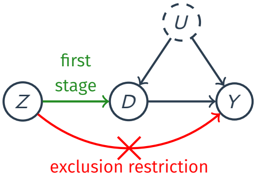
9. Two Stage Least Squares

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Where are we? Where are we going?

- Last time:
  - **Instrumental variable** under noncompliance in randomized experiments.
    - The local ATE (or CACE): $\tau_{\text{LATE}} = \text{ITT}_{Y,\text{co}} = \frac{\text{ITT}_Y}{\text{ITT}_D}$
    - The Wald/IV estimator: $\widehat{\tau}_{\text{iv}} = \widehat{\text{ITT}}_Y / \widehat{\text{ITT}}_D$
    - Intent-to-treat analysis, compliance types, identification assumptions..



- Today:
  1. Two stage least squares (TSLS)
  2. Types/Applications of IVs, e.g.,:
     - Regional characteristics (e.g., cellphone signal strength for app adoption)
     - Peer's environment (e.g., weather in friend's city for 'my' running behavior)

Source: *Chapter* 4 of Mostly Harmless Econometrics (Textbook 1) by J. Angrist & J. Pischke

1/ Basic Two-Stage Least Squares

Two Stage Least Squares

- **Two-stage least squares** (TSLS) is the classical approach to IV which assumes two linear models:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \eta_i$$

Two Stage Least Squares

- **Two-stage least squares** (TSLS) is the classical approach to IV which assumes two linear models:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \eta_i$$

- Here, the treatment D_i is **endogenous** so $\mathbb{E}[\varepsilon_i | D_i] \neq 0$.
- But we have an **instrument** Z_i that is exogenous $\mathbb{E}[\varepsilon_i | Z_i] = 0$.
    - It's also exogenous for treatment uptake, so $\mathbb{E}[\eta_i | Z_i] = 0$.

Two Stage Least Squares

- **Two-stage least squares** (TSLS) is the classical approach to IV which assumes two linear models:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \eta_i$$

- Here, the treatment D_i is **endogenous** so $\mathbb{E}[\varepsilon_i|D_i] \neq 0$.
- But we have an **instrument** Z_i that is exogenous $\mathbb{E}[\varepsilon_i|Z_i] = 0$.
  - It's also exogenous for treatment uptake, so $\mathbb{E}[\eta_i|Z_i] = 0$.
- This implies the following CEF form for Y_i conditional on Z_i:

$$\mathbb{E}[Y_i|Z_i] = \alpha + \tau\mathbb{E}[D_i|Z_i] = \alpha + \tau \cdot (\gamma Z_i)$$

TSLS Estimands

- Under the model, we have the following CEF: $\mathbb{E}[Y_i|Z_i] = \alpha + \tau \cdot (\gamma Z_i)$
  - $\rightsquigarrow$ A regression of Y_i on γZ_i would have τ as the slope.

TSLS Estimands

- Under the model, we have the following CEF: $\mathbb{E}[Y_i|Z_i] = \alpha + \tau \cdot (\gamma Z_i)$
  - $\rightsquigarrow$ A regression of Y_i on γZ_i would have τ as the slope.

- If the CEF is linear, then we have this simple relationship slopes:

$$\mathbb{E}[D_i|Z_i] = \delta + \gamma Z_i \qquad \rightsquigarrow \qquad \gamma = \frac{\text{cov}(D_i, Z_i)}{\mathbb{V}(Z_i)}$$

- Applying this to above CEF, we have:

$$\tau = \frac{\text{cov}(Y_i, \gamma Z_i)}{\mathbb{V}(\gamma Z_i)} = \frac{\text{cov}(Y_i, Z_i)}{\gamma \mathbb{V}(Z_i)} = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)}$$

TSLS Estimands

- Under the model, we have the following CEF: $\mathbb{E}[Y_i|Z_i] = \alpha + \tau \cdot (\gamma Z_i)$
    - $\rightsquigarrow$ A regression of Y_i on γZ_i would have τ as the slope.
- If the CEF is linear, then we have this simple relationship slopes:

$$\mathbb{E}[D_i|Z_i] = \delta + \gamma Z_i \qquad \rightsquigarrow \qquad \gamma = \frac{\text{cov}(D_i, Z_i)}{\mathbb{V}(Z_i)}$$

- Applying this to above CEF, we have:

$$\tau = \frac{\text{cov}(Y_i, \gamma Z_i)}{\mathbb{V}(\gamma Z_i)} = \frac{\text{cov}(Y_i, Z_i)}{\gamma \mathbb{V}(Z_i)} = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)}$$

- TSLS estimator:
    - Estimate $\widehat{\gamma}$ from regression of treatment D_i on instrument Z_i
    - Estimate $\widehat{\tau}_{2SLS}$ as the slope of a regression of Y_i on $\widehat{\gamma} Z_i$.
    - Under this model, $\widehat{\tau}_{2SLS} \xrightarrow{p} \tau$ (but don't use SEs from second stage; see MHE section 4.6.1. *2SLS Mistakes*)

Binary Treatment and Instrument

- Under binary treatment/instrument, TSLS estimand is the LATE:

$$\tau = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \frac{\text{ITT}_Y}{\text{ITT}_D} = \tau_{\text{LATE}}$$

Binary Treatment and Instrument

- Under binary treatment/instrument, TSLS estimand is the LATE:

$$\tau = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \frac{\text{ITT}_Y}{\text{ITT}_D} = \tau_{\text{LATE}}$$

- And that the TSLS estimator is the Wald estimator:

$$\widehat{\tau}_{\text{2SLS}} = \frac{\widehat{\text{cov}}(Y_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_0} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_D} = \widehat{\tau}_{\text{iv}}$$

$\rightsquigarrow$ Constant effects model is not required for TSLS in this setting.

