When Should We (Not) Interpret Linear IV Estimands as LATE?

Tymon Słoczyński

Brandeis University

ASSA 2022

Introduction

In economics, many policy variables of interest are potentially endogenous and, as a result, applied work often uses instrumental variables (IV) methods.

In this paper, I am primarily interested in a common situation where **(i)** both the endogenous variable ("treatment") and the instrument are binary; and **(ii)** the instrument is only valid conditional on additional covariates.

Selected applications:

- Angrist (1990) studies the effect of veteran status on earnings, with draft eligibility as instrument. This is only valid conditional on cohort.
- Card (1995) studies the effect of education on earnings, with college proximity as instrument. This is only valid conditional on appropriate student and/or parental characteristics.
- Stratified RCTs with varying assignment probability & noncompliance.

Introduction

In practice, many papers use the simplest linear model:

$$Y = D\beta + X\gamma + \upsilon, \tag{1}$$

where D is an endogenous treatment while X and the instrument(s) are assumed to be uncorrelated with υ. Also, β is the main coefficient of interest. Estimation is typically carried out using linear IV or 2SLS.

At the same time, few researchers would be willing to rule out the existence of treatment effect heterogeneity — so the model in equation (1) is implicitly treated as misspecified. **With treatment effect heterogeneity, how should we interpret the resulting IV and 2SLS estimands?**

IV=LATE?

With both D and the instrument, Z, being binary, it is common to interpret the IV estimand as **the local average treatment effect (LATE), i.e. the average treatment effect for individuals whose treatment status is affected by the instrument**. However, the usual references do not *really* support this interpretation:

- Imbens and Angrist (*ECTA* 1994) implicitly assume that additional covariates are irrelevant. This is often not the case.
- Angrist and Imbens (*JASA* 1995) allow for the possibility that Z is only valid after conditioning on X. However:
  - They also restrict their attention to saturated models with discrete covariates . . .
  - and implicitly require that the researcher estimates a separate first-stage coefficient on Z for every value of X.
  - In this case, the 2SLS estimand is **a convex combination of conditional LATEs** $\neq$ unconditional LATE.
  - "Heterogeneous" first stages are also generally absent from empirical work — see, e.g., Mogstad, Torgovitsky and Walters (2021) — so it is unclear how often this interpretation is applicable in practice.

Recent Extensions

Kolesár (2013) and Evdokimov and Kolesár (2019) relax many of the practical limitations of Angrist and Imbens (1995)'s result — and support the view that linear IV and 2SLS estimands can generally be written as a convex combination of conditional LATEs.

However, Evdokimov and Kolesár (2019) assume that the reduced-form and first-stage regressions are correctly specified, which is often implausible.

Kolesár (2013) allows for misspecification while also replacing Imbens and Angrist (1994)'s monotonicity assumption[1] with an assumption that I refer to as **weak monotonicity**: there are compliers but no defiers at some covariate values and defiers but no compliers elsewhere.

Still, Kolesár (2013) concludes that the interpretation of linear IV and 2SLS estimands as a convex combination of conditional LATEs is generally correct.

[1] There are no defiers, i.e. Z influences D in only one direction.

This Paper

In this paper I present **a more pessimistic view of the causal interpretability of linear IV and 2SLS estimands**.

- When IV is applied in the usual way, with a first stage that is separable in X and Z, the weights on some conditional LATEs are negative under Kolesár (2013)'s assumptions.
- Unlike in previous papers, I compare the weights in the usual application of IV and in Angrist and Imbens (1995)'s specification with the "desired" weights that recover the unconditional LATE parameter.
  - Angrist and Imbens (1995)'s specification guarantees to produce positive weights even under weak monotonicity.
  - Under strong monotonicity, Angrist and Imbens (1995)'s specification produces less desirable weights than the standard specification.
- I propose a new method, termed "reordered IV," which remains just identified but still promises no negative weights.
- The weights in the usual application of IV are problematic for interpretation even under strong monotonicity — unless the groups with different values of the instrument are roughly equal sized.

Notation

Y – outcome
D – binary treatment
Z – binary instrument
$X = (1, X_1, \ldots, X_J)$ – additional covariates

$Y = Y(D)$
$Y(1)$ and $Y(0)$ – potential outcomes

$D = D(Z)$
$D(1)$ and $D(0)$ – potential treatment statuses

If the observed outcome were to depend directly on Z, we would write $Y = Y(Z, D)$.

