ECON 626: Applied Microeconomics

Lecture 4:

Instrumental Variables

Professors: Pamela Jakiela and Owen Ozier

Compliance with Treatment

How High Is Take-Up?

Even "free" programs are costly for participants, and take-up is often low

Intervention	Take-Up	Source
Job training	61% – 64%	Abadie, Angrist, Imbens (2002)
Business training	65%	McKenzie & Woodruff (2013)
Deworming medication	75%	Kremer & Miguel (2007)
Microfinance	13% – 31%	JPAL & IPA (2015)

Only people who do a program can be impacted by the program*

⇒ We might like to know how much a program impacted participants (it depends on our notion of treatment)

*Some restrictions apply

Imperfect Compliance

True model when outcomes are impacted by program participation (P_i):

$$Y_i = \alpha + \beta P_i + \varepsilon_i$$

- Program take-up is endogenous conditional on treatment
- Only those randomly assigned to treatment ($T_i = 1$) are eligible

We estimate standard regression specification:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

What do we get?

Imperfect Compliance

Modifying our standard OLS equation, we get:

$$\hat{\beta} = E\left[Y_i \middle| T_i = 1\right] - E\left[Y_i \middle| T_i = 0\right]$$

$$= \alpha + \beta E\left[P_i \middle| T_i = 1\right] + \varepsilon_i - \left(\alpha + \beta E\left[P_i \middle| T_i = 0\right] + \varepsilon_i\right)$$

$$= \beta E\left[P_i \middle| T_i = 1\right]$$

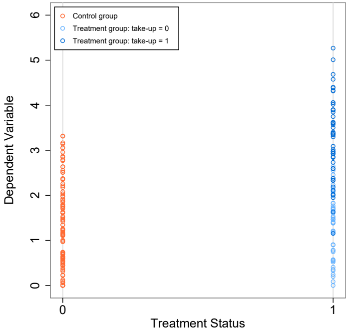
$$= \beta\lambda$$

where $\lambda < 1$ is the take-up rate in the treatment group.

$\beta\lambda$ is called the **intention to treat (ITT)** estimate.

$\Rightarrow$ Low compliance scales down the estimated treatment effect

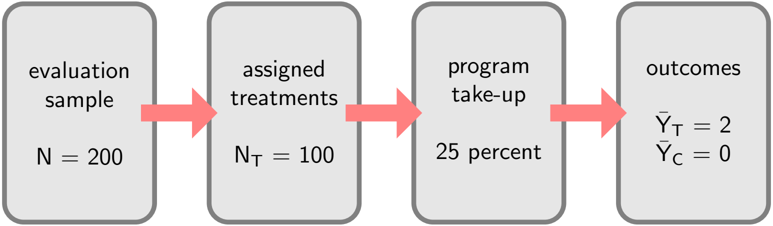
Treatment on the Treated



Your colleague suggests comparing the **compliers** to the control group

$\Rightarrow$ Is this a good idea?

Treatment on the Treated: A Thought Experiment



Questions:

- What was the average outcome among those assigned to the program?
- What does this suggest about the impact of treatment?

Treatment on the Treated: Intuition

The **treatment on the treated (TOT)** estimator:

$$\hat{\beta}_{tot} = \frac{E\left[Y_i|T_i = 1\right] - E\left[Y_i|T_i = 0\right]}{E\left[P_i|T_i = 1\right] - E\left[P_i|T_i = 0\right]}$$

Intuitively, the TOT scales up the ITT effect to reflect imperfect take-up

(Called TOT when one-sided noncompliance: compliers and never-takers, but no always-takers or defiers; see MH 4.4.3)

- Assumption: treatment only works through program take-up
  - (the "exclusion restriction")
  - Not always obvious whether this is true

Treatment on the Treated: Implementation

Estimated via two-stage least squares (2SLS):

$$Y_i = \alpha_1 + \beta_1 \hat{P}_i + \varepsilon_i \quad \text{[IV regression]}$$

$$P_i = \alpha_2 + \beta_2 T_i + \nu_i \quad \text{[first stage]}$$

Easy to implement using Stata's `ivregress 2sls` command

What Does Treatment on the Treated Measure?

