DS-UA 201: Causal Inference: More Instrumental Variables

Parijat Dube

New York University Center for Data Science

August 15, 2022

Acknowledgement: Slides including material from DS-US 201 Fall 2021 offered by Marco Morucci.

Experiments with non-compliance

Last lecture we talked about randomized experiments where units **do not comply** with the treatment.

- E.g., randomizing students to attend a class: some don't go.

We would like to estimate the **treatment effect**, However unobserved confounders could be affecting both compliance and outcome.

The setup:

- Z_i is our binary instrument
- $D_i(z)$ is our binary treatment, which now has potential outcomes
- $Y_i(d, z)$ are potential outcomes, defined in terms of both treatment and instrument

The IV assumptions

$$\begin{tikzcd} & U \arrow[dl, dashed] \arrow[dr, dashed] & \\ Z \arrow[r] & D \arrow[r] & Y \end{tikzcd}$$

We discussed the following assumptions:

- Randomization of instrument: $Z_i \perp\!\!\!\perp D_i(z)$, $Z_i \perp\!\!\!\perp Y_i(d, z)$
- Exclusion restriction: $Y_i(d, z) = Y_i(d)$
- First-stage relationship: $E[D_i(1) - D_i(0)] \neq 0$
- Monotonicity: $D_i(1) > D_i(0)$

The LATE

We showed that under the **IV assumptions** the **Local Average Treatment Effect** (LATE) is identified:

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

- This is the treatment effect on the **compliers**
- Likely not representative of the entire sample, but still useful

Today:

- How do we estimate the LATE?
- What are weak instruments?

Binary instrument

With a **binary instrument**, we can use the sample analog of the LATE:

$$\hat{\tau}_{\text{IV}} = \frac{\hat{\mathbb{E}}[Y_i|Z_i = 1] - \hat{\mathbb{E}}[Y_i|Z_i = 0]}{\hat{\mathbb{E}}[D_i|Z_i = 1] - \hat{\mathbb{E}}[D_i|Z_i = 0]},$$

Where:

$$\hat{E}[D_i|Z_i = z] = \frac{1}{N_{Z_i=z}} \sum_{i=1}^{n} D_i \mathbb{1}(Z_i = z),$$

and $\hat{E}[Y_i|Z_i = z]$ is defined in an analogous way.

- The numerator is an estimate of the **intent-to-treat** effect $\hat{\mathbb{E}}[Y_i|Z_i = 1] - \hat{\mathbb{E}}[Y_i|Z_i = 0]$
- The denominator is an estimate of the **first-stage** effect (effect of instrument on the treatment) $\hat{\mathbb{E}}[D_i|Z_i = 1] - \hat{\mathbb{E}}[D_i|Z_i = 0]$

The Wald estimator

The estimator we just saw can be **generalized** to settings where Z_i can have many different values:

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

Where $\widehat{Cov}(Y_i, Z_i)$ is the **sample covariance** of Y_i and Z_i and $\widehat{Cov}(D_i, Z_i)$ is the sample covariance of Z_i and D_i.

The Wald Estimator as a ratio of regressions

Recall that when we only have one regressor, i.e.: $Y_i = X_i\beta + \epsilon_i$, then the estimated regression coefficient can be written as:

$$\hat{\beta} = \frac{\widehat{Cov}(Y_i, X_i)}{\widehat{Var}(X_i)}$$

Because of this, we can write the **Wald estimator** as a ratio of two regression coefficients:

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

- The numerator is the regression coefficient from a regression of Y on Z.
- The denominator is the regression coefficient from a regression of D on Z.

Properties of the Wald estimator

The **Wald estimator** satisfies two of the three usual statistical properties we like:

- It is **consistent** for the LATE
- and it is **asympotically normal**
  - There exists an analytical formula for the variance
  - But the bootstrap also works

But it is **biased** in small samples.

- Bias inversely depends on $Cov(Z_i, D_i)$: the smaller this covariance, the larger the bias.

Regardless, it only requires **mild assumptions**:

- No need to assume we know the form of Y_i or D_i.

Regression and IV

Another way of thinking about IV is in terms of the linear model framework.

This has the useful function of allowing us to include **additional covariates**, X_i in our instrumental variables estimator.

