-F(y|1,A_0^1),
\end{aligned}$$

where the first equality follows from Equation (A.2), the second from the above equality, and the last holds from (A.1). Thus,

$$F_1(y|1,0) = F(y|1,A_0^1) + \frac{1}{1-\pi^1(A_1^1,A_0^1)}\left[F(y|1,A_1^1)-F(y|1,A_0^1)\right].$$

The reasoning is similar for $F_0(y|1,0)$. This completes the proof. □

Appendix B. Relaxing Monotonicity

Under Assumption 1, we can write the identified distribution $F(y|1,w)$ as a mixture of the potential outcome distributions of Y_1 for the compliers, the defiers and the always-takers as follows:

$$F(y|1,w) = \lambda_1^1(w)F_1(y|1,0) + \lambda_2^1(w)F_1(y|-1,1) + (1-\lambda_1^1(w)-\lambda_2^1(w))F_1(y|0,1). \quad \text{(B.1)}$$

Similarly, we have

$$F(y|0,w) = \lambda_1^0(w)F_0(y|1,0) + \lambda_2^0(w)F_0(y|-1,1) + (1-\lambda_1^0(w)-\lambda_2^0(w))F_0(y|0,1). \quad \text{(B.2)}$$

Following the results of Henry, Kitamura, and Salanié (2014) on finite mixture models, we can show under some relevance condition that the potential outcome distributions of Y_1 and Y_0 are partially identified for the compliers and the defiers, respectively. We can show that each of these distributions is function of six parameters that are partially identified. From there, we can obtain bounds on the local average / quantile treatment effects for the compliers and the defiers. To the best of my knowledge, this result is new in the literature.

Below, I derive the results for Equation (B.1). Results for Equation (B.2) can be obtained in a similar way. Equation (B.1) implies

$$\begin{aligned}
F(y|1,w)-F(y|1,A_0^1) &= [\lambda_1^1(w)-\lambda_1^1(A_0^1)][F_1(y|1,0)-F_1(y|0,1)] \\
&\quad +[\lambda_2^1(w)-\lambda_2^1(A_0^1)][F_1(y|-1,1)-F_1(y|0,1)], \\
&= \boldsymbol{\psi}(w)^t\boldsymbol{m}(y),
\end{aligned}$$

where

$$\boldsymbol{\psi}(w) \equiv \begin{bmatrix} \lambda_1^1(w) - \lambda_1^1(A_0^1) \\ \lambda_2^1(w) - \lambda_2^1(A_0^1) \end{bmatrix}, \quad \text{and} \quad \boldsymbol{m}(y) \equiv \begin{bmatrix} F_1(y|1,0) - F_1(y|0,1) \\ F_1(y|-1,1) - F_1(y|0,1) \end{bmatrix}.$$

I now state a relevance assumption similar to Assumption 3 in the main text.

Assumption 5. *There exist three subsets $A_0^1,\ A_1^1,\ A_2^1 \subset \mathcal{W}$ such that the matrix*

$$\boldsymbol{\Psi} \equiv \begin{bmatrix} \lambda_1^1(A_1^1) - \lambda_1^1(A_0^1) & \lambda_1^1(A_2^1) - \lambda_1^1(A_0^1) \\ \lambda_2^1(A_1^1) - \lambda_2^1(A_0^1) & \lambda_2^1(A_2^1) - \lambda_2^1(A_0^1) \end{bmatrix}$$

is invertible, i.e., the determinant $|\boldsymbol{\Psi}| = [\lambda_1^1(A_1^1) - \lambda_1^1(A_0^1)] * [\lambda_2^1(A_2^1) - \lambda_2^1(A_0^1)] - [\lambda_1^1(A_2^1) - \lambda_1^1(A_0^1)] * [\lambda_2^1(A_1^1) - \lambda_2^1(A_0^1)] \neq 0$.

