Gov 2002 - Causal Inference II: Instrumental Variables

Matthew Blackwell Arthur Spirling

October 2nd, 2014

Instrumental Variables

- Last week we talked about how to make progress when you have randomization or selection on the observables.

Instrumental Variables

- Last week we talked about how to make progress when you have randomization or selection on the observables.
- But what if you have neither of those two for your treatment variable? Are you doomed?

Instrumental Variables

- Last week we talked about how to make progress when you have randomization or selection on the observables.
- But what if you have neither of those two for your treatment variable? Are you doomed?
- Maybe.

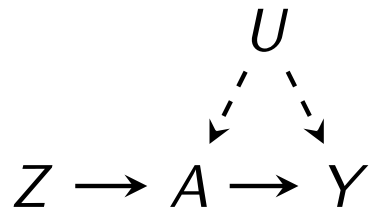
Instrumental Variables

- Last week we talked about how to make progress when you have randomization or selection on the observables.
- But what if you have neither of those two for your treatment variable? Are you doomed?
- Maybe.
- But if you can identify some exogenous sources of variation that drive the treatment, even if the treatment was not randomly assigned, you may be able to make headway.

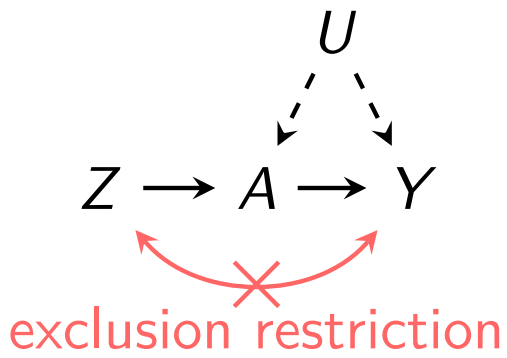
Instrumental Variables

- Last week we talked about how to make progress when you have randomization or selection on the observables.
- But what if you have neither of those two for your treatment variable? Are you doomed?
- Maybe.
- But if you can identify some exogenous sources of variation that drive the treatment, even if the treatment was not randomly assigned, you may be able to make headway.
- The basic idea behind instrumental variables is that we have a treatment with unmeasured confounding, but that we have another variable, called the instrument, that affects the treatment, but not the outcome, and thus give us that exogenous variation.

Basic IV setup with DAGs

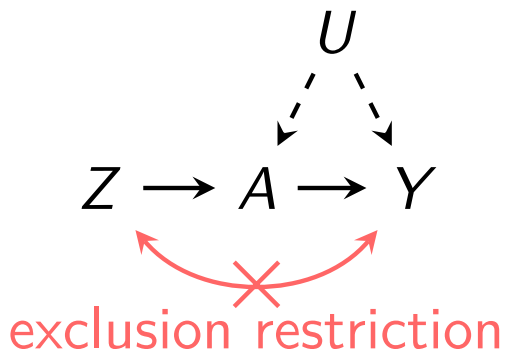


Basic IV setup with DAGs



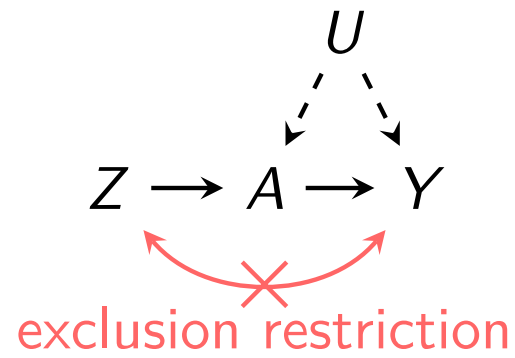
- Z is the instrument, A is the treatment, and U is the unmeasured confounder

Basic IV setup with DAGs



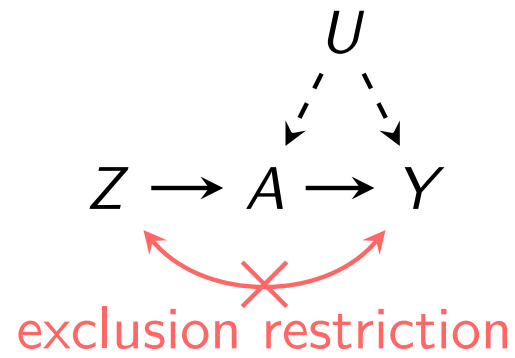
- Z is the instrument, A is the treatment, and U is the unmeasured confounder
- Exclusion restriction