Binary Treatment and Instrument

- Under binary treatment/instrument, TSLS estimand is the LATE:

$$\tau = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \frac{\text{ITT}_Y}{\text{ITT}_D} = \tau_{\text{LATE}}$$

- And that the TSLS estimator is the Wald estimator:

$$\widehat{\tau}_{\text{2SLS}} = \frac{\widehat{\text{cov}}(Y_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_0} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_D} = \widehat{\tau}_{\text{iv}}$$

$\rightsquigarrow$ Constant effects model is not required for TSLS in this setting.

- We need constant effects when we add covariates:

$$Y_i = \alpha + \tau D_i + \mathbf{X}_i'\beta_y + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \mathbf{X}_i'\beta_d + \eta_i$$

  - Otherwise, τ is an odd weighted function of causal effects and $\tau \neq \tau_{\text{LATE}}$.

Weak Instruments

- IV is unstable if instrument weakly affects treatment; $\text{cov}(D_i, Z_i) \approx 0$.
- **Example** completely irrelevant instrument:

$$\begin{aligned} Y_i &= \tau D_i + \varepsilon_i, & \mathbb{E}[\varepsilon_i | D_i] \neq 0 \\ D_i &= 0 \times Z_i + \eta_i, & \mathbb{E}[\epsilon_i | Z_i] = \mathbb{E}[\eta_i | Z_i] = 0 \end{aligned}$$

  - Note that we only assume mean independence, so $\text{cov}(D_i, Z_i)$ could be non-zero.

Weak Instruments

- IV is unstable if instrument weakly affects treatment; $\text{cov}(D_i, Z_i) \approx 0$.
- **Example** completely irrelevant instrument:

$$\begin{aligned} Y_i &= \tau D_i + \varepsilon_i, & \mathbb{E}[\varepsilon_i | D_i] \neq 0 \\ D_i &= 0 \times Z_i + \eta_i, & \mathbb{E}[\epsilon_i | Z_i] = \mathbb{E}[\eta_i | Z_i] = 0 \end{aligned}$$

  - Note that we only assume mean independence, so $\text{cov}(D_i, Z_i)$ could be non-zero.

- The bias of the Wald estimator:

$$\widehat{\tau}_{\text{iv}} - \tau = \frac{\widehat{\text{cov}}(\tau D_i + \varepsilon_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} - \tau = \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^n \eta_i Z_i} \xrightarrow{d} \underbrace{\frac{\text{cov}(\varepsilon_i, \eta_i)}{\mathbb{V}[\varepsilon_i]}}_{\text{bias}} + \underbrace{W_i}_{\text{Cauchy r.v.}}$$

Weak Instruments

- IV is unstable if instrument weakly affects treatment; $\text{cov}(D_i, Z_i) \approx 0$.
- **Example** completely irrelevant instrument:

$$\begin{aligned} Y_i &= \tau D_i + \varepsilon_i, & \mathbb{E}[\varepsilon_i | D_i] \neq 0 \\ D_i &= 0 \times Z_i + \eta_i, & \mathbb{E}[\epsilon_i | Z_i] = \mathbb{E}[\eta_i | Z_i] = 0 \end{aligned}$$

  - Note that we only assume mean independence, so $\text{cov}(D_i, Z_i)$ could be non-zero.

- The bias of the Wald estimator:

$$\widehat{\tau}_{\text{iv}} - \tau = \frac{\widehat{\text{cov}}(\tau D_i + \varepsilon_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} - \tau = \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} \eta_i Z_i} \xrightarrow{d} \underbrace{\frac{\text{cov}(\varepsilon_i, \eta_i)}{\mathbb{V}[\varepsilon_i]}}_{\text{bias}} + \underbrace{W_i}_{\text{Cauchy r.v.}}$$

- Inconsistent and asymptotically heavy tails (b/c of Cauchy)
  - When $Z \longrightarrow D$ effect is small but non-zero, we see similar behavior.

What to Do About Weak Instruments?

- Detecting weak instruments:
  - F-test on instruments (excluded from second stage): H_0: $\gamma = 0$.
  - Rule of thumb: bias is small when F-stat ≥ 10 (Stock & Yogo, 2005).
  - Correct coverage may require cutoff $F \geq 104.7$ (Lee et al. 2022).
  - The latter is a worst-case, typical data maybe okay with 10 cutoff.
  - With HC or cluster-robust SEs, report the Kleibergen–Paap F-stat. (e.g., see Barron et al. 2021)

What to Do About Weak Instruments?

- Detecting weak instruments:
  - F-test on instruments (excluded from second stage): H_0: $\gamma = 0$.
  - Rule of thumb: bias is small when F-stat ≥ 10 (Stock & Yogo, 2005).
  - Correct coverage may require cutoff $F \geq 104.7$ (Lee et al. 2022).
  - The latter is a worst-case, typical data maybe okay with 10 cutoff.
  - With HC or cluster-robust SEs, report the Kleibergen–Paap F-stat. (e.g., see Barron et al. 2021)
- Anderson-Rubin (1949) test (simplified setting, binary Z/D)
  - $H_0 : \tau = \tau_0$ equivalent to $H_0 : \text{ITT}_Y - \text{ITT}_D \cdot \tau_0$
  - Under the null, asymptotically we have:

$$g(\tau_0) = \widehat{\text{ITT}}_Y - \widehat{\text{ITT}}_D \tau_0 \sim N(0, \Omega(\tau_0))$$

$$\Omega(\tau_0) = \mathbb{V}[\widehat{\text{ITT}}_Y] + \tau_0^2 \mathbb{V}[\widehat{\text{ITT}}_D] - 2\tau_0 \text{cov}(\widehat{\text{ITT}}_Y, \widehat{\text{ITT}}_D)$$

  - AR test statistic: $g(\tau_0)^2 / \Omega(\tau_0) \sim \chi^2$ no matter first-stage effect.
  - Can invert (analytically!) to get confidence intervals.