Sometimes, additional instruments will be created by interacting Z with all elements of X; then, $Z_{\mathrm{C}} = (Z, ZX_1, \ldots, ZX_J)$.

Estimands

The linear IV estimand:

$$\beta_{\text{IV}} = \left[\left(\mathrm{E}\left[Q'W\right]\right)^{-1}\mathrm{E}\left[Q'Y\right]\right]_1,$$

where $W = (D, X)$, $Q = (Z, X)$, and $[\cdot]_k$ denotes the kth element of the corresponding vector. $\Rightarrow$ **"Homogeneous" first stage**

The 2SLS estimand:

$$\beta_{\text{2SLS}} = \left[\left(\mathrm{E}\left[W'Q_{\text{C}}\right]\left(\mathrm{E}\left[Q_{\text{C}}'Q_{\text{C}}\right]\right)^{-1}\mathrm{E}\left[Q_{\text{C}}'W\right]\right)^{-1}\mathrm{E}\left[W'Q_{\text{C}}\right]\left(\mathrm{E}\left[Q_{\text{C}}'Q_{\text{C}}\right]\right)^{-1}\mathrm{E}\left[Q_{\text{C}}'Y\right]\right]_1,$$

where $Q_{\text{C}} = (Z_{\text{C}}, X)$. $\Rightarrow$ **"Heterogeneous" first stage**

Conditional Estimands

True first stage:

$$\mathrm{E}\left[D \mid X, Z\right] = \psi(X) + \omega(X) \cdot Z,$$

where

$$\omega(x) = \mathrm{E}\left[D \mid Z = 1, X = x\right] - \mathrm{E}\left[D \mid Z = 0, X = x\right]$$

is the conditional first-stage slope coefficient.

The conditional IV (or Wald) estimand:

$$\beta(x) = \frac{\mathrm{E}\left[Y \mid Z = 1, X = x\right] - \mathrm{E}\left[Y \mid Z = 0, X = x\right]}{\mathrm{E}\left[D \mid Z = 1, X = x\right] - \mathrm{E}\left[D \mid Z = 0, X = x\right]}.$$

Local Average Treatment Effects

Four latent groups:

- always-takers, with $D(1) = 1$ and $D(0) = 1$;
- never-takers, with $D(1) = 0$ and $D(0) = 0$;
- compliers, with $D(1) = 1$ and $D(0) = 0$;
- defiers, with $D(1) = 0$ and $D(0) = 1$.

The unconditional LATE parameter:

$$\tau_{\text{LATE}} = \mathrm{E}\left[Y(1) - Y(0) \mid D(1) \neq D(0)\right].$$

Local Average Treatment Effects

We can also write:

$$\tau_{\text{LATE}} = \frac{\mathrm{E}\left[\pi(X) \cdot \tau(X)\right]}{\mathrm{E}\left[\pi(X)\right]},$$

where

$$\tau(x) = \mathrm{E}\left[Y(1) - Y(0) \mid D(1) \neq D(0), X = x\right]$$

is the conditional LATE and

$$\pi(x) = \mathrm{P}\left[D(1) \neq D(0) \mid X = x\right]$$

is the conditional proportion of compliers and defiers.

Assumptions

Assumption IV

(i) $\big(Y(0,0), Y(0,1), Y(1,0), Y(1,1), D(0), D(1)\big) \perp Z \mid X$;

(ii) $\mathrm{P}\left[Y(1,d) = Y(0,d) \mid X\right] = 1$ *for* $d \in \{0,1\}$ *a.s.*;

(iii) $0 < \mathrm{P}\left[Z = 1 \mid X\right] < 1$ *and* $\mathrm{P}\left[D(1) = 1 \mid X\right] \neq \mathrm{P}\left[D(0) = 1 \mid X\right]$ *a.s.*

(i) Conditional independence; (ii) Exclusion restriction; (iii) Relevance

Assumption SM (Strong monotonicity)

$\mathrm{P}\left[D(1) \geq D(0) \mid X\right] = 1$ *a.s.*

If Assumption SM is too restrictive and some defiers actually exist, an alternative assumption is necessary for point identification. We can put restrictions on treatment effects (Heckman and Vytlacil, 2005; Mogstad and Torgovitsky, 2018) or the first stage (Kolesár, 2013; Semenova, 2020).