Basic IV setup with DAGs



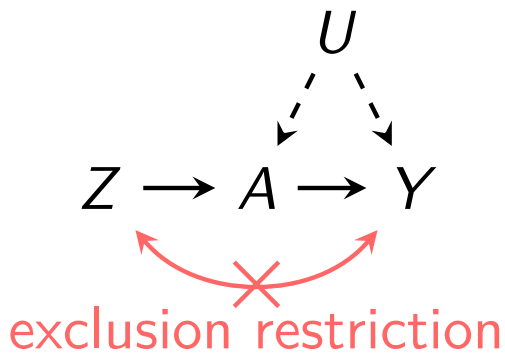
- Z is the instrument, A is the treatment, and U is the unmeasured confounder
- Exclusion restriction
  - no common causes of the instrument and the outcome

Basic IV setup with DAGs



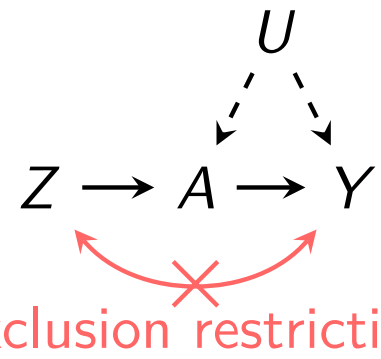
- Z is the instrument, A is the treatment, and U is the unmeasured confounder
- Exclusion restriction
  - no common causes of the instrument and the outcome
  - no direct or indirect effect of the instrument on the outcome not through the treatment.

Basic IV setup with DAGs



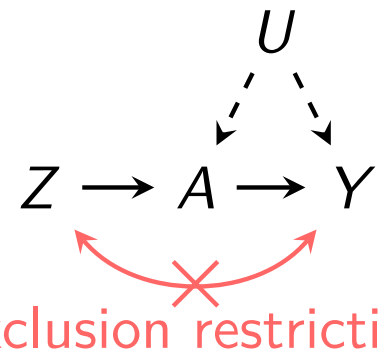
- Z is the instrument, A is the treatment, and U is the unmeasured confounder
- Exclusion restriction
  - no common causes of the instrument and the outcome
  - no direct or indirect effect of the instrument on the outcome not through the treatment.
- First-stage relationship: Z affects A

An IV is only as good as its assumptions



- Finding a believable instrument is incredibly difficult and some people never believe any IV setups.

An IV is only as good as its assumptions



- Finding a believable instrument is incredibly difficult and some people never believe any IV setups.
- We will see that even if all of the untestable assumptions are met, the IV approach estimates a "local" ATE. That is, local to this particular case/instrument.

IVs in the field

- Angrist (1990): Draft lottery as an IV for military service (income as outcome)

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)

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)

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)

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)

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 (2014): cotton suitability as IV for proportion slave in 1860 (outcome is white attitudes today)

IV with constant effects

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

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

IV with constant effects

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

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

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

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

IV with constant effects

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

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

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

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

- Here we assume that $E[A_i \eta_i] = 0$, so if we measured U_i, then we would be able to estimate τ.

IV with constant effects

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

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

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

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

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

The role of the instrument

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

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

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.

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.

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.

$$\text{cov}(Y_i, Z_i) = \text{cov}(\alpha + \tau A_i + \gamma U_i + \eta_i, Z_i)$$

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.

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

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.

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

IV estimator with constant effects

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

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

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

IV estimator with constant effects

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

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

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

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

IV estimator with constant effects

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

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

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

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

IV estimator with constant effects

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

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

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

- Reduced form coefficient: $\text{Cov}(Y_i, Z_i)/V[Z_i]$
- First stage coefficient: $\text{Cov}(A_i, Z_i)/V[Z_i]$
- What happens with a weak first stage?

Wald Estimator

- With a binary instrument, there is a simple estimator based on this formulation called the Wald estimator. It is easy to show that:

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

Wald Estimator

- With a binary instrument, there is a simple estimator based on this formulation called the Wald estimator. It is easy to show that:

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

- Intuitively, the effects of Z_i on Y_i divided by the effect of Z_i on A_i

What about covariates?

- No covariates up until now. What if we have a set of covariates X_i that we are also conditioning on?

