Webinar Series in Applied Quantitative Analysis - Updated

Date	Topic	
	Session One	
February 29	Potential Outcomes and Omitted Variable Bias I (Theory)	✓
March 7	Potential Outcomes and Omitted Variable Bias II (Application)	
	Session Two	
March 21	Difference-in-differences I (Theory)	✓
March 28	Difference-in-differences II (Application)	
	Session Three	
April 25	Power analysis, clustering and sample size calculations I (Theory)	✓
May 2	Power analysis, clustering and sample size calculations II (Application)	
	Session Four	
May 23	Propensity score matching (Theory)	✓
May 30	Propensity score matching (Application)	
	Session Five	
June 20	Fixed-effects I (Theory)	✓
June 27	Fixed-effects II (Application)	
	Session Six	
July 25	Instrumental variables I (Theory)	
August 1	Instrumental Variables II (Application)	
	Session Seven	
August 22	Lagged dependent variables and the Arellano-Bond Estimator I (Theory)	
August 29	Lagged dependent variables and the Arellano-Bond Estimator II (Application)	

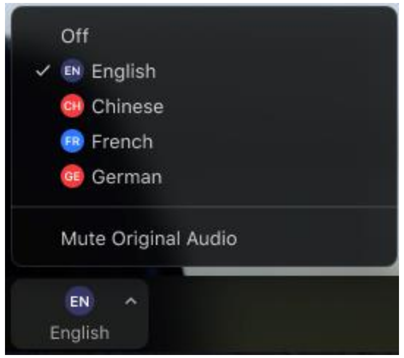
AIR®
Advancing Evidence.
Improving Lives.

https://www.airp3-africa.org/applied-quantitative-analysis-webinar-series

AERC
CREA

Écoute de l'interprétation d'une langue
Windows | macOS

1. Dans les contrôles de votre réunion/webinaire, cliquez sur Interprétation .
2. Cliquez sur la langue que vous souhaitez entendre. (Nous aurons le français) Pas besoin de choisir l'anglais, c'est la langue de la salle Zoom principale



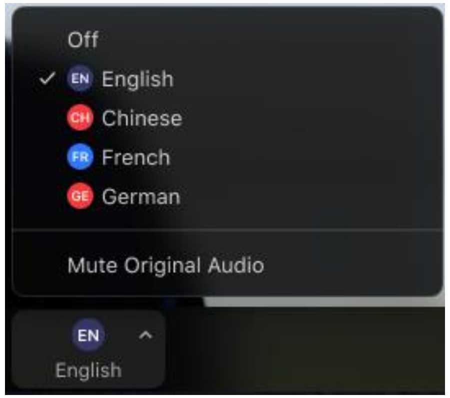
3. (Facultatif) Pour entendre uniquement la langue interprétée, cliquez sur Couper le son original.

Remarques:

- Vous devez rejoindre l’audio de la réunion via l’audio/VoIP de votre ordinateur. Vous ne pouvez pas écouter l’interprétation linguistique si vous utilisez les fonctions audio de connexion ou d’appel téléphonique.
- En tant que participant rejoignant une chaîne linguistique, vous pouvez retransmettre sur le canal audio principal canal si vous réactivez votre audio et parlez.

Listening to language interpretation
Windows | macOS

1. In your meeting/webinar controls, click Interpretation .
2. Click the language that you would like to hear. (We will have French) No need to choose English, that is the language in the main Zoom room



3. (Optional) To hear the interpreted language only, click Mute Original Audio.

Notes:

- You must join the meeting audio through your computer audio/VoIP. You cannot listen to language interpretation if you use the dial-in or call me phone audio features.
- As a participant joining a language channel, you can broadcast back into the main audio channel if you unmute your audio and speak.

Instrumental Variables Application

Professor Jeremy Moulton

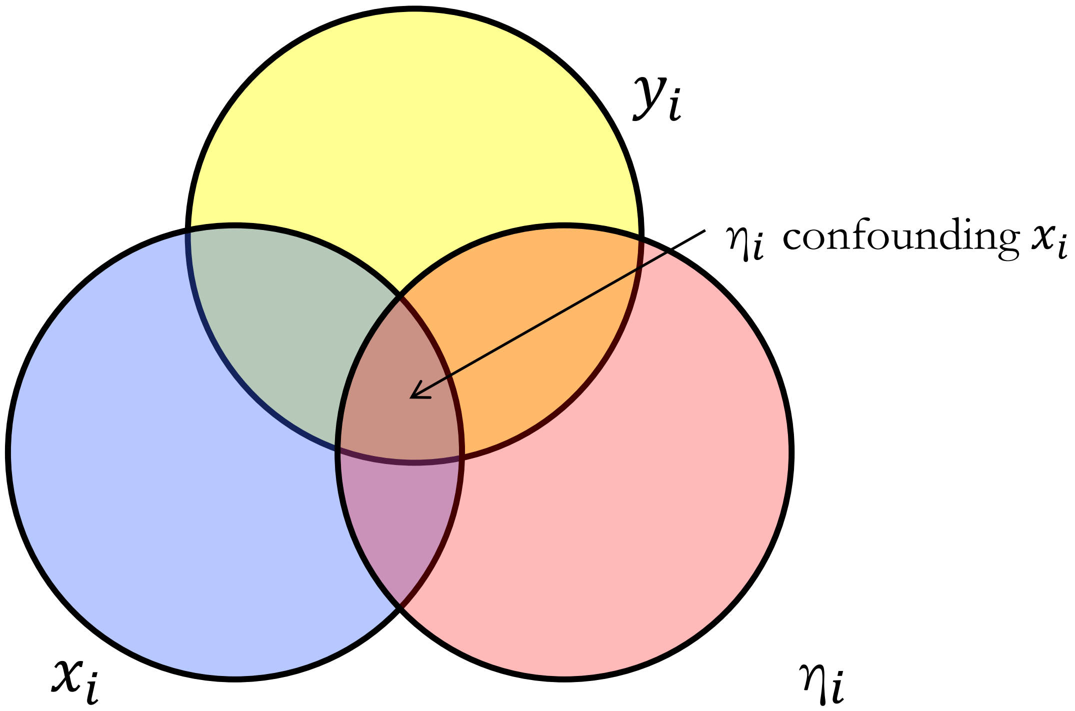
Department of Public Policy

University of North Carolina at Chapel Hill

July 25, 2024

The New Set Up

- $y_i = \alpha + \beta x_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$



- x_i = independent variable of interest
- η_i = error that is correlated with x_i
  - Example: innate ability

- **Main point: We no longer need panel data to "fix" the problem in our error term, but you can use one.**