Multi-Valued Treatments

- Generalization of these ideas:
  - Multi-valued treatment: $D_i \in \{0, 1, \ldots, K-1\}$
  - Binary instrument: $Z_i \in \{0, 1\}$

Multi-Valued Treatments

- Generalization of these ideas:
  - Multi-valued treatment: $D_i \in \{0, 1, \ldots, K-1\}$
  - Binary instrument: $Z_i \in \{0, 1\}$
- Assumptions:
  - Randomization: $[\{Y_i(z,d), \forall z, d\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$
  - Monotonicity: $D_i(1) \geq D_i(0)$ (instrument only increases treatment)
  - Exclusion restriction: $Y_i(1,d) = Y_i(0,d)$ for all $d = 0, 1, \ldots, K-1$

Multi-Valued Treatments

- Generalization of these ideas:
  - Multi-valued treatment: $D_i \in \{0, 1, \ldots, K-1\}$
  - Binary instrument: $Z_i \in \{0, 1\}$
- Assumptions:
  - Randomization: $[\{Y_i(z,d), \forall z,d\},\ D_i(1),\ D_i(0)] \perp\!\!\!\perp Z_i$
  - Monotonicity: $D_i(1) \geq D_i(0)$ (instrument only increases treatment)
  - Exclusion restriction: $Y_i(1,d) = Y_i(0,d)$ for all $d = 0, 1, \ldots, K-1$
- Can't identify the proportion of all compliance types here.

Multi-Valued Treatments

- Generalization of these ideas:
  - Multi-valued treatment: $D_i \in \{0, 1, \ldots, K-1\}$
  - Binary instrument: $Z_i \in \{0, 1\}$
- Assumptions:
  - Randomization: $[\{Y_i(z,d), \forall z, d\},\ D_i(1),\ D_i(0)] \perp\!\!\!\perp Z_i$
  - Monotonicity: $D_i(1) \geq D_i(0)$ (instrument only increases treatment)
  - Exclusion restriction: $Y_i(1,d) = Y_i(0,d)$ for all $d = 0, 1, \ldots, K-1$
- Can't identify the proportion of all compliance types here.
- Example: $K = 3 \rightsquigarrow 9$ principal strata
  - Affected: $(D_i(0), D_i(1)) \in \{(0,1), (0,2), (1,2)\}$
  - Unaffected: $(D_i(0), D_i(1)) \in \{(0,0), (1,1), (2,2)\}$
  - Negatively affected: $(D_i(0), D_i(1)) \in \{(1,0), (2,0), (2,1)\}$
  - Last ruled out by monotonicity.
  - 5 unknowns and 4 knowns under monotonicity.

TSLS with Multivalued Treatments

- Let $C_i = jk$ be an indicator for compliance type $D_i(1) = j$ and $D_i(0) = k$.
    - People that are moved from k to j by the instrument.
    - Let $\rho_{jk} = \mathbb{P}(D_i(1) = j, D_i(0) = k)$ be the strata size.

TSLS with Multivalued Treatments

- Let $C_i = jk$ be an indicator for compliance type $D_i(1) = j$ and $D_i(0) = k$.
  - People that are moved from k to j by the instrument.
  - Let $\rho_{jk} = \mathbb{P}(D_i(1) = j, D_i(0) = k)$ be the strata size.

- We can show that the 2SLS estimator converges to:

$$\widehat{\tau}_{2SLS} \xrightarrow{p} \sum_{k=0}^{K-1} \sum_{j=k+1}^{K-1} \omega_{jk} \mathbb{E}\left(\frac{Y_i(1) - Y_i(0)}{j - k} \,\middle|\, C_i = jk \right)$$

$$\omega_{jk} = \frac{(j-k)\rho_{jk}}{\sum_{s=0}^{K-1} \sum_{t=s+1}^{K-1} (s-t)\rho_{st}}$$

TSLS with Multivalued Treatments

- Let $C_i = jk$ be an indicator for compliance type $D_i(1) = j$ and $D_i(0) = k$.
  - People that are moved from k to j by the instrument.
  - Let $\rho_{jk} = \mathbb{P}(D_i(1) = j, D_i(0) = k)$ be the strata size.
- We can show that the 2SLS estimator converges to:

$$\widehat{\tau}_{2SLS} \xrightarrow{p} \sum_{k=0}^{K-1} \sum_{j=k+1}^{K-1} \omega_{jk} \mathbb{E}\left(\frac{Y_i(1) - Y_i(0)}{j - k} \,\middle|\, C_i = jk \right)$$

$$\omega_{jk} = \frac{(j-k)\rho_{jk}}{\sum_{s=0}^{K-1} \sum_{t=s+1}^{K-1} (s-t)\rho_{st}}$$

- Intuition: a weighted average of effects per dose for each affected type.
  - Weights are proportional to size of the strata and how big the effect of the instrument is for that strata.
  - If instrument can only increase by 1 dose, then simplifies to weighted average of principal strata effects.

