Gov 2002: 10. Instrumental Variables

Matthew Blackwell

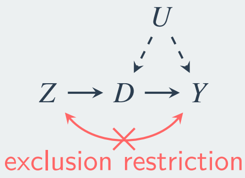
November 5, 2015

1/ IV setup

Where are we? Where are we going?

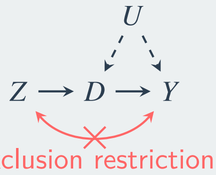
- We saw how to identify and estimate effects under no unmeasured confounding and with repeated measurements
- What if we have neither? Are we doomed?
- Not necessarily if you can identify some exogenous sources of variation that drives the treatment.
- Instrumental variables allows for unmeasured confounding on the the treatment-outcome relationship.
- Use the unconfounded variation in the instrument to help identify treatment effects.

Basic IV setup with DAGs



- Z is the instrument, D is the treatment, and U is the unmeasured confounder
- Exclusion restriction
  - no common causes of the instrument and the outcome
  - no direct or indirect effect of the instrument on the outcome not through the treatment.
- First-stage relationship: Z affects D

An IV is only as good as its assumptions



- Finding a believable instrument is incredibly difficult and some people never believe any IV setups.
- When effects vary, the IV approach estimates a "local" ATE that is local to this particular instrument.

IVs in the field

- Angrist (1990): Draft lottery as an IV for military service (income as outcome)
- Acemoglu et al (2001): settler mortality as an IV for institutional quality (GDP/capita as outcome)
- Levitt (1997): being an election year as IV for police force size (crime as outcome)
- Kern & Hainmueller (2009): having West German TV reception in East Berlin as an instrument for West German TV watching (outcome is support for the East German regime)
- Nunn & Wantchekon (2011): historical distance of ethnic group to the coast as a instrument for the slave raiding of that ethnic group (outcome are trust attitudes today)
- Acharya, Blackwell, Sen (2015): cotton suitability as IV for proportion slave in 1860 (outcome is white attitudes today)

2/ IV with constant treatment effects

IV with constant effects

- Let's write down a causal model for Y_i with constant effects and an unmeasured confounder, U_i:

$$Y_i(d, u) = \alpha + \tau d + \gamma u + \eta_i$$

- If we connect this with a consistency assumption, we get the this regression form:

$$Y_i = \alpha + \tau D_i + \gamma U_i + \eta_i$$

- Here we assume that $\mathbb{E}[D_i \eta_i] = 0$, so if we measured U_i, then we would be able to estimate τ.
- But $\text{Cov}(\gamma U_i + \eta_i, D_i) \neq 0$ because U is a common cause of D and Y.

The role of the instrument

- If we have an instrument, Z_i, that satisfies the exclusions restriction, then

$$\text{Cov}(\gamma U_i + \eta_i, Z_i) = 0$$

- It must be independent of U_i and it has no correlation with η_i because neither does the treatment.

$$\begin{aligned}
\text{Cov}(Y_i, Z_i) &= \text{Cov}(\alpha + \tau D_i + \gamma U_i + \eta_i, Z_i) \\
&= \text{Cov}(\alpha, Z_i) + \text{Cov}(\tau D_i, Z_i) + \text{Cov}(\gamma U_i + \eta_i, Z_i) \\
&= 0 + \tau \text{Cov}(D_i, Z_i) + 0
\end{aligned}$$

IV estimator with constant effects

$$Y_i = \alpha + \tau D_i + \gamma U_i + \eta_i$$

- With this in hand, we can formulate an expression for the average treatment effect here:

$$\tau = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)} = \frac{\text{Cov}(Y_i, Z_i) / \mathbb{V}[Z_i]}{\text{Cov}(D_i, Z_i) / \mathbb{V}[Z_i]}$$

- Reduced form coefficient: $\text{Cov}(Y_i, Z_i) / \mathbb{V}[Z_i]$
- First stage coefficient: $\text{Cov}(D_i, Z_i) / \mathbb{V}[Z_i]$

Weak instruments

- Natural estimator:

$$\widehat{\tau}_{IV} = \frac{\widehat{\text{Cov}}(Y_i, Z_i)}{\widehat{\text{Cov}}(D_i, Z_i)}$$

- What happens with a weak first stage? Can show that this estimator converges to:

$$\widehat{\tau}_{IV} \xrightarrow{p} \tau + \frac{\text{Cov}(Z_i, U_i)}{\text{Cov}(Z_i, D_i)}$$

- If $\text{Cov}(Z_i, D_i)$ is small, then even very small violations of the exclusion restriction $\text{Cov}(Z_i, U_i) \neq 0$ can lead to large inconsistencies and finite sample bias.
- Important to convey the strength of the first-stage via t-test or F-test with multiple instruments.