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:

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

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:

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

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

$$E[Z_i\nu_i] = 0 \quad E[Z_i\varepsilon_i] = 0$$

$$E[X_i\nu_i] = 0 \quad E[X_i\varepsilon_i] = 0$$

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:

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

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

$$E[Z_i\nu_i] = 0 \quad E[Z_i\varepsilon_i] = 0$$

$$E[X_i\nu_i] = 0 \quad E[X_i\varepsilon_i] = 0$$

- . . . but A_i is **endogenous**: $E[A_i\varepsilon_i] \neq 0$

Getting the reduced form

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

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

Getting the reduced form

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

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

- Red value in the brackets is the population fitted value of the treatment, $E[A_i|X_i, Z_i]$

Getting the reduced form

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

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

- Red value in the brackets is the population fitted value of the treatment, $E[A_i|X_i, Z_i]$
- Because Z_i and X_i are uncorrelated with ν_i and ε_i, then this fitted value is also independent of ε_i^*.

Getting the reduced form

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

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

- Red value in the brackets is the population fitted value of the treatment, $E[A_i|X_i, Z_i]$
- 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

- In practice, we estimate the first stage from a sample and calculate OLS fitted values:

$$\hat{A}_i = X_i'\hat{\alpha} + \hat{\gamma}Z_i.$$

Two-stage least squares

- In practice, we estimate the first stage from a sample and calculate OLS fitted values:

$$\hat{A}_i = X_i'\hat{\alpha} + \hat{\gamma}Z_i.$$

- Here, $\hat{\alpha}$ and $\hat{\gamma}$ are estimates from OLS. Then, we estimate a regression of Y_i on X_i and $\hat{A}_i$. We plug this into our equation for Y_i and note that the error for A_i is now a residual:

$$Y_i = X_i'\beta + \tau\hat{A}_i + [\varepsilon_i + \tau(A_i - \hat{A}_i)]$$

Two-stage least squares

- In practice, we estimate the first stage from a sample and calculate OLS fitted values:

$$\hat{A}_i = X_i'\hat{\alpha} + \hat{\gamma}Z_i.$$

- Here, $\hat{\alpha}$ and $\hat{\gamma}$ are estimates from OLS. Then, we estimate a regression of Y_i on X_i and $\hat{A}_i$. We plug this into our equation for Y_i and note that the error for A_i is now a residual:

$$Y_i = X_i'\beta + \tau\hat{A}_i + [\varepsilon_i + \tau(A_i - \hat{A}_i)]$$

- Key question: is $\hat{A}_i$ uncorrelated with the error?

Two-stage least squares

- In practice, we estimate the first stage from a sample and calculate OLS fitted values:

$$\hat{A}_i = X_i'\hat{\alpha} + \hat{\gamma}Z_i.$$

- Here, $\hat{\alpha}$ and $\hat{\gamma}$ are estimates from OLS. Then, we estimate a regression of Y_i on X_i and $\hat{A}_i$. We plug this into our equation for Y_i and note that the error for A_i is now a residual:

$$Y_i = X_i'\beta + \tau\hat{A}_i + [\varepsilon_i + \tau(A_i - \hat{A}_i)]$$

- Key question: is $\hat{A}_i$ uncorrelated with the error?
- $\hat{A}_i$ is just a function of X_i and Z_i so it is uncorrelated with ε_i.

Two-stage least squares

- In practice, we estimate the first stage from a sample and calculate OLS fitted values:

$$\hat{A}_i = X_i'\hat{\alpha} + \hat{\gamma}Z_i.$$

- Here, $\hat{\alpha}$ and $\hat{\gamma}$ are estimates from OLS. Then, we estimate a regression of Y_i on X_i and $\hat{A}_i$. We plug this into our equation for Y_i and note that the error for A_i is now a residual:

$$Y_i = X_i'\beta + \tau\hat{A}_i + [\varepsilon_i + \tau(A_i - \hat{A}_i)]$$

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

Two-stage least squares

- Heuristic procedure:

Two-stage least squares

- Heuristic procedure:
  1. Run regression of treatment on covariates and instrument

Two-stage least squares

- Heuristic procedure:
  1. Run regression of treatment on covariates and instrument
  2. Construct fitted values of treatment

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

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.

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^*, but what we really want is the standard errors based on ε_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

- To save on notation, we'll roll all the variables in the structural model in one vector, X_i, of size k, some of which may be endogenous.