$T = 0$	$T = 1$
always takers	always takers
compliers	compliers
never takers	never takers

TOT estimates local average treatment effect (LATE) on **compliers**.
Under homogeneous treatment effects (same for everyone), this is also the average treatment effect (ATE) for any population.
But: Under heterogeneous treatment effects (not the same for everyone), the LATE is particular to the compliers. It also requires...

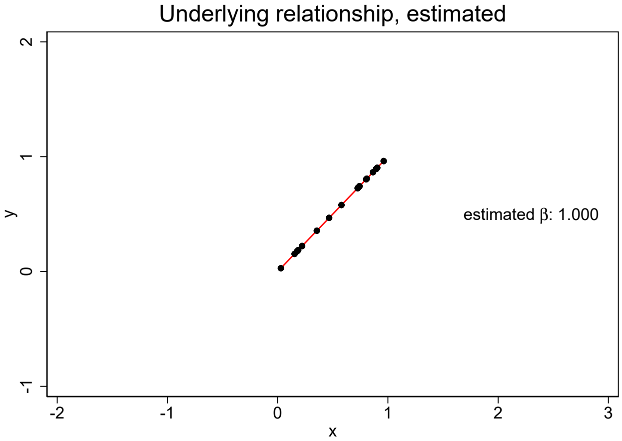
- Monotonicity assumption: there are no **defiers**
- When violated, TOT tells us about weighted difference between treatment effects on compliers and defiers... but it gets complicated

History and mechanics of instrumental variables

Wald

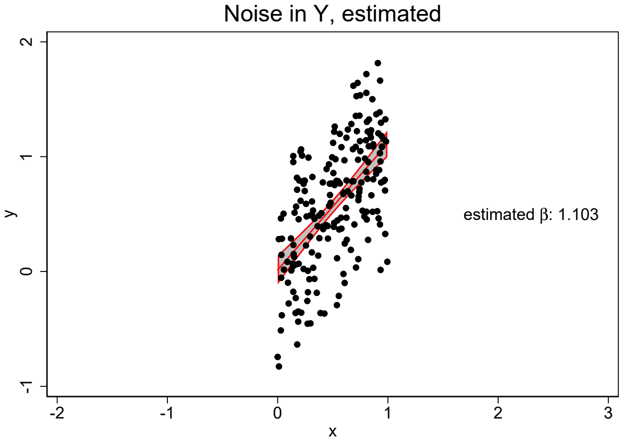
When two variables are measured with error,
how do we estimate their true relationship?