Assumptions

I follow Kolesár (2013) and Semenova (2020) in allowing for the existence of defiers but only in the following way:

Assumption WM (Weak monotonicity)

There exists a partition of the covariate space such that $\mathrm{P}[D(1) \geq D(0) \mid X] = 1$ *a.s. on one subset and* $\mathrm{P}[D(1) \leq D(0) \mid X] = 1$ *a.s. on its complement.*

Angrist and Imbens (1995), Revisited

Focus on a special case of the model in equation (1); namely,

$$Y = D\beta + X\gamma + \upsilon,$$

where all elements of X are binary and represent membership in disjoint groups or strata.

All original covariates need to be discrete, in which case we can divide the population into K groups, where K corresponds to the number of possible combinations of values of these variables. We have $G \in \{1, \ldots, K\}$ and $G_k = 1[G = k]$.

So, Angrist and Imbens (1995) study the case where $X = (1, G_1, \ldots, G_{K-1})$ and $Z_{\mathrm{C}} = (Z, ZG_1, \ldots, ZG_{K-1})$.

Angrist and Imbens (1995), Revisited

Lemma 3.1 (Angrist and Imbens, 1995; Kolesár, 2013)

Under Assumptions IV and either SM or WM, and with $X = (1, G_1, \ldots, G_{K-1})$ *and* $Z_C = (Z, ZG_1, \ldots, ZG_{K-1})$,

$$\beta_{2SLS} = \frac{\mathrm{E}\left[\sigma^2(X) \cdot \tau(X)\right]}{\mathrm{E}\left[\sigma^2(X)\right]},$$

where $\sigma^2(X) = \mathrm{E}\left[\left(\mathrm{E}\left[D \mid X, Z\right] - \mathrm{E}\left[D \mid X\right]\right)^2 \mid X\right]$.

All weights are positive regardless of monotonicity.

But how does β_{2SLS} differ from τ_{LATE} exactly?

Angrist and Imbens (1995), Revisited

Theorem 3.2

Under Assumptions IV and either SM or WM, and with $X = (1, G_1, \ldots, G_{K-1})$ *and* $Z_C = (Z, ZG_1, \ldots, ZG_{K-1})$,

$$\beta_{2\text{SLS}} = \frac{\mathrm{E}\left[[\pi(X)]^2 \cdot \mathrm{Var}\left[Z \mid X\right] \cdot \tau(X)\right]}{\mathrm{E}\left[[\pi(X)]^2 \cdot \mathrm{Var}\left[Z \mid X\right]\right]}.$$

Now, $\beta_{2\text{SLS}}$ and $\tau_{\text{LATE}} = \frac{\mathrm{E}[\pi(X)\cdot\tau(X)]}{\mathrm{E}[\pi(X)]}$ are directly comparable.

Major limitation: empirical applications of IV methods rarely consider fully heterogeneous first stages and saturated specifications with discrete covariates (see, e.g., Mogstad, Torgovitsky and Walters, 2021).

Results for Just Identified Models

Two cases considered jointly:

- A saturated model for covariates, as in Angrist and Imbens (1995), but without the fully heterogeneous first stage. No additional assumptions are necessary in this case.
- A nonsaturated model, subject to an additional assumption.

Define the instrument propensity score as

$$e(X) = \mathrm{E}\left[Z \mid X\right].$$

Assumption PS

$e(X) = X\alpha$.

Assumption PS has been used by Abadie (2003), Kolesár (2013), Lochner and Moretti (2015), and Evdokimov and Kolesár (2019), among others.