Wald Estimator

- Binary instrument leads to the Wald estimator:

$$\tau = \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(D_i, Z_i)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}$$

- Intuitively:

$$\frac{\text{effect of instrument on outcome}}{\text{effect of instrument on treatment}}$$

What about covariates?

- No covariates up until now. What if we have a set of covariates X_i that we are also conditioning on?
- Let's start with linear models for both the outcome and the treatment:

$$Y_i = X_i'\beta + \tau D_i + \varepsilon_i$$
$$D_i = X_i'\alpha + \gamma Z_i + \nu_i$$

- Now, we assume that X_i are **exogenous** along with Z_i:

$$\mathbb{E}[Z_i \nu_i] = 0 \quad \mathbb{E}[Z_i \varepsilon_i] = 0$$

$$\mathbb{E}[X_i \nu_i] = 0 \quad \mathbb{E}[X_i \varepsilon_i] = 0$$

- ...but D_i is **endogenous**: $\mathbb{E}[D_i \varepsilon_i] \neq 0$

Getting the reduced form

- We can plug the treatment equation into the outcome equation:

$$\begin{aligned} Y_i &= X_i'\beta + \tau[X_i'\alpha + \gamma Z_i + \nu_i] + \varepsilon_i \\ &= X_i'\beta + \tau[X_i'\alpha + \gamma Z_i] + [\tau\nu_i + \varepsilon_i] \\ &= X_i'\beta + \tau[X_i'\alpha + \gamma Z_i] + \varepsilon_i^* \\ &= X_i'\beta + \tau\mathbb{E}[D_i|X_i, Z_i] + \varepsilon_i^* \end{aligned}$$

- Red value in the brackets is the population fitted value of the treatment, $\mathbb{E}[D_i|X_i, Z_i]$
- Because Z_i and X_i are uncorrelated with ν_i and ε_i, then this fitted value is also independent of ε_i^*.
- Thus, the population regression coefficient of a Y_i on $[X_i'\alpha + \gamma Z_i]$ is the average treatment effect, τ.

Two-stage least squares

- Estimate $\widehat{\alpha}$ and $\widehat{\gamma}$ from OLS and form fitted values:

$$\widehat{\mathbb{E}}[D_i|X_i, Z_i] = \widehat{D}_i = X_i'\widehat{\alpha} + \widehat{\gamma}Z_i.$$

- Regress of Y_i on X_i and $\widehat{D}_i$. Add and subtract $\tau\widehat{D}_i$:

$$Y_i = X_i'\beta + \tau\widehat{D}_i + [\varepsilon_i + \tau(D_i - \widehat{D}_i)]$$

- Key question: is $\widehat{D}_i$ uncorrelated with the error?
- $\widehat{D}_i$ is just a function of X_i and Z_i so it is uncorrelated with ε_i.
- We also know that $\widehat{D}_i$ is uncorrelated with $(D_i - \widehat{D}_i)$?

Two-stage least squares

- Heuristic procedure:
  1. Run regression of treatment on covariates and instrument
  2. Construct fitted values of treatment
  3. Run regression of outcome on covariates and fitted values
- Note that this isn't how we actually estimate 2SLS because the standard errors are all wrong.
- Computer wants to calculate the standard errors based on ε_i^*:

$$\varepsilon_i^* = Y_i - X_i'\beta - \tau\widehat{D}_i$$

- but what we really want is the standard errors based on ε_i:

$$\varepsilon_i = Y_i - X_i'\beta - \tau D_i$$

Nunn & Wantchekon IV example

TABLE 5—IV ESTIMATES OF THE EFFECT OF THE SLAVE TRADE ON TRUST

	Trust of relatives (1)	Trust of neighbors (2)	Trust of local council (3)	Intragroup trust (4)	Intergroup trust (5)
Second stage: Dependent variable is an individual's trust					
ln (1 + exports/area)	−0.190***	−0.245***	−0.221***	−0.251***	−0.174**
	(0.067)	(0.070)	(0.060)	(0.088)	(0.080)
Hausman test (*p*-value)	0.88	0.53	0.09	0.44	0.41
R^2	0.13	0.16	0.20	0.15	0.12
First stage: Dependent variable is ln (1 + exports/area)					
Historical distance of ethnic group from coast	−0.0014***	−0.0014***	−0.0014***	−0.0014***	−0.0014***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Colonial population density	Yes	Yes	Yes	Yes	Yes
Ethnicity-level colonial controls	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	16,709	16,679	15,905	16,636	16,473
Number of clusters	147 / 1,187	147 / 1,187	146 / 1,194	147 / 1,186	147 / 1,184
F-stat of excl. instrument	26.9	26.8	27.4	27.1	27.0
R^2	0.81	0.81	0.81	0.81	0.81

Notes: The table reports IV estimates. The top panel reports the second-stage estimates, and the bottom panel reports first-stage estimates. Standard errors are adjusted for two-way clustering at the ethnicity and district levels. The individual controls, district controls, ethnicity-level colonial controls, and colonial population density measures are described in Table 3. The null hypothesis of the Hausman test is that the OLS estimates are consistent.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

General 2SLS

- Notational convenience: combine X_i and D_i into one matrix, X_i, of size k, where one column contains D_i.
- The structural model, then is:

$$Y_i = X_i' \beta + \varepsilon_i$$

- Z_i will be a vector of l exogenous variables that includes any exogenous variables in X_i plus any instruments.
- Key assumption on the instruments:

$$\mathbb{E}[Z_i \varepsilon_i] = 0$$

Nasty Matrix Algebra

- Projection matrix projects values from the columns of Z_i to the columns of X_i:

$$\Pi = (\mathbb{E}[Z_i Z_i'])^{-1}\mathbb{E}[Z_i X_i'] \qquad \text{(projection matrix)}$$

$$\tilde{X}_i = \Pi' Z_i \qquad \text{(fitted values)}$$

- To derive the 2SLS estimator, take the fitted values, $\Pi' Z_i$ and multiply both sides of the outcome equation by them:

$$\begin{aligned}
Y_i &= X_i'\beta + \varepsilon_i \\
\Pi' Z_i Y_i &= \Pi' Z_i X_i'\beta + \Pi' Z_i \varepsilon_i \\
\mathbb{E}[\Pi' Z_i Y_i] &= \mathbb{E}[\Pi' Z_i X_i']\beta + \mathbb{E}[\Pi' Z_i \varepsilon_i] \\
\mathbb{E}[\Pi' Z_i Y_i] &= \mathbb{E}[\Pi' Z_i X_i']\beta + \Pi' \mathbb{E}[Z_i \varepsilon_i] \\
\mathbb{E}[\Pi' Z_i Y_i] &= \mathbb{E}[\Pi' Z_i X_i']\beta \\
\mathbb{E}[\tilde{X}_i Y_i] &= \mathbb{E}[\tilde{X}_i X_i']\beta \\
\beta &= (\mathbb{E}[\tilde{X}_i X_i'])^{-1}\mathbb{E}[\tilde{X}_i Y_i]
\end{aligned}$$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $\mathbf{X} = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $\mathbf{Z} = (Z_1', \ldots, Z_n')$
- Let $\widehat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ be the matrix of fitted values for $\mathbf{X}$, then we have
- Matrix party trick: $\mathbf{X}'\mathbf{Z}/n = (1/n)\sum_i^N X_i Z_i' \overset{p}{\to} \mathbb{E}[X_i Z_i']$.
- Take the population formula for the parameters:

$$\beta = (\mathbb{E}[\tilde{X}_i X_i'])^{-1}\mathbb{E}[\tilde{X}_i Y_i]$$

- And plug in the sample values (the n cancels out):

$$\hat{\beta} = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

- This is how R/Stata estimates the 2SLS parameters

Asymptotics for 2SLS

$$\hat{\beta} = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

- We can insert the true model for $\mathbf{y}$:

$$\hat{\beta} = (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'(\mathbf{X}\beta + \varepsilon)$$

- Using the matrix party trick and that $\widehat{\mathbf{X}}'\mathbf{X} = \widehat{\mathbf{X}}'\widehat{\mathbf{X}}$, we have

$$\begin{aligned}\hat{\beta} &= (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\mathbf{X}\beta + (\widehat{\mathbf{X}}'\mathbf{X})^{-1}\widehat{\mathbf{X}}'\varepsilon \\ &= \beta + (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\varepsilon \\ &= \beta + \left[n^{-1}\sum_i \widehat{X}_i\widehat{X}_i'\right]^{-1} n^{-1}\sum_i \widehat{X}_i\varepsilon_i\end{aligned}$$

- Consistent because $n^{-1}\sum_i \widehat{X}_i\varepsilon_i \overset{p}{\to} \mathbb{E}[\widehat{X}_i\varepsilon_i] = 0$.

