Political Methodology II
TA Section: Instrumental Variables I

Apoorva Lal

March 4, 2021

Stanford University

Roadmap

Big Picture

Linear / Constant treatment effects

Big Picture

- IV methods can solve the problem of omitted variable bias when trying to isolate a causal effect using observational data.

Big Picture

- IV methods can solve the problem of omitted variable bias when trying to isolate a causal effect using observational data.
- The setting is selection on <u>un</u>observables. Originally developed to address simultaneity in the 1920s

Big Picture

- IV methods can solve the problem of omitted variable bias when trying to isolate a causal effect using observational data.
- The setting is selection on <u>un</u>observables. Originally developed to address simultaneity in the 1920s
- There are two broad types of IV frameworks: the (older) one that assumes constant treatment effects, and Potential-outcome-based one that assumes heterogeneous treatment effects.

Big Picture

- IV methods can solve the problem of omitted variable bias when trying to isolate a causal effect using observational data.
- The setting is selection on unobservables. Originally developed to address simultaneity in the 1920s
- There are two broad types of IV frameworks: the (older) one that assumes constant treatment effects, and Potential-outcome-based one that assumes heterogeneous treatment effects.
- The estimators we use in each framework are the same, but the assumptions and interpretation of the treatment effect are slightly different.

Big Picture

- IV methods can solve the problem of omitted variable bias when trying to isolate a causal effect using observational data.
- The setting is selection on unobservables. Originally developed to address simultaneity in the 1920s
- There are two broad types of IV frameworks: the (older) one that assumes constant treatment effects, and Potential-outcome-based one that assumes heterogeneous treatment effects.
- The estimators we use in each framework are the same, but the assumptions and interpretation of the treatment effect are slightly different.
- We'll start with the constant treatment effects framework to motivate the estimators, and then move on (next week) to the heterogeneous treatment effect framework, which is more common nowadays

Roadmap

Big Picture

Linear / Constant treatment effects

Setup

Suppose we are interested in the effect of schooling (s_i) on wages (Y_i). Using a selection-on-observables story with constant treatment effects, we know that conditional on a vector of control variables for "ability" (A_i), the causal model is

$$Y_i = \alpha + \tau s_i + A_i' \gamma + v_i$$

Setup

Suppose we are interested in the effect of schooling (s_i) on wages (Y_i). Using a selection-on-observables story with constant treatment effects, we know that conditional on a vector of control variables for "ability" (A_i), the causal model is

$$Y_i = \alpha + \tau s_i + A_i' \gamma + v_i$$

If we could observe A_i, we would just estimate this regression and be done. But what if we can't observe A_i? If we controlled for nothing, we would estimate

$$Y_i = \alpha + \tau s_i + e_i$$

What are we worried about in the short regression?

Setup

Suppose we are interested in the effect of schooling (s_i) on wages (Y_i). Using a selection-on-observables story with constant treatment effects, we know that conditional on a vector of control variables for "ability" (A_i), the causal model is

$$Y_i = \alpha + \tau s_i + A_i' \gamma + v_i$$

If we could observe A_i, we would just estimate this regression and be done. But what if we can't observe A_i? If we controlled for nothing, we would estimate

$$Y_i = \alpha + \tau s_i + e_i$$

What are we worried about in the short regression? Correlation between s_i and e_i. We know from our selection-on-observables story that this correlation is entirely captured by A_i.

Omitted Variable Bias

Formally, the naive OLS estimate of τ is:

$$\begin{aligned}\hat{\tau}_{ols} &= \frac{Cov(s_i, Y_i)}{Var(s_i)} = \frac{Cov(s_i, \alpha + \tau s_i + e_i)}{Var(s_i)} \\ &= \frac{\tau Cov(s_i, s_i) + Cov(s_i, e_i)}{Var(s_i)} = \tau + \frac{Cov(s_i, e_i)}{Var(s_i)} \\ &= \tau + \frac{Cov(s_i, A_i'\gamma + v_i)}{Var(s_i)} \\ &= \tau + \gamma' \frac{Cov(s_i, A_i)}{Var(s_i)} + \frac{Cov(s_i, v_i)}{Var(s_i)}\end{aligned}$$

Omitted Variable Bias

Formally, the naive OLS estimate of τ is:

$$
\begin{aligned}
\hat{\tau}_{ols} &= \frac{Cov(s_i, Y_i)}{Var(s_i)} = \frac{Cov(s_i, \alpha + \tau s_i + e_i)}{Var(s_i)} \\
&= \frac{\tau Cov(s_i, s_i) + Cov(s_i, e_i)}{Var(s_i)} = \tau + \frac{Cov(s_i, e_i)}{Var(s_i)} \\
&= \tau + \frac{Cov(s_i, A_i'\gamma + v_i)}{Var(s_i)} \\
&= \tau + \gamma' \frac{Cov(s_i, A_i)}{Var(s_i)} + \frac{Cov(s_i, v_i)}{Var(s_i)}
\end{aligned}
$$