Wald

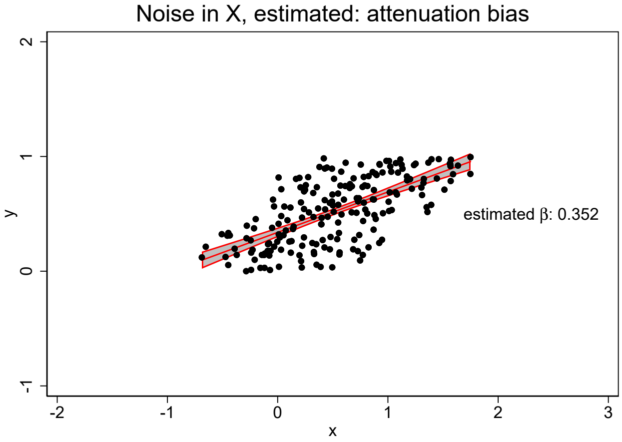


Underlying relationship, estimated
estimated β: 1.000
y
x
2
1
0
-1
-2
-1
0
1
2
3

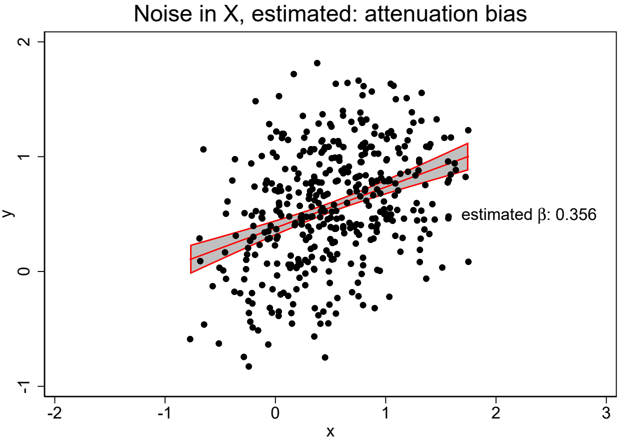
Wald



Wald - attenuation bias



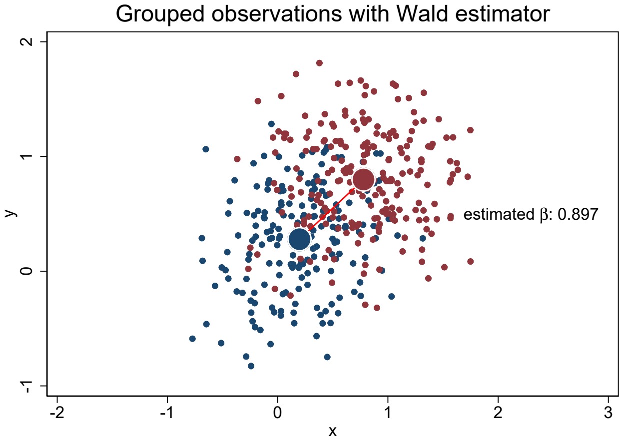
Wald - attenuation bias



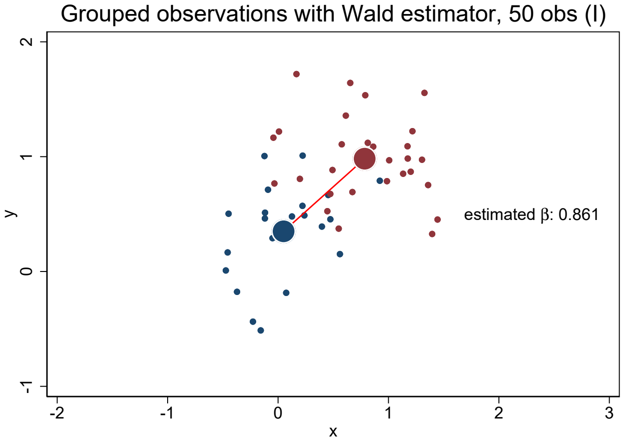
Wald - attenuation bias

Suppose we have one more piece of information: whether, for each observation, the underlying x value (without the measurement error) is above or below 0.5. This information will prove to be an "instrument."