General 2SLS

- To save on notation, we'll roll all the variables in the structural model in one vector, X_i, of size k, some of which may be endogenous.
- The structural model, then is:

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

General 2SLS

- To save on notation, we'll roll all the variables in the structural model in one vector, X_i, of size k, some of which may be endogenous.
- 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:

$$E[Z_i\varepsilon_i] = 0$$

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

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

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

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

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

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

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

$$\begin{aligned} Y_i &= X_i'\beta + \varepsilon_i \\ \Pi' Z_i Y_i &= \Pi' Z_i X_i'\beta + \Pi' Z_i \varepsilon_i \\ \Pi' E[Z_i Y_i] &= \Pi' E[Z_i X_i']\beta + \Pi' E[Z_i \varepsilon_i] \end{aligned}$$

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

$$\begin{aligned} Y_i &= X_i'\beta + \varepsilon_i \\ \Pi' Z_i Y_i &= \Pi' Z_i X_i'\beta + \Pi' Z_i \varepsilon_i \\ \Pi' E[Z_i Y_i] &= \Pi' E[Z_i X_i']\beta + \Pi' E[Z_i \varepsilon_i] \\ \Pi' E[Z_i Y_i] &= \Pi' E[Z_i X_i']\beta \end{aligned}$$

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

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

Nasty Matrix Algebra

- Useful quantities:

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

$$V_i = \Pi' Z_i \qquad \text{(fitted values)}$$

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

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

$$\beta = (E[X_i Z_i'](E[Z_i Z_i'])^{-1} E[Z_i X_i'])^{-1} E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i Y_i]$$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $Z = (Z_1', \ldots, Z_n')$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $Z = (Z_1', \ldots, Z_n')$
- Matrix party trick: $X'Z/n = (1/n)\sum_i^N X_i Z_i' \xrightarrow{p} E[X_i Z_i']$.

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $Z = (Z_1', \ldots, Z_n')$
- Matrix party trick: $X'Z/n = (1/n)\sum_i^N X_i Z_i' \xrightarrow{p} E[X_i Z_i']$.
- Take the population formula for the parameters:

$$\beta = (E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i X_i'])^{-1} E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i Y_i]$$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $Z = (Z_1', \ldots, Z_n')$
- Matrix party trick: $X'Z/n = (1/n)\sum_i^N X_i Z_i' \xrightarrow{p} E[X_i Z_i']$.
- Take the population formula for the parameters:

$$\beta = (E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i X_i'])^{-1} E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i Y_i]$$

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

$$\hat{\beta} = [(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}(Z'X)(Z'Z)^{-1}(Z'Y)$$

How to estimate the parameters

- Collect X_i into a $n \times k$ matrix $X = (X_1', \ldots, X_n')$
- Collect Z_i into a $n \times l$ matrix $Z = (Z_1', \ldots, Z_n')$
- Matrix party trick: $X'Z/n = (1/n)\sum_i^N X_i Z_i' \xrightarrow{p} E[X_i Z_i']$.
- Take the population formula for the parameters:

$$\beta = (E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i X_i'])^{-1} E[Z_i X_i'](E[Z_i Z_i'])^{-1} E[Z_i Y_i]$$

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

$$\hat{\beta} = [(X'Z)(Z'Z)^{-1}(Z'X)]^{-1}(Z'X)(Z'Z)^{-1}(Z'Y)$$

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

Asymptotics for 2SLS

- Let $V = Z(Z'Z)^{-1}Z'X$ be the matrix of fitted values for X, then we have

$$\hat{\beta} = (V'V)^{-1}V'Y$$

Asymptotics for 2SLS

- Let $V = Z(Z'Z)^{-1}Z'X$ be the matrix of fitted values for X, then we have
$$\hat{\beta} = (V'V)^{-1}V'Y$$
- We can insert the true model for Y:
$$\hat{\beta} = (V'V)^{-1}V'(X\beta + \varepsilon)$$

Asymptotics for 2SLS

- Let $V = Z(Z'Z)^{-1}Z'X$ be the matrix of fitted values for X, then we have
$$\hat{\beta} = (V'V)^{-1}V'Y$$
- We can insert the true model for Y:
$$\hat{\beta} = (V'V)^{-1}V'(X\beta + \varepsilon)$$
- Using the matrix party trick and that $V'X = V'V$, we have
$$\begin{aligned}\hat{\beta} &= (V'V)^{-1}V'X\beta + (V'V)^{-1}V'\varepsilon \\ &= \beta + \left[n^{-1}\sum_i V_i V_i'\right]^{-1} n^{-1}\sum_i V_i\varepsilon_i\end{aligned}$$

