Empirical Methods

Topic 2c:

Instrumental Variables Estimation

Instrumental Variables (IV)

Latex Color = "LightGoldenrodYellow"

Instrumental Variables Overview I

- We turn finally to Instrumental Variables estimation
  - ▶ This is a critically important topic
  - ▶ As it is the first and most important tool in handling the violation of the most important CLRM assumption

$$E(\epsilon_i|x_i) = 0$$

IV Basics

Instrumental Variables I

- Let's formalize how IV "works" using our familiar notation

$$y_i = x_i'\beta + \epsilon_i$$

- But assume $E(\epsilon_i|x_i) \neq 0$, in particular
  - Suppose some elements of x_i are *endogenous* while others are *exogenous*, i.e. $x_i' = \begin{bmatrix} x_{1i}' & x_{2i}' \end{bmatrix}$ with
    - ⋆ x_{1i} a $K_1 \times 1$ vector of included endogenous variables
    - ⋆ x_{2i} a $K_2 \times 1$ vector of included exogenous variables
    - ⋆ And $K_1 + K_2 = K$
  - Note difference in notation here compared to what we had in the last deck
    - ⋆ Where $\tilde{M}$ and $\tilde{K}$ were the numbers of included endogenous and exogenous variables

Instrumental Variables II

- To estimate our model when $E(\epsilon_i|x_i) \neq 0$, we need extra information
    - In particular, we need some variables to "move x_i" in ways that are uncorrelated with ϵ_i
- These variables are called Instrumental Variables
    - Let z_i be a $L \times 1$ vector of instruments with $z_i' = [z_{1i}' \quad x_{2i}']$ with
        - z_{1i} a $L_1 \times 1$ vector of excluded (from this equation) exogenous variables
        - (Note in our earlier notation, $L_1 = K^*$)
        - x_{2i} (still) is a $K_2 \times 1$ vector of included exogenous variables
        - And $L_1 + K_2 = L$

Instrumental Variables III

Instrumental Variables have two critical properties:

1. (IV1) <u>Instrument Exogeneity</u>

   ▶ $\frac{1}{N} Z'\epsilon = \frac{1}{N} \sum_{i=1}^{N} z_i \epsilon_i \xrightarrow{p} E(z_i \epsilon_i) = 0$

   ▶ In words: The instruments are uncorrelated with the error

     ★ Usually ensured by assuming $E(\epsilon_i | z_i) = 0$

     ★ And evaluated by thinking about $Cov(z_i, \epsilon_i)$

Instrumental Variables IV

- NOTE: Instrument Exogeneity also implies the instruments don't *themselves* belong in the structural equation.

- Suppose they did, i.e. suppose:

$$
\begin{aligned}
\text{True model: } y_i &= x_i'\beta + z_i'\gamma + \epsilon_i^* \\
\text{You estimate: } y_i &= x_i'\beta + \epsilon_i \\
&\quad \text{(using } z_i \text{ as an instrument)}
\end{aligned}
$$

Instrumental Variables V

$$
\begin{array}{rcl}
\text{True model: } y_i &=& x_i'\beta + z_i'\gamma + \epsilon_i^* \\
\text{You estimate: } y_i &=& x_i'\beta + \epsilon_i \\
&& (\text{using } z_i \text{ as an instrument})
\end{array}
$$

- Just because you don't include z_i in your model doesn't mean it's not actually *in there*

  ▶ It's clearly true that $\epsilon_i = z_i'\gamma + \epsilon_i^*$

- Thus *even if* $Cov(z_i, \epsilon_i^*) = 0$, you have...

$$
\begin{array}{rcl}
Cov(z_i, \epsilon_i) &=& Cov(z_i, z_i'\gamma + \epsilon_i^*) \\
&=& V(z_i)'\gamma \neq 0
\end{array}
$$

  ▶ ...and you've violated Instrument Exogeneity

Instrumental Variables VI

Two critical IV properties, cont.:

② (IV2) <u>Instrument Relevance</u>

   ▶ $\frac{1}{N}Z'X = \frac{1}{N}\sum_{i=1}^{N} z_i x_i' \xrightarrow{p} E(z_i x_i') \equiv \Sigma_{ZX} \neq 0$

   ▶ In words: The instruments are correlated with the RHS variables

      ★ Evaluated by estimating $Cov(z_i, x_i)$

      ★ (In a "first-stage" regression)

      ★ (To be introduced soon)

Instrumental Variables VII

- As well as one final regularity assumption,
  Assumption 6': Regular Z's:

  ▸ $\frac{1}{N} Z'Z = \frac{1}{N} \sum_{i=1}^{N} z_i z_i' \xrightarrow{p} E(z_i z_i') \equiv \Sigma_{ZZ}$

    ★ Where Σ_{ZZ} is a finite $L \times L$ matrix

    ★ (In words: A standard regularity condition analogous to OLS's (A6, Regular X's))

Instrumental Variables Intuition I

- OK, fine for the math. But what is the <u>intuition</u> for IV?

- It's a sequence of logic that goes like this:

  1. In essence, the OLS estimate $\hat{\beta}_k$ measures the correlation between y_i and x_{ik}

     - ★ (Controlling for the other x's)

  2. Under the CLRM Assumptions, this measures the *causal* effect of x_{ik} on y_i

     - ★ (Especially (A1, Linearity) and (A2, Mean-zero error))

Instrumental Variables Intuition II

Sequence of logic, cont.:

3. Sometimes, however, we have endogeneity and (A2, Mean-zero error) *doesn't* hold

   - $(E(\epsilon_i|x_i) \neq 0)$

   - (The bias formula shows that $E(\hat{\beta}_k) \neq \beta_k$)

   - (Thus OLS *cannot* recover the causal effect of x_{ik} on y_i)

Instrumental Variables Intuition III

Sequence of logic, cont.:

④ IV provides a solution: find an instrument that, as it varies...

- ...causes x_i to vary...
  - ⋆ [Assumption (IV2, Instrument relevance)]
- ...in a way that *isn't* correlated with ϵ_i...
  - ⋆ [Assumption (IV1, Instrument exogeneity)]
- ...and let that "clean" variation in x_{ik} measure the causal effect of β_k

That's It!

