Applied Microeconometrics I

Lecture 7: Instrumental variables (continued)

Tuomas Pekkarinen

Aalto University

October 5 2021

Lecture Slides

Instrumental variables in a regression: Schooling example

- Consider instrumental variables in a regression framework
- Our aim is to estimate the causal effect of schooling on wages
- Lets assume we do not observe everything that affects both selection into schooling and earnings (ablity)
- The relationship between earnings and schooling be

$$Y_i = \alpha_0 + \rho S_i + \eta_i$$

$$\eta_i = A_i' \gamma + v_i$$

- The variables A_i are assumed to be the only reason why η_i and S_i are correlated, i.e.

$$E[S_i v_i] = 0$$

Schooling example

- If we could observe the variables A_i we could simply include them to the regressions and estimate

$$Y_i = \alpha + \rho S_i + A_i'\gamma + v_i$$

- How to estimate ρ without observing A_i?
- Instrumental variable (IV) allows us to estimate ρ when A_i is unobserved

- With a valid instrumental variable we can consistently estimate ρ in

$$Y_i = \alpha + \rho S_i + A_i'\gamma + v_i$$

- We can write ρ in terms of the population moments

$$\text{Cov}(Z_i, Y_i) = \rho\text{Cov}(Z_i, S_i) + \text{Cov}(Z_i, \eta_i)$$

- Given the exclusion restriction, $\text{Cov}(Z_i, \eta_i) = 0$, it follows that

$$\rho = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,S_i)} = \frac{\frac{Cov(Z_i,Y_i)}{Var(Z_i)}}{\frac{Cov(Z_i,S_i)}{Var(Z_i)}}$$

- The coefficient of interest, ρ, is the ratio between regression of Y_i on Z_i (the reduced form) and regression of S_i on Z_i (the first stage).

What is a valid instrumental variable?

- Two main assumptions:
  1. **Relevance**: The instrument is correlated with the causal variable of interest, S_i,
     $\text{Cov}(Z_i, S_i) \neq 0$
  2. **Validity:** The instrument is uncorrelated with any other determinants of Y_i
     $\text{Cov}(Z_i, \eta_i) = 0$
     This requirement can be decomposed in two:
     - **2.1 Independence (or random assignment)**: Z is not correlated with any unobservable factors that affect Y
     - **2.2 Exclusion restriction**: Z_i only affects Y_i through its effect on S_i

Note: Some authors may refer to the validity assumption indifferently as the exogeneity condition or the exclusion restriction. However, it is useful to consider exogeneity and exclusion restriction as two distinct requirements.

Two-stage Least Squares (2SLS)

- In a model with a **single endogenous variable** and a **single instrument**, IV estimates are equivalent to a two stage procedure.
- Causal model with covariates

$$Y_i = X_i'\alpha + \rho S_i + \eta_i \quad (1)$$

- First-stage equation

$$S_i = X_i'\pi_{10} + \pi_{11} z_i + \epsilon_{1i} \quad (2)$$

Two-stage Least Squares (2SLS)

- Write the first stage as the sum of fitted values plus first stage residuals

$$S_i = X_i'\pi_{10} + \pi_{11} z_i + \epsilon_{1i} = \hat{s}_i + \epsilon_{1i}$$

- 2SLS estimates can be constructed by substituting the first-stage fitted values for S_i in (1)

$$Y_i = X_i'\alpha + \rho \hat{s}_i + [\eta_i + \rho \epsilon_{1i}]$$

- Estimate by OLS

Standard errors

Technical note

- With the manual two stage procedure, you do not get 'automatically' the correct standard errors
  - The residual that is used to calculate standard errors in second stage includes an extra error $Y_i - [X_i'\alpha - \rho\hat{s}_i] = [\rho\epsilon_{1i} + \eta_i]$
  - $\hat{x}$ is a generated regressor and inflates the variance
  - Stata `ivreg or ivreg2` fixes it: uses the original endogenous regressor to construct residuals: $Y_i - [X_i'\alpha - \rho s_i] = \eta_i$

The Wald estimator

- Consider the case when we have:
  - Model with one endogenous regressor and no covariates
  - Single binary instrument $[z_i \in \{0, 1\}]$
- If z_i equals 1 with probability p, it is easy to show that IV estimator is:

$$\rho = \frac{Cov(y_i, z_i)}{Cov(s_i, z_i)} = \frac{E[y_i|z_i=1] - E[y_i|z_i=0]}{E[s_i|z_i=1] - E[s_i|z_i=0]}$$

since $Cov(y_i, z_i) = E[y_i z_i] - E[y_i]E[z_i]$
$= E[y_i|z_i = 1]p - [E[y_i|z_i = 1]p + E[y_i|z_i = 0](1 - p)]p$
$= [E[y_i|z_i = 1] - E[y_i|z_i = 0]]p(1 - p)$
and $Cov(s_i, z_i) = [E[s_i|z_i = 1] - E[s_i|z_i = 0]]p(1 - p)$

