Review: Linear just-identified IV

Two equations:

$$Y = X\beta + u$$
$$X = Z\pi + \nu$$

Review: Linear just-identified IV

Two equations:

$$Y = X\beta + u$$
$$X = Z\pi + \nu$$

Two key assumptions for IV validity:

Review: Linear just-identified IV

Two equations:

$$Y = X\beta + u$$
$$X = Z\pi + \nu$$

Two key assumptions for IV validity:

- Instrument relevance: $\text{Cov}(Z, X) \neq 0$ (strong first stage)

Review: Linear just-identified IV

Two equations:

$$Y = X\beta + u$$
$$X = Z\pi + \nu$$

Two key assumptions for IV validity:

- Instrument relevance: $\text{Cov}(Z, X) \neq 0$ (strong first stage)
- Exclusion restriction: $\text{Cov}(Z, u) = 0$ (instrument only affects Y through X)

Review: Linear just-identified IV

Two equations:

$$Y = X\beta + u$$
$$X = Z\pi + \nu$$

Two key assumptions for IV validity:

- Instrument relevance: $\text{Cov}(Z, X) \neq 0$ (strong first stage)
- Exclusion restriction: $\text{Cov}(Z, u) = 0$ (instrument only affects Y through X)

Recent literature has raised the bar for what counts as "plausible"

Three critical papers in early 2020s:

Three critical papers in early 2020s:

1. Mellon (2025, *AJPS*): rainfall used as IV in many papers $\Rightarrow$ violates exclusion restriction

Three critical papers in early 2020s:

1. Mellon (2025, *AJPS*): rainfall used as IV in many papers $\Rightarrow$ violates exclusion restriction

2. Lee et al. (2022, *AER*): valid inference requires $F > 105$ rather than $F > 10$

Three critical papers in early 2020s:

1. Mellon (2025, *AJPS*): rainfall used as IV in many papers $\Rightarrow$ violates exclusion restriction

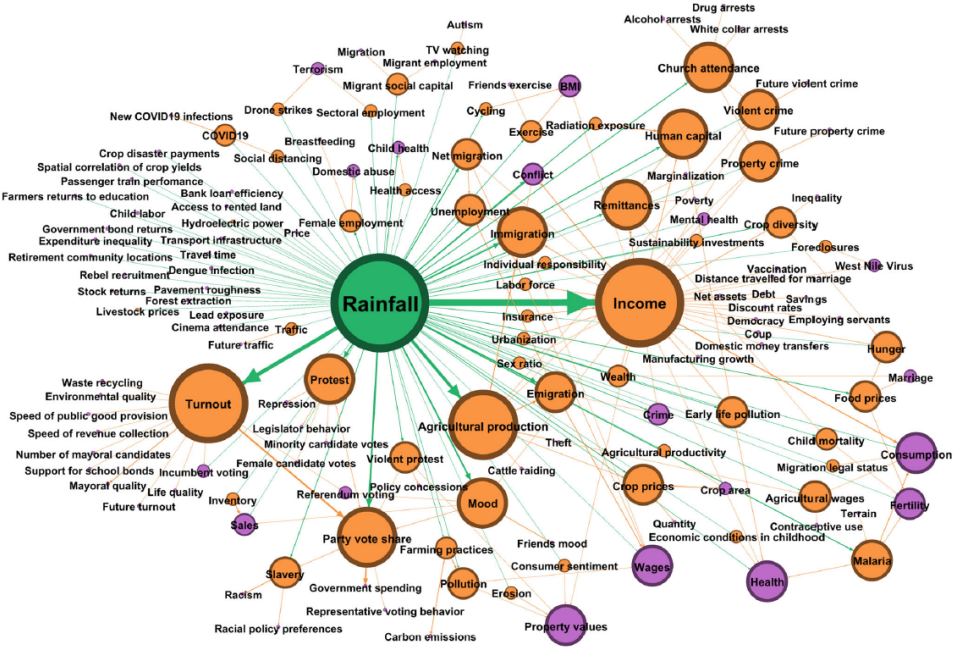
2. Lee et al. (2022, *AER*): valid inference requires $F > 105$ rather than $F > 10$

