Instrumental Variables II: 2SLS

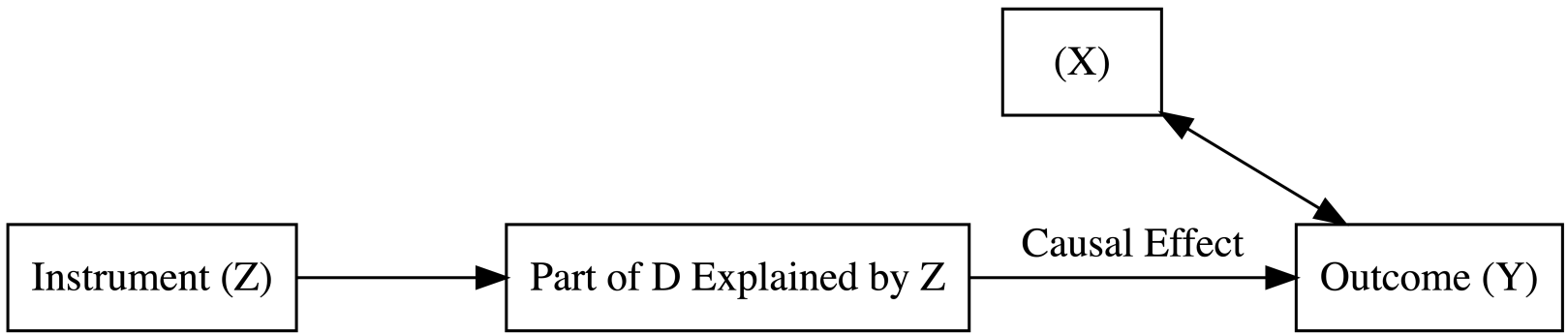
Lecture 10 - Introduction to Causal Inference

Kevin Li

Review: Instrumental Variables

We introduced IV in the last lecture. Brief review:

By using an instrument Z and $\hat{D}$, X is no longer a confounder between $\hat{D}$ and Y. Thus, $\hat{D}$ is exogenous (if Z is exogenous).



Since $\hat{D}$ is exogenous and there are no more confounders between $\hat{D}$ and Y, we can calculate the causal effect of $\hat{D}$ on Y.

- This is the Local Average Treatment Effect of $\hat{D}$ on Y.

Two-Stage Least Squares

The way to estimate the LATE causal effect is with 2-stage least squares (2SLS) estimator.

1st stage: Regress D on Z to estimate $\hat{D}$.

$$D_i = \delta + Z_i\beta + \varepsilon_i$$

- Significance of β can be used to test relevance.

2nd stage: Regression Y on $\hat{D}$ to estimate the causal effect:

$$Y_i = \alpha + \hat{D}_i\tau_{\text{LATE}} + u_i$$

The OLS estimate for τ_{LATE} in the second stage is our causal estimate.

Controlling for Exogeneity

We know from last lecture, Z must be exogenous to D and Y if it is a valid instrument.

- Thus, we might need to control for potential confounders between Z and Y, and Z and D.

Let us say we want to control for confounder X. We would include it in both stages of the regression:

- 1st stage: $D_i = \delta_0 + Z_i\delta_1 + X_i\delta_2 + \varepsilon_i$
- 2nd stage: $Y_i = \alpha + \hat{D}_i\tau_{\text{LATE}} + X_i\beta + u_i$

In panel data, we can do the same for fixed effects to control for differences between time periods and units. We add fixed effects to both the 1st and 2nd stages.

Multiple Instruments

You do not need to stick with just one valid instrument. Including more instruments has a few advantages:

1. Having multiple valid instruments provides more variation in $\hat{D}$, allowing for more precise (and less variance) estimates.
2. Multiple instruments can also be a solution for weak-instruments (see later slides).

First Stage (p number of instruments):

$$D_i = \delta_0 + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \cdots + \delta_p Z_{pi} + \varepsilon_i$$

Second Stage: remains the same.

$$Y_i = \alpha + \hat{D}_i \tau_{\text{LATE}} + u_i$$

Multiple Instrumented Regressors

So far, we have focused on using an instrument for D. However, if we have any control variables X for $D \to Y$ that themselves are not exogenous, that can cause issues.

- Endogeneity in just one regressor makes **all** OLS estimates biased (remember - second stage is OLS).

So we might want to use instruments to make both D and X exogenous. We will have **2 first stages**:

- D first stage: $D_i = \delta_0 + Z_{1i}\delta_1 + \varepsilon_i$
- X first stage: $X_i = \gamma_0 + Z_{2i}\gamma_1 + \epsilon_i$

Second Stage: $Y_i = \alpha + \hat{D}_i\tau_{\text{LATE}} + \hat{X}_i\beta + u_i$

Properties of 2SLS

Like OLS, two-stage least squares is asymptotically consistent.

- In other words - in infinitely large sample sizes, 2SLS is an unbiased estimator.

However, unlike OLS, 2SLS is **biased** in finite sample sizes, even if all assumptions are met.

- In other words - in small sample sizes, 2SLS might not produce accurate estimates.

When including confounders, 2SLS also only assumes homogeneous treatment effects, i.e. all units have the same treatment effect sizes. 2SLS can be inaccurate even in infinite sample sizes if there is heterogeneity.

- This is not a huge concern for most purposes.

Standard Errors

So far, I have portrayed 2SLS as essentially 2 regressions - the 1st stage, then the 2nd stage.

- This is true for estimating τ_{LATE}. However, for the standard errors of τ_{LATE}, this will not be accurate.
- Note: in causal inference, we use heteroscedasticity-robust standard errors by default.
- Recommendation: use statistical software (such as R), with packages that can automatically estimate correct IV standard errors (such as fixest).

Note: the stronger the correlation between Z and D, the smaller the variance becomes, and the more accurate/precise our IV estimator becomes.

Reduced Form

The reduced form regression is the regression of Y on Z:

$$Y_i = \gamma_0 + Z_i\gamma_1 + \epsilon_i$$

Assuming Z is exogenous (one of the assumptions of IV), that means the estimate γ_1 is the causal effect of Z on Y.

- This effect is also called the **intent-to-treat effect** τ_{ITT}.

If Z and D are both binary (quite common in causal inference), our LATE can also be estimated as:

$$\tau_{\text{LATE}} = \frac{\tau_{\text{ITT}}}{Pr(\text{compliers})} = \frac{\tau_{\text{ITT}}}{\widehat{Cov}(D_i, Z_i)}$$

- This is where the interpretation of LATE as the causal effect of compliers comes from.

Weak Instruments

We know that relevance is an assumption - Z needs to be correlated with D.

However, not only does Z need to be correlated with D, it needs to be moderately/strongly correlated. When there is a weak correlation, we run into the problem of **weak instruments**.

- Weak instruments have very biased τ_{LATE} estimates, especially in small samples.
- Weak instruments also have very high variance, making it very difficult to get a significant causal effect.

Why? Well if Z and D are weakly correlated, then what is $\hat{D}$, the part of D explained by Z? Is it even useful to interpret?

How to see if instrument is weak? Generally - a first stage with F-statistic lower than 10 is considered weak.