Instrumental variables and control variables

- When he have a set of controls X in the second stage, we also need to control for them also in the first stage

$$Y_i = \beta S_i + \gamma X_i + u_i$$

$$S_i = \pi_1 Z_i + \pi_2 X_i + v_i$$

Instrumental variables and control variables

- To see this, first note that we can substitute the first stage to the second stage:

$$\begin{aligned} Y &= \beta[\pi_1 Z_i + \pi_2 X_i] + \gamma X_i + \beta v_i + u_i \\ &= \beta \hat{S}_i + \gamma X_i + \beta v_i + u_i \\ &= \beta \hat{S}_i + \gamma X_i + \beta(S_i - \hat{S}_i) + u_i \end{aligned}$$

  where $\hat{S}_i = \pi_1 Z_i + \pi_2 X_i$

- Here both $\hat{S}_i$ and X_i are uncorrelated with $S_i - \hat{S}_i$ by construction and all the coefficients will be consistently estimated

Instrumental variables and control variables

- Suppose instead that our first stage is:

$$S_i = \pi_3 Z_i + \eta_i$$

- Substituting this to the second stage gives us:

$$\begin{aligned} Y &= \beta[\pi_3 Z_i + \eta_i] + \gamma X_i + u_i \\ &= \beta \hat{S}_i + \gamma X_i + \beta \eta_i + u_i \\ &= \beta \hat{S}_i + \gamma X_i + \beta(S_i - \hat{S}_i) + u_i \end{aligned}$$

  where $\hat{S}_i = \pi_3 Z_i$

- Since X is no longer in the first stage, there's no gurantee that it won't be correlated with $S_i - \hat{S}_i$
- As a result all the coefficients in the second stage will be inconsistent

Instrumental variables and control variables

- Another way to think about this is to consider the assumption underlying the validity of the instrument:

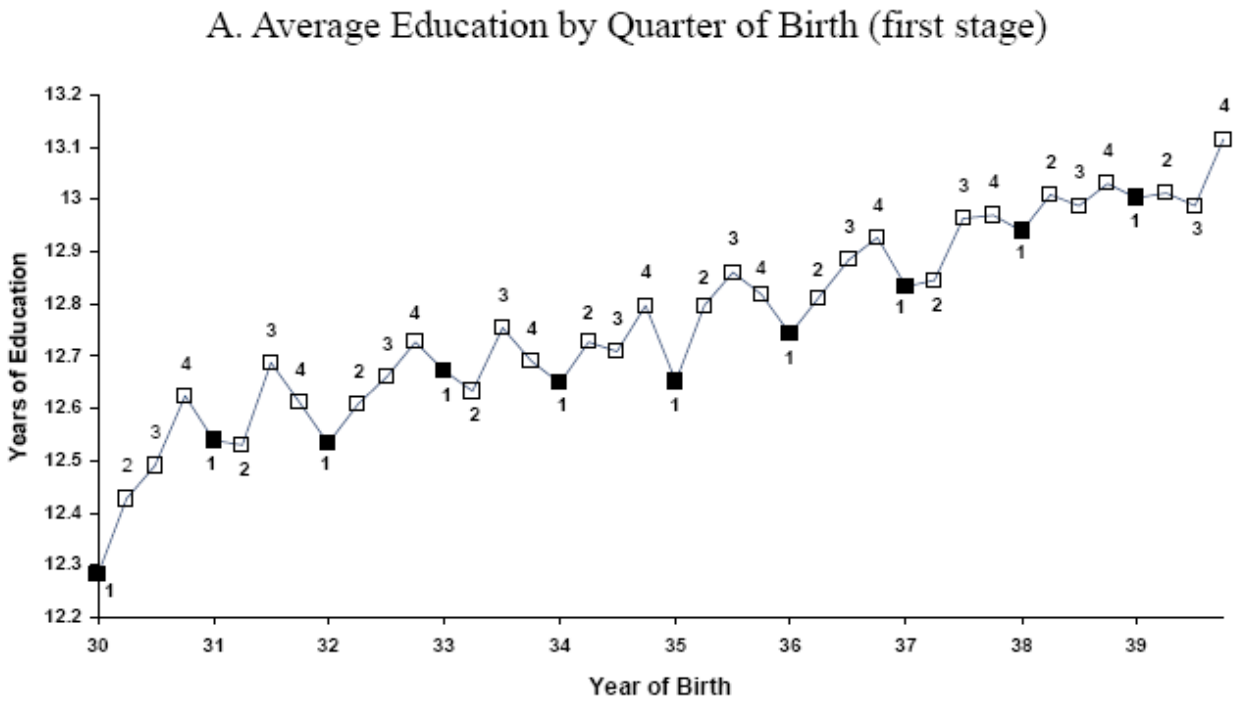
$$Cov(Z_i, u_i) = 0$$

- If we have controls X_i in the second stage, then this condition only holds conditional on X
- If Z was generated by a randomized experiment, then it would be unnecessary to condition on X in both the first and the second stage!

