Identification of Treatment Effects with Mismeasured Imperfect Instruments

Désiré Kédagni

Iowa State University
Department of Economics

ASSA 2020 Conference - Econometric Society
San Diego

Analytical Framework

- Consider this IV model

$$\begin{cases} Y = Y_1 D + Y_0(1 - D) \\ D = D_1 Z + D_0(1 - Z) \\ W = \varphi(Z, \epsilon) \end{cases} \tag{1}$$

  - $Y \in \mathcal{Y} \subset \mathbb{R}$, $D \in \{0, 1\}$, $W \in \mathcal{W}$ are observed data;
  - Y_0 and Y_1 are potential outcomes, D_0 and D_1 are potential treatments;
  - $Z \in \{0, 1\}$ is unobserved.
- Application: returns to college
  - Y is earnings, D is a college degree (at least 16 years of schooling);
  - Z is an indicator for low college cost (depends on financial cost, opportunity cost, psychological cost);
  - W is college proximity. It can be seen as a proxy for Z (Card 1995, 2001).

Analytical Framework

- The variables D and Z partition the population into 4 unobserved groups: *types* (Angrist, Imbens and Rubin, 1996) or *strata* (Frangakis and Rubin, 2002).
  - $D_0 = D_1 = 1$: always-takers (a) $D_0 = D_1 = 0$: never-takers (n)
  - $D_0 = 0,\ D_1 = 1$: compliers (c) $D_0 = 1,\ D_1 = 1$: defiers (df)
- Let $T \in \{a, c, n, df\}$ denote the random type of an individual.
- Main Assumption

$$(Z, W) \perp\!\!\!\perp Y_d | T$$

State of the art

- LATE Assumptions
  - **Selection on Types (ST)**: $Z \perp\!\!\!\perp Y_d | T$ for all $d \in \{0, 1\}$.
  - **Unconfounded Type (UT)**: $Z \perp\!\!\!\perp T$.
  - **Monotonicity (M)**: No-defiers, i.e., $T \in \{a, c, n\}$.
- Under ST, UT and M, the standard IV estimand identifies $\mathbb{E}[Y_1 - Y_0 | T = c]$ when Z is observed, which the literature calls **local average treatment effect (LATE)**. See Imbens and Angrist (1994).
- In my framework, the instrument Z violates UT and is mismeasured.

Contribution

In this paper, I allow for ***confounded types*** and ***mismeasured*** instruments.

- I show that with the help of a proxy for the instrument, the potential outcome distributions are partially identified for the compliers.
  - Under some tail restrictions, these distributions are point-identified.
- I provide an easy-to-implement inference procedure.
- I illustrate my methodology on the NLSYM data and find that getting a college degree increases the average hourly wage by **17 – 35%** for the compliers.
  - I use college proximity as a proxy for low college cost.

Identification

Assumption (Selection on Types: ST)

There exists W s.t. $(Z, W) \perp\!\!\!\perp Y_d | T$ for each $d \in \{0, 1\}$.

Assumption (Monotonicity: M)

There exist no defiers, i.e., $T \in \{a, c, n\}$.

Notation

$\alpha^d(w) \equiv \mathbb{P}(T = c | D = d, W = w)$, $F(y|d, w) \equiv \mathbb{P}(Y \leq y | D = d, W = w)$, $F_{1a}(y) \equiv \mathbb{P}(Y_1 \leq y | T = a)$, *and* $F_{1c}(y) \equiv \mathbb{P}(Y_1 \leq y | T = c)$.

Identification

Under Assumptions ST and M, we have the following mixture models

$$F(y|1,w) = \alpha^1(w)F_{1c}(y) + (1-\alpha^1(w))F_{1a}(y),$$

and

$$F(y|0,w) = \alpha^0(w)F_{0c}(y) + (1-\alpha^0(w))F_{0n}(y).$$

Identification

By differencing $F(y|1,w)$ w.r.t. w, we can write

$$\underbrace{F(y|1,1) - F(y|1,0)}_{\text{identified from data}} = \overbrace{\left(\alpha^1(1) - \alpha^1(0)\right)}^{\text{within group difference}} \underbrace{\left[F_{1c}(y) - F_{1a}(y)\right]}_{\text{between group difference}},$$

which implies under the assumption that $\alpha^1(1) \neq \alpha^1(0)$ that

$$F_{1c}(y) = F_{1a}(c) + \frac{1}{\alpha^1(1) - \alpha^1(0)} \left[F(y|1,1) - F(y|1,0)\right].$$