Asymptotics for 2SLS

- Let $V = Z(Z'Z)^{-1}Z'X$ be the matrix of fitted values for X, then we have
$$\hat{\beta} = (V'V)^{-1}V'Y$$
- We can insert the true model for Y:
$$\hat{\beta} = (V'V)^{-1}V'(X\beta + \varepsilon)$$
- Using the matrix party trick and that $V'X = V'V$, we have
$$\begin{aligned}\hat{\beta} &= (V'V)^{-1}V'X\beta + (V'V)^{-1}V'\varepsilon \\ &= \beta + \left[n^{-1}\sum_i V_i V_i'\right]^{-1} n^{-1}\sum_i V_i\varepsilon_i\end{aligned}$$
- Consistent because $n^{-1}\sum_i V_i\varepsilon_i \xrightarrow{p} E[V_i\varepsilon_i] = 0$.

Asymptotic variance for 2SLS

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

- By the CLT, $n^{-1/2} \sum_i V_i \varepsilon_i$ converges in distribution to $N(0, B)$, where $B = E[V_i' \varepsilon_i' \varepsilon_i V_i]$.

Asymptotic variance for 2SLS

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

- By the CLT, $n^{-1/2} \sum_i V_i \varepsilon_i$ converges in distribution to $N(0, B)$, where $B = E[V_i' \varepsilon_i' \varepsilon_i V_i]$.
- By the LLN, $n^{-1} \sum_i V_i V_i' \xrightarrow{p} E[V_i' V_i]$.

Asymptotic variance for 2SLS

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

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

$$(E[V_i' V_i])^{-1} E[V_i' \varepsilon_i' \varepsilon_i V_i] (E[V_i' V_i])^{-1}$$

Asymptotic variance for 2SLS

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

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

$$(E[V_i' V_i])^{-1} E[V_i' \varepsilon_i' \varepsilon_i V_i] (E[V_i' V_i])^{-1}$$

- Replace with the sample quantities to get estimates:

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

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

Overidentification

- What if we have more instruments than endogenous variables?

Overidentification

- What if we have more instruments than endogenous variables?
- When there are more instruments than causal parameters ($l > k$), the model is **overidentified**.

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**.

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.

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 test, Hansen test, J-test, etc.

Overidentification tests

- Sargan test, Hansen test, J-test, etc.
- Basic idea: under null that all instruments are good, running it with different subset of the instruments should only differ due to sampling noise.

Overidentification tests

- Sargan test, Hansen test, J-test, etc.
- Basic idea: under null that all instruments are good, running it with different subset of the instruments should only differ due to sampling noise.
- Identify the distribution of that noise under the null to develop a test.

Overidentification tests

- Sargan test, Hansen test, J-test, etc.
- Basic idea: under null that all instruments are good, running it with different subset of the instruments should only differ due to sampling noise.
- Identify the distribution of that noise under the null to develop a test.
- 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.

Overidentification tests

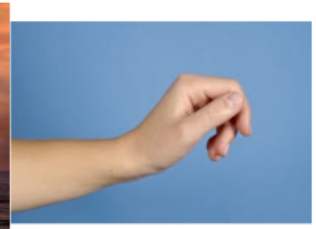
- Sargan test, Hansen test, J-test, etc.
- Basic idea: under null that all instruments are good, running it with different subset of the instruments should only differ due to sampling noise.
- Identify the distribution of that noise under the null to develop a test.
- If we reject the null hypothesis 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

Overidentification tests

- Sargan test, Hansen test, J-test, etc.
- Basic idea: under null that all instruments are good, running it with different subset of the instruments should only differ due to sampling noise.
- Identify the distribution of that noise under the null to develop a test.
- 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
- Once we move away from constant effects, we no longer can generally pool multiple instruments together in this way.

Reading

Reading



Instrumental Variables and Potential Outcomes

- The basic idea behind instrumental variable approaches is that we do not have ignorability for A_i, but we do have a variable, Z_i, that affects A_i, but only affects the outcome through A_i.