2/ General Two-Stage Least Squares

General 2SLS

- Linear model for each i:

$$Y_i = \mathbf{X}_i'\beta + \varepsilon_i$$

  - $\mathbf{X}_i$ is $k \times 1$ and now includes D_i and any pretreatment covariates.
  - Parts of $\mathbf{X}_i$ are endogenous so that $\mathbb{E}[\varepsilon_i \mid \mathbf{X}_i] \neq 0$

- Instruments $\mathbf{Z}_i$ that is $\ell \times 1$ vector such that $\mathbb{E}[\varepsilon_i \mid \mathbf{Z}_i] = 0$.
  - $\mathbf{Z}_i$ might include exogenous/pretreatment variables from $\mathbf{X}_i$ as well.
  - Rank condition: $\mathbb{E}[\mathbf{Z}_i\mathbf{Z}_i']$ and $\mathbb{E}[\mathbf{X}_i\mathbf{Z}_i']$ have full rank.

- Identification:
  - $k = \ell$: just-identified.
  - $k < \ell$: over-identified (can test the exclusion restriction, kinda)
  - $k > \ell$: unidentified (fails rank condition)

Nasty Matrix Algebra

- Projection matrix projects values of $\mathbf{X}_i$ onto $\mathbf{Z}_i$:

$$\mathbf{\Pi} = \left(\mathbb{E}[\mathbf{Z}_i\mathbf{Z}_i']\right)^{-1}\mathbb{E}[\mathbf{Z}_i\mathbf{X}_i'] \quad \text{(projection matrix, i.e., pi/}\mathbf{\Pi}\text{)}$$
$$\tilde{\mathbf{X}}_i = \mathbf{\Pi}'\mathbf{Z}_i \quad \text{(projected values)}$$

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

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

How to Estimate the Parameters

- Collect $\mathbf{X}_i$ into an $n \times k$ matrix $\mathbb{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_n')$
- Collect $\mathbf{Z}_i$ into an $n \times \ell$ matrix $\mathbb{Z} = (\mathbf{Z}_1', \ldots, \mathbf{Z}_n')$

- In-sample projection matrix produces fitted values:

$$\widehat{\mathbb{X}} = \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{X}$$

  - Fitted values of the regression of $\mathbb{X}$ on $\mathbb{Z}$.
  - Matrix party trick: $\mathbb{X}'\mathbb{Z}/n = (1/n)\sum_i^n \mathbf{X}_i\mathbf{Z}_i' \overset{p}{\to} \mathbb{E}[\mathbf{X}_i\mathbf{Z}_i']$.

- Take the population formula for the parameters:

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

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

$$\widehat{\beta}_{2SLS} = (\widehat{\mathbb{X}}'\mathbb{X})^{-1}\widehat{\mathbb{X}}'\mathbf{y} \overset{p}{\to} \beta$$

- This is how R/`Stata` estimate the 2SLS parameters.

Asymptotic Variance for 2SLS

- We can write the centered, normalized TSLS estimator as:

$$\sqrt{n}\left(\widehat{\beta}_{2\text{SLS}} - \beta\right) = \underbrace{\left(n^{-1}\sum_i \widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'\right)^{-1}}_{\xrightarrow{p} (\mathbb{E}[\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'])^{-1}} \underbrace{\left(n^{-1/2}\sum_i \widehat{\mathbf{X}}_i\varepsilon_i\right)}_{\xrightarrow{d} N\left(0, \mathbb{E}[\widehat{\mathbf{X}}_i'\varepsilon_i'\varepsilon_i\widehat{\mathbf{X}}_i]\right)}$$

- Thus, $\sqrt{n}\left(\widehat{\beta}_{2\text{SLS}} - \beta\right)$ has asymptotic variance:

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

- **Robust 2SLS variance estimator** with residuals $\widehat{u}_i = Y_i - \mathbf{X}_i'\widehat{\beta}$:

$$\widehat{\text{Var}}\left(\widehat{\beta}_{2\text{SLS}}\right) = (\widehat{\mathbb{X}}'\widehat{\mathbb{X}})^{-1}\left(\sum_i \widehat{u}_i^2\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'\right)(\widehat{\mathbb{X}}'\widehat{\mathbb{X}})^{-1}$$

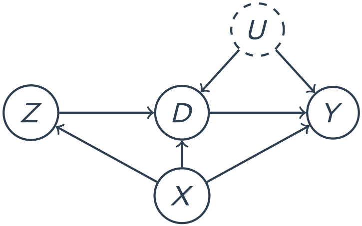
- HC2, clustering, and autocorrelation versions exist.

IV in Observational Studies and TSLS

- Motivation: what if there is unmeasured confounding?

- In observational studies where:
    1. treatment is not randomized and there exist unmeasured confounders;
    2. can find instrumental variable;
    3. exogenous covariates ($\mathbf{X}_i$): may exist observable confounders between Z_i, D_i, and Y_i

- DAG example:



Property Rights & Economic Development

Created using DALL·E 3 motivated by Ratna Sagar Shrestha

Recognize the person on the right?

- Q: Do property rights (i.e., institutions) promote economic development?
  - Famous paper on this: Acemoglu, Johnson, and Robinson (2001) AER
  - Relationship between strength of property rights in a country and GDP.