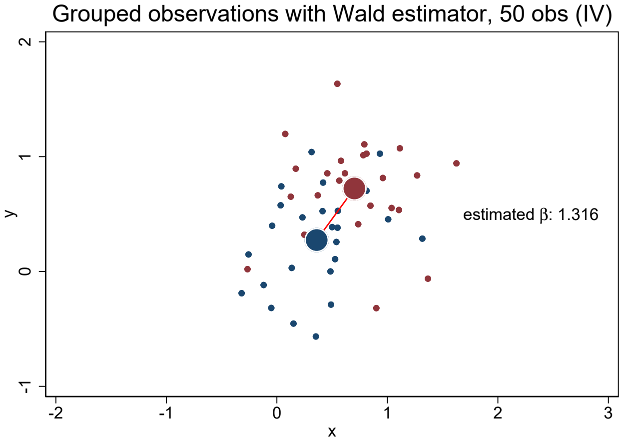
Wald - overcoming attenuation bias



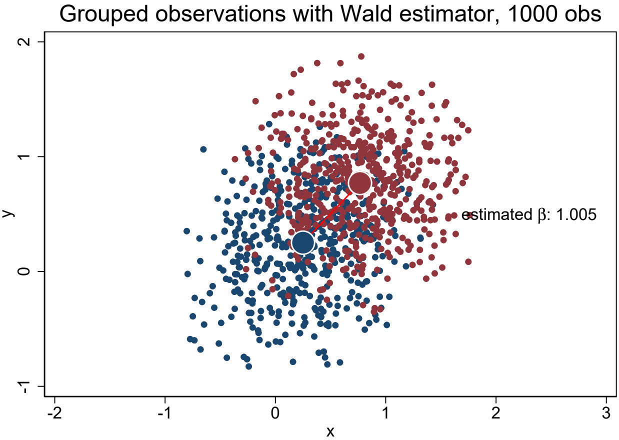
Wald - overcoming attenuation bias



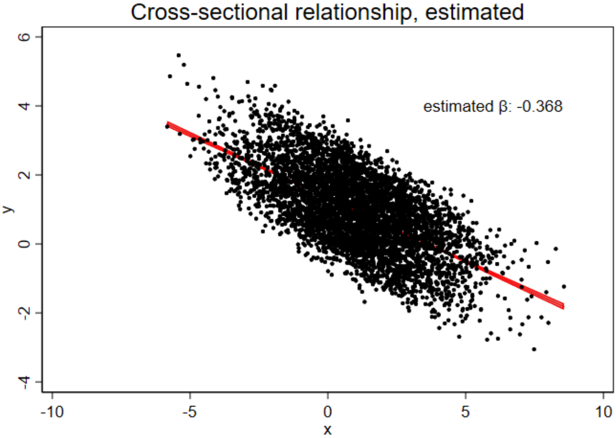
Wald - overcoming attenuation bias



Wald - overcoming attenuation bias



Wald - extending to endogeneity



Wald - extending to endogeneity

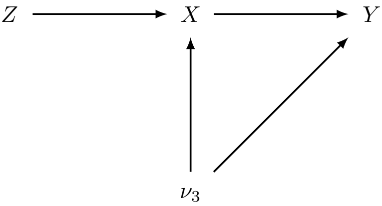
Data generating process:

$Z \sim \mathcal{U}(0,2)$
$\nu_1, \nu_2, \nu_3 \sim \mathcal{N}(0,1)$ *i.i.d.*

$\xi = 2\nu_3 + 0.2\nu_1$
$\eta = -3\nu_3 + 0.2\nu_2$

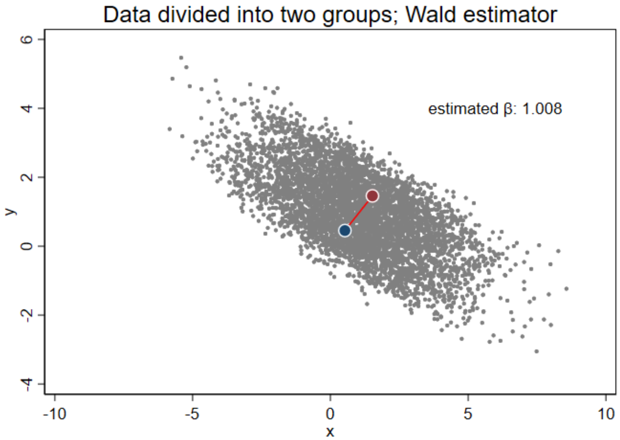
ξ and η **not** independent;
strongly negatively correlated.