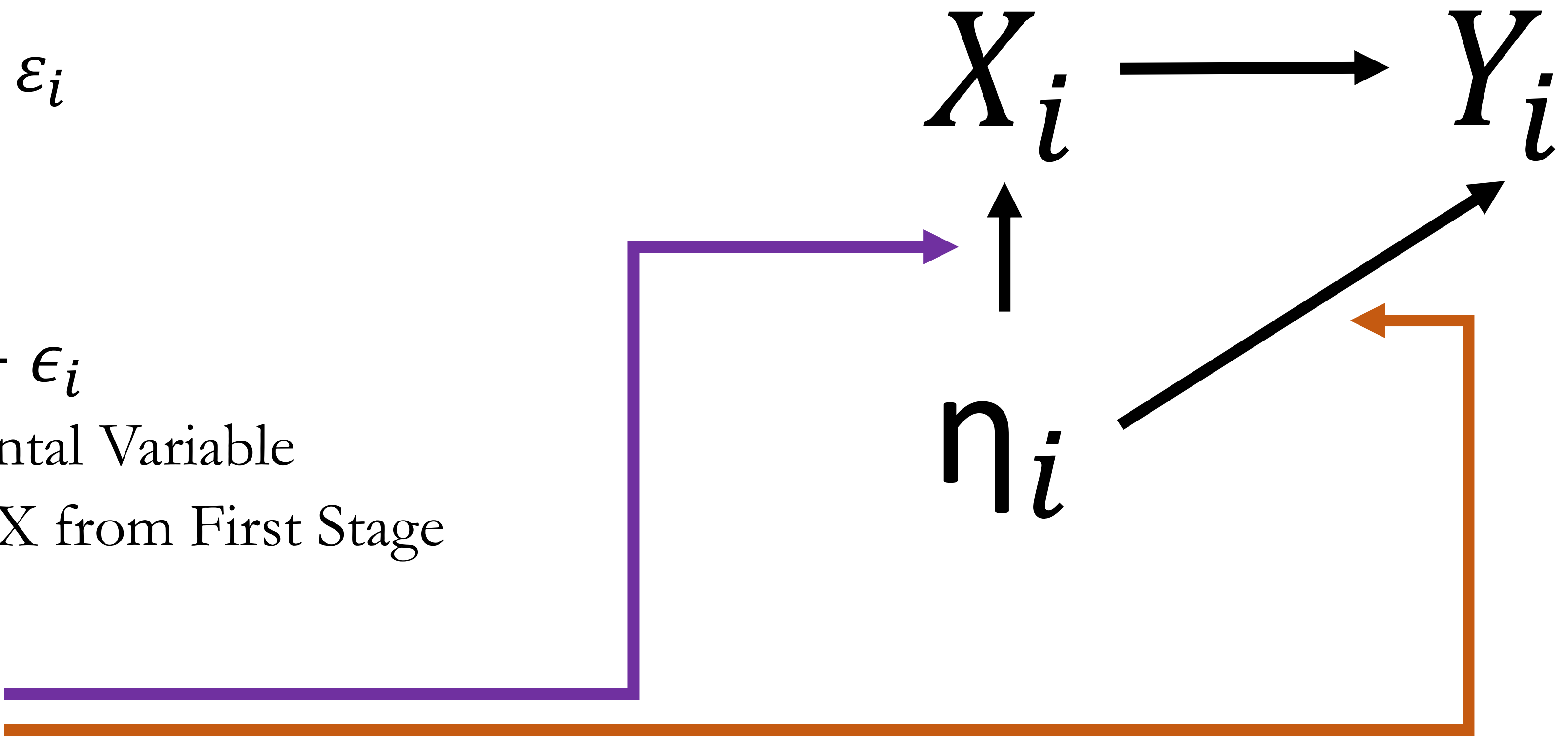
The New Solution (Instrumental Variables)

Second Stage

- $y_i = \alpha + \beta\hat{x}_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$

First Stage

- $x_i = \delta + \gamma IV_i + \epsilon_i$
  - IV_i = Instrumental Variable
  - $\hat{x}_i$ = Predicted X from First Stage
- **Requires:**
  - **cov(X, $\boldsymbol{\varepsilon}$) ≠ 0**



The New Solution (Instrumental Variables)

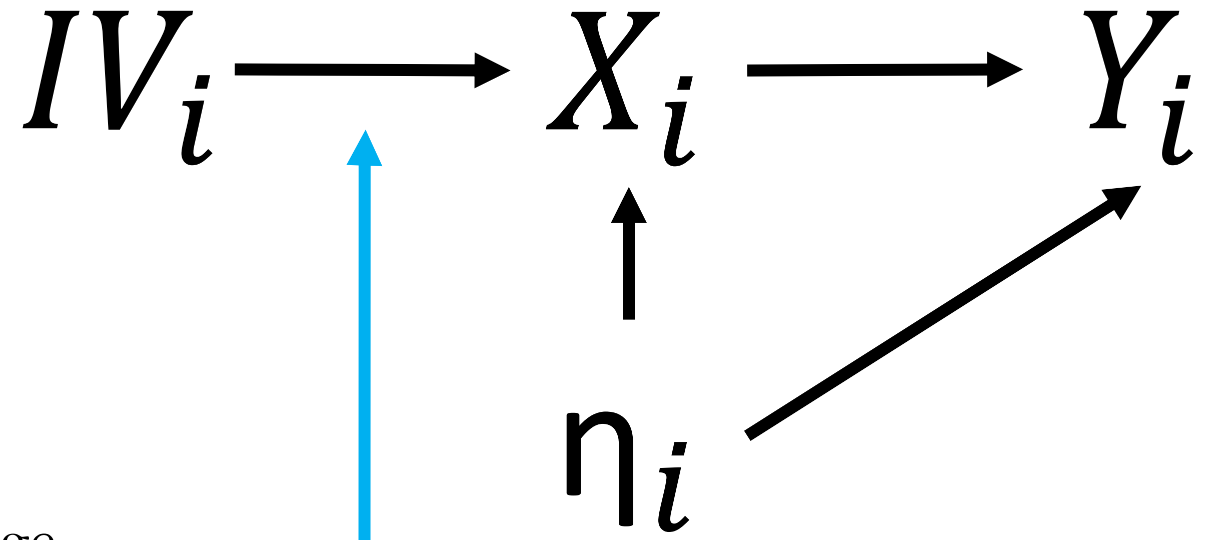
Second Stage

- $y_i = \alpha + \beta \hat{x}_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$

First Stage

- $x_i = \delta + \gamma IV_i + \epsilon_i$
  - IV_i = Instrumental Variable
  - $\hat{x}_i$ = Predicted X from First Stage

- **Requires:**
  - **cov(X, $\boldsymbol{\varepsilon}$) ≠ 0**
  - **cov(IV, X) ≠ 0**



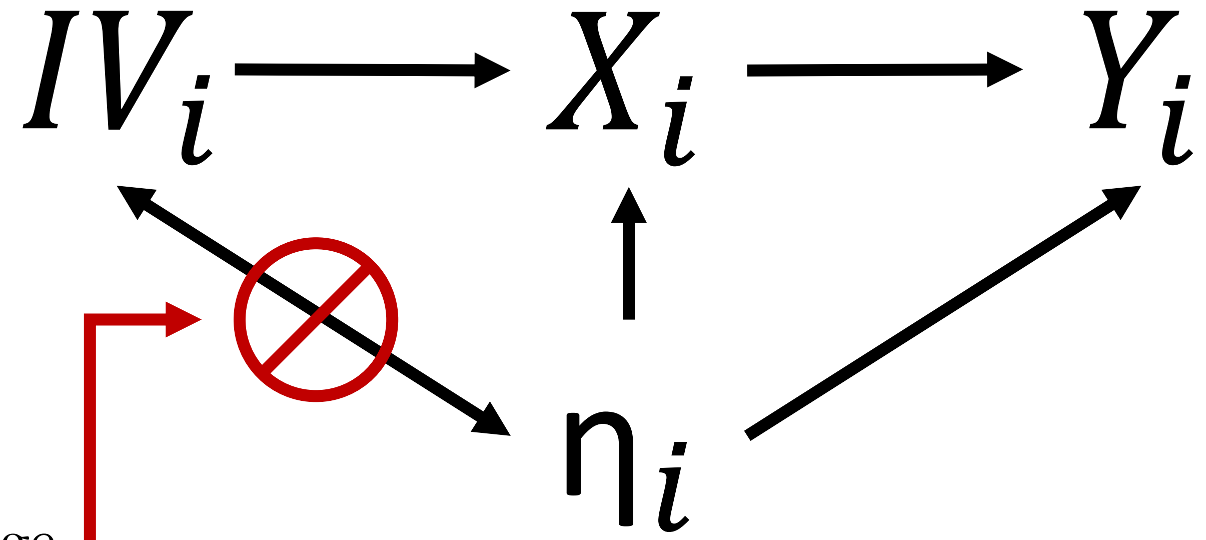
The New Solution (Instrumental Variables)