The AJR Data

Name	Description
shortnam	three-letter country code
africa	indicator for if the country is in Africa
asia	indicator for if country is in Asia
logem4	log mortality rates faced by European settlers (IV)
avexpr	strength of property rights (protection against expropriation)
logpgp95	log GDP per capita

```
> ajr <- read_csv("https://bit.ly/3RUJDWK"); ajr

# A tibble: 163 × 15
   shortnam africa lat_abst malfal94 avexpr logpgp95 logem4  asia yellow baseco leb95
   <chr>     <dbl>    <dbl>    <dbl>  <dbl>    <dbl>  <dbl> <dbl>  <dbl>  <dbl> <dbl>
 1 AFG           0   0.367   0.00372 NA       NA       4.54     1      0     NA  NA
 2 AGO           1   0.137   0.950    5.36     7.77    5.63     0      1      1  46.5
 3 ARE           0   0.267   0.0123   7.18     9.80   NA        1      0     NA  NA
 4 ARG           0   0.378   0        6.39     9.13    4.23     0      0      1  72.9
 5 ARM           0   0.444   0       NA        7.68   NA        1      0     NA  NA
 6 AUS           0   0.300   0        9.32     9.90    2.15     0      1      1  78.2
 7 AUT           0   0.524   0        9.73     9.97   NA        0      0     NA  NA
 8 AZE           0   0.448   0       NA        7.31   NA        1      0     NA  NA
 9 BDI           1   0.0367  0.950   NA        6.57    5.63     0      1     NA  NA
10 BEL           0   0.561   0        9.68     9.99   NA        0      0     NA  NA
#  153 more rows
#  4 more variables: imr95 <dbl>, meantemp <dbl>, lt100km <dbl>, latabs <dbl>
#  Use `print(n = ...)` to see more rows
```

In R: Example Code Using ajr Data

```r
# Center (i.e., demeaning) the variables
> ajr <- ajr |>
    mutate(
      D_cnt = avexpr - mean(avexpr, na.rm = TRUE),
      Y_cnt = logpgp95 - mean(logpgp95, na.rm = TRUE),
      Z_cnt = logem4 - mean(logem4, na.rm = TRUE)
    ) |>
    na.omit()

# Compute the ITTs on D and on Y:
> ITT_D <- cov(ajr$Z_cnt, ajr$D_cnt)
> ITT_Y <- cov(ajr$Z_cnt, ajr$Y_cnt)
> wald_estimate <- ITT_Y / ITT_D; round(wald_estimate, digits = 4)
[1] 0.9242

# Same as reg Y~Z / reg D~Z:
> ITT_Y <- coef(lm(Y_cnt ~ Z_cnt, data = ajr))[[2]]
> ITT_D <- coef(lm(D_cnt ~ Z_cnt, data = ajr))[[2]]
> wald_estimate <- ITT_Y / ITT_D; round(wald_estimate, digits = 4)
[1] 0.9242
```

- The Wald estimate from manual calculation: 0.9242

In R: Example Code Using ajr Data

```
# Compare with ivreg
> ivreg_result <- AER::ivreg(Y_cnt ~ D_cnt | Z_cnt, data = ajr); summary(ivreg_result)

Call:
AER::ivreg(formula = Y_cnt ~ D_cnt | Z_cnt, data = ajr)

Residuals:
     Min       1Q   Median       3Q      Max
-2.40175 -0.54950  0.01792  0.68944  1.67361

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.031e-16  1.231e-01   0.000        1
D_cnt       9.242e-01  1.547e-01   5.974 1.59e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9458 on 57 degrees of freedom
Multiple R-Squared: 0.1107,        Adjusted R-squared: 0.09506
Wald test: 35.68 on 1 and 57 DF,  p-value: 1.589e-07

> round(coef(ivreg_result), digits = 4)
(Intercept)       D_cnt
     0.0000      0.9242
```

- The 2sls estimate from `AER::ivreg()`: 0.9242
    - Caveat: compute robust SE separately! (don’t use SE from 2nd stage)

In R: Example Code Using ajr Data

```
# Estimate robust SE with estimatr::iv_robust()
> iv_rob <- estimatr::iv_robust(Y_cnt ~ D_cnt | Z_cnt, data = ajr, se_type = "HC2"); iv_rob
               Estimate Std. Error      t value     Pr(>|t|)   CI Lower CI Upper DF
(Intercept) 4.148820e-16  0.1234908 3.359619e-15 1.000000e+00 -0.2472860 0.247286 57
D_cnt       9.242173e-01  0.1791511 5.158871e+00 3.262823e-06  0.5654735 1.282961 57
```

- Robust SE with HC2 using `estimatr::iv_robust()`: 0.1791

```
> iv_rob_cov <- estimatr::iv_robust(Y_cnt ~ D_cnt + lat_abst + asia + africa |
                                    Z_cnt  + lat_abst + asia + africa, data = ajr,
                                  se_type = "HC2")
> var_labels <- c(
   "D_cnt" = "Avg. Expropriation Risk (D_cnt)", "lat_abst" = "Abs. Value of Latitude",
   "asia" = "Asian country", "africa" = "African country", "Z_cnt" = "log Mortality Rate (D_cnt)"
 )
```

- When adjusting for covariates in TSLS, include them both in the 1st and 2nd stage.

IV Regression Table with `modelsummary`

```
> modelsummary::modelsummary(
   models = list(ivreg_result, iv_robust, iv_rob_cov),
   coef_map = var_labels, gof_map = c("nobs", "r.squared", "adj.r.squared"), stars = T,
   notes = "Note: See appendix for other model statistics.", output = "modelsummary_tab.tex")
```

	(1)	(2)	(3)
Avg. Expropriation Risk (D_cnt)	0.924***	0.924***	1.015**
	(0.155)	(0.179)	(0.351)
Abs. Value of Latitude			-1.596
			(1.529)
Asian country			-1.048*
			(0.425)
African country			-0.390
			(0.342)
Num.Obs.	59	59	59
R2	0.111	0.111	0.045
R2 Adj.	0.095	0.095	-0.026

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: See appendix for other model statistics.