Example: (Angrist and Krueger, QJE 1991)

- Quarter of birth as an instrument for schooling
- Students enter schooling in the September of the calendar year in which they turn 6
- And compulsory school law requires them to remain in school until they become 16
- Hence people born late in the year are more likely to stay at school longer

Is the first stage right?



The reduced form for earnings

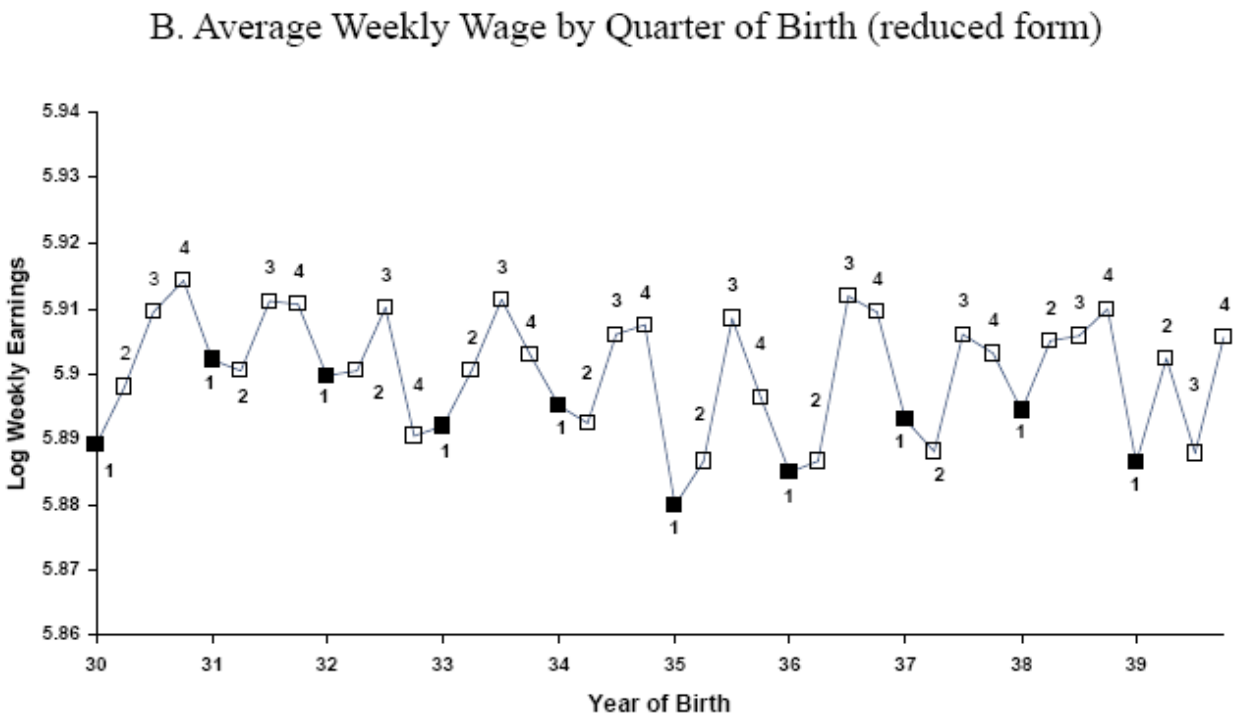


Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

Notes: Adapted from a re-analysis of Angrist and Krueger (1991) by Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930-39 birth cohorts in the 1980 Census 5 percent file. The sample size is 329,509.

Validity of the instrument

1. Power of the instrument?
2. Exogeneity?
3. Exclusion restriction?

- Overall, are the OLS estimates mostly larger than the corresponding IV estimates?
- What does this tell us about the omitted variable bias?
- How are the IV estimates "local" in this case?

Weak instruments

- The bias of 2SLS can be written as:

$$\hat{\rho} = \rho + \frac{Cov(z,\eta)}{Cov(S,z)}$$

- When the instrument is only weakly correlated with endogenous regressor and even slightly correlated with error term the bias in IV estimator can still be very large