Second Stage

- $y_i = \alpha + \beta \hat{x}_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$

First Stage

- $x_i = \delta + \gamma IV_i + \epsilon_i$
  - IV_i = Instrumental Variable
  - $\hat{x}_i$ = Predicted X from First Stage
- **Requires:**
  - $\mathbf{cov(X, \varepsilon) \neq 0}$
  - $\mathbf{cov(IV, \varepsilon) = 0}$
  - $\mathbf{cov(IV, X) \neq 0}$



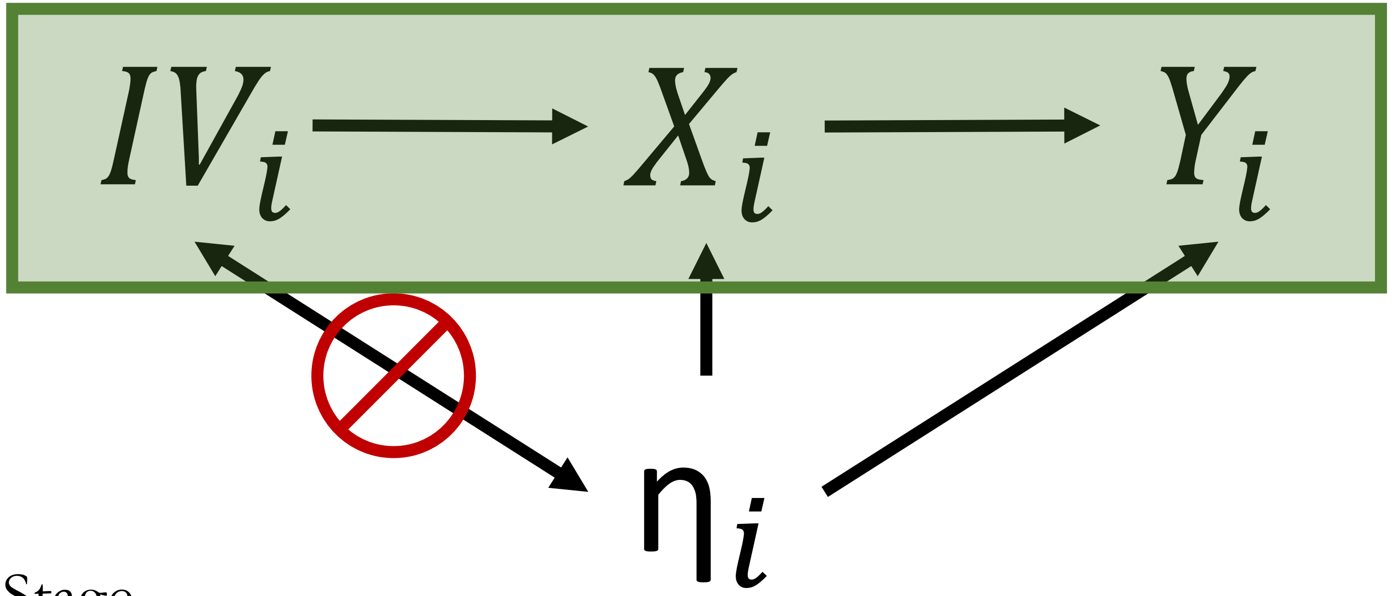
The New Solution (Instrumental Variables)

Second Stage

- $y_i = \alpha + \beta\hat{x}_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$

First Stage

- $x_i = \delta + \gamma IV_i + \epsilon_i$
  - IV_i = Instrumental Variable
  - $\hat{x}_i$ = Predicted X from First Stage
- **Requires:**
  - **cov(X, $\boldsymbol{\varepsilon}$) ≠ 0**
  - **cov(IV, $\boldsymbol{\varepsilon}$) = 0**
  - **cov(IV, X) ≠ 0**



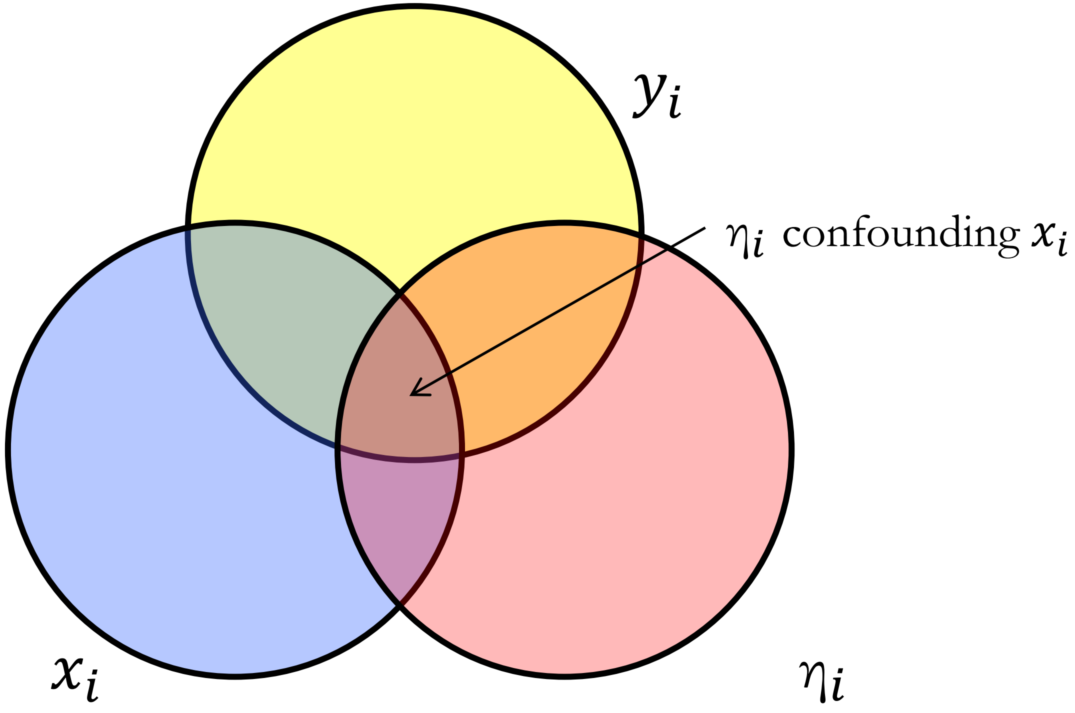
The New Solution (Instrumental Variables)

Second Stage

- $y_i = \alpha + \beta\hat{x}_i + \varepsilon_i$
  - $\varepsilon_i = \nu_i + \eta_i$

First Stage

- $x_i = \delta + \gamma IV_i + \epsilon_i$
  - IV_i = Instrumental Variable
  - $\hat{x}_i$ = Predicted X from First Stage



The New Solution (Instrumental Variables)