Fitting our example into this framework

- We can fit our earlier example into this framework

- Our demand curve was written as:

$$q_i = \alpha_0 + \alpha_1 p_i + \alpha_2 inc_i + \epsilon_{i1}$$

- Thus:

  ▶ $y_i = q_i$,

  ▶ $x'_{1i} = p_i$, $x'_{2i} = [1 \quad inc_i]$, thus $x'_i = [p_i \quad 1 \quad inc_i]$

  ▶ $z'_{1i} = w_i$, $x'_{2i} = [1 \quad inc_i]$, thus $z'_i = [w_i \quad 1 \quad inc_i]$

Instrumental Variables Notes I

Note:

1. Once we move into instrumental variables, we focus on large-sample properties, i.e.

   - Consistency rather than Unbiasedness

     - Indeed taking the expectation of the IV estimator is a *pain in the neck*

     - Whereas taking its probability limit is easy

   - Asymptotic Normality rather than Normality

   - We'll discuss the properties of the IV estimator after defining it

Instrumental Variables Notes II

Note:

② z_i is always constructed so that its exogenous components, x_{2i}, "instrument for themselves"

▸ So $z_i' = [z_{1i}' \quad x_{2i}']$

★ (where z_{1i} are the instruments for the RHS endog variables, x_{1i})

▸ The real trick is therefore to find the z_{1i},

★ In our example, $z_{1i}' = [w_i]$, $x_{2i}' = [1 \quad inc_i]$

★ And thus $z_i' = [w_i \quad 1 \quad inc_i]$

Instrumental Variables Notes III

Note, cont:

3. What might qualify as instruments? The two key IV conditions suggest the answer:

   1. <u>Instrument Exogeneity</u> $\Rightarrow$ variables uncorrelated with the error term

      * (i.e. Exogenous variables)

   2. <u>Instrument Relevance</u> $\Rightarrow$ variables *in the (some!) model*

      * (*Even if the full model isn't specified*)

      * (Without this, there is no justification for the claim that $\Sigma_{ZX} \neq 0$)

Instrumental Variables Notes IV

- All the exogenous variables (across all equations) are candidates.
  - ▶ Thus cost/competition shifters are good candidates in our example:
    - ★ They both plausibly influence our (unspecified) supply equation...
    - ★ (And - hopefully - can be assumed to be exogenous)
    - ★ (And don't belong in the demand curve directly)
  - ▶ What if we haven't written down a full system of simultaneous equations?
    - ★ Where do instruments come from in this case???
    - ★ Answer:* _____

IV & 2SLS: Just- versus Over-identified

- The exact form of the IV estimator depends on the number of
  - ▶ Instruments versus RHS variables (parameters)
- If $L = K$, we say we are just-identified.
  - ▶ We can estimate by (simple) IV
    - ⋆ Note the condition $L = K$ is the same as the condition $L_1 = K_1$
- If $L > K$, we say we are over-identified.
  - ▶ We can estimate by Two-Stage Least Squares (2SLS)
- And finally if $L < K$, we say we are* _____
  - ▶ * _____

OLS and IV as moment estimators

Aside: An alternative way to derive the OLS estimator I

- You've been taught OLS minimizes the sum of squared residuals
  - It can also be rationalized another - completely different - way

- Recall our key assumption that $E(\epsilon_i|x_i) = 0 \Rightarrow Cov(x_i, \epsilon_i) = 0$
  - This is a "population moment condition", i.e.
  - It is a *moment condition* that we assume *holds in the population*

- Moment conditions play an important role in econometrics
  - One of the most widely used estimation methods is the Generalized Method of Moments (GMM)

Aside: An alternative way to derive the OLS estimator II

- The general idea is to solve a system of equations for parameters that satisfy the moment condition in one's sample

  ▶ The sample analog to $Cov(x_i, \epsilon_i) = 0$ is $\frac{1}{N}X'e = 0 \iff X'e = 0$

- Suppose we imposed (A1, Linearity) and (A2, Mean-zero error)

$$
\begin{aligned}
y &= X\beta + \epsilon \qquad \text{is the PopRegFn} \\
&= X\tilde{\beta} + e \qquad \text{is the SampRegFn}
\end{aligned}
$$

  ▶ ...and decided to choose an estimator, $\tilde{\beta}$ to satisfy the moment condition, $X'e = 0$

Aside: An alternative way to derive the OLS estimator III

$$
\begin{aligned}
X'e &= 0 \\
X'(y - X\tilde{\beta}) &= 0 \\
X'y &= X'X\tilde{\beta} \\
\tilde{\beta} &= (X'X)^{-1}X'y
\end{aligned}
$$

- That looks familiar! $\tilde{\beta} = \hat{\beta}$, the OLS estimator of β!

  $\hat{\beta}^{OLS}$ solves the sample analog of the population moment condition, $Cov(x_i, \epsilon_i) = 0$

IV as the solution to a moment condition

- This same principle underlies all IV estimators
  - i.e., both the just- and over-identified cases
- Instead of imposing the OLS population moment condition, $Cov(x_i, \epsilon_i) = 0$,
  - IV does so with the instrument matrix, z_i, i.e. $Cov(z_i, \epsilon_i) = 0$
- Why using z_i and not x_i?
  - Because of endogeneity, $Cov(x_i, \epsilon_i) \neq 0$...
  - ... but with a valid instrument matrix $Cov(z_i, \epsilon_i) = 0$)

Just-identified IV

Just-identified IV Estimation I

- In the simplest case, we are just-identified, i.e.

    - $L = K$, i.e.

        - ⋆ the number of instruments we have is equal to the number of parameters, OR

    - $L_1 = K_1$, i.e.

        - ⋆ the number of instruments for our RHS endogenous variables equals the number of our RHS endogenous variables

    - (These are the same because the K_2 elements in x_{2i} are present in both X and Z...