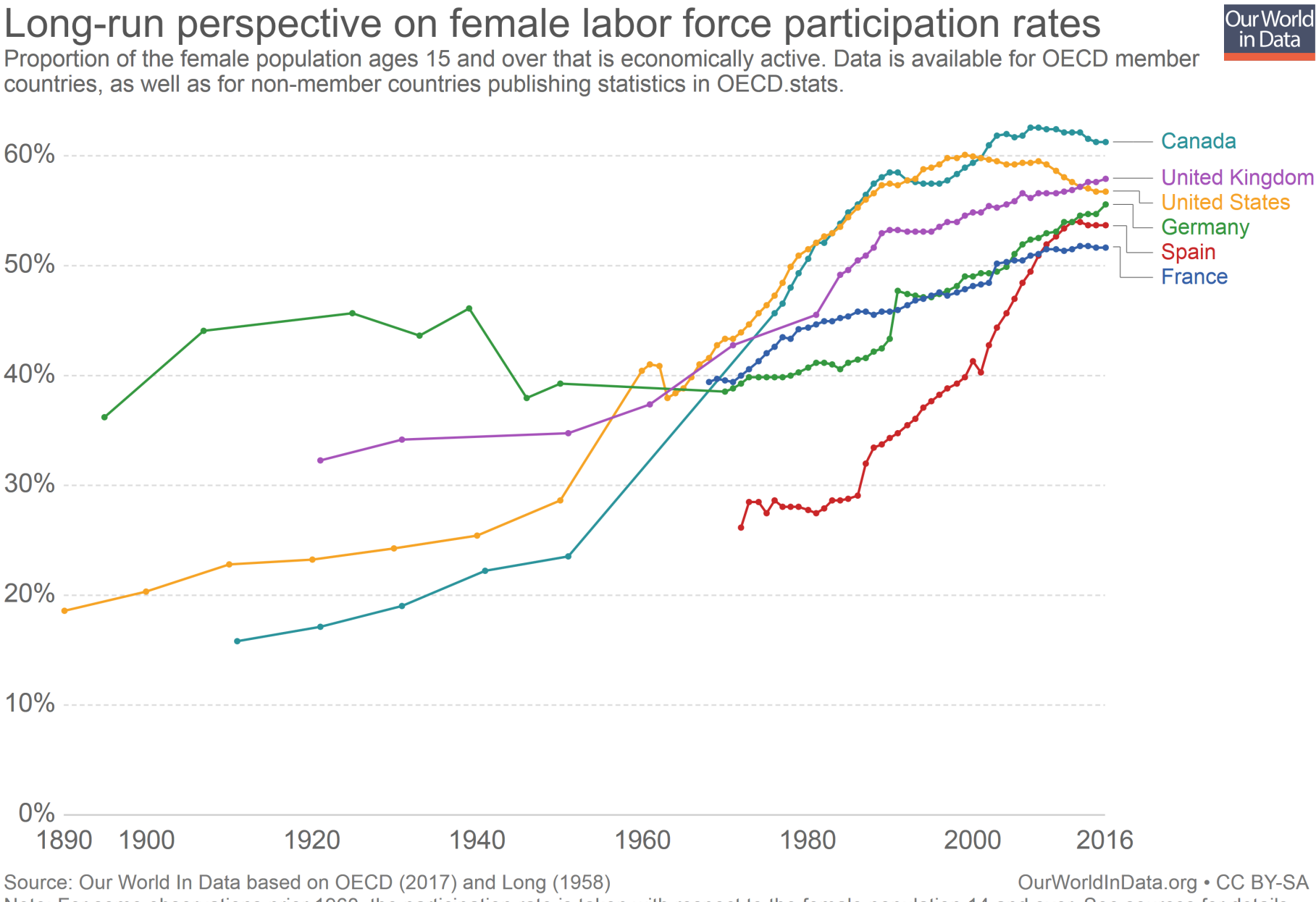
Local average treatment effects

- In the previous lecture we analyzed IV in the context of RCT's and acknowledged that the effects may be heterogeneous
- In this case IV estimates the Local Average Treatment Effect (LATE)
- Note that the LATE framework partitions any population with an instrument into a set of four instrument-dependent subgroups, defined by the manner in which members of the population react to the instrument:
  - Compliers: $D_{i1} = 1, D_{i0} = 0$
  - Always-takers: $D_{i1} = 1, D_{i0} = 1$
  - Never-takers: $D_{i1} = 0, D_{i0} = 0$
  - Defiers: $D_{i1} = 0, D_{i0} = 1$
- Who would be the compliers in the Angrist-Krueger setting?