First Stage

$x_i = \delta + \gamma IV_i + \epsilon_i$

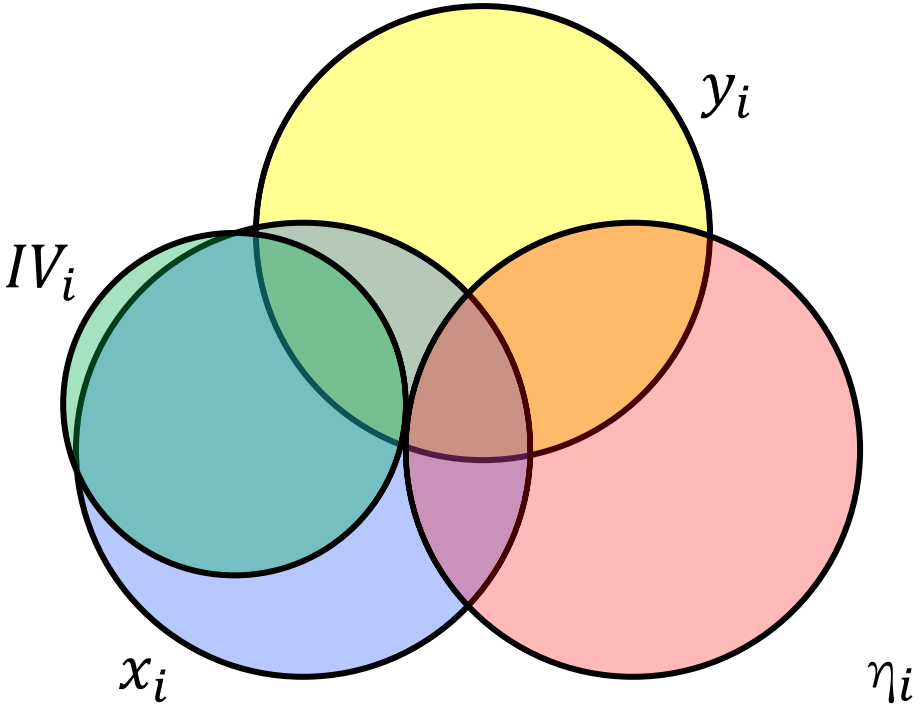
IV_i = Instrumental Variable

Requires:

cov(X, $\boldsymbol{\varepsilon}$) ≠ 0

cov(IV, $\boldsymbol{\varepsilon}$) = 0

cov(IV, X) ≠ 0



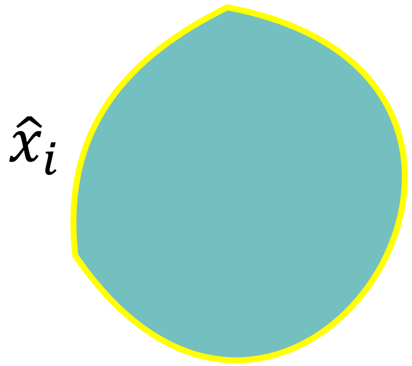
The New Solution (Instrumental Variables)

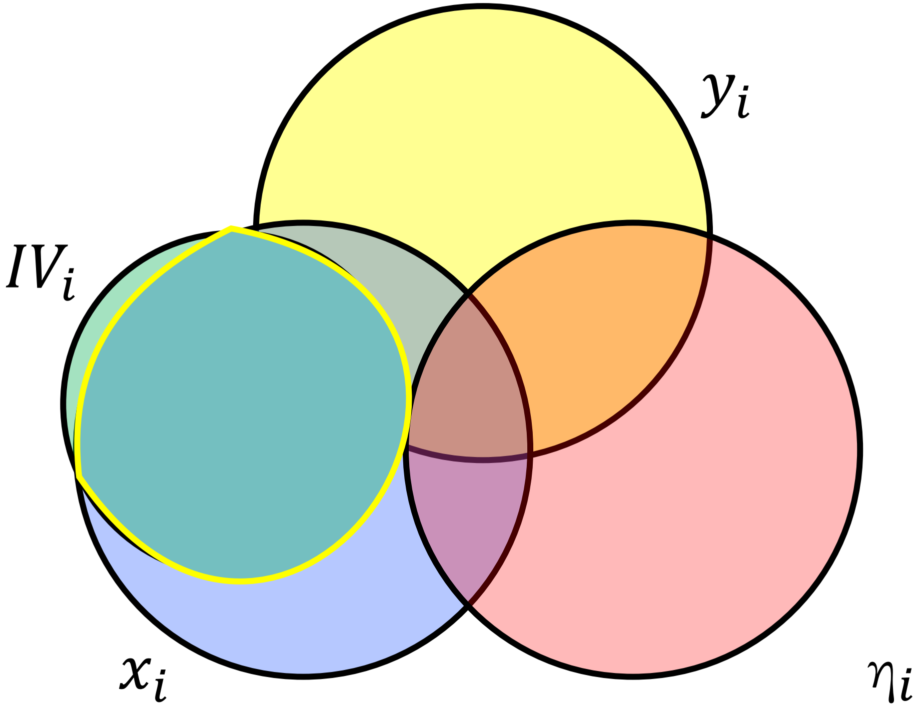
First Stage

$x_i = \delta + \gamma IV_i + \epsilon_i$

IV_i = Instrumental Variable

$\hat{x}_i$ = Predicted X from First Stage





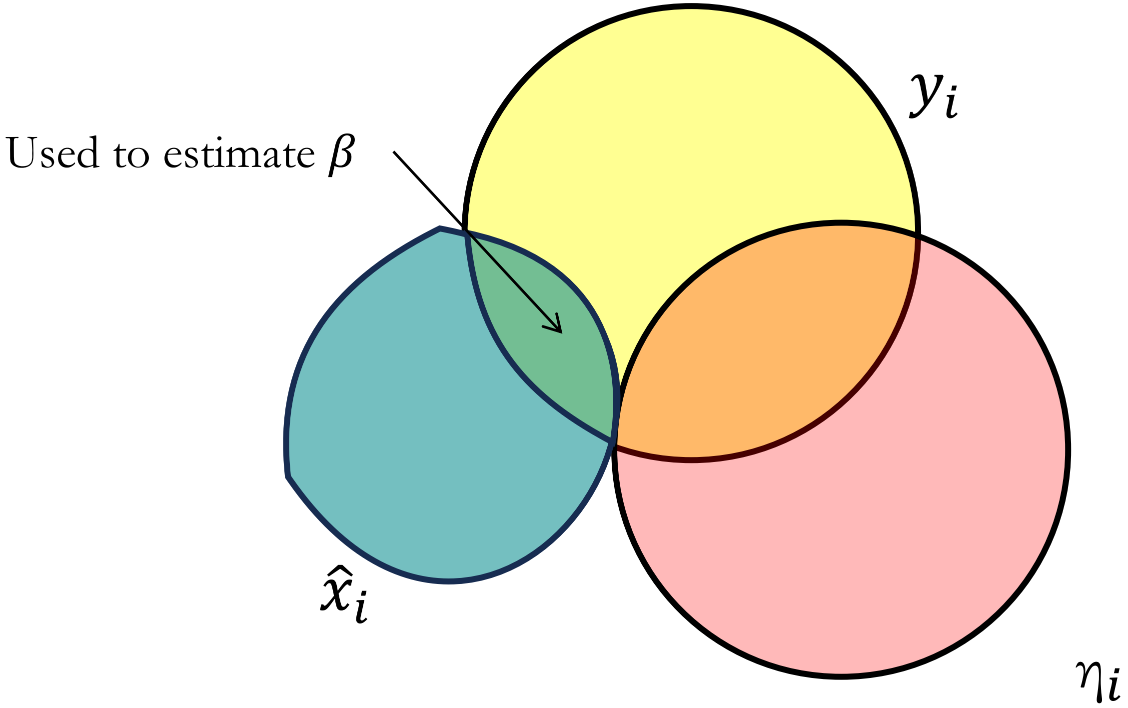
The New Solution (Instrumental Variables)