Asymptotic variance for 2SLS

$$\sqrt{n}(\hat{\beta} - \beta) = \left(n^{-1} \sum_i \widehat{X}_i \widehat{X}_i'\right)^{-1} \left(n^{-1/2} \sum_i \widehat{X}_i \varepsilon_i\right)$$

- By the CLT, $n^{-1/2} \sum_i \widehat{X}_i \varepsilon_i$ converges in distribution to $N(0, B)$, where $B = \mathbb{E}[\widehat{X}_i' \varepsilon_i' \varepsilon_i \widehat{X}_i]$.
- By the LLN, $n^{-1} \sum_i \widehat{X}_i \widehat{X}_i' \overset{p}{\to} \mathbb{E}[\widehat{X}_i \widehat{X}_i']$.
- Thus, we have that $\sqrt{n}(\hat{\beta} - \beta)$ has asymptotic variance:

$$(\mathbb{E}[\widehat{X}_i \widehat{X}_i'])^{-1} \mathbb{E}[\widehat{X}_i' \varepsilon_i' \varepsilon_i \widehat{X}_i] (\mathbb{E}[\widehat{X}_i \widehat{X}_i'])^{-1}$$

- Replace with the sample quantities to get estimate of the robust 2SLS variance estimator:

$$\widehat{\text{var}}(\hat{\beta}) = (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1} \left(\sum_i \hat{u}_i^2 \widehat{X}_i \widehat{X}_i' \right) (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}$$

where $\hat{u}_i = Y_i - X_i' \hat{\beta}$

Overidentification

- What if we have more instruments than endogenous variables?
- When there are more instruments than causal parameters ($l > k$), the model is overidentified.
- When there are as many instruments as causal parameters ($l = k$), the model is just identified.
- With more than one instrument and constant effects, we can test for the plausibility of the exclusion restriction(s) using an overidentification test.
- Is it plausible to find more than one instrument?

Overidentification tests

- Sargan-Hausman test:
  - Under the null of all valid instruments, using all instruments versus a subset should only differ by sampling variation.
  - Regress 2SLS residuals, $\hat{\varepsilon}_i$ on X_i and calculate R_u^2 from this regression.
  - Under the null (and homoskedasticity), $NR_u^2 \sim \mathrm{X}_{l-k}^2$.
  - Degrees of freedom depends on how many overidentifying restrictions there are.
- If we reject the null hypothesis in these overidentification tests, then it means that the exclusion restrcitions for our instruments are probably incorrect.
- Note that it won't tell us which of them are incorrect, just that at least one is.
- These overidentification tests depend heavily on the constant effects assumption

3/ IV with heterogenous treatment effects

Instrumental Variables and Potential Outcomes

- Basic idea of IV:
  - D_i not randomized, but Z_i is
  - Z_i only affects Y_i through D_i
- D_i now depends on $Z_i \rightsquigarrow$ potential treatments: $D_i(1) = D_i(z = 1)$ and $D_i(0)$.
- Consistency:

$$D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$$

- Outcome can depend on both the treatment and the instrument: $Y_i(d, z)$ is the outcome if unit i had received treatment $D_i = d$ and instrument value $Z_i = z$.