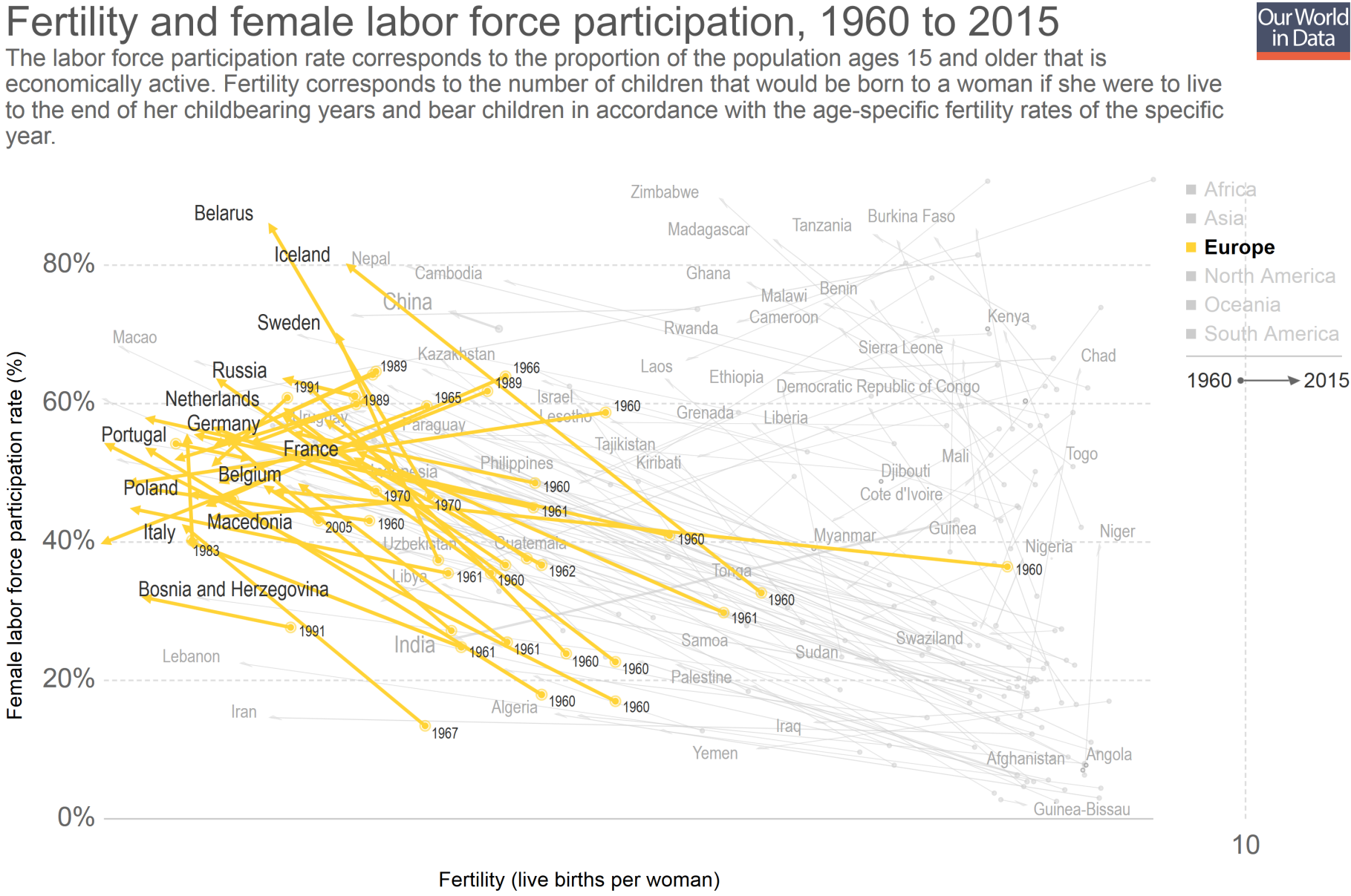
Example: Fertility and female labor supply

- Powerful global trend: Increasing female labor force participation
- Similarly powerful decline in fertility
- Are these trends linked? Does childbearing keep women from developing their careers?
- Huge number of studies that show a negative correlation
- Why wouldn't this correlation be causal?

Female labor force participation 1890-2016



Fertility and female labor force participation 1960-2015



Source: UN Population Division (2017 Revision), World Bank – WDI – OurWorldInData.org/female-labor-force-participation-key-facts • CC BY SA

Example: Fertility and female labor supply

- Strong theoretical reasons to believe that fertility and labor supply decisions are jointly determined
- Fertility is not allocated at random
- Unclear whether observed differences in labor market outcomes reflect the causal effects of having children

Example: Fertility and female labor supply

- Look for variation in the number of children that is as good as randomly assigned
- First attempts: twin births
  - Why could this work?
  - What could go wrong here?
- We cover:
  - Angrist & Evans (1998): Sibling sex mix in families with two or more children
  - Lundborg et al (2016): The success of IVF treatments

Example: Fertility and female labor supply

- Angrist & Evans (1998): Sibling sex mix of the first two children as an instrument for the decision to have a third child
- Based on two assumptions:
  - Parental preference for mixed sibling-sex composition
  - Sex mix is virtually randomly assigned
- Why does this mean that sex mix can work as an instrument?

Angrist & Evans: First stage

Sex of first two children in families with two or more children	All women				Married women			
	1980 PUMS (394,835 observations)		1990 PUMS (380,007 observations)		1980 PUMS (254,654 observations)		1990 PUMS (301,588 observations)	
	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
two girls	0.242	0.441 (0.002)	0.241	0.412 (0.002)	0.239	0.425 (0.002)	0.239	0.408 (0.002)
two boys	0.264	0.423 (0.002)	0.264	0.401 (0.002)	0.266	0.404 (0.002)	0.264	0.396 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
(2) both same sex	0.506	0.432 (0.001)	0.505	0.407 (0.001)	0.506	0.414 (0.001)	0.503	0.401 (0.001)
difference (2) − (1)	—	0.060 (0.002)	—	0.063 (0.002)	—	0.068 (0.002)	—	0.070 (0.002)