$X = Z + \xi$
$Y = X + \eta$

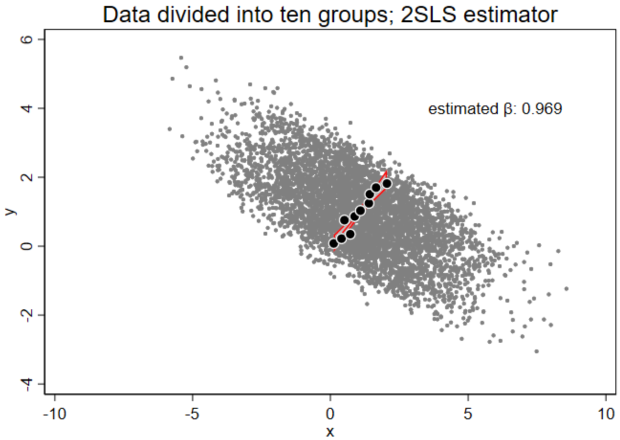


Begin Wald approach by considering a split based on whether $Z > 1$.

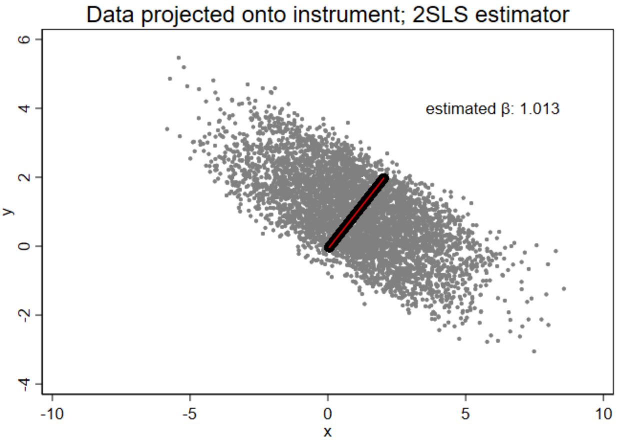
Wald - extending to endogeneity



Wald - extending to endogeneity



Wald - extending to endogeneity



Instrumental variables scenarios

Problem: measure the causal *casual* effect of X^{end} on Y.
Inconsistency of least-squares methods when: measurement error in regressors, simultaneity, or when causal equation (Y) error term is correlated with X^{end} (omitted variables). Discussion in Cameron and Trivedi, section 6.4, and Angrist and Pishke chapter 4.

Example: X^{end} is schooling; Y is wage;
"ability" drives both Y and X^{end}, so may bias cross-sectional regression of Y on X^{end}.

Example: X^{end} is number of children; Y is labor force participation;
"inclination to remain outside the formal labor force" drives Y down and X^{end} up, so may bias cross-sectional regression of Y on X^{end}.

Example: X^{end} is medical treatment; Y is health;
prior illness drives Y down and X^{end} up, so may bias cross-sectional regression of Y on X^{end}.

Instrumental variables basics

Terminology of Instrumental Variables ("IV") approach:

First stage: Z affects X^{end}
Exclusion restriction: Z ONLY affects Y via its effect on X^{end}

Z: "instrument(s)" or "excluded instrument(s)"
Y: "dependent variable" or "endogenous dependent variable"
X^{end}: "endogenous variable" or "endogenous regressor"

What about other covariates?

X^{ex}: "covariates" or "exogenous regressors"

(First stage and exclusion restriction now conditional on X^{ex}.)

Instrumental variables basics

$$X_i^{end} = \pi_{11} Z_i + \mathbf{X}_i^{ex\prime} \pi_{10} + \xi_{1i} \text{ (“First stage”)}$$

$$Y_i = \rho X_i^{end} + \mathbf{X}_i^{ex\prime} \alpha + \eta_i \text{ (causal model)}$$

$$E[\eta_i | X_i^{ex}] = 0;\ E[\xi_{1i} | X_i^{ex}] = 0;\ E[\eta_i \xi_{1i} | X_i^{ex}] \neq 0;\ E[\eta_i | Z_i, X_i^{ex}] = 0;$$

$$Y_i = \rho(\pi_{11} Z_i + \mathbf{X}_i^{ex\prime} \pi_{10} + \xi_{1i}) + \mathbf{X}_i^{ex\prime} \alpha + \eta_i$$

$$Y_i = \rho \pi_{11} Z_i + \mathbf{X}_i^{ex\prime} (\rho \pi_{10} + \alpha) + (\rho \xi_{1i} + \eta_i)$$