        - ⋆ ...and therefore don't influence whether or not you're just-identified)

Just-identified IV Estimation II*

- In the just-identified case, the IV estimator just sets the sample analog of the (identifying) moment condition ($Z'\epsilon = 0$) to 0

$$Z'e_{IV} = 0 \Rightarrow$$
$$Z'(y - X\hat{\beta}_{IV}) = 0 \Rightarrow$$

$$\boxed{\hat{\beta}_{IV} = (Z'X)^{-1}Z'y}$$

   ▸ (where $e_{IV} = y - X\hat{\beta}_{IV}$ is the residual from the IV estimator)

- Being just-identified is important, as it ensures that $(Z'X)$ is square

   ▸ (Assumption (IV2) plus a final assumption that rank(Z) = rank(X) = K ensures $Z'X$ is invertible)

   ▸ (This is just the IV-analog to "No Perfect Multicollinearity")

Just-identified IV Properties: Consistency

- As suggested, we focus on the consistency rather than the unbiasedness of the IV estimator

- By the properties of the probability limit operator:

$$
\begin{aligned}
\hat{\beta}_{IV} &= \beta + (Z'X)^{-1}Z'\epsilon \\
\text{plim } \hat{\beta}_{IV} &= \text{plim } \beta + \text{plim } (Z'X)^{-1}\text{plim } Z'\epsilon \\
\text{plim } \hat{\beta}_{IV} &= \beta + \text{plim } (\frac{1}{N}Z'X)^{-1}\text{plim } \frac{1}{N}Z'\epsilon \\
&= \beta + \Sigma_{ZX}^{-1} \; 0 \quad \text{by (IV1)} + \text{(IV2)} \\
&= \beta
\end{aligned}
$$

  ▸ Thus the IV estimator is consistent

Just-identified IV Properties: Asymptotic Normality I

- We turn next to the asymptotic distribution of $\hat{\beta}_{IV}$

- From the last slide, we can write

$$
\begin{aligned}
\hat{\beta}_{IV} &= \beta + (Z'X)^{-1}Z'\epsilon \\
\Rightarrow \sqrt{N}(\hat{\beta}_{IV} - \beta) &= (\tfrac{1}{N}Z'X)^{-1}\tfrac{1}{\sqrt{N}}Z'\epsilon
\end{aligned}
$$

  ▶ (This should look familiar - it's very similar to the OLS formulas...

    ★ ...but with Z' in place of X')

Just-identified IV Properties: Asymptotic Normality II*

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) = (\tfrac{1}{N}Z'X)^{-1}\tfrac{1}{\sqrt{N}}Z'\epsilon$$

- We know from earlier that

$$(\tfrac{1}{N}Z'X)^{-1} \xrightarrow{p} [E(Z'X)]^{-1} = \Sigma_{ZX}^{-1} \qquad \text{and}$$

$$\tfrac{1}{N}Z'Z \xrightarrow{p} E(Z'Z) = \Sigma_{ZZ}$$

- What of $\tfrac{1}{\sqrt{N}}Z'\epsilon$?

Just-identified IV Properties: Asymptotic Normality III

- We could (but won't) show that

$$\frac{1}{\sqrt{N}} Z'\epsilon \xrightarrow{d} N(E(x_i\epsilon_i), V(z_i\epsilon_i))$$

where

- ► $E(z_i\epsilon_i) = \underline{\hspace{2cm}}$

- ► The form of $V(z_i\epsilon_i) = E(\epsilon_i^2 z_i z_i') = E(\epsilon_i^2)E(z_i z_i')$ depends on further assumptions about the variance-covariance matrix of ϵ

Just-identified IV Properties: Asymptotic Normality IV

- If we're willing to make our earlier assumptions (A3, Homoskedasticity) and (A4, No Correlation), then

$$
\begin{aligned}
V(z_i \epsilon_i) &= E(\epsilon_i)^2 E(z_i z_i') \\
&= \sigma^2 E(z_i z_i') \qquad \text{under (A3) and (A4)} \\
&= \sigma^2 \Sigma_{zz} \qquad \text{under (our new) (A6')}
\end{aligned}
$$

- By Slutsky's Theorem, then:

$$
\begin{aligned}
\sqrt{N}(\hat{\beta}_{IV} - \beta) &\overset{d}{\to} \Sigma_{ZX}^{-1} N(0, \sigma^2 \Sigma_{ZZ})) \\
&= N(0, \sigma^2 \Sigma_{ZX}^{-1} \Sigma_{ZZ} \Sigma_{ZX}^{-1})
\end{aligned}
$$

Just-identified IV Properties: Asymptotic Normality V

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{ZX}^{-1} \Sigma_{ZZ} \Sigma_{ZX}^{-1})$$

- We can approximate this in small samples with

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \overset{a}{\sim} N(0, s^2(\tfrac{1}{N}Z'X)^{-1}(\tfrac{1}{N}Z'Z)(\tfrac{1}{N}Z'X)^{-1})$$

$$\hat{\beta}_{IV} \overset{a}{\sim} N(\beta, \tfrac{s^2}{N}(\tfrac{1}{N}Z'X)^{-1}(\tfrac{1}{N}Z'Z)(\tfrac{1}{N}Z'X)^{-1})$$

$$\hat{\beta}_{IV} \overset{a}{\sim} N(\beta, s^2(Z'X)^{-1}(Z'Z)(Z'X)^{-1})$$

  - where $e_i = y_i - x_i'\hat{\beta}_{IV}$ and $s^2 = \tfrac{1}{N}\sum_{i=1}^{N} e_i^2$

    ★ (Note for IV, we always get the "sandwich formula", even if (A3) and (A4) hold)

Just-identified IV Properties: Summary

- Based on these two properties, we can say that the IV estimator is consistent and asymptotically normally distributed:

$$\hat{\beta}_{IV} \quad \overset{a}{\sim} \quad N(\beta, s^2(Z'X)^{-1}(Z'Z)(Z'X)^{-1})$$

- If the variance-covariance matrix of the errors is heteroskedastic,* then the asymptotic distribution of $\hat{\beta}_{IV}$ is