Angrist & Evans: Wald estimates

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Angrist & Evans: Comparison OLS and 2SLS

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.176	−0.120	−0.113	−0.167	−0.120	−0.113	−0.008	0.004	0.001
	(0.002)	(0.025)	(0.025)	(0.002)	(0.028)	(0.028)	(0.001)	(0.009)	(0.008)
			[0.013]			[0.013]			[0.013]
Weeks worked	−8.97	−5.66	−5.37	−8.05	−5.40	−5.16	−0.82	0.59	0.45
	(0.07)	(1.11)	(1.10)	(0.09)	(1.20)	(1.20)	(0.04)	(0.60)	(0.59)
			[0.017]			[0.071]			[0.030]
Hours/week	−6.66	−4.59	−4.37	−6.02	−4.83	−4.61	0.25	0.56	0.50
	(0.06)	(0.95)	(0.94)	(0.08)	(1.02)	(1.01)	(0.05)	(0.70)	(0.69)
			[0.030]			[0.049]			[0.71]
Labor income	−3768.2	−1960.5	−1870.4	−3165.7	−1344.8	−1321.2	−1505.5	−1248.1	−1382.3
	(35.4)	(541.5)	(538.5)	(42.0)	(569.2)	(565.9)	(103.5)	(1397.8)	(1388.9)
			[0.126]			[0.703]			(0.549)
ln(*Family income*)	−0.126	−0.038	−0.045	−0.132	−0.051	−0.053	—	—	—
	(0.004)	(0.064)	(0.064)	(0.004)	(0.056)	(0.056)			
			[0.319]			[0.743]			
ln(*Non-wife income*)	—	—	—	−0.053	0.023	0.016	—	—	—
				(0.005)	(0.066)	(0.066)			
						[0.297]			

Validity of the instrument

1. Power of the instrument?
2. Exogeneity?
3. Exclusion restriction?

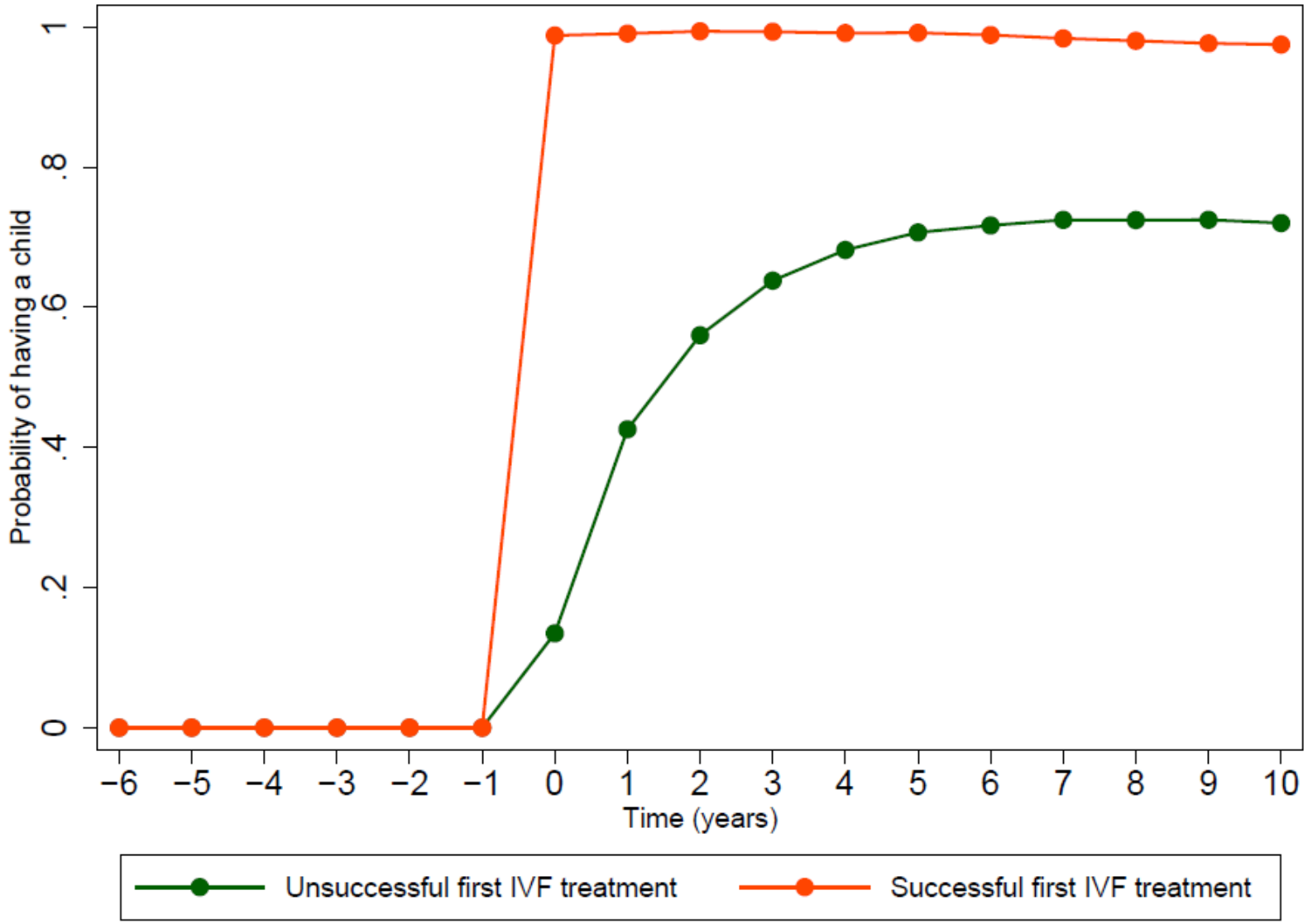
- Overall, are the OLS estimates mostly smaller than the corresponding IV estimates?
- What does this tell us about the omitted variable bias?
- Based on this, what would you say about the effect of childbearing on female labor market outcomes?
- How are the IV estimates local in this case?