Key assumptions

1. Randomization
2. Exclusion Restriction
3. First-stage relationship
4. Monotonicity

Randomization

- Need the instrument to be randomized:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability
- But why believe conditional ignorability for the instrument but not the treatment?
- Best instruments are truly randomized.
- Identifies the intent-to-treat (ITT) effect:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]$$

Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d, 1) = Y_i(d, 0) \quad \text{for } d = 0, 1$$

- Given this exclusion restriction, we know that the potential outcomes for each treatment status only depend on the treatment, not the instrument:

$$Y_i(1) \equiv Y_i(1, 1) = Y_i(1, 0)$$
$$Y_i(0) \equiv Y_i(0, 1) = Y_i(0, 0)$$

- NOT A TESTABLE ASSUMPTION

The linear model with heterogeneous effects

- As usual, rewrite Y_i using consistency:

$$\begin{aligned} Y_i &= Y_i(0) + (Y_i(1) - Y_i(0))D_i \\ &= \alpha_0 + \tau_i D_i + \eta_i \end{aligned}$$

- Here, we have $\alpha_0 = E[Y_i(0)]$ and $\tau_i = Y_i(1) - Y_i(0)$.

First Stage

- This next assumption is a little mundane, but turns out to be very important: the instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

- Otherwise, what would we be doing? The instrument wouldn't affect anything.
- Implies that $\text{Cov}(D_i, Z_i) \neq 0$

Monotonicity

- Lastly, we need to make another assumption about the relationship between the instrument and the treatment.
- Monotonicity says that the presence of the instrument never dissuades someone from taking the treatment:

$$D_i(1) - D_i(0) \geq 0$$

- Note if this holds in the opposite direction $D_i(1) - D_i(0) \leq 0$, we can always rescale D_i to make the assumption hold.

Monotonicity means no defiers

- This is sometimes called no defiers.
- With a binary treatment and a binary instrument, there are four groups:

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- These compliance groups are sometimes called principal strata.
- The monotonicity assumption remove the possibility of there being defiers in the population.
- Anyone with $D_i = 1$ when $Z_i = 0$ must be an always-taker and anyone with $D_i = 0$ when $Z_i = 1$ must be a never-taker.

Local Average Treatment Effect (LATE)

- Under these four assumptions, the Wald estimator is equal what we call Local average treatment effect (LATE) or the complier average treatment effect (CATE).
- This is is the ATE among the compliers: those that take the treatment when encouraged to do so.
- That is, the LATE theorem, states that:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

- This fact was a massive intellectual jump in our understanding of IV.

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

$$\begin{aligned} E[Y_i|Z_i = 1] &= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1] \\ &= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)] \qquad \text{(randomization)} \end{aligned}$$

- The same applies to when $Z_i = 0$, so we have

$$E[Y_i|Z_i = 0] = E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)]$$

- Thus, $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$\begin{aligned} &E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\ =&E[(Y_i(1) - Y_i(0))(1)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)] \\ +&E[(Y_i(1) - Y_i(0))(-1)|D_i(1) < D_i(0)] \Pr[D_i(1) < D_i(0)] \\ =&E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)] \end{aligned}$$

- The third equality comes from monotonicity: with this assumption, $D_i(1) < D_i(0)$ never occurs.

Proof (continued)

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)]$$

- We can use the same argument for the denominator:

$$\begin{aligned} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] &= E[D_i(1) - D_i(0)] \\ &= \Pr[D_i(1) > D_i(0)] \end{aligned}$$

- Dividing these two expressions through gives the LATE.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is a unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{LATE} \neq \tau_{ATE}$.
- Complier group depends on the instrument $\rightsquigarrow$ different IVs will lead to different estimands.
- 2SLS "cheats" by assuming that the effect is constant, so it is the same for compliers and non-compliers.

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.
- Think of a randomized experiment:
  - Randomized treatment assignment = instrument (Z_i)
  - Non-randomized actual treatment taken = treatment (D_i)
- One-sided noncompliance: only those assigned to treatment (control) can actually take the treatment (control). Or

$$D_i(0) = 0 \forall i \quad \rightsquigarrow \quad \Pr[D_i = 1 | Z_i = 0] = 0$$

- Maybe this is because only those treated actually get pills or only they are invited to the job training location.

Benefits of one-sided noncompliance

- One-sided noncompliance $\rightsquigarrow$ no "always-takers" and since there are no defiers,
  - Treated units must be compliers.
  - ATT is the same as the LATE.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$\mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)|Z_i = 0]$$

(exclusion restriction + one-sided noncompliance)

$$= \mathbb{E}[Y_i(0)|Z_i = 1] + E[(Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)|Z_i = 0]$$

$$= \mathbb{E}[Y_i(0)] + \mathbb{E}[(Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)]$$

(randomization)

$$= \mathbb{E}[(Y_i(1) - Y_i(0))D_i|Z_i = 1]$$

$$= \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1, Z_i = 1] \Pr[D_i = 1|Z_i = 1]$$