Second Stage

$$y_i = \alpha + \beta \hat{x}_i + \varepsilon_i$$

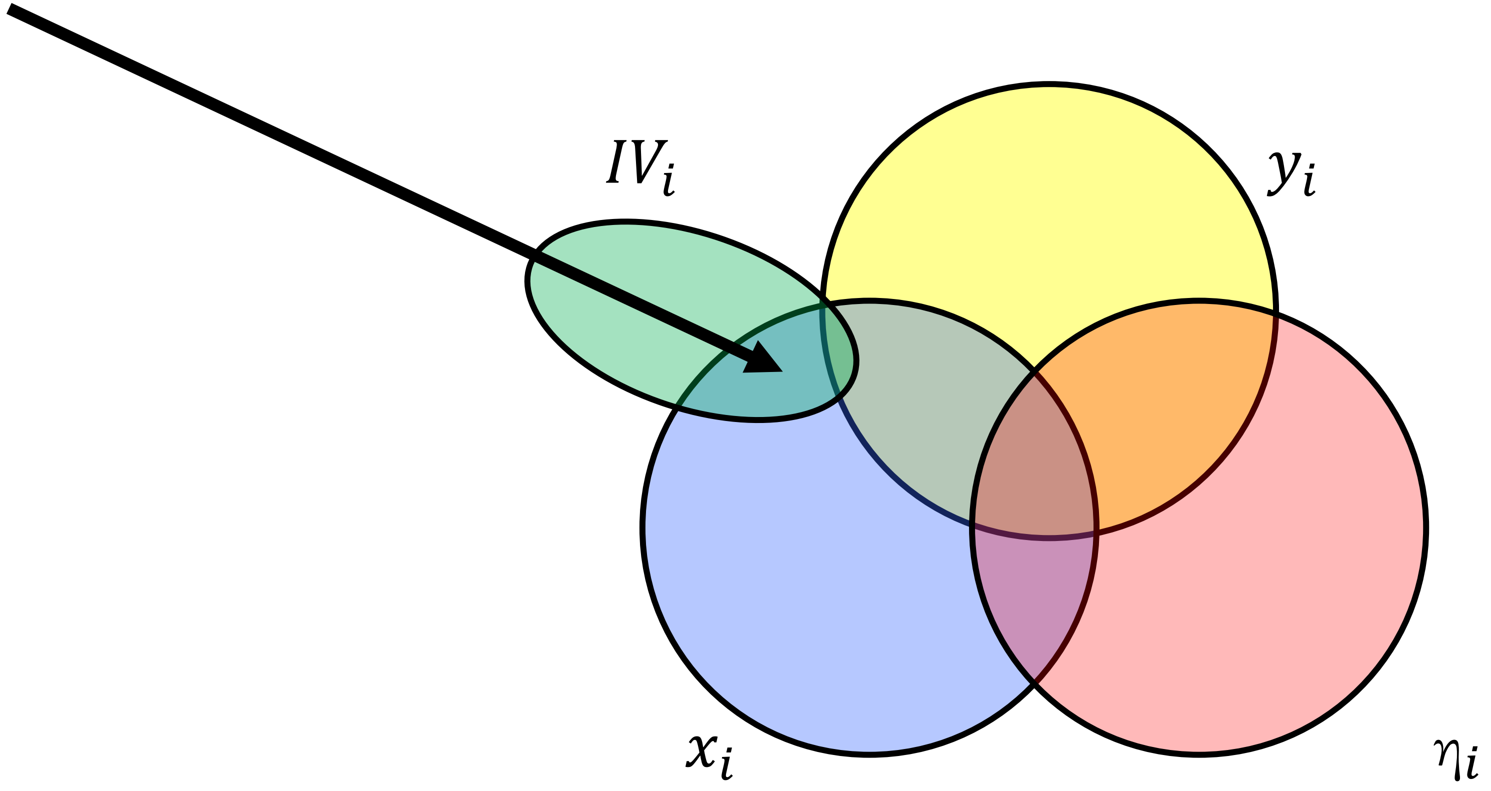
$$\varepsilon_i = \nu_i + \eta_i$$

Note that $\mathrm{cov}(\hat{x}_i, \varepsilon) = 0$, so we solved the problem.



Can There Be Any Problems?

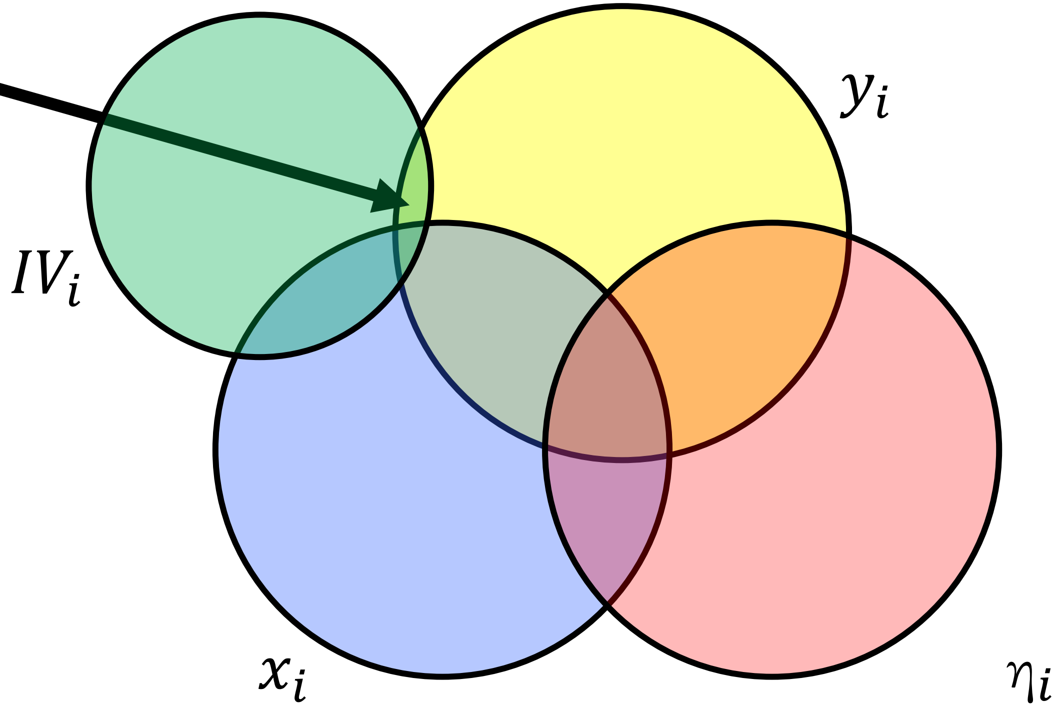
Small cov(IV, x)
Larger standard errors



Can There Be Any Problems?