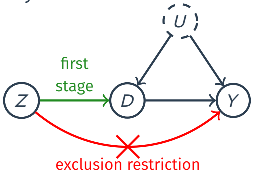
3/ Applications of IV

Source: Causal Inference for the Brave and True By Matheus Facure Alves

Treatment Assignments as IVs

Bloom et al. (*QJE* 2015)[1]

- Context: 9-month RCT in Ctrip's Shanghai call-centre; 249 volunteers entered the lottery.



Instrument Z: even-birthday lottery eligibility.

Treatment D: actually working from home ≥4 days/week.

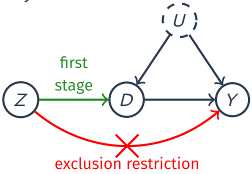
Outcome Y: weekly productivity index (calls, minutes logged-in, calls/minute).

[1] Bloom, N., Liang, J., Roberts, J. & Ying, Z. "Does Working from Home Work?" *QJE* 2015. 130(1): 165–218.

Treatment Assignments as IVs

Bloom et al. (*QJE* 2015)[1]

- Context: 9-month RCT in Ctrip's Shanghai call-centre; 249 volunteers entered the lottery.



Instrument Z: even-birthday lottery eligibility.

Treatment D: actually working from home ≥4 days/week.

Outcome Y: weekly productivity index (calls, minutes logged-in, calls/minute).

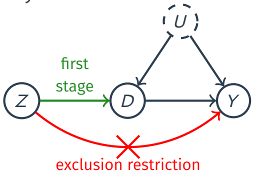
- Identification issue: Endogenous uptake; one-sided non-compliance (some winners revert, losers cannot WFH).

[1] Bloom, N., Liang, J., Roberts, J. & Ying, Z. "Does Working from Home Work?" *QJE* 2015. 130(1): 165–218.

Treatment Assignments as IVs

Bloom et al. (*QJE* 2015)[1]

- Context: 9-month RCT in Ctrip's Shanghai call-centre; 249 volunteers entered the lottery.



Instrument Z: even-birthday lottery eligibility.

Treatment D: actually working from home ≥4 days/week.

Outcome Y: weekly productivity index (calls, minutes logged-in, calls/minute).

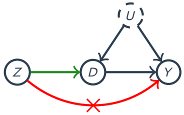
- Identification issue: Endogenous uptake; one-sided non-compliance (some winners revert, losers cannot WFH).
- IV logic:
  1. Randomization: Balance tests on 18 baseline characteristics show no joint differences; p=0.466.
  2. Relevance: 80-90% of winners comply (1st-stage F $\approx$ 23)
  3. Exclusion: tasks, pay, IT identical – only location changes.

[1] Bloom, N., Liang, J., Roberts, J. & Ying, Z. "Does Working from Home Work?" *QJE* 2015. 130(1): 165–218.

Regional Characteristics as IVs

Narang and Shankar (*Marketing Sci.* 2019)[2]

- Context: Observational panel on 32 mil. customers of a U.S. video-game / electronics retailer, Jan 2013 – Dec 2015; firm's mobile-shopping app launched July 2014.



Instrument Z: number of cell towers in shopper's ZIP.

Treatment D: adoption of the retailer's mobile app.

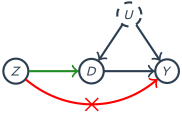
Outcome Y: Monthly on/offline purchase & return amounts.

[2] Narang & Shankar. "Mobile App Introduction and On- /Off-line Purchases and Product Returns." Marketing Sci. 2019. 38(5): 756 – 772.

Regional Characteristics as IVs

Narang and Shankar (*Marketing Sci.* 2019)[2]

- Context: Observational panel on 32 mil. customers of a U.S. video-game / electronics retailer, Jan 2013 – Dec 2015; firm's mobile-shopping app launched July 2014.



Instrument Z: number of cell towers in shopper's ZIP.

Treatment D: adoption of the retailer's mobile app.

Outcome Y: Monthly on/offline purchase & return amounts.

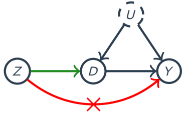
- Identification issue: Self-selection—tech-savvy or high-value shoppers more likely to install the app.

[2] Narang & Shankar. "Mobile App Introduction and On- /Off-line Purchases and Product Returns." Marketing Sci. 2019. 38(5): 756 – 772.

Regional Characteristics as IVs

Narang and Shankar (*Marketing Sci.* 2019)[2]

- Context: Observational panel on 32 mil. customers of a U.S. video-game / electronics retailer, Jan 2013 – Dec 2015; firm's mobile-shopping app launched July 2014.



Instrument Z: number of cell towers in shopper's ZIP.

Treatment D: adoption of the retailer's mobile app.

Outcome Y: Monthly on/offline purchase & return amounts.

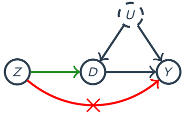
- Identification issue: Self-selection—tech-savvy or high-value shoppers more likely to install the app.
- IV logic:
    1. Randomization: 90% of cell-towers built before 2010 and shows only weak correlations with 2010–15 population growth and store count, included in $\mathbf{X}_i$.
    2. Relevance: Each cell tower raises the log-odds of adoption by 0.0022 (z=3.7).
    3. Exclusion: controlled for ZIP income, store distance, and competitor presence to block direct demand channels.
    4. Monotonicity: Better signal only increases the prob. of downloading the app (no defiers).