(law of iterated expectations + binary treatment)

$$= \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] \Pr[D_i = 1|Z_i = 1]$$

(one-sided noncompliance)

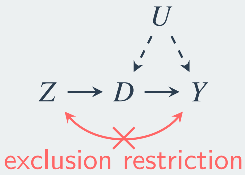
- Noting that $\Pr[D_i = 1|Z_i = 0] = 0$, then the Wald estimator is just the ATT:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{\Pr[D_i = 1|Z_i = 1]} = E[Y_i(1) - Y_i(0)|D_i = 1]$$

- Thus, under the additional assumption of one-sided compliance, we can estimate the ATT using the usual IV approach

4/ IV extensions

Falsification tests



- The exclusion restriction cannot be tested directly, but it can be falsified.
- Falsification test Test the reduced form effect of Z_i on Y_i in situations where it is impossible or extremely unlikely that Z_i could affect D_i.
- Because Z_i can't affect D_i, then the exclusion restriction implies that this falsification test should have 0 effect.
- Nunn & Wantchekon (2011): use distance to coast as an instrument for Africans, use distance to the coast in an Asian sample as falsification test.

Nunn & Wantchekon falsification test

TABLE 7—REDUCED FORM RELATIONSHIP BETWEEN THE DISTANCE FROM THE COAST AND TRUST WITHIN AFRICA AND ASIA

	Trust of local government council			
	Afrobarometer sample		Asiabarometer sample	
	(1)	(2)	(3)	(4)
Distance from the coast	0.00039***	0.00031***	−0.00001	0.00001
	(0.00009)	(0.00008)	(0.00010)	(0.00009)
Country fixed effects	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	Yes
Number of observations	19,913	19,913	5,409	5,409
Number of clusters	185	185	62	62
R^2	0.16	0.18	0.19	0.22

Notes: The table reports OLS estimates. The unit of observation is an individual. The dependent variable in the Asiabarometer sample is the respondent's answer to the question: "How much do you trust your local government?" The categories for the answers are the same in the Asiabarometer as in the Afrobarometer. Standard errors are clustered at the ethnicity level in the Afrobarometer regressions and at the location (city) level in the Asiabarometer and the WVS samples. The individual controls are for age, age squared, a gender indicator, education fixed effects, and religion fixed effects.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Size, characteristics of the compliers

- While we cannot identify who is a complier and who is not a complier in general, we can estimate the size of the complier group:

$$\Pr[D_i(1) > D_i(0)] = E[D_i(1) - D_i(0)] = E[D_i | Z_i = 1] - E[D_i | Z_i = 0]$$

- Can extend this to calculate features of the complier group:
  - Covariate means, variances, etc.
  - Abadie (2003) shows how to weight the data to estimate these quantities.

Multiple instruments

- Different instruments $\rightsquigarrow$ different LATEs
    - Instrument 1, Z_{i1} with LATE τ_1
    - Instrument 2, Z_{i2} with LATE τ_2
- Use both in the first stage:

$$\widehat{D}_i = \pi_1 Z_{1i} + \pi_2 Z_{2i}.$$

2SLS as weighted average

- MHE shows that the 2SLS estimator using these two instruments is a weighted sum of the two component LATEs:

$$\rho_{2SLS} = \psi \tau_1 + (1 - \psi)\tau_2,$$

where the weights are:

$$\psi = \frac{\pi_1 \text{Cov}(D_i, Z_{1i})}{\pi_1 \text{Cov}(D_i, Z_{1i}) + \pi_2 \text{Cov}(D_i, Z_{2i})}$$

- Thus, the 2SLS estimate is a weighted average of causal effects for each instrument, where the weights are related to the strength of first-stage.

Covariates and heterogeneous effects

- It might be the case that the above assumptions only hold conditional on some covariates, X_i. That is, instead of randomization, we might have conditional ignorability:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i | X_i$$

- We would also have exclusion conditional on the covariates:

$$\Pr[Y_i(d,0) = Y_i(d,1)|X_i] = 1 \quad \text{for } d = 1, 0$$

- Under these assumptions, with fully saturated first and second stages, then 2SLS estimates a weighted average of the covariates-specific LATEs (very similar to regression).
- Abadie (2003) shows how to estimate the overall LATE using a weighting approach based on a "propensity score" for the instrument.