Instrumental Variables

Caleb Leedy

March 29, 2023

Ordinary Least Squares

- In OLS, we consider the model,

$$Y = X\beta + \varepsilon \tag{1}$$

  where $E[\varepsilon \mid X] = 0$ and $X \perp \varepsilon$.
- However, what if $\mathrm{Cov}(X, \varepsilon) \neq 0$?
- We say a variable X_k is **endogenous** if $\mathrm{Cov}(X_k, \varepsilon) \neq 0$.
- A variable X_k is **exogenous** if $X_k \perp \varepsilon$.

Modifying Previous Assumptions

- Last week we discussed the assumptions of the potential outcomes framework. One of them was: No Unmeasured Confounders (NUC),

$$Y(1), Y(0) \perp A \mid X.$$

- If a variable X_k is endogenous, then the model does *not* satisfy the NUC condition.

Parametric Models

- Consider the following linear model:[1]

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

  with $x_1, x_2 \perp \varepsilon$ but $x_3 \not\perp \varepsilon$
- To estimate β_3 we need an instrumental variable.

[1] Example taken from (Wooldridge 2010).

Instrumental Variables

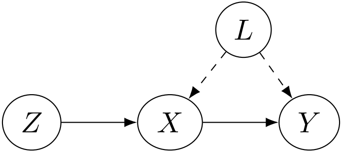
- A variable z_1 is an **instrumental variable** (IV) if it satisfies:

$$\text{Cov}(z_1, \varepsilon) = 0 \tag{2}$$

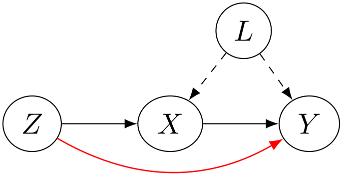
$$\text{Cov}(z_1, x_3) \neq 0 \tag{3}$$

- This makes sense because we want it to be exogenous with respect to Equation 1, yet we need it to influence x_3 if we are going to measure β_3.
- Note, that Equation 2 *cannot* be tested but Equation 3 can and should be tested.

Graphical Model



Graphical Model



Reduced Form Equations

- When we have an instrument z_1, we can estimate:

$$\hat{x}_3 = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \hat{\theta} z_1$$
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 \hat{x}_3$$

- This framework is called **two-stage least squares** (2SLS).
- This can be generalized to have K exogenous x_i variables and L instruments z_j.

Identification

- Then the IV solves the identification problem.
- Let $z = (x_1, x_2, z_1)$.
- Equation 2 implies that $E[z'\varepsilon] = 0$.
- The normal equations for the IV estimator are:

$$E[z'x]\beta = E[z'y].$$

- This has a unique solution if $E[z'x]$ has full rank, which happens if Equation 3 is satisfied.

Results for 2SLS

Under regularity conditions, 2SLS is

- consistent,
- asymptotically normal, and
- asymptotically efficient.

See (Wooldridge 2010), Chapter 5 for these proofs.

Problems with IVs

- Bias
- Weak instruments: In the linear model,

$$\operatorname{plim} \hat{\beta}_3 = \beta_3 + \frac{\mathrm{Cov}(z_1, \varepsilon)}{\mathrm{Cov}(z_1, x_3)}.$$

Causal Models with No Unconfoundedness

- Suppose that we have the model,[2]

$$Y_i(a) = Y_i(0) + \tau a_i.$$

We can also express this as

$$Y_i = \alpha + A_i\tau + \varepsilon_i$$

- We do *not* use the NUC. So

$$Y(1), Y(0) \perp A \mid X.$$

- Notice that OLS does not work because

$$\tau_{OLS} = \frac{\mathrm{Cov}(Y_i, A_i)}{\mathrm{Var}\, A_i} = \frac{\mathrm{Cov}(\tau A_i + \varepsilon, A_i)}{\mathrm{Var}\, A_i} = \tau + \frac{\mathrm{Cov}(\varepsilon, A_i)}{\mathrm{Var}\, A_i}$$

[2] The rest of the slides were based off of Stefan Wager's S361 Causal Inference Notes (Wager 2020).

Causal Models with IVs

- We can add an instrument and have something similar to 2SLS,

$$
\begin{aligned}
Y_i &= \alpha + A_i\tau + \varepsilon_i \\
A_i &= Z_i\gamma + \eta_i \qquad \varepsilon_i \perp Z_i.
\end{aligned}
$$

- Then

$$
\mathrm{Cov}(Y_i, Z_i) = \mathrm{Cov}(A_i\tau + \varepsilon_i, Z_i) = \tau\,\mathrm{Cov}(A_i, Z_i).
$$

- Hence,

$$
\tau = \frac{\mathrm{Cov}\left(Y_i, Z_i\right)}{\mathrm{Cov}(A_i, Z_i)}.
$$

Optimal Instruments

- If Z is a d-dimensional vector then we have

$$\tau = \frac{\mathrm{Cov}\left(Y_i, w(Z_i)\right)}{\mathrm{Cov}(A_i, w(Z_i))}$$

  where $w : \mathbb{R}^d \to \mathbb{R}$.
- The optimal choice of $w(\cdot)$ that minimizes the variance of τ, is

$$w^*(Z) \propto E[A \mid Z].$$

Estimation

The previous slide suggests the following estimation strategy:

1. Estimate $\hat{w}(\cdot) = E[A \mid Z]$ nonparametrically, and then
2. Estimate the covariances using $\hat{w}$,

$$\hat{\tau} = \frac{\hat{\mathrm{Cov}}(Y_i, \hat{w}(Z_i))}{\hat{\mathrm{Cov}}(A_i, \hat{w}(Z_i))}$$

However, this can fail from overfitting with weak instruments.

Cross Fitting

A better strategy is to use cross-fitting, and solve

$$\hat{\tau} = \frac{\hat{\text{Cov}}(Y_i, \hat{w}^{k(-i)}(Z_i))}{\hat{\text{Cov}}(A_i, \hat{w}^{k(-i)}(Z_i))}$$

where $\hat{w}^{k(-i)}$ is the estimation of $\hat{w}$ on the k-th fold in which element i is missing.

Extension to Nonparametric Regression

- Suppose we have the model:

$$Y_i = g(A_i) + \varepsilon_i, \quad Z_i \perp \varepsilon_i$$

- Then,

$$E[Y_i \mid Z_i] = \int g(a) f(a \mid z) da.$$

- This can be estimated using basis splines (or other nonparametric techniques).

Local Average Treatment Effects

- Consider the model:

$$
\begin{aligned}
Y_i &= \alpha + \tau A_i + \varepsilon_i \\
A_i &= \gamma Z_i + \eta_i \quad Z_i \perp \varepsilon_i.
\end{aligned}
$$

- Then we can identify τ with

$$
\begin{aligned}
\tau &= \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(A_i, Z_i)} \\
&= \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[A_i \mid Z_i = 1] - E[A_i \mid Z_i = 0]}
\end{aligned}
$$

where the second equation holds because Z_i is binary.

Strata

	$A_i(1) = 1$	$A_i(1) = 0$
$A_i(0) = 1$	Always taker	Denier
$A_i(0) = 0$	Complier	Never taker

Result

- Assuming that there exist some compliers, the **local average treatment effect** is

$$\tau_{LATE} = E[Y_i(1) - Y_i(0) \mid i \text{ is a complier}].$$

References I

Wager, Stefan (2020). *Stats 361: Causal inference*.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.