Our SOO story assumes $\mathbb{E}[s_i v_i] = 0$, so the expected value of $\hat{\tau}_{ols}$ is

$$
\mathbb{E}[\hat{\tau}_{ols}] = \tau + \gamma' \mathbb{E}\left[\frac{Cov(s_i, A_i)}{Var(s_i)}\right]
$$

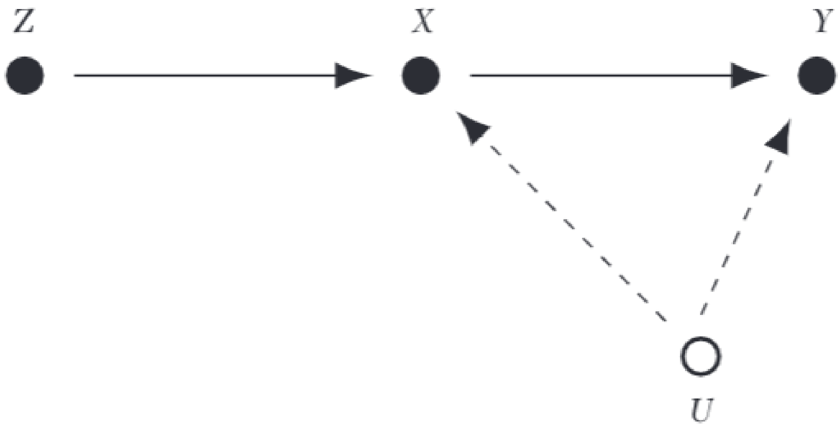
Finding an instrument

- We can see that the problem with using the schooling variable as measured in the wild is that it produces a non-zero $Cov(s_i, A_i)$. But what if we could find another variable, z_i, that is correlated with s_i but not with A_i or v_i? In other words, z_i produces as-if randomized variation in schooling, and is only correlated with wages through schooling. What could be some possibilities for this kind of instrument?

Finding an instrument

- We can see that the problem with using the schooling variable as measured in the wild is that it produces a non-zero $Cov(s_i, A_i)$. But what if we could find another variable, z_i, that is correlated with s_i but not with A_i or v_i? In other words, z_i produces as-if randomized variation in schooling, and is only correlated with wages through schooling. What could be some possibilities for this kind of instrument?
- One example used by Angrist and Krueger (1991) is the variation induced in years of schooling by the fact that most states require students to start school in the calendar year that they turn 6 years old. This means that kids born at the beginning of the calendar year are older when they start school than kids born at the end of the year, and the two groups will have had different amounts of time in school when they reach the legal dropout age at 16.

IV DAG



IV Setup

- We suppress controls without loss-of-generality since, by the FWL, one can eliminate the controls $\boldsymbol{c}_i$ in the structural equation by regressing Y, X, and Z on $\boldsymbol{c}_i$ and using the residuals $\widetilde{Y}$, $\widetilde{X}$, $\widetilde{Z}$ for all subsequent computation.

IV Setup

- We suppress controls without loss-of-generality since, by the FWL, one can eliminate the controls $\boldsymbol{c}_i$ in the structural equation by regressing Y, X, and Z on $\boldsymbol{c}_i$ and using the residuals $\widetilde{Y}, \widetilde{X}, \widetilde{Z}$ for all subsequent computation.

$$\begin{aligned} \text{Structural Equation} &: Y = \alpha_0 + \beta X + \varepsilon \\ \text{First Stage} &: X = \pi_0 + \pi Z + \eta \\ \text{Reduced Form} &: Y = \gamma_0 + \gamma Z + \upsilon \end{aligned}$$

$$\begin{aligned} Y &= \alpha_0 + \beta X + \varepsilon \\ &= \alpha_0 + \beta(\pi_0 + \pi Z + \eta) + \varepsilon && \text{substitute in X from first-stage} \\ &= \underbrace{(\alpha_0 + \beta\pi_0)}_{\gamma_0} + \underbrace{(\beta\pi)}_{\gamma} Z + (\beta\eta + \varepsilon) \implies \gamma = \beta\pi \quad \to \beta = \frac{\gamma}{\pi} \end{aligned}$$

IV Assumptions

Assumption (Exclusion Restriction / Validity)

X: $\varepsilon_i \perp (Z_i, \boldsymbol{c}_i)$

This requires that Z has no direct effect on Y except through X, where ε_i is the residual in the strictural equation. The instrument needs to be uncorrelated with unobservables in the structural equation, potentially conditional on controls $\boldsymbol{c}_i$.