Identification

After some manipulations, we obtain that

$$\begin{aligned} F_{1a}(y) &= F(y|1,0) - \delta^1 \left[F(y|1,1) - F(y|1,0) \right], \\ F_{1c}(y) &= F(y|1,0) + \left(\gamma^1 - \delta^1 \right) \left[F(y|1,1) - F(y|1,0) \right], \\ \alpha^1(w) &= \frac{1}{\gamma^1} \left(\delta^1 + \Delta^1(w) \right), \end{aligned} \tag{2}$$

where

$$\Delta^1(w) = \frac{F(y^1|1,w) - F(y^1|1,0)}{F(y^1|1,1) - F(y^1|1,0)}$$

for some $y^1 \in \mathcal{Y}$.

Identification

Assumption (Relevance: REL)

There exist w_0^1 *and* w_1^1 *such that* $\alpha^1(w_0^1) \neq \alpha^1(w_1^1)$.

Theorem

Under Assumptions ST, M and REL, the distribution of Y_1 *is set-identified for the always-takers and compliers:*

$$
\begin{aligned}
F_{1a}(y) &= F(y|1,0) - \delta^1 \left[F(y|1,1) - F(y|1,0)\right], \\
F_{1c}(y) &= F(y|1,0) + \left(\gamma^1 - \delta^1\right) \left[F(y|1,1) - F(y|1,0)\right].
\end{aligned}
$$

Moreover, $\theta^1 \equiv \left(\gamma^1, \delta^1\right)$ *is set-identified:* $\theta^1 \in \Theta^1$. *The set* Θ^1 *is sharp.*

Sharp bounds on LATE

- $\mu_{dc}^{\theta^d}$ the expectation of Y_d for compliers for a given value of θ^d.

Proposition

Under Assumptions CI, M and REL, the LATE is set-identified:

$$\inf_{\theta^1 \in \Theta^1} \mu_{1c}^{\theta^1} - \sup_{\theta^0 \in \Theta^0} \mu_{0c}^{\theta^0} \leq \mathbb{E}\left[Y_1 - Y_0 | T = c\right] \leq \sup_{\theta^1 \in \Theta^1} \mu_{1c}^{\theta^1} - \inf_{\theta^0 \in \Theta^0} \mu_{0c}^{\theta^0}.$$

These bounds are sharp.

Point-identification

Assumption (TR)

$\lim_{y \downarrow y^{\ell}} \frac{F_{0c}(y)}{F_{0n}(y)} = 0$ *and* $\lim_{y \uparrow y^{u}} \frac{1-F_{1c}(y)}{1-F_{1a}(y)} = 0.$

Proposition

Under Assumptions ST, M, REL and TR, the distributions $F_1(y|1,0)$ *and* $F_0(y|1,0)$ *are point-identified as follows:*

$$
\begin{aligned}
F_{0c}(y) &= F(y|0,w_0^0) + \frac{1}{1-\zeta^0(w_1^0,w_0^1)}\left[F(y|0,w_1^0) - F(y|0,w_0^0)\right], \\
F_{1c}(y) &= F(y|1,w_0^1) + \frac{1}{1-\pi^1(w_1^1,w_0^1)}\left[F(y|1,w_1^1) - F(y|1,w_0^1)\right],
\end{aligned}
$$

where

$$
\zeta^0(w_1^0,w_0^0) = \lim_{y \downarrow y^{\ell}} \frac{F(y|0,w_1^0)}{F(y|0,w_0^0)}, \quad \textit{and} \quad \pi^1(w_1^1,w_0^1) = \lim_{y \uparrow y^{u}} \frac{1-F(y|1,w_1^1)}{1-F(y|1,w_0^1)}.
$$

