Instrumental Variable III

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- The previous lecture focuses on the economic tradition of using IVs.
- It is seen as an exogenous factor that may affect an agent's rational choice in the triangular system.
- Common approaches to use an IV include 2SLS and control function.
- There exists a deep connection between an IV and non-compliance conceptually.
- Results from 2SLS can be identical to that from the Wald estimator under certain conditions.

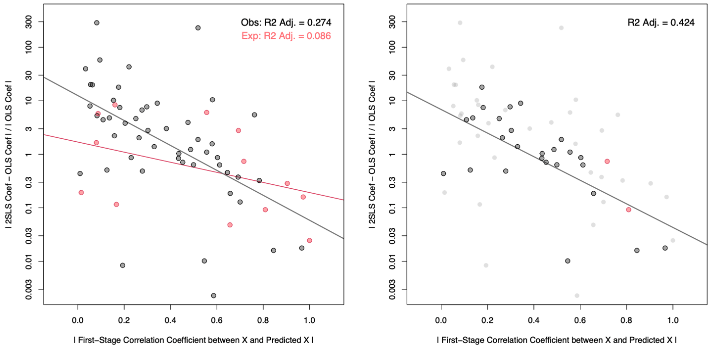
Testing assumptions in IV estimation

- We have seen that the LATE framework and the triangular system rely on similar assumptions:
  1. random assignment of the instrument.
  2. exclusion restriction.
  3. first stage.
  4. monotonicity.
- It is easier to ensure that all these assumptions are satisfied in experiments.
- In observational studies, these assumptions are less plausible.
- It explains why IV is becoming less popular in observational studies.
- Testing their validity is necessary.

Weak instruments

- Let's start from an assumption that seems easy to be satisfied: first stage.
- Remember that $\hat{\tau} = \frac{\hat{\xi}}{\hat{\delta}}$.
- If Z_i and D_i are only weakly correlated, $\hat{\delta}$ will be close to zero and the finite sample bias of $\hat{\tau}$ will be huge.
- Staiger and Stock (1997): If $\delta = \frac{c}{\sqrt{N}}$, then $\hat{\tau} \to \tau + \eta$, where η is the ratio between two normally distributed variables.
- The bias may even larger than that from OLS.
- Lal et al. (2023) show that this is a common problem in political science.
- The difference between 2SLS and OLS is larger when the IV is weaker.

Weak instruments



(a) Full Sample (b) Subsample with Significant OLS Results

Weak instruments

- A well-known example is Angrist and Krueger (1991).
- They instrument students' schooling years with their season of birth.
- But the correlation between the two variables is very weak.
- Bound, Jaeger, and Baker (1995) show that replacing the actual season of birth with randomly assigned ones does not change the estimate.

Weak instruments

- The confidence intervals of $\hat{\tau}$ will not have the correct coverage when the IV is weak.
- A traditional rule of thumb for a strong IV is that $F = \frac{\hat{\delta}^2}{\widehat{V}(\hat{\delta})}$ is larger than 10 (Staiger and Stock 1997).
- This is not the overall F-statistic from the first stage regression.
- If there are multiple instruments, we should test the joint null hypothesis that all these coefficients are zero.
- Yet Lee et al. (2022) show that this is not sufficient.
- If we impose no restrictions on the DGP, F must be larger than 104.7 to ensure the correct coverage.
- Or we use the critical value of 3.43.

Weak instruments

- Intuitively, the distribution of $\hat{\tau}$ deviates from normality significantly if $\hat{\delta}$ is not distinguishable from zero.
- The degree of deviation hinges on the value of F.
- There are several solutions to this problem.
- Anderson and Rubin (1949) propose to test the null hypothesis $\xi = \xi_0 = \delta\tau_0$ instead of $\tau = \tau_0$.
- We can see that $\hat{\xi} - \hat{\delta}\tau_0 \rightarrow \mathcal{N}(0, \Sigma)$ under the null, where

$$\Sigma = Var[\hat{\xi}] - 2\tau_0 Cov[\hat{\delta}, \hat{\xi}] + \tau_0^2 Var[\hat{\delta}].$$

- Such a test is valid even when $\hat{\delta}$ is zero.

Weak instruments

- The Anderson-Rubin (AR) test avoids the issue of taking the ratio but the result is hard to interpret.
- Lee et al. (2022) note the following relationship

$$t^2 = \frac{t_{AR}^2}{1 - 2\rho \frac{t_{AR}}{F} + \frac{t_{AR}^2}{F^2}},$$

  where ρ is the correlation between ν_i and ε_i.
- The relative performance of the t-test to the AR test is decided by the value of F.
- They hence suggest that we should combine the tests from the first stage and the second stage.
- For each value of F, they provide the corresponding critical value in the second stage.
- Or, we can impose more restrictions on the DGP, such as ρ.