Assumption (Relevance)

Z affects X, i.e. $\text{Cov}\left[Z, X\right] \neq 0$ or $\pi_1 \neq 0$.

Which of these is testable in our sample data?

IV Assumptions

Assumption (Exclusion Restriction / Validity)

X: $\varepsilon_i \perp (Z_i, \boldsymbol{c}_i)$

This requires that Z has no direct effect on Y except through X, where ε_i is the residual in the strictural equation. The instrument needs to be uncorrelated with unobservables in the structural equation, potentially conditional on controls $\boldsymbol{c}_i$.

Assumption (Relevance)

Z affects X, i.e. $\text{Cov}[Z, X] \neq 0$ or $\pi_1 \neq 0$.

Which of these is testable in our sample data? Non-zero first stage. Can you think of any critiques of the quarter of birth instrument based on these assumptions?

Two-stage least squares

The 2SLS coefficient is equivalent to the IV estimator:

$$\begin{aligned}
\hat{\beta} &= \frac{\text{Cov}(Y_i, \hat{x}_i)}{\text{Var}(\hat{x}_i)} = \frac{\text{Cov}(Y_i, \widehat{\pi_0} + \hat{\pi} z_i)}{\text{Var}(\widehat{\pi_0} + \hat{\pi} z_i)} \\
&= \frac{\hat{\pi}\text{Cov}(Y_i, z_i)}{\hat{\pi}^2 \text{Var}(z_i)} = \frac{\text{Cov}(Y_i, z_i)}{\hat{\pi}\text{Var}(z_i)} \\
&= \frac{\text{Cov}(Y_i, z_i)}{\frac{\text{Cov}(x_i, z_i)}{\text{Var}(z_i)} \cdot \text{Var}(z_i)} = \frac{\text{Cov}(Y_i, z_i)}{\text{Cov}(x_i, z_i)} = \frac{\text{Reduced Form}}{\text{First Stage}}
\end{aligned}$$

$$\widehat{\beta}_{IV} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\boldsymbol{y}$$

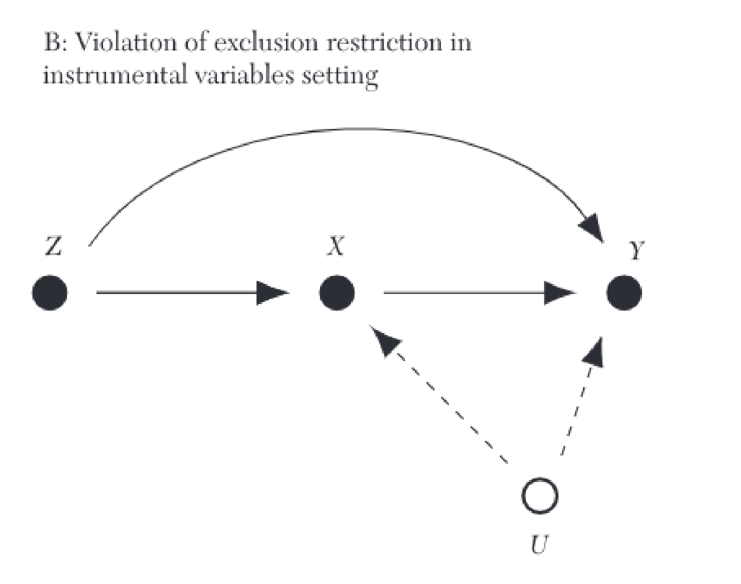
Multiple Instruments

If you have multiple valid instruments (pigs may fly), the matrix analogue is

$$\beta_{2SLS} = \left(\mathbf{X}'\mathbf{P}_z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_z\boldsymbol{y}$$

- where $\mathbf{P}_z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ is the hat-maker matrix from the first-stage which projects the endogenous variables $\mathbf{X}$ into the column space of $\mathbf{Z}$
- this preserves only the ‘clean’ variation that is uncorrelated with ε.