$$Y_i = \pi_{21} Z_i + \mathbf{X}_i^{ex\prime} \pi_{20} + \xi_{2i} \text{ (“Reduced form”)}$$

$$\hat{X}_i^{end} = \mathbf{Z}_i' \hat{\pi}_{11} + \mathbf{X}_i^{ex\prime} \hat{\pi}_{10} \text{ (Estimated first stage)}$$

$$Y_i = \rho(\hat{X}_i^{end} + (X_i^{end} - \hat{X}_i^{end})) + \mathbf{X}_i^{ex\prime} \alpha + \eta_i$$

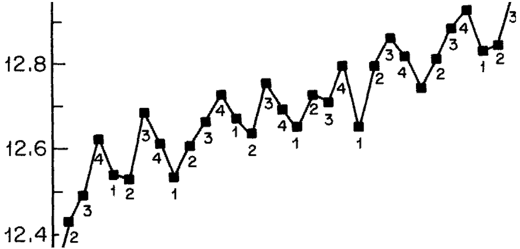
$$Y_i = \rho \hat{X}_i^{end} + \mathbf{X}_i^{ex\prime} \alpha + (\eta_i + \rho(X_i^{end} - \hat{X}_i^{end})) \text{ (“Second stage”)}$$

Hence: “Two-stage least squares,” “2SLS” or “TSLS”

Instrumental variables scenarios

Example: quarter of birth / compulsory schooling instrument
X^{end} is schooling (endogenous regressor); Y is wage (dependent var.); how do we find variation in education that is not driven by the common (unobserved) causes of education and wage ("ability")?
Z is quarter of birth (instrument). Exclusion restriction? First stage?

Born in Q4: start school just before you turn 6. At age 16, you have completed 10+ years of school.
Born in Q1: start school September after you turn 6. At age 16, you have completed 9 years and a few months of school.



Finding: wage returns to education via 2SLS slightly larger than OLS.

Instrumental variables scenarios

Example: same-sex and twins instruments
X^{end} is number of children (endogenous regressor);
Y is labor force participation (dependent variable);
how do we find variation in family size that is not driven by the common (unobserved) causes of family size and labor force participation ("inclination to remain outside the formal labor force")?
$\mathbf{Z}$ = two indicators: twins at second birth; first two children same sex (instruments). Exclusion restriction? First stage?

Finding: family size decreases women's labor force participation, but not by as much as OLS would suggest. (Angrist and Evans 1998, Mostly Harmless Table 4.1.4)

Instrumental variables scenarios

Likely source of OLS bias? Exclusion restriction? First stage?

- Vietnam draft lottery
- Job Training Partnership Act (JTPA) randomized trial
- Ocean weather
- **Rainfall!** (Paxson 1992; Miguel et al 2004: Maccini and Yang 2009; Madestam et al 2013; etc.)
- Electrification... slope of land (Dinkelman 2011)

Instrumental variables scenarios

Likely source of OLS bias? Exclusion restriction? First stage?
Other kinds of scenarios

- Y = Child IQ; X^{end} = growing cotton; Z = born in US south
- Y = "Happiness, 1-5;" X^{end} = "Fair workplace, 1-5;" Z = variation in when a pay raise is announced to individuals
- Y = "Satisfied w/ govt services;" X^{end} = city pruned tree branches over sidewalk recently; Z = city repaved street recently

Instrumental variables: LATE (MHE Chapter 4.4)

Consider a randomized trial with imperfect compliance (as in JTPA).

Terminology:

- *Always-takers* $D_{0i} = D_{1i} = 1$, so $D_i = 1$ regardless of Z_i
- *Never-takers* $D_{0i} = D_{1i} = 0$, so $D_i = 0$ regardless of Z_i
- *Compliers* $D_{0i} = 0$; $D_{1i} = 1$, so $D_i = Z_i$

Under heterogeneous treatment effects, having not only *compliers* but also *defiers* would cause a problem.