[2] Narang & Shankar. "Mobile App Introduction and On- /Off-line Purchases and Product Returns." Marketing Sci. 2019. 38(5): 756 – 772.

Peers' Environments as IVs

Aral and Nicolaides (*Nature Comm.* 2017)[3]

- Context: Global fitness-tracking network, 1.1 Mil. runners, 3.4 Mil. ties, 350 Mil. km logged over 5 years.



Instrument Z: daily weather (rain/extreme temperature) in friend's city.

Treatment D: friend's running distance that day.

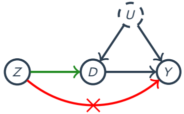
Outcome Y: user/ego's running distance (same or next day).

[3] Sinan A. & Nicolaides C. "Exercise contagion in a global social network." Nature Communications. 2017. 8: 14753.

Peers' Environments as IVs

Aral and Nicolaides (*Nature Comm.* 2017)[3]

- Context: Global fitness-tracking network, 1.1 Mil. runners, 3.4 Mil. ties, 350 Mil. km logged over 5 years.



Instrument Z: daily weather (rain/extreme temperature) in friend's city.

Treatment D: friend's running distance that day.

Outcome Y: user/ego's running distance (same or next day).

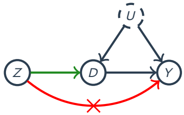
- Identification issue: Homophily & shared shocks—similar friends exercise together or face the same local weather.

[3] Sinan A. & Nicolaides C. "Exercise contagion in a global social network." Nature Communications. 2017. 8: 14753.

Peers' Environments as IVs

Aral and Nicolaides (*Nature Comm.* 2017)[3]

- Context: Global fitness-tracking network, 1.1 Mil. runners, 3.4 Mil. ties, 350 Mil. km logged over 5 years.



Instrument Z: daily weather (rain/extreme temperature) in friend's city.

Treatment D: friend's running distance that day.

Outcome Y: user/ego's running distance (same or next day).

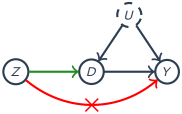
- Identification issue: Homophily & shared shocks—similar friends exercise together or face the same local weather.
- IV logic:
  1. Randomization: Use only friend pairs in different cities whose weather paths are uncorrelated; ego's own weather + date FE included.
  2. Relevance: first-stage F = 216–430 well above Stock-Yogo cutoff.
  3. Exclusion: Weather in friend's city is uncorrelated with ego's weather by construction.
  4. Monotonicity: bad weather never makes the friend run more.

[3] Sinan A. & Nicolaides C. "Exercise contagion in a global social network." Nature Communications. 2017. 8: 14753.

Bartik (Shift-Share) IVs

Barron et al. (*Marketing Sci.* 2020)[4]

- Context: Panel of 43,000 ZIP codes (100 largest U.S. metro areas), monthly 2011-2016. Airbnb listings scraped; Zillow rent & price indices matched at ZIP-month.



Instrument Z: Google-Trends "Airbnb" (global shock) × 2010 ZIP "touristiness" (# food/lodging firms)

Treatment D: $\ln(1 + \text{Airbnb listings})$ in ZIP-month.

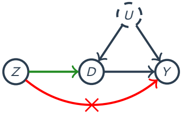
Outcome Y: Zillow rent index, house price index , price-to-rent.

[4] Barron, K., Kung, E., & Proserpio, D. "The Effect of Home-Sharing on House Prices and Rents" Marketing Sci. 2020. 40(2): 283–304.

Bartik (Shift-Share) IVs

Barron et al. (*Marketing Sci.* 2020)[4]

- Context: Panel of 43,000 ZIP codes (100 largest U.S. metro areas), monthly 2011-2016. Airbnb listings scraped; Zillow rent & price indices matched at ZIP-month.



Instrument Z: Google-Trends "Airbnb" (global shock) × 2010 ZIP "touristiness" (# food/lodging firms)

Treatment D: $\ln(1 + \text{Airbnb listings})$ in ZIP-month.

Outcome Y: Zillow rent index, house price index , price-to-rent.

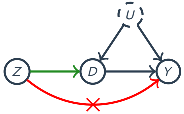
- Identification issue: Hot ZIPs attract both Airbnb supply and rising housing costs $\rightsquigarrow$ upward bias.

[4] Barron, K., Kung, E., & Proserpio, D. "The Effect of Home-Sharing on House Prices and Rents" Marketing Sci. 2020. 40(2): 283–304.

Bartik (Shift-Share) IVs

Barron et al. (*Marketing Sci.* 2020)[4]

- Context: Panel of 43,000 ZIP codes (100 largest U.S. metro areas), monthly 2011-2016. Airbnb listings scraped; Zillow rent & price indices matched at ZIP-month.



Instrument Z: Google-Trends "Airbnb" (global shock) × 2010 ZIP "touristiness" (# food/lodging firms)

Treatment D: $\ln(1 + \text{Airbnb listings})$ in ZIP-month.

Outcome Y: Zillow rent index, house price index , price-to-rent.

- Identification issue: Hot ZIPs attract both Airbnb supply and rising housing costs $\rightsquigarrow$ upward bias.
- IV logic:
  1. Relevance: first-stage F = 650–820 well above Stock-Yogo cutoff.
  2. Exclusion: global search shocks unrelated to local housing; touristiness fixed in 2010. City×month FE + placebo ZIPs (no listings) show no direct price effect.
  3. Monotonicity: more global Airbnb awareness cannot lower listings in touristy ZIPs.