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \overset{d}{\to} N(0, \Sigma_{ZX}^{-1} \lim_{N \to \infty} E\left[\left(\frac{1}{N}\sum_{i=1}^{N} \epsilon_i^2 z_i z_i'\right)\right] \Sigma_{ZX}^{-1})$$

and we can approximate it with

$$\hat{\beta}_{IV} \quad \overset{a}{\sim} \quad N(\beta, (Z'X)^{-1}(\sum_{i=1}^{N} e_i^2 z_i z_i')(Z'X)^{-1})$$

Over-identified IV / 2SLS

Latex Color = "LightSeaGreen"

Two-Stage Least Squares (2SLS)

- If one has more instruments than parameters $(L > K)$, then our moment conditions are:

$$Z'\epsilon = Z'(y - X\beta) = 0$$

- This is of dimension $L \times 1$, but we only have K unknowns in β

  ▸ There generally will not be any way to exactly satisfy *all* the L moments

- Q: So what do we do? Which $\hat{\beta}_{IV}$ should we choose?*

  ▸ _____

  ▸ _____

  ▸ _____

2SLS - Two-Stage Least Squares I

If interpreted *literally*, Two-Stage Least Squares proposes:

1. <u>Stage 1</u>: Regress the included RHS endogenous variables, x_{1i}, on all the instruments, z_i

$$
\begin{aligned}
X_1 &= Z\pi + v & \Rightarrow \\
\hat{\pi} & & \Rightarrow \\
\hat{X}_1 &= Z\hat{\pi}
\end{aligned}
$$

- ▶ Where we assume for simplicity that we have just a single endogenous variable, i.e. X_1 is $N \times 1$
  - ★ (If we had more, we would run regress each column of X_1 on the full matrix of instruments, Z, getting a $\hat{\pi}$ for each column)
- ▶ This is (These are) called the "First-Stage Regression(s)"

2SLS: The First Stage I*

- Economists play close attention to the results of the first stage when evaluating an IV regression

- They pay particular attention to the joint significance of the instruments for X_1, Z_1, in $Z = [Z_1 \quad X_2]$

  ▶ i.e. in the reduced-form equation,

  $$x_{1i} = z_{1i}'\pi_1 + x_{2i}'\pi_2 + \nu_{i1}$$

  ▶ We calculate the F-statistic for the hypothesis, $H_0 : \pi_1 = 0$.

    ⋆ This has L_1 numerator degree of freedom and $N - L - 1$ denominator degrees of freedom

2SLS: The First Stage II

- A typical rule of thumb is that the F-test on the instruments *should exceed 10*.
  - ▶ (If you have just one instrument, the F-statistic is the square of the t-statistic...
  - ▶ ...so this means a t-statistic in excess of $\sqrt{10} = 3.16$)
    - ★ [This is from the paper by Staiger and Stock (1997)]

- We'll discuss why once I give the intuition of the IV estimator

2SLS - Two-Stage Least Squares II

② <u>Stage 2</u>: "Regress" y on $\hat{X}_1$ and X_2, i.e.

$$
\begin{aligned}
y &= \hat{X}_1\beta_1 + X_2\beta_2 + \epsilon \\
&= \hat{X}\beta + \epsilon \qquad \text{where} \quad \hat{X} = [\hat{X}_1 \quad X_2]
\end{aligned}
$$

③ That's it!

④ Those are the two stages in Two-Stage Least Squares, but ...

   ▸ How does that compare with the intuition we covered earlier?

2SLS Intuition I

No surprise: it's the same intuition:

- We'd like to estimate

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

  - but X_1 is endogenous, i.e. $E(\epsilon|X_1) \neq 0$

  - (And even if $E(\epsilon|X_2) = 0$, we showed that correlation between ϵ and *any* X's biases *all* coefficients)

- We estimate instead:

$$y = \hat{X}_1\beta_1 + X_2\beta_2 + \epsilon$$

  - where $\hat{X}_1 = Z\hat{\pi}$ from the First-Stage regression

    - ★ Critically: $\hat{X}_1$ is that part of X_1 that *can be explained by Z*

2SLS Intuition II

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon$$
$$y = \hat{X}_1\beta_1 + X_2\beta_2 + \epsilon$$

2SLS Intuition, cont.

- We've essentially replaced the "problematic" (i.e. endogenous) X_1 with its prediction as a function of Z

  - Since it's a function of Z, *all of which are exogenous*, $\hat{X}_1$ isn't correlated with ϵ... even if X_1 is

    - (Because under Assumption (IV1), $E(\epsilon|Z) = 0$)

- Thus no more endogeneity problem!

  - $\Rightarrow$ consistent estimation of β_1 and β_2!

IV Intuition Redux I

- So the intuition of 2SLS is:
  - ▶ "Replace the problematic variables with their fitted values (which are a function of exogenous instruments)"

- Another way to say the same thing goes like this:

- Imagine taking one of the elements of the problematic (endogenous) X_1

- And dividing it into two parts:
  1. A problematic part
     - ★ Which is correlated with ϵ_i
  2. A non-problematic part
     - ★ Which is uncorrelated with ϵ_i

IV Intuition Redux II

- We'd like to rely only on the *non-problematic* variation in X_1 to identify β
  - Leaving aside the problematic variation
- This is imperfect of course
  - As relying on less than the full variation in any variable means we'll get noisier estimates of $\hat{\beta}$
  - But imprecision is better than bias!

IV Intuition Redux III

- This is exactly what IV does:
  - ▶ It relies on variation in X_1 *induced by variation in the instruments, Z*
  - ▶ Since these instruments are uncorrelated with ϵ, this is "non-problematic" variation
    - ★ You can imagine...