Weak instruments: application

```
## The 2SLS estimate is 2.987609
## The first-stage F-statistic value is: 197.268
## The p-value from the Anderson-Rubin test is: 2.220446e-1
```

Test exclusion restriction

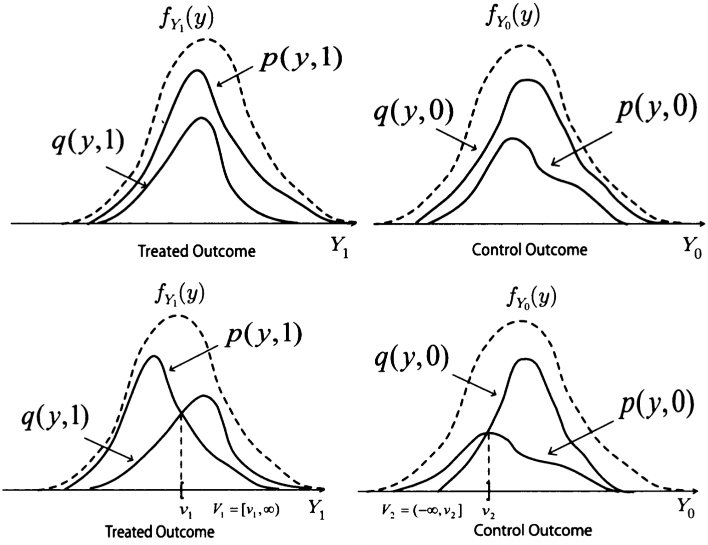
- Kitagawa (2015) and Huber and Mellace (2015) suggest that we may test exclusion restriction and monotonicity altogether.
- These two assumptions imply that

$$
\begin{aligned}
P(y, D=1|Z=1) &> P(y, D=1|Z=0) \\
P(y, D=0|Z=1) &< P(y, D=0|Z=0)
\end{aligned}
$$

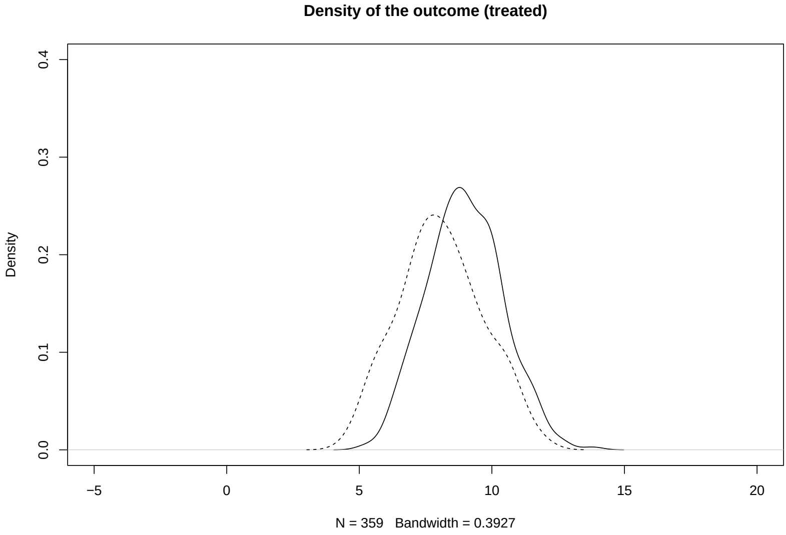
  for all y.
- We can test these two inequalities with data.
- To obtain critical values, we need to construct a variance-weighted Kolmogorov-Smirnov test statistic and apply bootstrap to it.
- But visualization alone can be helpful.

Test exclusion restriction

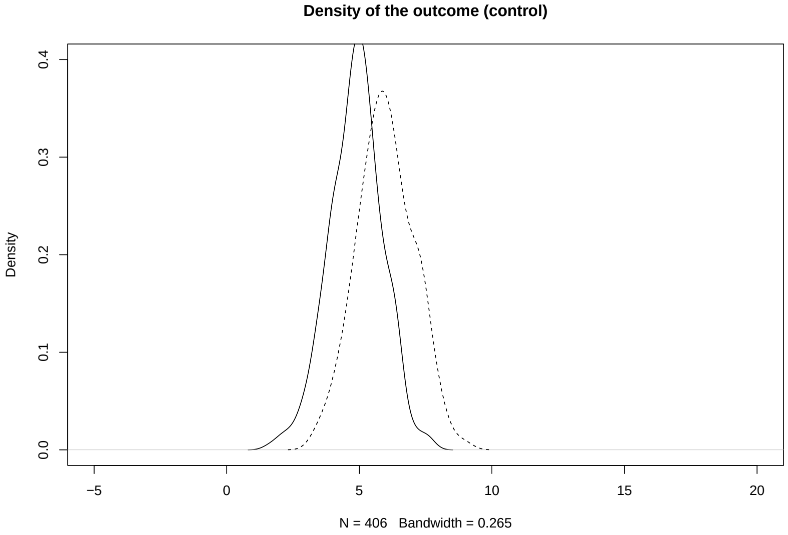
- We want to see the plots on the top rather than the ones on the bottom:



Test exclusion restriction: application



Test exclusion restriction: application



Exclusion restriction in practice

- It is common that exclusion restriction is violated in observational studies.
- Two popular instrumental variables in practice are rainfall and distance to the location where an event happened.
- Sarsons (2015) demonstrates that in Indian areas with dams, rainfall still affects conflicts even though it is not correlated with income.
- Zhao (2023) shows that using distance to any German city as the instrument can replicate the results in Becker and Woessmann (2009).
- The problem is that excluding all the other channels is more difficult than we thought.
- Glaeser et al. (2004) suggest that settlers' mortality may also influence the level of human capital in former colonies.

Generalizing the LATE

- It is reasonable to assume that the principal strata are determined by observable attributes of units.
- Abadie (2003) proposes a statistic (Abadie's κ):

$$\kappa_i = 1 - \frac{D_i(1 - Z_i)}{P(Z_i = 0|\mathbf{X}_i)} - \frac{(1 - D_i)Z_i}{P(Z_i = 1|\mathbf{X}_i)}.$$

- He shows that for any function $g(Y_i, D_i, \mathbf{X}_i)$,

$$\begin{aligned} & E[g(Y_i, D_i, \mathbf{X}_i)|D_i(1) > D_i(0)] \\ = & \frac{1}{P(D_i(1) > D_i(0))} E[\kappa_i g(Y_i, D_i, \mathbf{X}_i)]. \end{aligned}$$

- It means that we can estimate the average of any statistic over compliers by calculating its average weighted by κ_i.

Compliance score

- Under the same assumption, we can explicitly model the relationship between principal strata and covariates.
- We define $q_{ic} = q_c(\mathbf{X}_i)$ as the compliance score (Aronow and Carnegie 2013).
- We need to impose assumptions on the form of $q_c(\mathbf{X}_i)$.
- To estimate the ATE, we weight each observation with $\frac{1}{q_{ic}}$ and apply the Wald estimator.
- The intuition resembles the IPW estimator.

Compliance score

- Let's assume that

$$q_a(\mathbf{X}_i) + q_c(\mathbf{X}_i) = h(\mathbf{X}_i\beta_{a,c}),$$
$$\frac{q_a(\mathbf{X}_i)}{q_a(\mathbf{X}_i) + q_c(\mathbf{X}_i)} = h(\mathbf{X}_i\beta_{a|(a,c)}),$$

- We observe (D_i, Z_i) in the data with the likelihood

$$\prod_{i=1}^{N}(h(\mathbf{X}_i\beta_{a,c})(1 - h(\mathbf{X}_i\beta_{a|(a,c)}))Z_i + h(\mathbf{X}_i\beta_{a,c})h(\mathbf{X}_i\beta_{a|(a,c)}))^{D_i} *$$
$$(1 - h(\mathbf{X}_i\beta_{a,c})(1 - h(\mathbf{X}_i\beta_{a|(a,c)}))Z_i - h(\mathbf{X}_i\beta_{a,c})h(\mathbf{X}_i\beta_{a|(a,c)}))^{(1-D_i)}$$

- We estimate $\beta_{a,c}$ and $\beta_{a|(a,c)}$ via MLE and obtain the compliance score.

References I

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–63.

Anderson, Theodore W, and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics* 20 (1): 46–63.

Angrist, Joshua D, and Alan B Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4): 979–1014.

Aronow, Peter M, and Allison Carnegie. 2013. "Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable." *Political Analysis*, 492–506.

Becker, Sascha O, and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *The Quarterly Journal of Economics* 124 (2): 531–96.

References II

Bound, John, David A Jaeger, and Regina M Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90 (430): 443–50.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9: 271–303.

Huber, Martin, and Giovanni Mellace. 2015. "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints." *Review of Economics and Statistics* 97 (2): 398–411.

Kitagawa, Toru. 2015. "A Test for Instrument Validity." *Econometrica* 83 (5): 2043–63.

References III

Lal, Apoorva, Mac Lockhart, Yiqing Xu, and Ziwen Zu. 2023. "How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice Based on over 60 Replicated Studies." *arXiv Preprint arXiv:2303.11399*.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. "Valid t-Ratio Inference for IV." *American Economic Review* 112 (10): 3260–90.

Sarsons, Heather. 2015. "Rainfall and Conflict: A Cautionary Tale." *Journal of Development Economics* 115: 62–72.

Staiger, Douglas, and James H Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica: Journal of the Econometric Society*, 557–86.

Zhao, Qiyi C. 2023. "Rethinking'distance from': Lessons from Wittenberg and Mainz." *Available at SSRN 4495711*.