Results for Just Identified Models

Theorem 3.3

Under Assumptions IV and WM, and additionally (i) with $X = (1, G_1, \ldots, G_{K-1})$ *or (ii) under Assumption PS,*

$$\beta_{\mathrm{IV}} = \frac{\mathrm{E}\left[c(X) \cdot \pi(X) \cdot \mathrm{Var}\left[Z \mid X\right] \cdot \tau(X)\right]}{\mathrm{E}\left[c(X) \cdot \pi(X) \cdot \mathrm{Var}\left[Z \mid X\right]\right]}.$$

$\Rightarrow$ some weights may be negative, in which case the IV estimand may no longer be interpretable as a causal effect for any subpopulation

Results for Just Identified Models

Corollary 3.4

Under Assumptions IV and SM, and additionally (i) with $X = (1, G_1, \ldots, G_{K-1})$ *or (ii) under Assumption PS,*

$$\beta_{\text{IV}} = \frac{\text{E}\left[\pi(X) \cdot \text{Var}\left[Z \mid X\right] \cdot \tau(X)\right]}{\text{E}\left[\pi(X) \cdot \text{Var}\left[Z \mid X\right]\right]}.$$

The problem of negative weights disappears when we impose strong monotonicity.

Under strong monotonicity, the weights in Corollary 3.4 are perhaps more desirable than those in Angrist and Imbens (1995)'s specification (recall that $\tau_{\text{LATE}} = \frac{\text{E}[\pi(X) \cdot \tau(X)]}{\text{E}[\pi(X)]}$).

Testable Implication of Strong Monotonicity

A testable implication of strong monotonicity is that $\omega(X)$ is nonnegative for all values of X.

This implication can be used to construct **a specification test**, similar to Semenova (2020). Can we reject the null that the fraction of units with a negative first stage is zero?

Of course, this is similar to what some papers are already doing — reporting first-stage estimates for a number of subsamples (e.g., Maestas, Mullen and Strand, 2013; Autor, Kostøl, Mogstad and Setzler, 2019).

Alternatively, we can **construct an alternative, "reordered" instrument**.

Reordered IV

Define an alternative, "reordered" instrument as

$$Z_{\mathrm{R}} \;=\; 1[\omega(X) > 0] \cdot Z \;+\; 1[\omega(X) < 0] \cdot (1 - Z)\,.$$

The corresponding estimand is:

$$\beta_{\mathrm{RIV}} = \left[\left(\mathrm{E}\left[Q_{\mathrm{R}}' W \right] \right)^{-1} \mathrm{E}\left[Q_{\mathrm{R}}' Y \right] \right]_1,$$

where $W = (D, X)$ and $Q_{\mathrm{R}} = (Z_{\mathrm{R}}, X)$.

Theorem 3.5

Under Assumptions IV and SM or WM, and additionally (i) with $X = (1, G_1, \ldots, G_{K-1})$ *or (ii) with* $\mathrm{E}[Z_{\mathrm{R}} \mid X] = X\alpha_{\mathrm{R}}$,

$$\beta_{\mathrm{RIV}} = \frac{\mathrm{E}\left[\pi(X) \cdot \mathrm{Var}\left[Z \mid X\right] \cdot \tau(X)\right]}{\mathrm{E}\left[\pi(X) \cdot \mathrm{Var}\left[Z \mid X\right]\right]}.$$

Under strong monotonicity, $Z_{\mathrm{R}} = Z$ and $\beta_{\mathrm{RIV}} = \beta_{\mathrm{IV}}$.

"Reversed" Weights

Theorem 4.1

Under Assumptions IV, SM, PS, and LN,

$$\beta_{\text{IV}} = w_{\text{LATT}} \cdot \tau_{\text{LATT}} + w_{\text{LATU}} \cdot \tau_{\text{LATU}},$$

where

$$w_{\text{LATT}} = \frac{(1-\theta) \cdot \text{Var}\left[e(X) \mid Z=0\right] \cdot \pi_1}{\theta \cdot \text{Var}\left[e(X) \mid Z=1\right] \cdot \pi_0 + (1-\theta) \cdot \text{Var}\left[e(X) \mid Z=0\right] \cdot \pi_1}$$

and

$$w_{\text{LATU}} = \frac{\theta \cdot \text{Var}\left[e(X) \mid Z=1\right] \cdot \pi_0}{\theta \cdot \text{Var}\left[e(X) \mid Z=1\right] \cdot \pi_0 + (1-\theta) \cdot \text{Var}\left[e(X) \mid Z=0\right] \cdot \pi_1}.$$

Recall that $\tau_{\text{LATE}} = \frac{\theta \cdot \pi_1}{\theta \cdot \pi_1 + (1-\theta) \cdot \pi_0} \cdot \tau_{\text{LATT}} + \frac{(1-\theta) \cdot \pi_0}{\theta \cdot \pi_1 + (1-\theta) \cdot \pi_0} \cdot \tau_{\text{LATU}}.$

Importance of $\theta = 0.5$

I skip Corollary 4.2, which expresses $\beta_{\text{IV}} - \tau_{\text{LATE}}$ as a product of a simple function of weights and $\tau_{\text{LATT}} - \tau_{\text{LATU}}$.