  See Figure in Class

2SLS - Two-Stage Least Squares III

- Important note:
    - ▸ You should <u>never</u> literally run the 2 regressions in 2SLS as it gives the wrong standard errors for $\hat{\beta}_{2SLS}$
        - ⋆ Why? Because we are using a <u>fitted value</u> in the second stage regression ($\hat{X}_1$)
        - ⋆ This introduces additional error into $V(\hat{\beta}_{2SLS})$, but Stata and R don't know you are using a fitted value
        - ⋆ (and therefore don't account for it)
- Of course, you could bootstrap the standard errors, but
    - ▸ I didn't teach you how to bootstrap standard errors.   [:-)]
    - ▸ There is an easier solution

2SLS - Two-Stage Least Squares IV

- Instead: Simply use $\hat{X}_1$ as part of a just-identified IV specification,

  ▶ i.e. instrument for $X = [X_1 \quad X_2]$ with $Z_{2SLS} = [\hat{X}_1 \quad X_2]$

    ★ This is $N \times K$ no matter how many instruments were in Z_1

$\Rightarrow$ Thus, from our earlier equation: $\hat{\beta}_{2SLS} = \hat{\beta}_{IV}$ using $Z_{2SLS} = [\hat{X}_1 \quad X_2]$ as the instrument matrix

$$\hat{\beta}_{2SLS} = (Z'_{2SLS}X)^{-1}Z'_{2SLS}y$$

- We're all set to show the properties of the 2SLS estimator...

  ▶ Except to do so we need a convenient way to represent $\hat{X}_1$

2SLS using P_Z I

- Recall our projection matrix from earlier, $P_X = X(X'X)^{-1}X'$
  - Which gives the fitted value of a regression of <anything> on X

- Earlier, we said the first-stage regression is the regression of x_{1i} on z_i:

$$
\begin{aligned}
X_1 &= Z\pi + v &\Rightarrow \\
\hat{\pi} &= (Z'Z)^{-1}Z'X_1 &\Rightarrow \\
\hat{X}_1 &= Z\hat{\pi} \\
&= Z(Z'Z)^{-1}Z'X_1 \\
\Rightarrow \hat{X}_1 &= P_Z X_1
\end{aligned}
$$

2SLS using P_Z II

- Recall that a general instrument vector is written as $Z = [Z_1 \quad X_2]$
  - So it includes the instrument(s) for X_1 (Z_1) as well as the exogenous variables, X_2
- We can use this fact and the properties of the projection matrix to write Z_{2SLS} in a convenient form:
  - (1) $P_Z X_2 = ???$
  - (2) Last slide we showed that $P_Z X_1 = \hat{X}_1$
  - (1) + (2) $\Rightarrow Z_{2SLS} \equiv [\hat{X}_1 \quad X_2] = P_Z[X_1 \quad X_2] = P_Z X$
- Thus:

$$\boxed{Z_{2SLS} = P_Z X \qquad \text{and} \qquad Z'_{2SLS} = X'P_Z}$$

2SLS using P_Z III

$$Z_{2SLS} = P_Z X \qquad \text{and} \qquad Z'_{2SLS} = X'P_Z$$

- These allow us to write the formula for the 2SLS estimator in terms of its fundamental elements

$$
\begin{aligned}
\hat{\beta}_{2SLS} &= (Z'_{2SLS}X)^{-1}Z'_{2SLS}y \\
&= (X'P_ZX)^{-1}X'P_Zy
\end{aligned}
$$

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y$$

  ▶ (This representation is much easier to work with to show consistency and asymptotic normality)

2SLS Properties: Summary I

- One can derive the properties of the 2SLS estimator in much the same way as we did for the just-identified IV estimator
  - ▸ So I'll just summarize the formulas for you here:
    - ★ The 2SLS estimator is consistent:

      $$\hat{\beta}_{2SLS} \xrightarrow{p} \beta$$

    - ★ Under (A3) and (A4) - homoscedasticity and no serial correlation - the 2SLS estimator is asymptotically normal with approximate distribution:

      $$\hat{\beta}_{2SLS} \overset{a}{\sim} N(\beta, \sigma^2 (X' P_Z X)^{-1})$$

      $$\overset{a}{\sim} N(\beta, \sigma^2 (X' Z (Z' Z)^{-1} Z' X)^{-1})$$

2SLS Properties: Summary II

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'P_Z X)^{-1} X' P_Z y \\
\hat{\beta}_{2SLS} &\overset{a}{\sim} N(\beta, \sigma^2 (X' P_Z X)^{-1})
\end{aligned}$$

- Compare the formulas above to the OLS formulas:
  - $\hat{\beta} = (X'X)^{-1} X'y$ and $\hat{\beta} \overset{a}{\sim} N(\beta, \sigma^2 (X'X)^{-1})$

- Endogeneity requires us to instrument for "problematic" X_1 with Z_1, with two effects:
  1. We can only rely on that variation in X induced by Z (in blue)
     - (i.e. $X'P_Z X$ instead of $X'X$)
  2. And this induces *higher variance* in $V(\hat{\beta}_{2SLS})$ relative to $V(\hat{\beta})$ (in red)
     - Due to relying only on that part of the variation in X induced by Z, $X'P_Z X$

Instrumental Variables Comments I

- As mentioned earlier, in the over-identified case, we have more moments than parameters to estimate
    - So we may not be able to find a single β that solves all of them
- The 2SLS estimator is the most efficient one
    - (Under our assumptions, particularly (A3) and (A4))
- More generally, one must weight the L moments to solve for the K parameters in β
    - The next - optional - slides show the form of the IV estimator for an arbitrary weighting matrix, W
    - (Where the 2SLS weighting matrix is $W = (Z'Z)^{-1}$)

Instrumental Variables Comments II*

- Formally, let M be an $(L \times L)$ matrix of weights to give to each of our moments

- Let Q be the quadratic form (the matrix analog to a sum of squares & cross products) in the sample moments:

$$Q = \underbrace{(y - X\beta)'Z}_{(1 \times L)} \underbrace{W}_{(L \times L)} \underbrace{Z'(y - X\beta)}_{(L \times I)}$$

Note:

- ▸ Q is a scalar

- ▸ The optimal weighting matrix in general is proportional to the inverse of the variance of the moment conditions

    - ★ Under (A3) and (A4), $V(Z'\epsilon) = E(Z'\epsilon\epsilon'Z) = (Z'(\sigma^2 I_N)Z) = \sigma^2(Z'Z)$