- But it comes at a cost of having to make stronger assumptions about Y.

IV with constant effects

Let's write down a model for Y_i with treatment D_i and an unobserved confounder U_i

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

Important: We assume that $Cov(D_i, \eta_i) = 0$ – in other words, if we knew U_i, we'd be able to estimate this directly and get τ:

- Same as assuming $E[\eta_i | D_i] = 0$, as we usually do in regression.

The instrumental variable

However, what we actually have is

$$Y_i = \alpha + \tau D_i + \epsilon_i$$

where:

$$\epsilon_i = \gamma U_i + \eta_i$$

Our assumption is violated! Since U_i is a confounder, $Cov(D_i, \gamma U_i + \eta_i) \neq 0$ and the bivariate regression of Y_i on D_i will not identify the causal effect.

- η_i is just statistical noise with $E[\eta_i|D_i] = 0$
- But $E[\epsilon_i|D_i] = \gamma E[U_i|D_i] \neq 0$.

The instrumental variable

Our model is:

$$Y_i = \alpha + \tau D_i + \epsilon_i$$

- Note that Z_i does not appear directly in the equation for Y_i (exclusion restriction).

Furthermore, if instrument Z_i satisfies exogeneity and the exclusion restriction, we can say:

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

Unlike D_i, Z_i is not correlated with the error term ϵ_i.

We can use the properties of covariances to get an expression for τ in terms of $Cov(Y_i, Z_i)$

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

Therefore:

$$\tau = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)},$$

This is the **LATE**!

- And we can estimate it with the Wald estimator we saw before.

IV with covariates

What if we have covariates that we want to include in both the instrument and outcome regressions?

$$\begin{tikzcd}
X \arrow[d] \arrow[dr] \arrow[drr] & & U \arrow[dl, dashed] \arrow[d, dashed] \\
Z \arrow[r] & D \arrow[r] & Y
\end{tikzcd}$$

- The instrument is valid only *conditional* on X_i: $Y_i(d, z) \perp\!\!\!\perp D_i | X_i$
- We want to improve the precision of our estimates (maybe Z_i is a stronger instrument conditional on X_i): $Cov(Z_i, D_i | X_i) \geq Cov(Z_i, D_i)$
- Or we have multiple instruments: $Z_i = [A_i, B_i]$.

Regrssion offers a simple way to include covariates in IV estimators.

IV with covariates

We can generalize the IV estimator with two linear equations – one for the outcome Y_i and the other for the treatment D_i

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

X_i goes into both equations (they're neither instruments nor treatments).

Plug one into the other to get the "reduced form" equation

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

- So the LATE, τ is the coefficient on the term: $(X_i'\alpha + \gamma Z_i)$,
- And $[\tau\nu_i + \epsilon_i]$ is a combined statistical error term

IV with covariates

Our outcome model is:

$$Y_i = X_i'\beta + \tau[X_i'\alpha + \gamma Z_i] + [\tau\nu_i + \epsilon_i]$$

If:

- We can **estimate** $X_i'\alpha + \gamma Z_i$, and
- $Cov(X_i'\alpha + \gamma Z_i, \tau\nu_i + \epsilon_i | X_i) = 0$ ...

Then we can estimate τ by:

1. First estimating $X_i'\alpha + \gamma Z_i$,
2. Then running a regression of Y_i on X_i and $X_i'\alpha + \gamma Z_i$

Can we estimate $X_i'\alpha + \gamma Z_i$?

Back to our assumption on D_i:

$$D_i = X_i'\alpha + \gamma Z_i + \nu_i,$$

With $E[\nu_i | X_i, Z_i] = 0$, therefore:

$$E[D_i | X_i, Z_i] = X_i'\alpha + \gamma Z_i$$

It's the predicted value from the regression of D_i on X_i and Z_i!

- We can estimate it consistently by regressing D_i on X_i and Z_i to obtain coefficient estimates $\hat{\alpha}$ and $\hat{\gamma}$
- And then obtaining predicted values: $\hat{D}_i = X_i'\hat{\alpha} + \hat{\gamma} Z_i$

Is $X_i'\alpha + \gamma Z_i$ independent of the error term?