Assumption EV (Equality of variances)

$\text{Var}\left[e(X) \mid Z = 1\right] = \text{Var}\left[e(X) \mid Z = 0\right]$.

Corollary 4.3

Under Assumptions IV, SM, PS, LN, and EV,

$$\beta_{\text{IV}} = \tau_{\text{LATE}} \quad \textit{if and only if} \quad \tau_{\text{LATT}} = \tau_{\text{LATU}} \ \textit{or} \ \theta = 0.5.$$

Empirical Application

A replication of Card (1995)'s study of returns to schooling with the college proximity instrument.

Data from the National Longitudinal Survey of Young Men (NLSYM). Outcome is log wages in 1976. Instrument is whether an individual grew up in the vicinity of a four-year college.

Two modifications:

- Following Kitagawa (2015), it may be sufficient to control for five binary covariates: whether Black, whether lived in a metropolitan area in 1966 and 1976, and whether lived in the South in 1966 and 1976.
- The binary treatment is "some college attendance," defined as having strictly more than twelve years of schooling.
  - Kitagawa (2015) focuses on having at least sixteen years of schooling but the instrument is stronger for the treatment margin that I consider.

Just Identified Specifications

	$\hat{\beta}_{\text{IV}}$			
	(1)	(2)	(3)	(4)
College attendance	0.661**	0.575*	0.610*	0.570*
	(0.294)	(0.308)	(0.354)	(0.343)
Sample	Full	Full	Full	Restricted
Covariates	Full	Discrete	Saturated	Saturated
Robust F	12.46	8.97	7.27	7.48
Observations	3,010	3,010	3,010	2,988

Negative First Stage

	(1)	(2)
$\hat{\mathrm{P}}\left[\omega(X) < 0\right]$	0.178**	0.177**
	(0.086)	(0.087)
Sample	Full	Restricted
Covariates	Saturated	Saturated
Observations	3,010	2,988