    - ★ Thus the 2SLS estimator uses the weighting matrix $W = (Z'Z)^{-1}$

Instrumental Variables Comments III*

- For an arbitrary W, taking derivatives w.r.t β and solving yields

$$\hat{\beta} = (X'ZWZ'X)^{-1}X'ZWZ'y$$

- You can also show:

$$
\begin{aligned}
\hat{\beta} &\overset{a}{\sim} N(\beta, AV_\beta) \\
AV_\beta &\approx (X'ZWZ'X)^{-1}X'ZW\Omega WZ'X(X'ZWZ'X)^{-1} \\
\Omega &= V(Z'\epsilon) = E(Z'\epsilon\epsilon'Z)
\end{aligned}
$$

Show yourself you get the 2SLS formula if $\Omega = \sigma^2 Z'Z$ and $W = (Z'Z)^{-1}$

  ▶ (What a complicated mess!)

  ▶ (But useful if you take further econometrics...)

Instrumental Variables Comments IV

- While these formulas are superficially complicated, efficient weighting mechanisms pop up frequently in econometrics:
  - ▶ Weighting moments for efficient estimation (GMM)
    - ⋆ What we've done here
  - ▶ Weighting observations for heteroscedasticity correction (GLS)
  - ▶ Weighting equations for cross-equation correlations (SUR)

IV: Practical Considerations

IV: Practical considerations

- We close our discussion of IV with some practical considerations:

  1. Finding instruments

  2. Avoiding the need for instruments

  3. IV standard errors

  4. Weak Instruments

  5. Testing

     ⋆ Endogeneity

     ⋆ Overidentification

Finding instruments I

- One of the greatest challenges with IV estimation is *finding instruments*

- We need something that's both exogenous w.r.t the error in the equation of interest *and* correlated with an endogenous variable of interest

- This can be very, very challenging

Finding instruments II

- For example, imagine modeling a family's decisions:

  ▶ Whether and how much to work, whether and for how long to go to school, whether and how many children to have, where to live, etc.

  ▶ Each plausibly impacts each other (structural equations), requiring instruments for each in the appropriate equation

  ▶ Good luck!!!

- In practice, researchers usually focus on a subset of the decisions

  ▶ And/or try to jointly model them all as a function of (hopefully!) exogenous "initial conditions"

  ▶ (Where born, parent's characteristics, "innate" attributes, etc.)

- We'll go through many examples of finding instruments in class and on your problem set(s)

Example I

- Indeed, consider a recent student's bachelor's thesis seeking to measure the impact of cross-border workers on anti-immigrant vote shares among a panel of $100+$ Swiss cities in the canton of Ticino over 10 years:

$$v_{it} = \beta_1 + \beta_2 CBW_{it} + \tilde{x}_i' \tilde{\beta} + \epsilon_{it}$$

- First/Most important question: *Is this an interesting research question???*

  ▶ _____

Example II

$$v_{it} = \beta_1 + \beta_2 CBW_{it} + \tilde{x}_i' \tilde{\beta} + \epsilon_{it}$$

- Swiss context:

  ▸ Thesis has interesting description of history of Swiss attitudes/policies towards immigration

    ⋆ Most recent: 2014 vote seeking to set quotas

    ⋆ (Would violate EU/CH agreement on free movement of labor)

    ⋆ Resolved 12/2016 requiring Swiss employers to give preference to Swiss-based job seekers

- Next question: *What do we think is the sign of the causal effect of CBW_{it} on v_{it}*

  ▸ _____

Example III

- Thesis analyzes 115 municipalities in the canton of Ticino
  - ▸ One of the 3 cantons (with Geneva, Basel) with highest number CBWs
  - ▸ Anyone care to guess what share of workers are CBW?
- Ticino often votes most strongly against foreigners in immigration-related referenda
  - ▸ (See FT article, 9/25/16)
  - ▸ (Symbolic - can't be implemented at cantonal level)

Example IV

$$v_{it} = \beta_1 + \beta_2 CBW_{it} + \tilde{x}_i' \tilde{\beta} + \epsilon_{it}$$

- Might there be sources of bias in this specification? If so, which kind(s)?

  ▶ _____

  ▶ _____

Example IV.5*

$$v_{it} = \beta_1 + \beta_2 CBW_{it} + \tilde{x}_i'\tilde{\beta} + \epsilon_{it}$$

- What sign(s) do we anticipate for the bias based on the stories we've told so far?

  ▶ Negative for my stories...

  ▶ (Where "story" = "model of behavior generating the data")

- Is it possible to tell a story where the bias is *positive*?*

  ▶ _____

Example V*

- OK, let's now consider finding an instrument for the number of CBWs

- What would be the qualitative features of a good instrument?
  - ▸ (Hint: Relevance:)   Something correlated with the number of cross-border workers...

  - ▸ (Hint: Exogeneity:)   ... that is uncorrelated with unobserved determinants of anti-immigrant vote shares (ϵ_{it})

- Any ideas???
  - ▸ _____

  <p style="text-align:center; color:red;">Go Through Paper Briefly in Class</p>

Example VI*

Let's explore some of the student's results:

- Skim/Replicate the results in the thesis (WON'T SHOW)

- Data review

- Pooled OLS - what are the right standard errors?

- Baseline: Pooled OLS and IV - ~~Puzzling!~~

- Check alternative specifications (they don't matter)

    ▶ Split out urbanization

    ▶ Drop foreign population share in 2000

Example VII*

More:

- Check data for errors

  - $\rightarrow$ Even "stronger" (Pooled OLS) results!

- More specs:

  - Nonlinear effects of CBW? Vote (time) dummies? Neither matters...

- Data are telling a pretty consistent story...

  - ... it's just not the one I was initially expecting!

  - ~~What's going on???~~ Results now make sense if consider Swiss unemployment as an omitted variable

    - ⋆ Maybe panel data methods "can help" - let's come back to this thesis/these results after I go through those lecture notes

    - ⋆ ("can help" = provide results consistent with our priors - tho be careful! Your priors can be wrong! Ultimately your job as an econometrician is to let the data tell its story!)