[4] Barron, K., Kung, E., & Proserpio, D. "The Effect of Home-Sharing on House Prices and Rents" Marketing Sci. 2020. 40(2): 283–304.

Bartik (Shift–Share) IVs

Barron et al. (*Marketing Sci.* 2020)

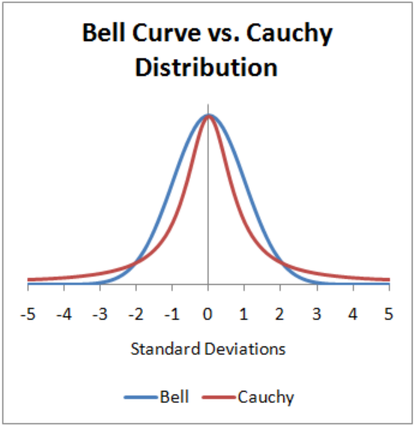
- IV logic: 4. conditional unconfoundedness/randomization
  - Parallel pre-trends: prior to 2012, rents and prices evolve identically across touristiness quartiles (Fig. 5) $\rightsquigarrow$ no pre-existing divergence.
  - Touristiness $\times$ time test: adding a ZIP-specific trend term $(h_{i,2010} \times t)$ yields an insignificant coefficient; IV effect unchanged.
  - Rich time-varying controls: results robust after adding ZIP income, population, hotel rooms, TripAdvisor reviews, and airport arrivals $\rightsquigarrow$ IV not picking up gentrification or tourism shocks.

Onto the presentations & discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

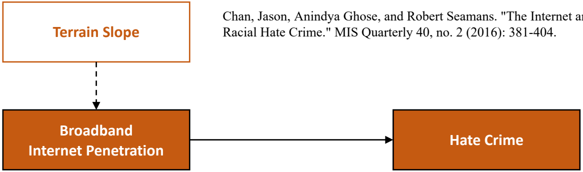
Cauchy vs. Normal Distribution



Source: https://stats.stackexchange.com/questions/36027/why-does-the-cauchy-distribution-have-no-mean

Regional Characteristics as IVs

- Chan, Ghose, Seamans, MISQ, 2016:

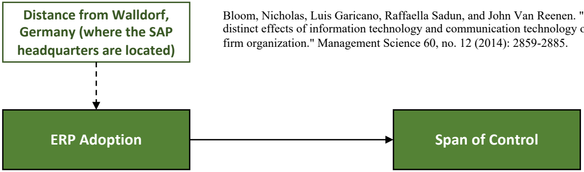


Chan, Jason, Anindya Ghose, and Robert Seamans. "The Internet and Racial Hate Crime." MIS Quarterly 40, no. 2 (2016): 381-404.

  - Observational study using *Hate Crime Statistics* from FBI.
    - RQ: Does the spread of Internet increase racial hate crime?
  - Issue? Confounding
  - IV? terrain slope/steepness (i.e., how many hills in a given region?)

Geographical Proximity-Based IVs

- RQ: Does the adoption of an ERP system impact plant manager autonomy (i.e., span of control)?
  - Span of control: the number of employees managed by supervisors or managers in an organization (high SoC = centralized).

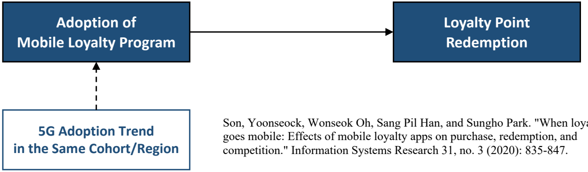


Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. "The distinct effects of information technology and communication technology on firm organization." Management Science 60, no. 12 (2014): 2859-2885.

  - Observational study: the CEP management and organization survey and the Harte-Hanks ICT panel (Bloom, Garicano, Sadun, Van Reenen, MgmtSci, 2014).
  - IV: Distance from the ERP market leader (i.e., SAP) w/ 25%+ market share ⇝ likely more established connections with SAP (German firms vs. French, England firms).

Macro/Cohort Trends as IVs

- RQ: How does loyalty app adoption affect customer point redemption behavior?

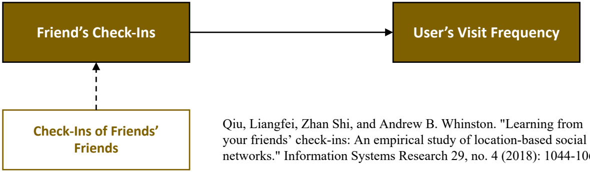


Son, Yoonseock, Wonseok Oh, Sang Pil Han, and Sungho Park. "When loyalty goes mobile: Effects of mobile loyalty apps on purchase, redemption, and competition." Information Systems Research 31, no. 3 (2020): 835-847.

  - Observational study: data on loyalty app adoption status, loyalty point redemption patterns, and purchase behaviors in multivendor loyalty program (MVLP) context (Son, Oh, Han, Park, ISR, 2020).
  - IV: 5G adoption rate in the same cohort (e.g., age group).

Peers’ Environments as IVs

- Qiu, Shi, Whinston, ISR, 2018:



Qiu, Liangfei, Zhan Shi, and Andrew B. Whinston. "Learning from your friends’ check-ins: An empirical study of location-based social networks." Information Systems Research 29, no. 4 (2018): 1044-1061.

  - Observational study using data on restaurant check-in information and the users’ social network ties from a major SNS in China.
    - RQ: Is there observational learning/herding effect for restaurant discovery?
  - Issue? Homophily.
  - IV: Check-in activities of friend’s friends.