cov(IV, $\mathcal{E}$) ≠ 0

IV variable belongs in second stage
Biased coefficient

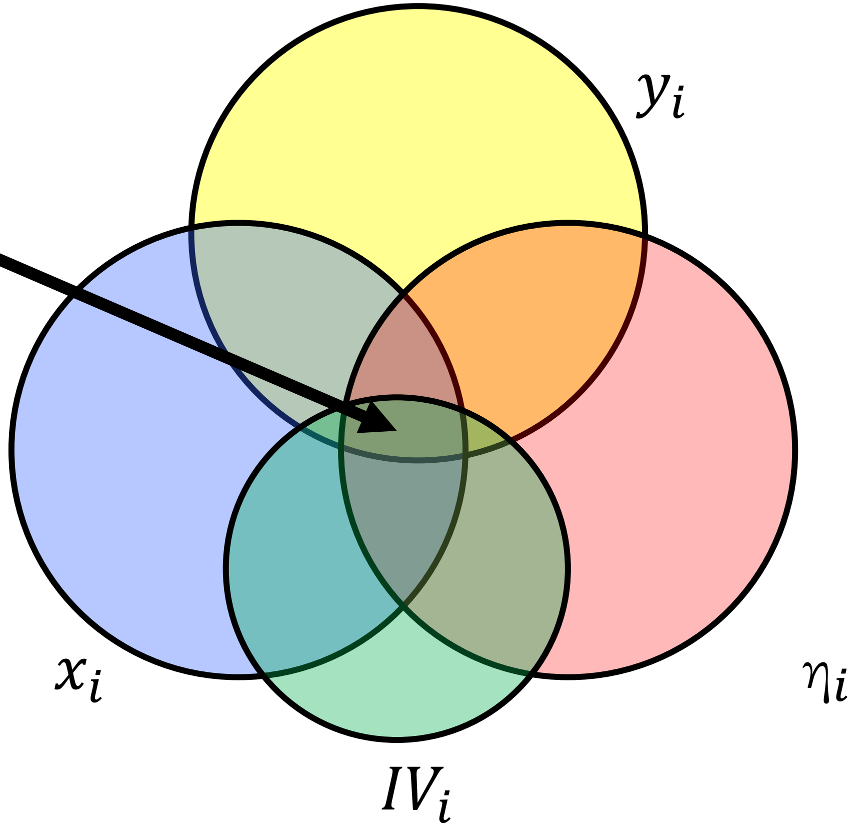


Can There Be Any Problems?

cov(IV, $\mathcal{E}$) ≠ 0

Note that η is part of $\boldsymbol{\varepsilon}$

Biased coefficient



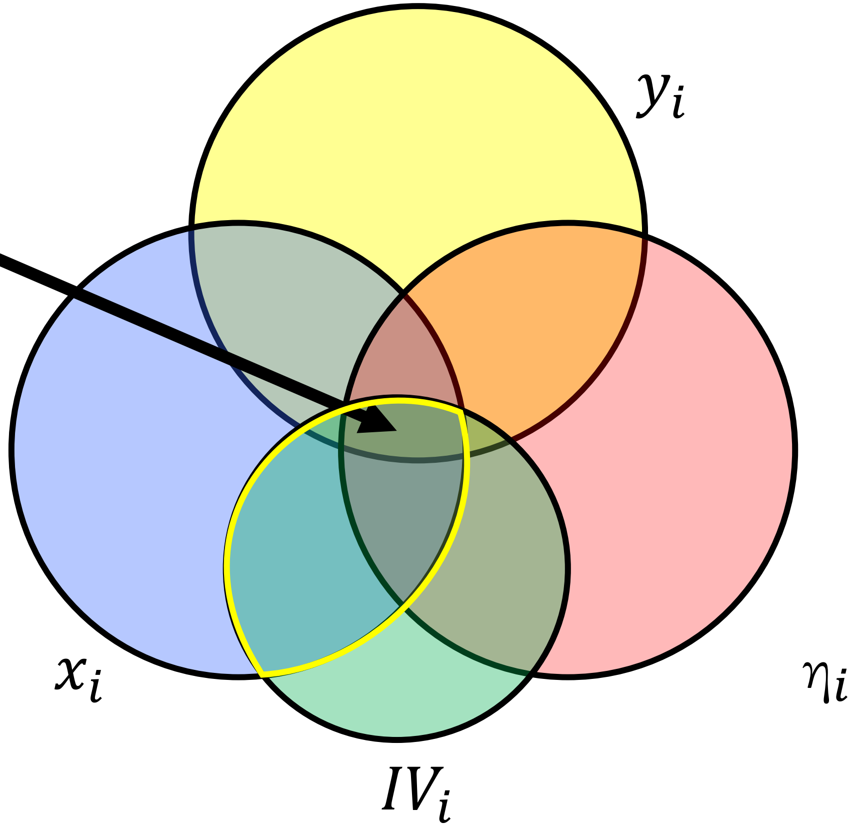
Second Stage

$$y_i = \alpha + \beta \hat{x}_i + \varepsilon_i$$

$$\varepsilon_i = \nu_i + \eta_i$$

Can There Be Any Problems?

cov(IV, $\mathcal{E}$) ≠ 0
Note that η is part of $\boldsymbol{\varepsilon}$
Biased coefficient

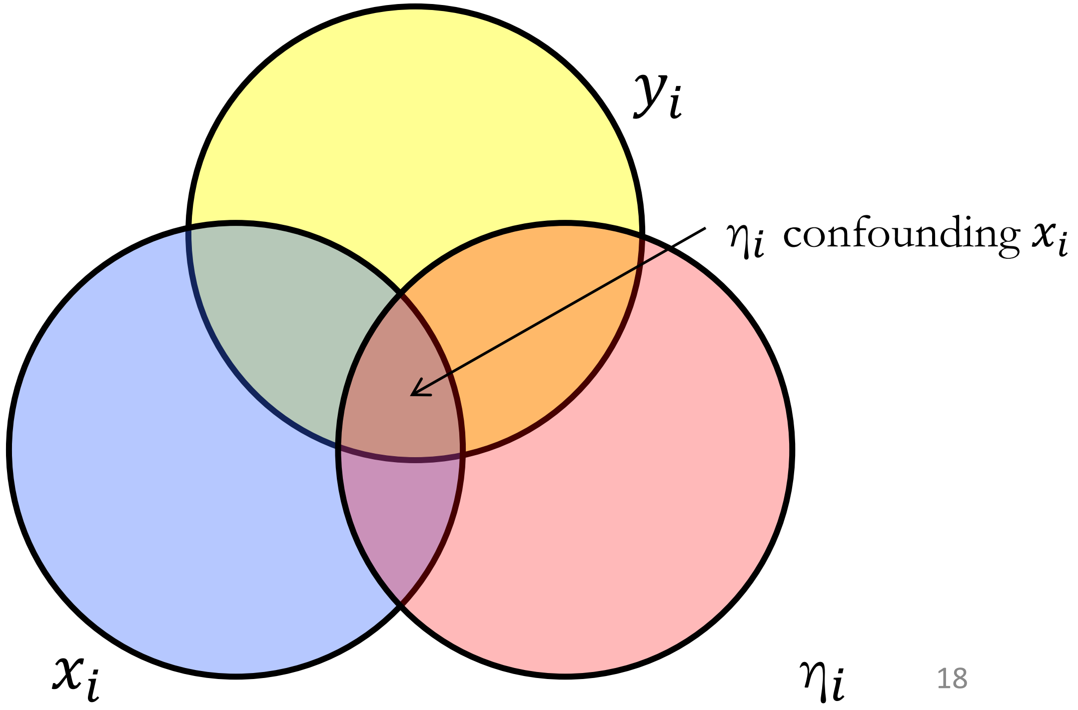


Second Stage

$$y_i = \alpha + \beta \hat{x}_i + \varepsilon_i$$
$$\varepsilon_i = \nu_i + \eta_i$$