3. Young (2022, *EER*): AEA papers show IV results highly sensitive to outliers

Exclusion restriction violations

Exclusion restriction violations

- Mellon (2025) documents 192 papers using weather as IV


Rainfall
Income
Turnout
Protest
Agricultural production
Mood
Party vote share
Human capital
Remittances
Immigration
Unemployment
Net migration
Conflict
Child health
Breastfeeding
Sectoral employment
Drone strikes
New COVID19 infections
COVID19
Crop disaster payments
Spatial correlation of crop yields
Passenger train perfomance
Farmers returns to education
Bank loan efficiency
Access to rented land
Child labor
Hydroelectric power
Government bond returns
Expenditure inequality
Transport infrastructure
Retirement community locations
Travel time
Dengue infection
Rebel recruitment
Stock returns
Pavement toughness
Forest extraction
Livestock prices
Lead exposure
Cinema attendance
Traffic
Future traffic
Female employment
Price
Social distancing
Domestic abuse
Health access
Autism
Migration
TV watching
Terrorism
Migrant employment
Migrant social capital
Friends exercise
BMI
Cycling
Exercise
Radiation exposure
Church attendance
Drug arrests
Alcohol arrests
White collar arrests
Future violent crime
Violent crime
Future property crime
Property crime
Marginalization
Poverty
Inequality
Mental health
Crop diversity
Sustainability investments
Foreclosures
Vaccination
West Nile Virus
Distance travelled for marriage
Net assets
Debt
Savings
Discount rates
Democracy
Employing servants
Coup
Domestic money transfers
Hunger
Manufacturing growth
Marriage
Food prices
Wealth
Emigration
Crime
Early life pollution
Child mortality
Consumption
Migration legal status
Agricultural productivity
Theft
Crop prices
Crop area
Agricultural wages
Terrain
Fertility
Quantity
Economic conditions in childhood
Contraceptive use
Malaria
Health
Wages
Friends mood
Consumer sentiment
Farming practices
Pollution
Erosion
Slavery
Racism
Government spending
Representative voting behavior
Racial policy preferences
Property values
Carbon emissions
Cattle raiding
Policy concessions
Violent protest
Referendum voting
Female candidate votes
Incumbent voting
Minority candidate votes
Legislator behavior
Repression
Waste recycling
Environmental quality
Speed of public good provision
Speed of revenue collection
Number of mayoral candidates
Support for school bonds
Mayoral quality
Life quality
Future turnout
Inventory
Sales
Insurance
Urbanization
Sex ratio
Individual responsibility
Labor force

Exclusion restriction violations

- Mellon (2025) documents 192 papers using weather as IV
- Each pathway represents a potential exclusion-restriction violation

Exclusion restriction violations

- Mellon (2025) documents 192 papers using weather as IV
- Each pathway represents a potential exclusion-restriction violation
- For weather IVs, scholars must assume hundreds of additional identifying assumptions

Exclusion restriction violations

- Mellon (2025) documents 192 papers using weather as IV
- Each pathway represents a potential exclusion-restriction violation
- For weather IVs, scholars must assume hundreds of additional identifying assumptions
- Similar concerns likely apply to other commonly-used instruments

Problems in IV when using t-test for $H_0 : \beta = 0$:

Problems in IV when using t-test for $H_0 : \beta = 0$:

- Stock and Yogo (2005): $F > 10$ became standard pre-test for IV relevance

Problems in IV when using t-test for $H_0 : \beta = 0$:

- Stock and Yogo (2005): $F > 10$ became standard pre-test for IV relevance
- Lee et al. (2022): conventional t-ratio over-rejects even with $F > 10$

Problems in IV when using t-test for $H_0 : \beta = 0$:

- Stock and Yogo (2005): $F > 10$ became standard pre-test for IV relevance
- Lee et al. (2022): conventional t-ratio over-rejects even with $F > 10$
- Introduce *tF* procedure: adjust critical values for 2nd-stage t based on 1st-stage F

Problems in IV when using t-test for $H_0 : \beta = 0$:

- Stock and Yogo (2005): $F > 10$ became standard pre-test for IV relevance
- Lee et al. (2022): conventional t-ratio over-rejects even with $F > 10$
- Introduce *tF* procedure: adjust critical values for 2nd-stage t based on 1st-stage F
- For 25% of AER specifications, corrected standard errors 49% larger at 5% level

Statistical power in practice

Statistical power in practice

- Young (2022): IV has little power in published work

Statistical power in practice

- Young (2022): IV has little power in published work
- Despite producing different estimates, IV rarely rejects OLS point estimate

Statistical power in practice

- Young (2022): IV has little power in published work
- Despite producing different estimates, IV rarely rejects OLS point estimate
- Rarely rejects null that OLS is unbiased

Statistical power in practice

- Young (2022): IV has little power in published work
- Despite producing different estimates, IV rarely rejects OLS point estimate
- Rarely rejects null that OLS is unbiased
- Non-*iid* errors and high leverage worsen both size and power

Sensitivity to observations

Sensitivity to observations

- Young (2022): with deletion of one cluster, 39% of significant results become insignificant

Sensitivity to observations

- Young (2022): with deletion of one cluster, 39% of significant results become insignificant
- With deletion of two clusters, 58% become insignificant

Sensitivity to observations

- Young (2022): with deletion of one cluster, 39% of significant results become insignificant
- With deletion of two clusters, 58% become insignificant
- First-stage F-statistics similarly sensitive to outliers

Sensitivity to observations

- Young (2022): with deletion of one cluster, 39% of significant results become insignificant
- With deletion of two clusters, 58% become insignificant
- First-stage F-statistics similarly sensitive to outliers
- Problems concentrated in high-leverage specifications

Given these limitations, maybe we shouldn't do IV?

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

- Do OLS with lots and lots of covariates to eliminate OVB

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

- Do OLS with lots and lots of covariates to eliminate OVB
- Walk away and admit that it's impossible to obtain a causal estimate

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

- Do OLS with lots and lots of covariates to eliminate OVB
- Walk away and admit that it's impossible to obtain a causal estimate
- Use partial identification methods to put bounds on causal effects:

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

- Do OLS with lots and lots of covariates to eliminate OVB
- Walk away and admit that it's impossible to obtain a causal estimate
- Use partial identification methods to put bounds on causal effects:
    - "Upper" bound: biased OLS

Given these limitations, maybe we shouldn't do IV?

And if we're not going to do IV, what else can we do?

- Do OLS with lots and lots of covariates to eliminate OVB
- Walk away and admit that it's impossible to obtain a causal estimate
- Use partial identification methods to put bounds on causal effects:
  - "Upper" bound: biased OLS
  - "Lower" bound: plausible treatment effect