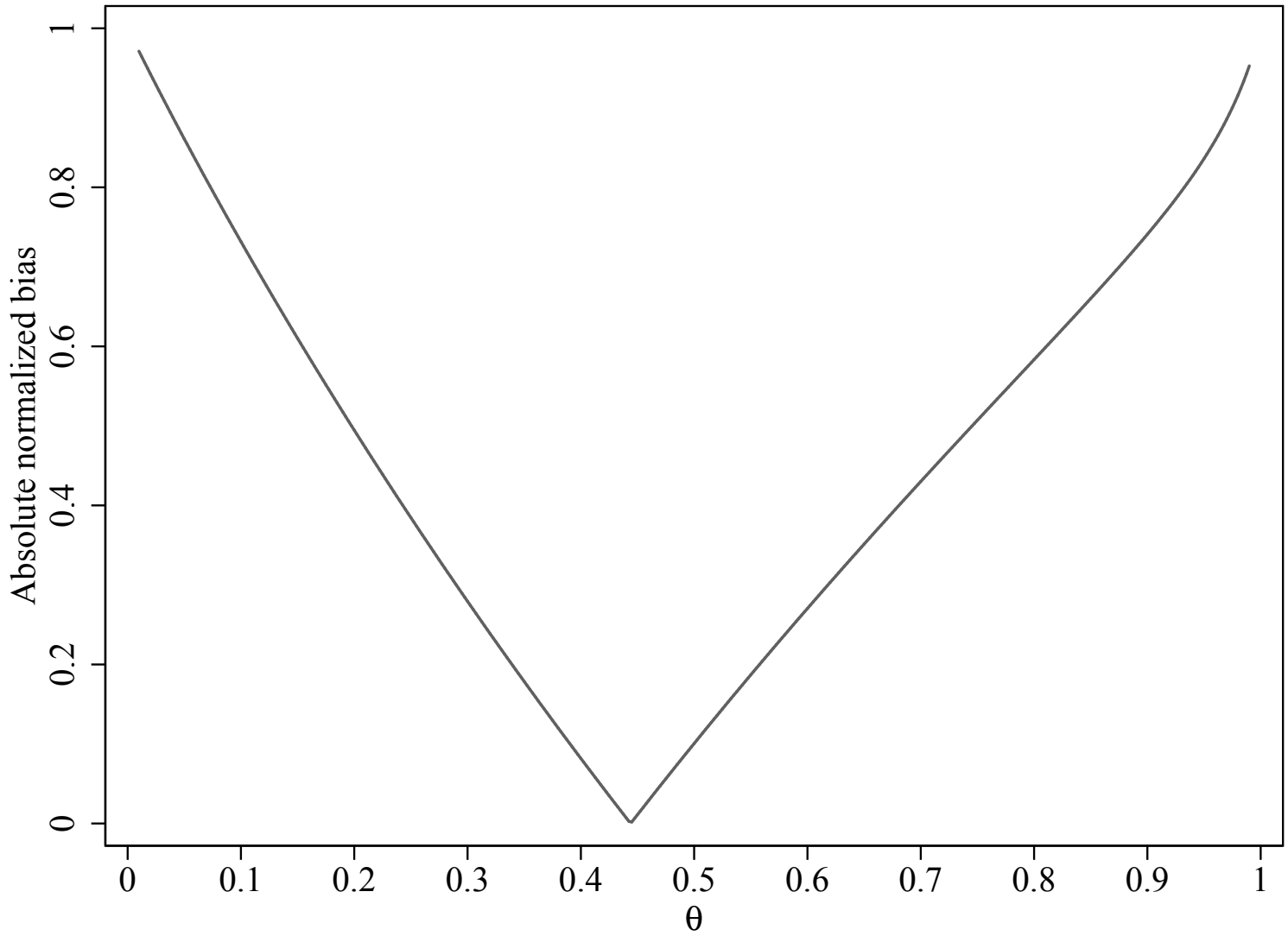
Correcting for Negative Weights

	$\hat{\beta}_{\text{IV}}$	$\hat{\beta}_{\text{2SLS}}$	$\hat{\beta}_{\text{RIV}}$	$\hat{\tau}_{\text{LATE}}$
	(1)	(2)	(3)	(4)
College attendance	0.570*	0.156	0.289	0.192
	(0.343)	(0.138)	(0.196)	(0.174)
Sample	Restricted	Restricted	Restricted	Restricted
Covariates	Saturated	Saturated	Saturated	Saturated
Robust F	7.48	3.11	24.21	N/A
Observations	2,988	2,988	2,988	2,988

Pseudo-Simulations

Reestimate $\hat{\beta}_{\mathrm{RIV}}$ and $\hat{\tau}_{\mathrm{LATE}}$ using weights of 1 for units with $Z_{\mathrm{R}} = 1$ and w for units with $Z_{\mathrm{R}} = 0$. Variation in w leads to variation in $\hat{\theta}$ and allows us to examine the dependence of IV bias on the proportion of individuals that are encouraged to get treated.

Normalized Bias of $\hat{\beta}_{\mathrm{RIV}}$



Conclusion

This paper studies the interpretation of IV and 2SLS estimands with binary D and Z, and when additional covariates are required for identification.

I conclude that the usual practice of interpreting these estimands as "LATE," or even a convex combination of conditional LATEs, is substantially more problematic than previously thought.

At the very least, empirical researchers should:

- either strengthen Kolesár (2013)'s assumption of weak monotonicity, and rule out the existence of defiers altogether;
- or account for heterogeneity in the reduced-form and first-stage regressions, as in Angrist and Imbens (1995)'s specification;
- or use the "reordered IV" procedure that this paper also proposes.

Still, this is not going to be sufficient to recover the unconditional LATE parameter unless the groups with $Z = 1$ and $Z = 0$ are roughly equal sized.

Consistent estimators of the unconditional LATE parameter under strong monotonicity are available (e.g., Abadie, 2003; Frölich, 2007). Formal treatment of estimation of LATE under weak monotonicity is left for future work.