- *Defiers*: $D_{0i} = 1$; $D_{1i} = 0$, so $D_i = (1 - Z_i)$.

We need **monotonicity** for an interpretable *Local Average Treatment Effect* when there are heterogeneous treatment effects: either $D_{1i} \geq D_{0i} \forall i$, or $D_{1i} \leq D_{0i} \forall i$.

Instrumental variables: Overidentification

Terminology:

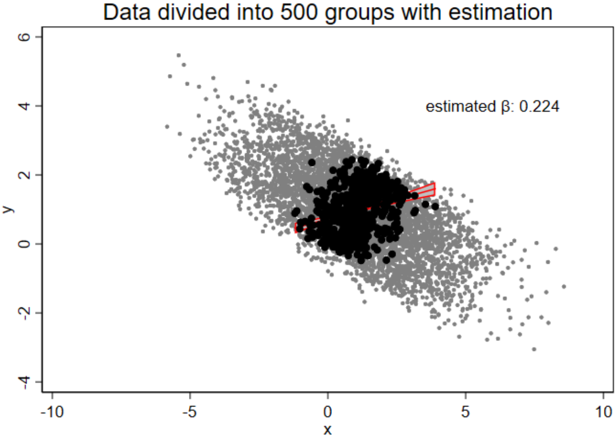
- Exactly as many linearly independent instruments as endogenous regressors?
  Just identified.
- More linearly independent instruments than endogenous regressors?
  Overidentified.

Overidentification, exogeneity, and heterogeneous effects:

- Suppose we have two instruments, one endogenous regressor, and there are statistically significant differences between the 2SLS estimates given by one instrument as compared to the other. What does it mean? *(at least two possibilities)*
- Suppose we have two instruments, one endogenous regressor, and there are **not** statistically significant differences between the 2SLS estimates given by one instrument as compared to the other. What does it mean?*(at least two possibilities)*

Weak Instruments

Instrumental variables: Weak instruments



Instrumental variables: Weak instruments

2SLS bias towards OLS (MHE 4.6.21):

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \frac{1}{F+1}$$

F =F statistic for the joint significance of **the excluded instruments**.

Just-identified 2SLS median-unbiased even with weak first stage, but many weak instruments can lead to bias.

Note: other IV estimators exist (and are implemented in Stata), including LIML. LIML may be less biased than 2SLS w/ weak instruments, but imposes distributional assumptions; less to gain under heteroskedasticity. See discussion: end of Chapter 4 of MHE; Cameron and Trivedi section 6.4. Also note: 2SLS confidence intervals may be incorrect for weak instruments, but heteroskedasticity-robust Anderson-Rubin confidence intervals can be constructed via user-written Stata routines.

Instrumental variables: Weak instruments

2SLS bias towards OLS (MHE 4.6.21):

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \frac{1}{F+1}$$

F =F statistic for the joint significance of **the excluded instruments in the first stage**.

Note that this is the "population" F statistic. We will return to this point in the context of Alwyn Young's paper. What is an F statistic?

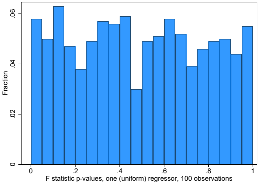
$$\frac{Explained\ variation/regressors}{Residual\ variation/residual\ d.o.f.}$$

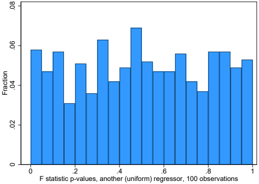
Young: Consistency without Inference

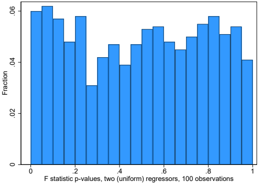
Experiment 1:

$x_1, x_2 \sim$ *iid* $\mathcal{U}(0,1)$
$\varepsilon_1, \varepsilon_2 \sim$ *iid* $\mathcal{N}(0,1)$
$y = \varepsilon_1 + \varepsilon_2$
$N = \mathbf{100}$ observations
`regress y x1 x2`
(1,000 times)