Avoiding the need for instruments I

- We've talked to this point as if the presence of a "problematic" RHS variable is a given

    ▶ Induced by there being correlation between at least one x_i and ϵ_i

    ▶ And focused on finding an instrument(s) z_i to shift x_i in a way that's uncorrelated with ϵ_i

- But I have to emphasize that there is another - indeed easier - solution if the endogeneity is due to a correlated unobservable:

    ▶ *Take the problematic variation out of the error term*

    ▶ (if it's possible!)

Avoiding the need for instruments II

- In particular, suppose there is some *potentially measurable* economic variable that
    - Isn't currently in the econometric model
        - And is thus in ϵ_i
    - Is correlated with one of our x's
        - Causing endogeneity
- If we can just include that variable in the econometric model, all is good:
    - We'll be picking up its effect (taking it *out* of ϵ_i)
    - (Hopefully) leaving whatever is left in ϵ_i uncorrelated with x_i
        - Permitting consistent estimation with OLS

Avoiding the need for instruments III*

- Some support for this line of argument in the academic literature
    - e.g. Rossi (2014, *Marketing Science*), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications"

- Thus always keep this strategy in the back of your mind!
    - If it's possible, it is very likely to be a better strategy than IV as retains all the variation in x_i to estimate β

    - (Example: Online Word-of-mouth paper - if time/after panel data methods)

IV Standard Errors I

- It's very common for standard errors on IV estimates to be larger - and perhaps much larger - than their OLS counterparts

- We've hinted at why this is a few slides ago

IV Standard Errors II

- IV "works" by having an instrument, z_{1i}, "move around" the "problematic"/included RHS endogenous variable, x_{1i}

    ► Recall $\hat{\beta}_{2SLS} = (X'P_Z X)^{-1} X'P_Z y$

- This is necessarily imperfect, as z_{1i} isn't perfectly correlated with x_{1i}

    ► (And if it was, then it *wouldn't* satisfy (IV1), Instrument Exogeneity)

- As a result, there is less instrument-induced-variation in x_{1i} than x_{1i} could provide by itself (if it weren't endogenous)

    ► And less x-variation means higher standard errors

Weak Instruments I

- Many applied researchers often worry about *Weak Instruments*
- What does this mean?
    - Weak instruments have no precise definition
    - A useful working definition is instruments for which the F-test on the joint significance of the instruments in the first-stage regression is "small"
    - (Recall we wanted an F-test of about 10 when we had a single RHS endogenous variable)

Weak Instruments II

- Weak instruments are problematic because they can exacerbate *finite-sample bias*
  - i.e. IV may be consistent in large samples, but is in general still biased
    - ⋆ (The bias is generally towards the OLS estimate)
  - And this bias increases the weaker is the correlation between instruments and included endogenous variables

Weak Instruments III

- With a single endogenous regressor, Staiger and Stock (1997, *Econometrica*) estimate the size of the bias as $1/F$

    - Where F is the value of the first-stage F-statistic introduced earlier

- With multiple endogenous regressors, Stock, Wright, and Yogo (2002) calculate the minimum eigenvalue of a matrix-analogue to the first-stage F-statistic

    - And provide critical values for this eigenvalue for different bias tolerances

IV Tradeoffs I*

- Larger standard errors and bias from weak IVs are key parts in the thesis of the Rossi paper "Even the Rich Can Make Themselves Poor" paper

- IV comes with significant costs, both because:

  1. You need to rationalize that your IV satisfies (IV2) Relevance and (especially) (IV1) Exogeneity

     ★ Noting that the former can be tested but the latter usually not

  2. Even if these are satisfied, IV gives

     ★ (Possibly much) larger standard errors

     ★ Parameters affected by small sample bias

- But... without an IV can you argue that you've identified causal effects???

  ▸ ⇒ Hard Tradeoffs!!!

IV Tradeoffs II*

- Well, this is all quite depressing
  - ▶ What should you do???

- No great options, but here's what we have:

  1 Try to include covariates that soak up the source of endogeneity

  2 If the problem is due to too many instruments, some of which are weak, then just drop the weak ones and keep the strong ones

     ★ (Can test with individual t-values in the first-stage regression(s))

  3 If problem is small-sample bias, can use estimators other than 2SLS

     ★ e.g, LIML, Split-sample IV, Jackknife IV

     ★ But all have downsides (harder to implement, less efficiency, etc.)

  4 Consider alternative research designs

     ★ Particularly (field) experiments

Testing

Testing*

- There are two types of tests that often arise in the context of IV estimation

  1. Testing endogeneity

  2. Testing overidentifying restrictions

     ⋆ (a.k.a. testing the validity of (some of) your instruments)

Endogeneity Test I

- Since we've spent so much time worrying about endogeneity...
  - ▶ Finding them, higher standard errors, weak instruments, etc.
- It would be nice if we could test for it!
- Indeed, there is a test as long as you have at least one potential instrument
  - ▶ (So we're not completely out of the woods)
  - ▶ (And note: we are testing for endogeneity...
  - ▶ ... *conditional* on the validity of our instrument)
- How?

Endogeneity Test II

- Consider both structural and reduced-form equations with scalar x_{1i}:

$$y_i = x_{1i}\beta_1 + x'_{2i}\beta_2 + \epsilon_{i1} \qquad (1)$$

$$x_{1i} = z'_{1i}\pi_1 + x'_{2i}\pi_2 + \nu_{i1} \qquad (2)$$

- We'd like to test whether x_{1i} is correlated with ϵ_{i1} in (1)

  ▸ Perhaps we can use equation (2) to do so!

  ▸ Since ϵ_{i1} is (by assumption) uncorrelated with both z_{1i} and x_{2i}...

  ▸ ...then the only way x_{1i} could be correlated with ϵ_{i1} is if ν_{i1} is

    ★ Why is this true?