BAD IV DAG


B: Violation of exclusion restriction in instrumental variables setting
Z
X
Y
U

Exclusion Restriction Violations

- If instrument is 'imperfect' (exclusion restriction violation) : $\text{Cov}\left[\varepsilon_i, Z_i\right] \neq 0$
  - Then $\text{Cov}\left[Y_i, Z_i\right] = \beta\text{Cov}\left[X_i, Z_i\right] + \text{Cov}\left[\varepsilon_i, Z_i\right]$.
  - Then the ratio of RF/FS is

$$\frac{\text{Cov}\left[Y_i, Z_i\right]}{\text{Cov}\left[X_i, Z_i\right]} = \beta + \frac{\text{Cov}\left[\varepsilon_i, Z_i\right]}{\text{Cov}\left[X_i, Z_i\right]} = \beta + \underbrace{\frac{\text{Cor}(\varepsilon_i, Z_i)}{\text{Cor}(X_i, Z_i)} \frac{\sigma_\varepsilon}{\sigma_x}}_{\text{Bias}}$$

  - Bias potentially very large if $\text{Cov}\left[X_i, Z_i\right] \approx 0 \implies$ problems compound each other
- To learn more about problems and potential fixes, come to Monday's departmental seminar [Lal, Lockhart, Xu, Zu 2021]

Weak Instruments

- If instrument is weak (i.e. $\text{Cov}[X_i, Z_i] \approx 0$), you're dividing by zero
- 2SLS is consistent, but is biased in small samples. The bias is worse with a weak instrument (even worse with multiple weak instruments).

Weak Instruments

- If instrument is weak (i.e. $\text{Cov}\left[X_i, Z_i\right] \approx 0$), you're dividing by zero
- 2SLS is consistent, but is biased in small samples. The bias is worse with a weak instrument (even worse with multiple weak instruments).
- In the worst-case scenario (multiple weak instruments that produce no first stage), 2SLS sampling distribution is centered on the probability limit of OLS.
- The first stage estimates reflect some of the randomness in the endogenous variable. If the population first stage is 0, then the resulting first stage coefficient just reflect randomness in the endogenous regressor.

Weak Instruments

- If instrument is weak (i.e. $\text{Cov}\left[X_i, Z_i\right] \approx 0$), you're dividing by zero
- 2SLS is consistent, but is biased in small samples. The bias is worse with a weak instrument (even worse with multiple weak instruments).
- In the worst-case scenario (multiple weak instruments that produce no first stage), 2SLS sampling distribution is centered on the probability limit of OLS.
- The first stage estimates reflect some of the randomness in the endogenous variable. If the population first stage is 0, then the resulting first stage coefficient just reflect randomness in the endogenous regressor.

$$\mathbb{E}\left[\widehat{\beta}_{2SLS} - \beta\right] \approx \overbrace{\frac{\sigma_{\eta\varepsilon}}{\sigma_{\varepsilon}^2}}^{\text{Bias of OLS}} \frac{1}{F+1}$$

- where F is the first-stage F statistic. As $F \to 0$ (i.e. the instrument is weak), the bias of the IV tends to the bias of the OLS
- Moral of the story: Always check the F-statistic <u>for the instrument</u> in your first stage (bigger than 10 is considered "safe" - this rule-of-thumb keeps growing; compute 'Effective F-stat').

Standard errors

- You can manually construct the 2SLS estimate using `lm`, but you can't use the OLS standard errors from the second stage regression.

Standard errors

- You can manually construct the 2SLS estimate using `lm`, but you can't use the OLS standard errors from the second stage regression.
- This is because the second stage OLS standard errors use the variance of the residuals against the fitted values from the first stage. What you really want is the variance of the residuals using the IV coefficients and the original endogenous regressor

Standard errors

- You can manually construct the 2SLS estimate using `lm`, but you can't use the OLS standard errors from the second stage regression.
- This is because the second stage OLS standard errors use the variance of the residuals against the fitted values from the first stage. What you really want is the variance of the residuals using the IV coefficients and the original endogenous regressor
- Moral of the story: use a canned routine like `ivreg` or `felm` to make the adjustment for you.

Standard errors

- You can manually construct the 2SLS estimate using `lm`, but you can't use the OLS standard errors from the second stage regression.
- This is because the second stage OLS standard errors use the variance of the residuals against the fitted values from the first stage. What you really want is the variance of the residuals using the IV coefficients and the original endogenous regressor
- Moral of the story: use a canned routine like `ivreg` or `felm` to make the adjustment for you.
- Analytic standard errors for IV are generally a mess (Alwyn Young - 'Consistency without Inference') - use the bootstrap whenever possible

Standard errors

- You can manually construct the 2SLS estimate using `lm`, but you can't use the OLS standard errors from the second stage regression.
- This is because the second stage OLS standard errors use the variance of the residuals against the fitted values from the first stage. What you really want is the variance of the residuals using the IV coefficients and the original endogenous regressor
- Moral of the story: use a canned routine like `ivreg` or `felm` to make the adjustment for you.
- Analytic standard errors for IV are generally a mess (Alwyn Young - 'Consistency without Inference') - use the bootstrap whenever possible
- More on this next week