Recall our models:

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

By **randomization** and **exclusion-restriction**:

$$Cov(Z_i, \nu_i | X_i) = 0, \text{ and } Cov(Z_i, \epsilon_i | X_i) = 0,$$

therefore:

$$\begin{aligned} Cov(X_i'\alpha + \gamma Z_i, \tau\nu_i + \epsilon_i | X_i) &= Cov(X_i'\alpha, \tau\nu_i + \epsilon_i | X_i) \\ &+ Cov(\gamma Z_i, \tau\nu_i | X_i) + Cov(\gamma Z_i, \epsilon_i | X_i) \\ &= 0 + \tau\gamma Cov(Z_i, \nu_i | X_i) + \gamma Cov(Z_i, \epsilon_i | X_i) \\ &= 0. \end{aligned}$$

So, **yes**!

Two-stage least squares

Stage 1: Regress treatment D_i on instrument Z_i and covariates X_i using an OLS estimator

$$E[D_i|X_i, Z_i] = X_i'\alpha + \gamma Z_i$$

Get the predicted values from the regression $\hat{D}_i = X_i'\hat{\alpha} + \hat{\gamma} Z_i$.

Stage 2: Regress outcome Y_i on the fitted values $\hat{D}_i$ and covariates X_i

$$E[Y_i|X_i, \hat{D}_i] = X_i'\beta + \tau \hat{D}_i$$

The coefficient on $\hat{D}_i$ is our estimate of the effect of D_i using IV.

Two-stage least squares SEs

- However, this isn't *exactly* what a canned 2SLS procedure (like `iv_robust` in `estimatr` will do). The point estimate will be correct, but the standard errors in that second regression will be wrong.
- Let $\mathbf{X}$ be the matrix of all second-stage regressors and D. Let $\mathbf{Z}$ be the matrix of all first stage regressors and Z (the covariates appear in both).
- We can write the linear projection from the first stage (the fitted values) as

$$\hat{\mathbf{X}} = Z(Z'Z)^{-1}Z'\mathbf{X}$$

- Then the second stage coefficients can be estimated by substituting the projection $Z(Z'Z)^{-1}Z'\mathbf{X}$ for $\mathbf{X}$ - 2SLS routines will do this all in one step (and get correct SEs).

Forbidden regressions

IV analysis with 2SLS is extremely popular in applied research, but there are some common pitfalls to avoid:

- Don't include covariates X_i only in one stage but not the other.
- Don't use non-linear transformations of the fitted values in the second stage (remember $E[X^2] \neq E[X]^2$)
- Don't use a non-linear first stage in 2SLS (expectations/linear projections don't propagate directly through non-linear functions).

Weak instruments

Recall that one of the IV assumptions is that:

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

i.e., the instrument **must** have some effect on the treatment.

- The **magnitude** of this effect influences the accuracy of estimates
- When this effect is small we say that our instrument is **weak**

We will now see how and why this happens.

Weak instrument problem

Our Wald estimator is the ratio of two covariances.

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

What happens when that denominator is really close to 0? Wildly variable estimates!

Weak instrument problem

Can show that the Wald estimator converges in probability to

$$\hat{\tau}_{\text{IV}} \to \tau + \frac{Cov(Z_i, U_i)}{Cov(Z_i, D_i)}$$

- Since $Cov(Z_i, U_i)$ is 0 by the exclusion restriction/exogeneity, IV is *consistent*.
- But in finite-samples, that bias term can be large. If the the instrument is weak ($Cov(Z_i, D_i) \approx 0$), the divergence from the true treatment effect can sometimes be worse than just the naive regression of Y_i on D_i!

Diagnosing weak instruments

Simulation evidence (Stock and Yogo, 2005) suggests thresholds for how "strong" our first-stage relationship should be to get valid t-ratio inference.

- Threshold criteria based on the F-test statistic of excluding the variables from the first-stage regression.

- Stock and Yogo (2005) suggest that when the first-stage F-statistic is 10 or more, the relative bias of IV is small (though stronger = better).
- However! Lee, McCrary, Moreira, Porter (2020) show that this is far too low – need closer to an F of 104.7! – in general, the usual asymptotic CIs of $\hat{\tau} \pm 1.96 \times \hat{SE}(\hat{\tau})$ will under-cover the true value.

- Weak-instrument-robust confidence intervals are increasingly more common.