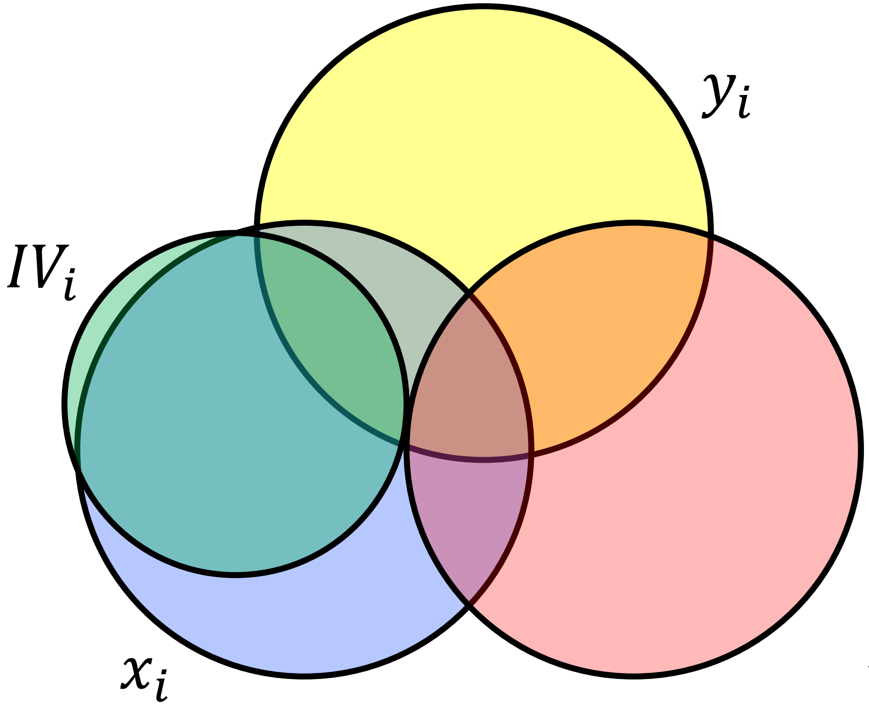
Tests for Instrumental Variables

- Test the strength of the instrument
- Test whether IV is necessary
- If multiple instruments (and willing to make assumptions) then you can also test exclusion restriction



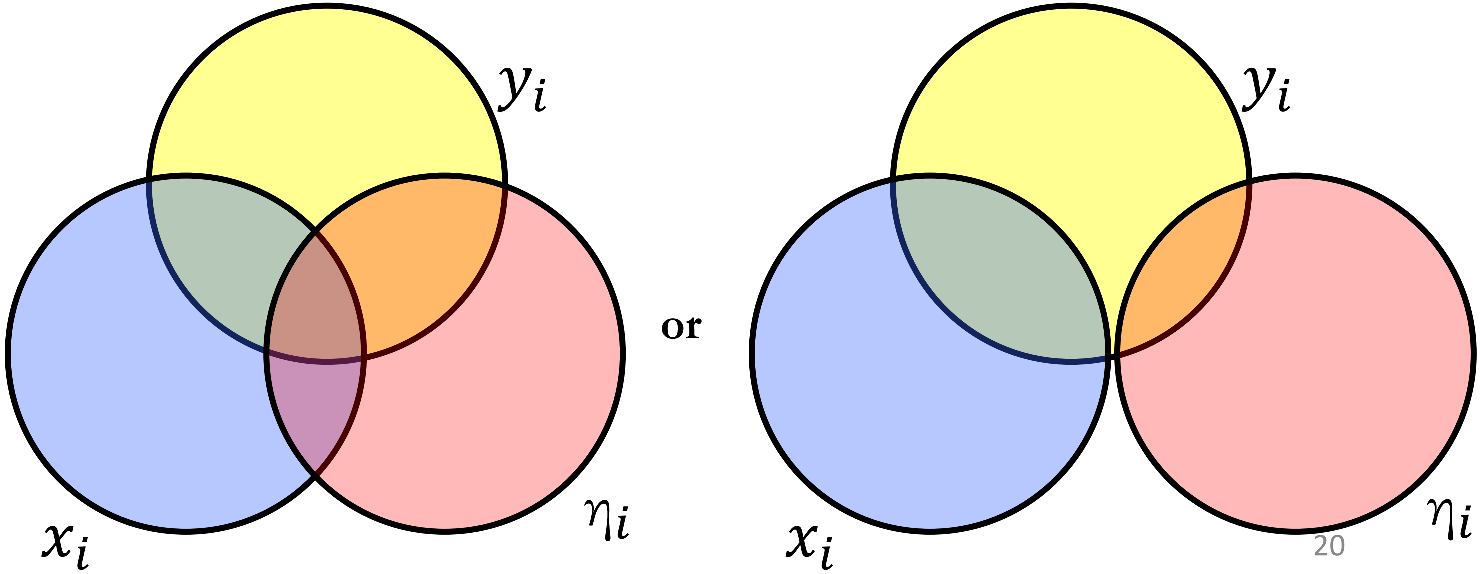
Strength of Instruments

- Bound et al (1995)
- Estimate First Stage and F-test the Instrumental Variables = 0
- Want an F-stat above 10.
- Also want large R^2 for First Stage



Endogeneity Tests

- Has endogeneity made OLS inconsistent?
  - Test if Cov(X,ε)≠0
  - If all the IV assumptions hold, then IV is consistent and OLS is efficient, but might be biased.
    - IV only uses the predicted X that is hopefully not correlated with the error
    - OLS uses all the variation in X
- Hausman Test
- Three Variants of Hausman
- **NOT A TEST OF IV SUITABILITY**, we have to assume that IV "works" to conduct these tests

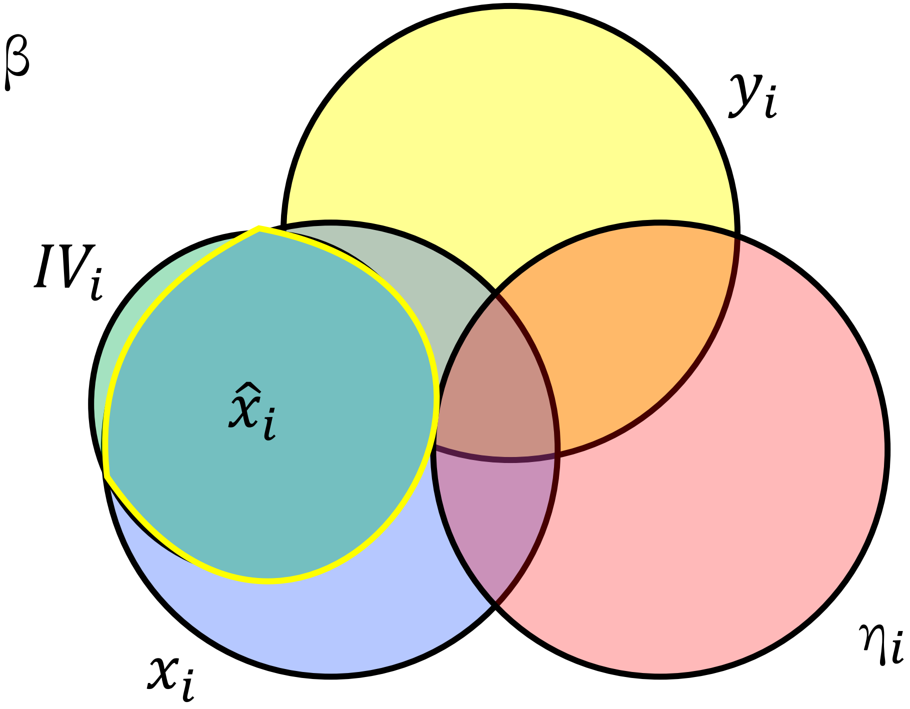


Durban-Wu-Hausman

- Consistent (IV) & Efficient (OLS)
- Null is β^{OLS} and β^{IV} are asymptotically equal and the alternative is that they differ.
- Assuming OLS is efficient, the test statistic is
- $(\beta^{IV}–\beta^{OLS})[V(\beta^{IV})-V(\beta^{OLS})]^{-1} (\beta^{IV}–\beta^{OLS})$
    - Same as the Hausman tests we have done before.
    - Sometimes $[V(\beta^{IV})-V(\beta^{OLS})]$ is the "wrong sign" and is non-invertible.