Inference

- The identified set for θ^1 is given by the following restrictions:

$$\inf_{(y,w)\in\mathcal{Y}\times\mathcal{W}} \beta^1(y,w;\theta^1) \geq 0, \tag{3}$$

where

$$\beta^1(y,w;\theta^1) = \begin{bmatrix} f(y|1,0) - \delta^1 \left[f(y|1,1) - f(y|1,0)\right] \\ f(y|1,0) + \left(\gamma^1 - \delta^1\right)\left[f(y|1,1) - f(y|1,0)\right] \\ \frac{1}{\gamma^1}\left(\delta^1 + \Delta^1(w)\right) \\ 1 - \frac{1}{\gamma^1}\left(\delta^1 + \Delta^1(w)\right) \end{bmatrix}.$$

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

Inference

- Assume that W is discrete.
- Let $f(y)$ denote the density (probability mass) function of Y. Using Bayes' rule, we have:

$$f(y|d,w) = \frac{\mathbb{P}(D=d, W=w|Y=y)f(y)}{\mathbb{P}(D=d, W=w)}.$$

- Then the first inequality becomes:

$$\frac{\mathbb{P}(D=1|Y=y)f(y)}{\mathbb{P}(D=1)} \\ -\delta^1 \left[\frac{\mathbb{P}(D=1, W=w_1^1|Y=y)f(y)}{\mathbb{P}(D=1, W=w_1^1)} - \frac{\mathbb{P}(D=1, W=w_0^1|Y=y)f(y)}{\mathbb{P}(D=1, W=w_0^1)} \right] \geq 0.$$

Inference

- Assume that W is discrete.
- Let $f(y)$ denote the density (probability mass) function of Y. Using Bayes' rule, we have:

$$f(y|d,w) = \frac{\mathbb{P}(D=d, W=w|Y=y)f(y)}{\mathbb{P}(D=d, W=w)}.$$

- Then the first inequality becomes:

$$\frac{\mathbb{P}(D=1|Y=y)\cancel{f(y)}}{\mathbb{P}(D=1)} - \delta^1 \left[\frac{\mathbb{P}(D=1, W=w_1^1|Y=y)\cancel{f(y)}}{\mathbb{P}(D=1, W=w_1^1)} - \frac{\mathbb{P}(D=1, W=w_0^1|Y=y)\cancel{f(y)}}{\mathbb{P}(D=1, W=w_0^1)} \right] \geq 0.$$

Inference

Theorem

The identified set for θ^1 is given by the following restrictions:

$$\begin{cases} \inf_{y \in \mathcal{Y}} \mathbb{E}[m_0^1(\theta^1, D, W) | Y = y] \geq 0 \\ \\ \inf_{w \in \mathcal{W}} \mathbb{E}[m_1^1(\theta^1, Y) | D = 1, W = w] \geq 0 \end{cases} \tag{4}$$

- So we have a standard conditional moment inequality model.
  - Use Chernozhukov, Kim, Lee and Rosen's (2015) or Andrews, Kim and Shi's (2016) stata packages.
- When W is continuous, replace $\{W = w_\ell^1\}$ by $\{W \in A_\ell^1\}$ where $\mathbb{P}(W \in A_\ell^1) > 0$ $(\ell = 0, 1)$.

Empirical illustration: returns to college

- NLSYM data: Card (1995).

Table 1: Summary statistics

	Total
Observations	3,010
log wage (in cents)	6.2618 (0.4438)
college degree	0.2714 (0.4448)
college proximity	0.6821 (0.4658)

Average and standard deviation (in the parentheses)

Empirical illustration: returns to college

Table 2: Confidence sets for parameters

Parameters	95% conf. LB	95% conf. UB
γ^1	-0.3	0.4
γ^0	-0.75	1.5
δ^1	-0.4	-0.1
δ^0	-0.9	-0.5
$\mathbb{E}[Y_1 \| T = c]$	6.3663	6.3953
$\mathbb{E}[Y_0 \| T = c]$	6.0960	6.2128
LATE	0.1534	0.2993
	17%	**35%**

conf.: confidence; LB: lower bound; UB: upper bound.

Summary

- This paper develops a new identification strategy when the LATE exogeneity assumption is violated and the instrument is mismeasured.
- I show that with the help of a proxy for the instrument, the potential outcome distributions are partially identified for the compliers.
    - Under some tail restrictions, these distributions are point-identfied.
- I apply the results to the NLSYM data and find that getting a college degree increases the average wage by 17 – 35% for the people who attend college only because they judge the cost low.

Thank you!!!