Example: Fertility and female labor supply

- Angrist & Evans only studied the effect of having a third child, why?
- Difference between the effects of childbearing at the *intensive* and *extensive* margin
- Lundborg et al (2016): The success of IVF treatments as an instrument for having a first child
  - Focus on women who are going through IVF treatment
  - The success in the first round of treatment is as good as random
- Why does this mean that the success of IVF can work as an instrument?

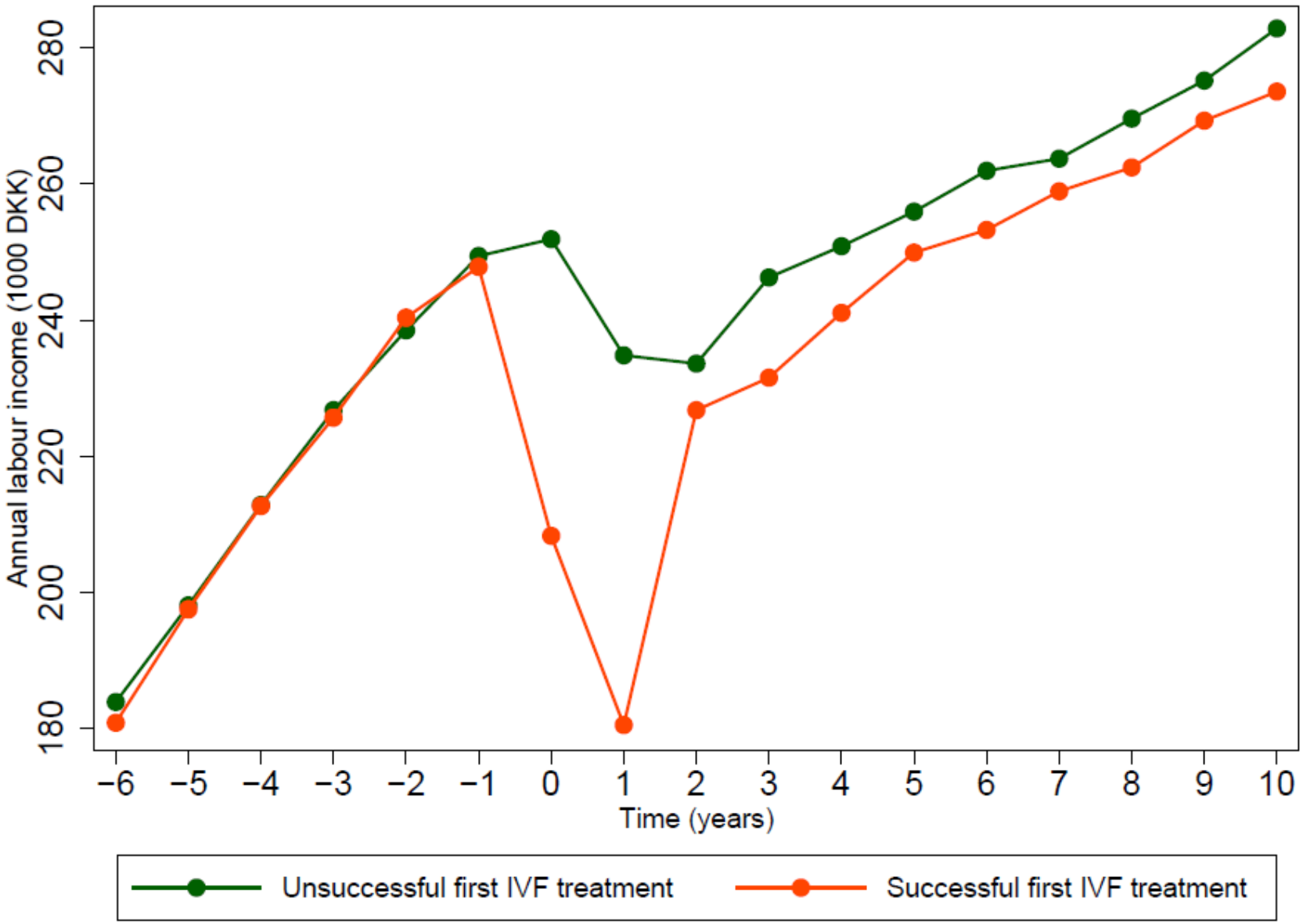
Lundborg et al: Fertility and IVF success

Figure 1: Fertility at the extensive margin before and after the first IVF treatment.