Young, Consistency without Inference: misleading F





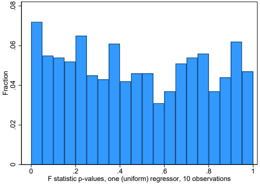


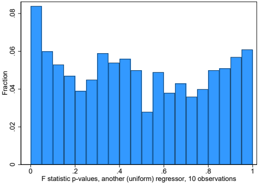
Young, Consistency without Inference: misleading F

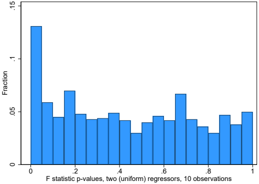
Experiment 2:

$x_1, x_2 \sim$ *iid* $\mathcal{U}(0,1)$
$\varepsilon_1, \varepsilon_2 \sim$ *iid* $\mathcal{N}(0,1)$
$y = \varepsilon_1 + \varepsilon_2$
$N = \mathbf{10}$ observations
`regress y x1 x2`
(1,000 times)

Young, Consistency without Inference: misleading F





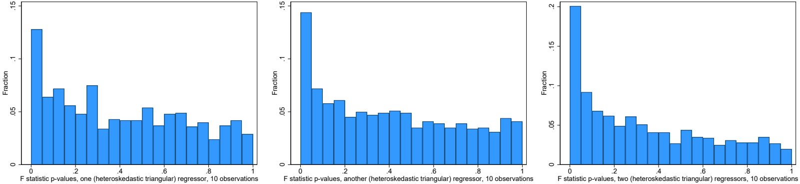


Young, Consistency without Inference: misleading F

Experiment 3:

$x_1, x_2 \sim$ *iid triangular* $f_x(x) = 2 - x$ *if* $0 \leq x \leq 1;\ 0$ *o.w.*
$\varepsilon_1, \varepsilon_2 \sim$ *iid* $\mathcal{N}(0,1)$
$\tilde{\varepsilon}_1 = \varepsilon_1 \cdot x_1$
$\tilde{\varepsilon}_2 = \varepsilon_2 \cdot x_2$
$y = \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2$
$N = \mathbf{10}$ observations
`regress y x1 x2`
(1,000 times)

Young, Consistency without Inference: misleading F

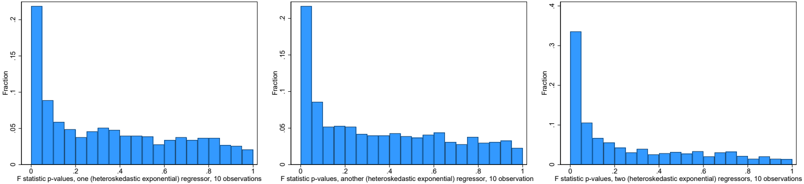


Young, Consistency without Inference: misleading F

Experiment 4:

$x_1, x_2 \sim$ *iid exponential* $f_x(x) = e^{-x}$ *if* $x > 0$; 0 *o.w.*
$\varepsilon_1, \varepsilon_2 \sim$ *iid* $\mathcal{N}(0, 1)$
$\tilde{\varepsilon}_1 = \varepsilon_1 \cdot x_1$
$\tilde{\varepsilon}_2 = \varepsilon_2 \cdot x_2$
$y = \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2$
$N = \mathbf{10}$ observations
`regress y x1 x2`
(1,000 times)

Young, Consistency without Inference: misleading F



Young, Consistency without Inference: misleading F

Some conclusions:

- Small number of observations? Not in *asymptopia*.
- Small number of clusters? Not in *asymptopia*.
- Small number of disproportionately large clusters (even if also a large number of small clusters)? Not in *asymptopia*.
- Important outliers? Not in *asymptopia*.
- Robustness: dropping any one observation/cluster should not change things much.
- (much more in the paper)