Denote $\boldsymbol{h}(y) \equiv \begin{bmatrix} F_1(y|1,0) - F_1(y|0,1) \\ F_1(y|-1,1) - F_1(y|0,1) \end{bmatrix}$. We have $\boldsymbol{h}(y) = \boldsymbol{\Psi}^t \boldsymbol{m}(y)$, which implies under Assumption 5 that $\boldsymbol{m}(y) = (\boldsymbol{\Psi}^t)^{-1}\boldsymbol{h}(y)$. By rewriting this last equality, we obtain:

$$\begin{aligned} F_1(y|1,0) &= F_1(y|0,1) + \gamma_1^1[F(y|1,A_1^1) - F(y|1,A_0^1)] - \gamma_3^1[F(y|1,A_2^1) - F(y|1,A_0^1)], \\ F_1(y|-1,1) &= F_1(y|0,1) - \gamma_2^1[F(y|1,A_1^1) - F(y|1,A_0^1)] + \gamma_4^1[F(y|1,A_2^1) - F(y|1,A_0^1)], \end{aligned}$$

where

$$\gamma_1^1 = \frac{\lambda_2^1(A_2^1) - \lambda_2^1(A_0^1)}{|\boldsymbol{\Psi}|}, \qquad \gamma_2^1 = \frac{\lambda_1^1(A_2^1) - \lambda_1^1(A_0^1)}{|\boldsymbol{\Psi}|},$$
$$\gamma_3^1 = \frac{\lambda_2^1(A_1^1) - \lambda_2^1(A_0^1)}{|\boldsymbol{\Psi}|}, \qquad \gamma_4^1 = \frac{\lambda_1^1(A_1^1) - \lambda_1^1(A_0^1)}{|\boldsymbol{\Psi}|}.$$

From Equation (B.1), we have

$$F(y|1,A_0^1) = \rho_1^1 F_1(y|1,0) + \rho_2^1 F_1(y|-1,1) + (1 - \rho_1^1 - \rho_2^1) F_1(y|0,1), \tag{B.3}$$

where $\rho_1^1 = \lambda_1^1(A_0^1)$ and $\rho_2^1 = \lambda_2^1(A_0^1)$. I can therefore plug the expressions for $F_1(y|1,0)$ and $F_1(y|-1,1)$ in (B.3), and obtain

$$\begin{aligned} F(y|1,A_0^1) &= F_1(y|1,0) + (\rho_1^1\gamma_1^1 - \rho_2^1\gamma_2^1)[F(y|1,A_1^1) - F(y|1,A^0)] \\ &\quad + (\rho_2^1\gamma_4^1 - \rho_1^1\gamma_3^1)[F(y|1,A_1^1) - F(y|1,A_0^1)]. \end{aligned}$$

From this last equation, we have

$$\begin{aligned} F_1(y|1,0) &= F(y|1,A_0^1) - (\rho_1^1\gamma_1^1 - \rho_2^1\gamma_2^1)[F(y|1,A_1^1) - F(y|1,A^0)] \\ &\quad - (\rho_2^1\gamma_4^1 - \rho_1^1\gamma_3^1)[F(y|1,A_2^1) - F(y|1,A_0^1)]. \end{aligned} \tag{B.4}$$

Thus,

$$\begin{aligned} F_1(y|0,1) &= F(y|1,A_0^1) + (\gamma_1^1 - \rho_1^1\gamma_1^1 + \rho_2^1\gamma_2^1)[F(y|1,A_1^1) - F(y|1,A^0)] \\ &\quad -(\gamma_3^1 - \rho_1^1\gamma_3^1 + \rho_2^1\gamma_4^1)[F(y|1,A_2^1) - F(y|1,A_0^1)]. \end{aligned} \tag{B.5}$$

$$\begin{aligned} F_1(y|-1,1) &= F(y|1,A_0^1) - (\gamma_2^1 + \rho_1^1\gamma_1^1 - \rho_2^1\gamma_2^1)[F(y|1,A_1^1) - F(y|1,A^0)] \\ &\quad +(\gamma_4^1 - \rho_2^1\gamma_4^1 + \rho_1^1\gamma_3^1)[F(y|1,A_2^1) - F(y|1,A_0^1)]. \end{aligned} \tag{B.6}$$

I have just shown that the potential outcome distributions of Y_1 are identified up to 6 scalar parameters. I now write the probability weights $\lambda_1^1(w)$ and $\lambda_2^1(w)$ as functions of those parameters. From there, using the monotonicity condition for cumulative distribution functions and the condition that probability weights lie between 0 and 1, I partially identify those 6 parameters.