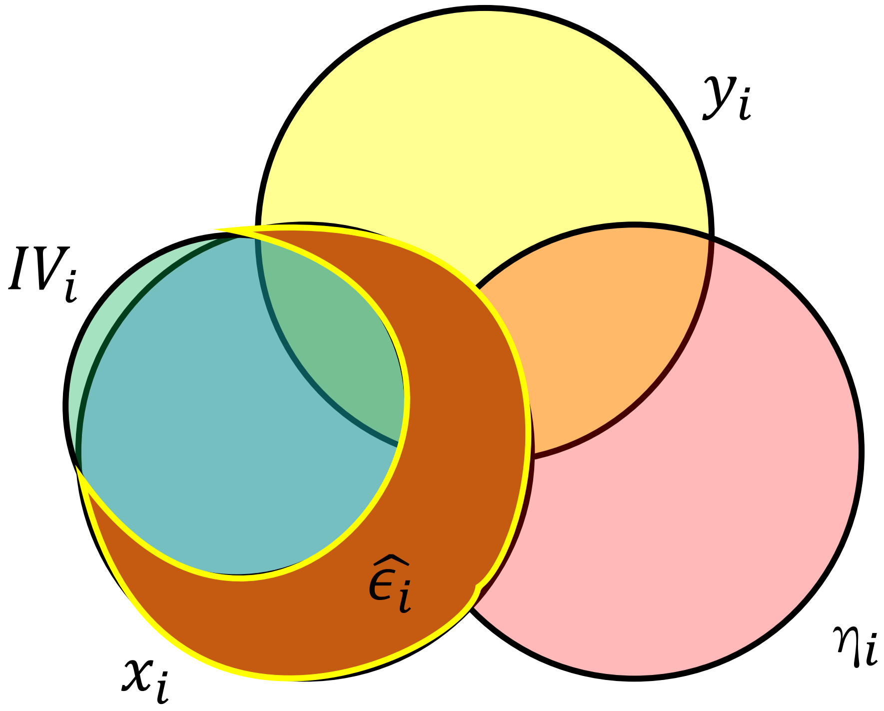
1^{st} Variant of Hausman

- First predict X from First Stage and include in OLS equation
- $Yi = \alpha + \beta^{FV}\hat{x}_i + \beta X_i + \varepsilon_i$
- Null is that $\beta^{FV} = 0$ (test with a t or F test)
  - If X is exogenous then β is the estimate of the true β
  - If X is not, then β^{FV} = True β - β^{FV}



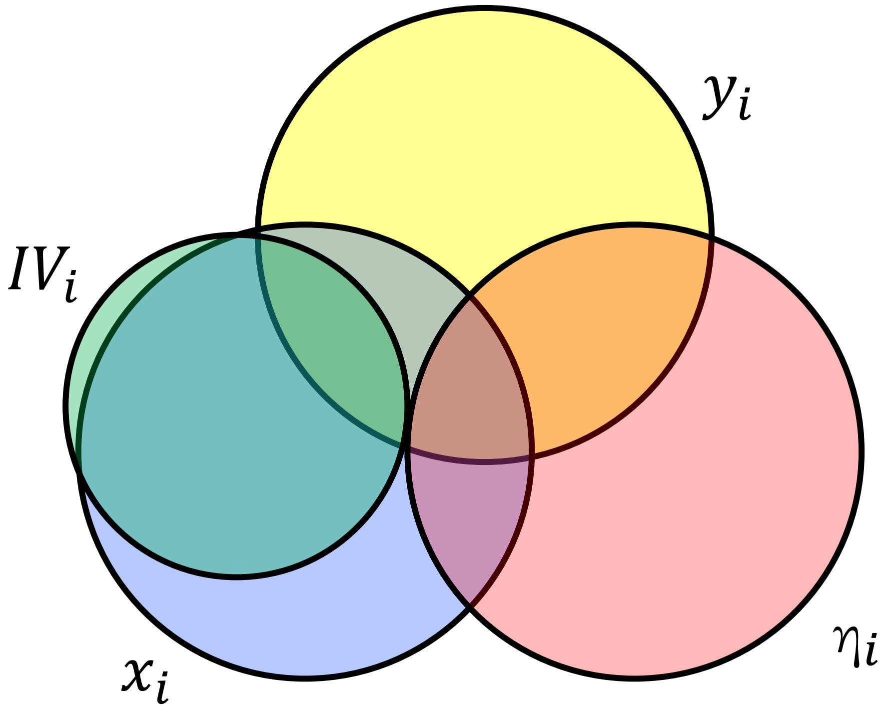
2^{nd} Variant of Hausman

- Predict residual from First Stage and include in OLS equation
- $Yi = \alpha + \beta^{SV}\widehat{\epsilon_i} + \beta X_i + \varepsilon_i$
- Null is that $\beta^{SV} = 0$ (test with a t or F test)
    - Same interpretation as before



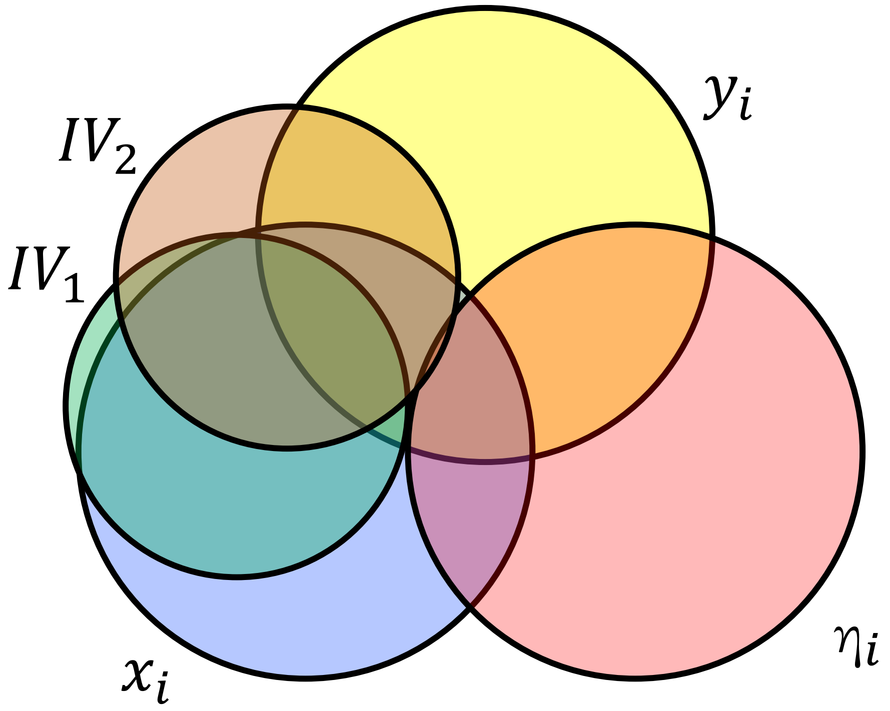
3^{rd} Variant of Hausman

- Do not need to estimate first stage, just include instrument in second stage.
- $Yi = \alpha + \beta^{TV} IV_i + \beta X_i + \varepsilon_i$
- Null is that $\beta^{TV} = 0$ (test with a t or F test)
  - Same interpretation as before



Multiple Instrument Tests

- If you assume that 1 instrument is legitimate, then you can test if other instruments are also good
- Hausman Test if coefficients using IV_1 are similar to those using IV_1-Iv_k
  - If the same then instruments are likely ok
  - If different then the extra instruments are correlated with error
  - BUT you have to assume at IV_1 is legit



Lagrange Multiplier TEST

- Estimate IV with all the instruments
- Predict residuals from second stage
- Null is that IV_1 to IV_k are valid instruments
- Regress residuals on controls, instruments (IV), and constant
- Calculate unadjusted R^2 (this is the Stata default)
- Test statistic is N^*R^2 and is distributed Chi-Squared with degrees of freedom = number of overidentified instruments (some of the extra instruments might affect Y directly)
- Rejecting means that some instruments are problematic (correlated with the error)