Lundborg at al: IVF success and annual earnings

Figure 2: Annual earnings before and after the (potential) birth of a child.



Lundborg et al: First stage, reduced form, and Wald estimates

Table 3: Fertility effects on female labor earnings: Results from first-stage, reduced form, and instrumental variable regressions.

Independent variable	(1) t=0	(2) t=1	(3) t=2	(4) t=3	(5) t=4	(6) t=5	(7) t=6	(8) t=7	(9) t=8	(10) t=9	(11) t=10
	Panel A: First stage regressions using any children (0/1) as dependent variable										
IVF success (0/1)	0.84***	0.54***	0.41***	0.33***	0.28***	0.26***	0.24***	0.23***	0.23***	0.23***	0.23***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
F-stat.	57424	13303	7844	5493	4055	3199	2482	1874	1520	1176	943
	Panel B: Reduced form regressions using annual earnings as dependent variable										
IVF success (0/1)	-43299***	-53082***	-6009***	-14340***	-9050***	-5269**	-8679***	-5992**	-7536**	-5566	-10975***
	(1541)	(1700)	(1830)	(1934)	(2055)	(2244)	(2463)	(2738)	(3038)	(3491)	(3981)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620
	Panel C: IV regressions using annual earnings as dependent variable										
Any children (0/1)	-51251***	-97914***	-14781***	-43798***	-32147***	-20493**	-35866***	-26064**	-32814**	-24338	-47079***
	(1826)	(3099)	(4474)	(5872)	(7275)	(8710)	(10168)	(11901)	(13237)	(15262)	(17114)
N	18538	18494	18435	18381	17404	15599	13779	11983	10173	8342	6620

Validity of the instrument

1. Power of the instrument?
2. Exogeneity?
3. Exclusion restriction?

- Compare the results to the ones in Angrist & Evans
- What does this comparison say about the effect of having the first vs. the third child?
- How are the IV estimates local in this case?

Let us wrap it up

- IV estimates are a powerful tool to identify causal links
- But IV power relies on the quality of the instruments
- Always discuss instrument plausibility
- Three dimensions:
  1. Power
     - Always report the first stage (F-test above 10)
     - Weak instruments have very unpleasant consequences
  2. Exogeneity
     - Does it make sense to believe that the instrument is randomly assigned?
     - To be sure: check if the instrument is correlated with predetermined variables
  3. Exclusion restriction
     - Cannot be tested, but discuss the possible links between z and u
- Specify the group which is affected by the instrument (LATE)

Wald estimator

- Since:

$$\begin{aligned} E[yz] &= E[y|z=1]p \\ E[y]E[z] &= \{E[y|z=1]p + E[y|z=0](1-p)\}\, p \end{aligned}$$

- We have that:

$$\begin{aligned} Cov(yz) &= E[y - E[y]][z - E[z]] \\ &= E[yz] - E[y]E[z] \\ &= E[y|z=1]p - \{E[y|z=1]p + E[y|z=0](1-p)\}\, p \\ &= p\{E[y|z=1] - E[y|z=1]p - E[y|z=0](1-p)\} \\ &= p(1-p)\{E[y|z=1] - E[y|z=0]\} \end{aligned}$$

- and similarly:

$$Cov(sz) = p(1-p)\{E[s|z=1] - E[s|z=0]\}$$

What did we do last time?

- Instrumental variables in a regression framework

$$\begin{aligned} Y_i &= \alpha_0 + \rho S_i + \eta_i \\ \eta_i &= \gamma A_i + v_i \end{aligned}$$

  with $Cov(S, v) = 0$
- No RCT, Can't observe A
- Come up with Z such that

$$Cov(S, Z) \neq 0 \text{ and } Cov(Z, \eta) = 0$$

What did we do last time?

- Write

$$\begin{aligned} Cov(Y,Z) &= Cov(\alpha_0 + \rho S_i + \eta_i, Z) \\ &= \rho Cov(Z,S) + Cov(Z,\eta) \end{aligned}$$

- It follows that:

$$\rho = \frac{Cov(Z,Y)}{Cov(Z,S)} = \frac{\frac{Cov(Z,Y)}{Var(Z)}}{\frac{Cov(Z,S)}{Var(Z)}}$$