Instrumental Variables and Potential Outcomes

- The basic idea behind instrumental variable approaches is that we do not have ignorability for A_i, but we do have a variable, Z_i, that affects A_i, but only affects the outcome through A_i.
- Note that we allow the instrument, Z_i to have an effect on A_i, so the treatment must have potential outcomes, $A_i(1)$ and $A_i(0)$, with the usual consistency assumption:

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

Instrumental Variables and Potential Outcomes

- The basic idea behind instrumental variable approaches is that we do not have ignorability for A_i, but we do have a variable, Z_i, that affects A_i, but only affects the outcome through A_i.
- Note that we allow the instrument, Z_i to have an effect on A_i, so the treatment must have potential outcomes, $A_i(1)$ and $A_i(0)$, with the usual consistency assumption:

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

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

Instrumental Variables and Potential Outcomes

- The basic idea behind instrumental variable approaches is that we do not have ignorability for A_i, but we do have a variable, Z_i, that affects A_i, but only affects the outcome through A_i.
- Note that we allow the instrument, Z_i to have an effect on A_i, so the treatment must have potential outcomes, $A_i(1)$ and $A_i(0)$, with the usual consistency assumption:

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

- Outcome can depend on both the treatment and the instrument: $Y_i(a, z)$ is the outcome if unit i had received treatment $A_i = a$ and instrument value $Z_i = z$.
- The effect of the treatment given the value of the instrument is

$$Y_i(1, Z_i) - Y_i(0, Z_i)$$

.

Key assumptions

1. Randomization

Key assumptions

1. Randomization
2. Exclusion Restriction

Key assumptions

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

Key assumptions

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

Randomization

- Need the instrument to be randomized:

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

Randomization

- Need the instrument to be randomized:

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

- We can weaken this to conditional ignorability

Randomization

- Need the instrument to be randomized:

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

- We can weaken this to conditional ignorability
- But why believe conditional ignorability for the instrument but not the treatment?

Randomization

- Need the instrument to be randomized:

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

- We can weaken this to conditional ignorability
- But why believe conditional ignorability for the instrument but not the treatment?
- Best instruments are truly randomized.

Randomization

- Need the instrument to be randomized:

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

- 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**:

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

Exclusion Restriction

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

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

Exclusion Restriction

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

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

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

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

Exclusion Restriction

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

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

- 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

Exclusion Restriction

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

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

- 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

Exclusion Restriction

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

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

- 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

Exclusion Restriction

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

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

- 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

Exclusion Restriction

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

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

- 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

- Rewriting the usual consistency assumption gives us a linear model with heterogeneous effects (we have seen this before in randomized experiments):

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

The linear model with heterogeneous effects

- Rewriting the usual consistency assumption gives us a linear model with heterogeneous effects (we have seen this before in randomized experiments):

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

- 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.

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

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.

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

- Otherwise, what would we be doing? The instrument wouldn't affect anything.

Monotonicity

- Lastly, we need to make another assumption about the relationship between the instrument and the treatment.

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:

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

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:

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

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

Monotonicity means no defiers

- This is sometimes called "no defiers". It turns out that with a binary treatment and a binary instrument, we can group units into four categories:

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

Monotonicity means no defiers

- This is sometimes called "no defiers". It turns out that with a binary treatment and a binary instrument, we can group units into four categories:

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

- These compliance groups are sometimes called "principal strata."

Monotonicity means no defiers

- This is sometimes called "no defiers". It turns out that with a binary treatment and a binary instrument, we can group units into four categories:

Name	$A_i(1)$	$A_i(0)$
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.

Monotonicity means no defiers

- This is sometimes called "no defiers". It turns out that with a binary treatment and a binary instrument, we can group units into four categories:

Name	$A_i(1)$	$A_i(0)$
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 $A_i = 1$ when $Z_i = 0$ must be an always-taker and anyone with $A_i = 0$ 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).

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.

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:

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

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:

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

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

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

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

$$E[(Y_i(1) - Y_i(0))(A_i(1) - A_i(0))]$$

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

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

$$E[(Y_i(1) - Y_i(0))(A_i(1) - A_i(0))]$$

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

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

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

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

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

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

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

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

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

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

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

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

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

Proof (continued)

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

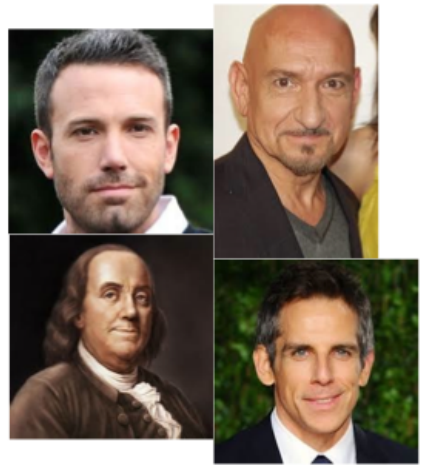
- We can use the same argument for the denominator:

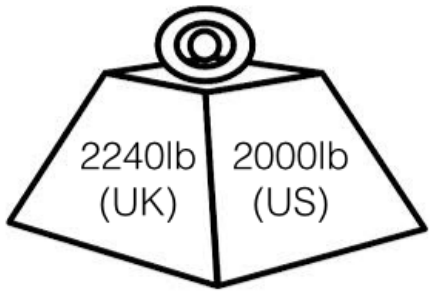
$$\begin{aligned} E[A_i|Z_i = 1] - E[A_i|Z_i = 0] &= E[A_i(1) - A_i(0)] \\ &= \Pr[A_i(1) > A_i(0)] \end{aligned}$$

- Dividing these two expressions through gives the LATE.

Reading

Reading





Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.

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. Among treated units with $Z_i = 1$, we cannot distinguish them from the always-takers and similarly for the control units with $Z_i = 0$.

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. Among treated units with $Z_i = 1$, we cannot distinguish them from the always-takers and similarly for the control units with $Z_i = 0$.
- Without further assumptions, this estimand is not equal to overall treatment effect or the treatment effect on the treated.

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. Among treated units with $Z_i = 1$, we cannot distinguish them from the always-takers and similarly for the control units with $Z_i = 0$.
- Without further assumptions, this estimand is not equal to overall treatment effect or the treatment effect on the treated.
- Furthermore, since the complier group depends on the instrument, an IV estimate with one instrument will generally estimate a different quantity than an IV estimate of the *same effect* with a different instrument.

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. Among treated units with $Z_i = 1$, we cannot distinguish them from the always-takers and similarly for the control units with $Z_i = 0$.
- Without further assumptions, this estimand is not equal to overall treatment effect or the treatment effect on the treated.
- Furthermore, since the complier group depends on the instrument, an IV estimate with one instrument will generally estimate a different quantity than an IV estimate of the *same effect* with a different instrument.
- 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?

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.

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 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)

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 (A_i)

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 (A_i)
- **One-sided noncompliance**: only those assigned to treatment (control) can actually take the treatment (control). Or

$$\Pr[A_i = 1 | Z_i = 0] = 0$$

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 (A_i)
- **One-sided noncompliance**: only those assigned to treatment (control) can actually take the treatment (control). Or

$$\Pr[A_i = 1 | Z_i = 0] = 0$$

- 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

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

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

  (exclusion restriction + one-sided noncompliance)

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

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

  (exclusion restriction + one-sided noncompliance)

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

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

(exclusion restriction + one-sided noncompliance)

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

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

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

(exclusion restriction + one-sided noncompliance)

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

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

(randomization)

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$
\begin{aligned}
& E[Y_i(0) + (Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
& \text{(exclusion restriction + one-sided noncompliance)} \\
=& E[Y_i(0)|Z_i = 1] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
=& E[Y_i(0)] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)] \\
& \text{(randomization)} \\
=& E[(Y_i(1) - Y_i(0))A_i|Z_i = 1]
\end{aligned}
$$

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

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

(exclusion restriction + one-sided noncompliance)

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

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

(randomization)

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

$$= E[Y_i(1) - Y_i(0)|A_i = 1, Z_i = 1] \Pr[A_i = 1|Z_i = 1]$$

(law of iterated expectations + binary treatment)

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$
\begin{aligned}
& E[Y_i(0) + (Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
& \quad \text{(exclusion restriction + one-sided noncompliance)} \\
=& E[Y_i(0)|Z_i = 1] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
=& E[Y_i(0)] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)] \\
& \quad \text{(randomization)} \\
=& E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] \\
=& E[Y_i(1) - Y_i(0)|A_i = 1, Z_i = 1] \Pr[A_i = 1|Z_i = 1] \\
& \quad \text{(law of iterated expectations + binary treatment)} \\
=& E[Y_i(1) - Y_i(0)|A_i = 1] \Pr[A_i = 1|Z_i = 1] \\
& \quad \text{(one-sided noncompliance)}
\end{aligned}
$$