For $y_1^1,\ y_2^1 \in \mathcal{Y}$, we have

$$\begin{aligned} \begin{bmatrix} F(y_1^1|1,w) - F(y_1^1|1,A_0^1) \\ F(y_2^1|1,w) - F(y_2^1|1,A_0^1) \end{bmatrix} &= \begin{bmatrix} F_1(y_1^1|1,0) - F_1(y_1^1|0,1) & F_1(y_1^1|-1,1) - F_1(y_1^1|0,1) \\ F_1(y_2^1|1,0) - F_1(y_2^1|0,1) & F_1(y_2^1|-1,1) - F_1(y_2^1|0,1) \end{bmatrix} \\ &\quad * \begin{bmatrix} \lambda_1^1(w) - \rho_1^1 \\ \lambda_2^1(w) - \rho_2^1 \end{bmatrix} \end{aligned}$$

If the matrix $\begin{bmatrix} F_1(y_1^1|1,0) - F_1(y_1^1|0,1) & F_1(y_1^1|-1,1) - F_1(y_1^1|0,1) \\ F_1(y_2^1|1,0) - F_1(y_2^1|0,1) & F_1(y_2^1|-1,1) - F_1(y_2^1|0,1) \end{bmatrix}$ is invertible then we can solve for $\lambda_1^1(w)$ and $\lambda_2^1(w)$. This condition holds if and only if $\boldsymbol{\Psi}$ and $\boldsymbol{H}$ are invertible, where

$$\boldsymbol{H} \equiv \begin{bmatrix} F(y_1^1|1,A_1^1) - F(y_1^1|1,A_0^1) & F(y_2^1|1,A_1^1) - F(y_2^1|1,A_0^1) \\ F(y_1^1|1,A_2^1) - F(y_2^1|1,A_0^1) & F(y_2^1|1,A_2^1) - F(y_2^1|1,A_0^1) \end{bmatrix}.$$

By adapting the proof of Proposition 1, we can show that the identified set for the 6 parameters is sharp.

References

Andrews, D. W. K., W. Kim, and X. Shi. 2016. "Stata Commands for Testing Conditional Moment Inequalities/Equalities." *Unpublished manuscript* .

Andrews, D. W. K. and X. Shi. 2013. "Inference Based on Conditional Moment Inequalities." *Econometrica* 81:609–666.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434):444–455.

Card, David. 2001. "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems." *Econometrica* 69:1127–1160.

Chalak, K. 2017. "Instrumental Variables Methods with Heterogeneity and Mismeasured Instruments." *Econometric Theory* 33:69—-104.

Chernozhukov, Victor, Wooyoung Kim, Sokbae Lee, and Adam M. Rosen. 2015. "Implementing Intersection Bounds in Stata." *Stata Journal* 15 (1):21–44.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica* 81 (2):667–737. URL http://dx.doi.org/10.3982/ECTA8718.

Heckman, James J., Justin L. Tobias, and Edward J. Vytlacil. 2001. "Four Parameters of Interest in the Evaluation of Social Programs." *Southern Economic Journal* 68 (2):210—-23.

Henry, M., Y. Kitamura, and B. Salanié. 2014. "Partial Identification of Finite Mixtures Econometric Models." *Quantitative Economics* 5:123–144.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475. URL http://www.jstor.org/stable/2951620.

Jochmans, Koen, Marc Henry, and Bernard Salanié. 2017. "Inference on Two Component Mixtures under Tail Restrictions." *Econometric Theory* 33:610–635.

Kédagni, D. 2018. "Identifying Treatment Effects in the Presence of Confounded Types." *Economics Working Papers, Iowa State University* 18014.

Kédagni, D. and I. Mourifié. 2015. "Generalized Instrumental Inequalities: Testing the IV Independence Assumption." *Unpublished manuscript* .

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe. 2012. "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of American Statistical Association* 107:958–975.

Lewbel, Arthur. 2007. "Estimation of Average Treatment Effects with Misclassification." *Econometrica* 75 (2):537–551.

Lubotsky, D. and M. Wittenberg. 2006. "Interpretation of Regressions with Multiple Proxies." *The Review of Economics and Statistics* 88 (3):549–562.

Mahajan, Aprajit. 2006. "Identification and Estimation of Regression Models with Misclassification." *Econometrica* 74 (3):631–665.

Mourifié, I. and Y. Wan. 2017. "Testing Local Average Treatment Effect Assumptions." *The Review of Economics and Statistics* 99 (2):305–313.

Ura, Takuya. 2018. "Heterogeneous Treatment Effects with Mismeasured Endogenous Treatment." *Quantitative Economics* 9 (3):1335–1370.