Endogeneity Test III

- We don't observe ν_{i1}, but we can estimate it as the residual in the equation (2), $\hat{\nu}_{i1}$

- We can test whether this correlated with ϵ_{1i} by simply putting it into Equation (1).
    - If the parameter on it is significant, then we have endogeneity

- Formally estimate the following *by OLS*:

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + \delta\hat{\nu}_{i1} + \epsilon_{i1}$$

  and test $H_0 : \delta = 0$

- This is called (Hausman) endogeneity test

Control Functions I*

- Note you might be surprised by the familiarity of the estimated coefficients (other than $\hat{\delta}$ in this regression)
    - ▶ They are the the 2SLS estimates!!!
- Why???
    - ▶ This specification is another way of seeing how 2SLS "works"
    - ▶ We talked about 2 sources of variation in our "problematic" x_{i1}:
        1. The "problematic" variation (correlated with ϵ_{i1})
        2. The "good" variation (uncorrelated with ϵ_{i1})
- (An IV induces only the good variation in x_{i1})

Control Functions II*

- But another way to "fix" the endogeneity problem caused by x_{i1} is to simply add the "problematic" variation *as it's own covariate*

    ▸ And that's exactly what $\hat{\nu}_{i1}$ is!

- Once we include $\hat{\nu}_{i1}$, we allow x_{i1} to vary *controlling for its bad variation*

    ▸ Thus only allowing it's good variation!

    ▸ $\Rightarrow$ the 2SLS estimates

- This general approach of including an estimate of the source of the endogeneity as its own covariate is called "Control Function Estimation"

    ▸ (And $\hat{\nu}_{i1}$ the "control function")

    ▸ As is evident, it can be equivalent to 2SLS

Overidentifying Test I

- In general, we would *love* to test the validity of an instrument
    - In other words, is it really the case that $Z'\epsilon = 0$???
- In principle, we have an estimate of ϵ_i, e_i, so can't we use this as a proxy?
    - Unfortunately not
    - We showed earlier that the $Z'e$ is *exactly* zero when we are just-identified
        - (Thus no test is possible)

Overidentifying Test II

- But... we can test whether $Z'\epsilon = 0$ when we are *over-identified*

  ▸ As $Z'\epsilon = 0$ in the population, but $Z'e$ won't (exactly) equal zero in any sample

  ▸ So the question is whether the deviation of $Z'e$ from zero is statistically large

- Tho note: this way we can test the the validity of *some* of the instruments

  ▸ But may not know which one(s) are causing trouble if we reject!

  ▸ (So a bit frustrating...)

Overidentifying Test III

- The formal test has three steps:

  1. Estimate (1) by 2SLS $\Rightarrow e_{i1}$

  2. Regress e_{i1} on z_{1i} and x_{2i} (i.e. all the exogenous variables)

     ★ And calculate the R^2

  3. Then, under H_0 : *Instrument validity*, $NR^2 \overset{a}{\sim} \chi^2_{L_1 - K_1}$

     ★ where the degrees of freedom equal the number of "extra" (i.e. overidentifying) instruments you have

- Last note: the overidentifying test is often also called the "J-Test" (from Sargan (1958))

Examples Redux

Examples Redux

- Recall that we said there are three common sources of endogeneity:

  1. Correlated unobservables

  2. Measurement error (in X)

  3. Reverse causality / Simultaneous equations

- I also showed three examples whose results suggested bias of the type we had anticipated

  - Can we "solve" these issues using IV estimation?

Correlated Unobservables Example Redux I

- For the first, correlated unobservables/wage-ability example, we use data on married women's wages

    ▶ And found (perhaps implausibly) high estimates of the returns to education (of 10.9%)

- Authors have proposed using as an instrument for a woman's education the *education of her parents*

- Do you think these are likely to satisfy the conditions for a good instrument?*

    ▶ Relevance: _____

    ▶ Exogeneity:

    _____

Correlated Unobservables Example Redux II

- Remember what we said was the likely sign of the bias?

  ▶ _____

- Let's see what happens when we instrument with father's education

  See Stata Example in Class

- Do these results seem (more) reasonable to you?*

  ▶ _____

  ▶ _____

Measurement Error Example Redux I

- For this second, measurement-error example, we used data from average health insurance coverage across the 50 US states (and the District of Columbia) in 2007

  ▶ And (perhaps) found evidence of attenuation bias (0.23% pretty small)

- The author of the study from which I grabbed the data suggested using average high-school completion rate as an IV for their permanent income

- Do you think these are likely to satisfy the conditions for a good instrument?*

  ▶ Relevance: _____

  ▶ Exogeneity:

  _____

Measurement Error Example Redux II*

- Let's see what happens when we instrument with whether someone got an advanced degree

<p align="center" style="color:red">See Stata Example in Class</p>

- Do these results seem (more) reasonable to you?*

    ▶ _____

    ▶ _____

- Which do you believe (most)?*

    ▶ _____

Simultaneous Equations Example Redux I

- For our final, simultaneous equations example, we use annual data on the purchase of truffles

- This is a classic Supply and Demand environment

  ▶ So it's natural to use a supply shifter as an instrument for a demand equation

  ▶ In this data we have the rental price of...

Simultaneous Equations Example Redux II

- Do you think this is likely to satisfy the conditions for a good instrument?*

    ▶ Relevance: _____

    ▶ Exogeneity:

        _____

- Remember what we said was the likely sign of the bias?

    ▶ _____

Simultaneous Equations Example Redux III

- Let's see what happens when we instrument with the rental price of truffle pigs

<div align="center" style="color:red">See Stata Example in Class</div>

- Do these results seem (more) reasonable to you?*

  ▶ _____

IV Conclusions I

- Understanding the potential sources of bias in regression...

  ▶ ...and how to resolve it using Instrumental Variables...

  ▶ ...*cannot* be overestimated as a key skill in your "Econometrician's toolkit"

- In my view it's what separates "Careful Econometricians" from "Regression Runners"

  ▶ Please be one of the former and not one of the latter!

IV Conclusions II

- In the remainder of the course, we will cover the second-most common method to address endogeneity bias

  ▶ Panel Data Methods

Coming Up Next!

Table of Contents