Benefits of one-sided noncompliance

- With this assumption, we know that there are no "always-takers" and since there are no defiers, anyone treated ($Z_i = 1$) that takes the treatment ($A_i = 1$) is a complier.
- Thus, we know that: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$
\begin{aligned}
& E[Y_i(0) + (Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
& \quad \text{(exclusion restriction + one-sided noncompliance)} \\
=& E[Y_i(0)|Z_i = 1] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)|Z_i = 0] \\
=& E[Y_i(0)] + E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] - E[Y_i(0)] \\
& \quad \text{(randomization)} \\
=& E[(Y_i(1) - Y_i(0))A_i|Z_i = 1] \\
=& E[Y_i(1) - Y_i(0)|A_i = 1, Z_i = 1] \Pr[A_i = 1|Z_i = 1] \\
& \quad \text{(law of iterated expectations + binary treatment)} \\
=& E[Y_i(1) - Y_i(0)|A_i = 1] \Pr[A_i = 1|Z_i = 1] \\
& \quad \text{(one-sided noncompliance)}
\end{aligned}
$$

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

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

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

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

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

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

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

- Thus, under the additional assumption of one-sided compliance, we can estimate the ATT using the usual IV approach .
- The ATT is a combination of the LATE and the ATE for the always-takers. If we remove the possibility of the always takers, then anyone who actually takes the treatment is a complier.

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

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

- Thus, under the additional assumption of one-sided compliance, we can estimate the ATT using the usual IV approach .
- The ATT is a combination of the LATE and the ATE for the always-takers. If we remove the possibility of the always takers, then anyone who actually takes the treatment is a complier.
- It's also easy to see that if we switch the direction of one-sided compliance, then we can esimate the average treatment effect for the controls.

Falsification tests

- The exclusion restriction cannot be tested directly, but it can be falsified.

Falsification tests

- The exclusion restriction cannot be tested directly, but it can be falsified.
- Under the exclusion restriction, Z_i only has an effect on Y_i because it has an effect on A_i.

Falsification tests

- The exclusion restriction cannot be tested directly, but it can be falsified.
- Under the exclusion restriction, Z_i only has an effect on Y_i because it has an effect on A_i.
- **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 A_i.

Falsification tests

- The exclusion restriction cannot be tested directly, but it can be falsified.
- Under the exclusion restriction, Z_i only has an effect on Y_i because it has an effect on A_i.
- **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 A_i.
- Because Z_i can't affect A_i, then the exclusion restriction implies that this falsification test should have 0 effect. If we find an effect, instrument is suspicious.

Falsification tests

- The exclusion restriction cannot be tested directly, but it can be falsified.
- Under the exclusion restriction, Z_i only has an effect on Y_i because it has an effect on A_i.
- **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 A_i.
- Because Z_i can't affect A_i, then the exclusion restriction implies that this falsification test should have 0 effect. If we find an effect, instrument is suspicious.
- 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:

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

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:

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

- Angrist and Pischke describe ways to calculate the difference between the compliers and overall population in terms of binary covariates.

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:

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

- Angrist and Pischke describe ways to calculate the difference between the compliers and overall population in terms of binary covariates.
- Abadie (2003) shows how to calculate the mean of any covariate in the complier group.

Multiple instruments

- Since each instrument implies a different complier group, each instrument estimates a causal effect for a different subset of the population.

Multiple instruments

- Since each instrument implies a different complier group, each instrument estimates a causal effect for a different subset of the population.
- Thus, if we had two instrument, then there would be two different LATEs, ρ_1 and ρ_2 for instruments Z_{1i} and Z_{2i}. We might try to use 2SLS to estimate an overall effect with these instruments with following first stage:

$$\hat{A}_i = \pi_1 Z_{1i} + \pi_2 Z_{2i}.$$

2SLS as weighted average

- In Angrist and Pischke, they show that the 2SLS estimator using these two instruments is a weighted sum of the two component LATEs:

$$\rho_{2SLS} = \psi\rho_1 + (1 - \psi)\rho_2,$$

where the weights are:

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

2SLS as weighted average

- In Angrist and Pischke, they show that the 2SLS estimator using these two instruments is a weighted sum of the two component LATEs:

$$\rho_{2SLS} = \psi\rho_1 + (1-\psi)\rho_2,$$

where the weights are:

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

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

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:

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

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:

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

- We would also have exclusion conditional on the covariates:

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

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:

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

- We would also have exclusion conditional on the covariates:

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

- Under these assumptions, Angrist and Pischke show that if you fully saturate the first stage and the second stage in the covariates, then 2SLS estimates a weighted average of the covariates-specific LATEs (very similar to regression).

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:

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

- We would also have exclusion conditional on the covariates:

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

- Under these assumptions, Angrist and Pischke show that if you fully saturate the first stage and the second stage in the covariates, 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.