ECON42720 Causal Inference and Policy Evaluation

5 Instrumental Variables

Ben Elsner (UCD)

Resources

Most **textbooks have a chapter on IV**

- ▶ In the applied world, Cunningham's Mixtape (Ch. 7) and Huntington-Klein's The Effect (Ch. 19) are good resources
- ▶ Angrist and Pischke's Mostly Harmless Econometrics (Ch. 4) is slightly more technical

IV: Starting Point

$$y_i = \alpha + \beta D_i + u_i$$

CIA $cov(D_i, u_i) = 0$ **often doesn't hold** $\Rightarrow$ **OLS estimates of** β **are biased**
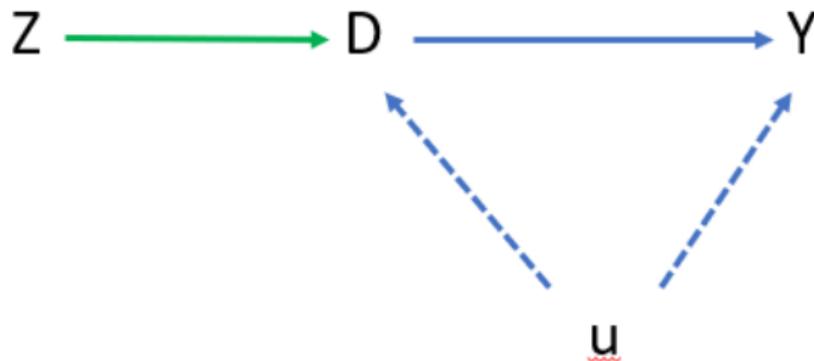
- ▶ **Unobserved heterogeneity**: we may not observe all confounding variables
- ▶ D_i may be **measured with error**
- ▶ Simultaneity or **reverse causality**

Instrumental Variables
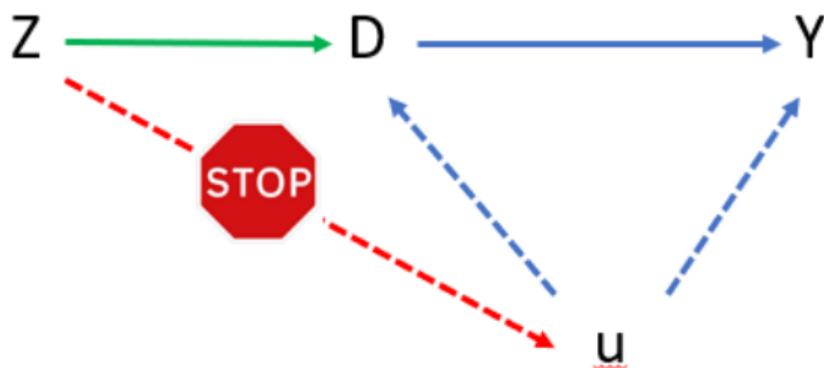
In theory, **instrumental variables** offer a way to

- ▶ break the correlation $cov(D_i, u_i)$
- ▶ and obtain a **consistent causal estimate of the treatment on** y_i

Idea behind an instrumental variable (Z):

Instrumental Variables

1) The IV **must not be correlated with unobservable characteristics** (conditional independence)

2) An IV **affects Y only through its effect on D**
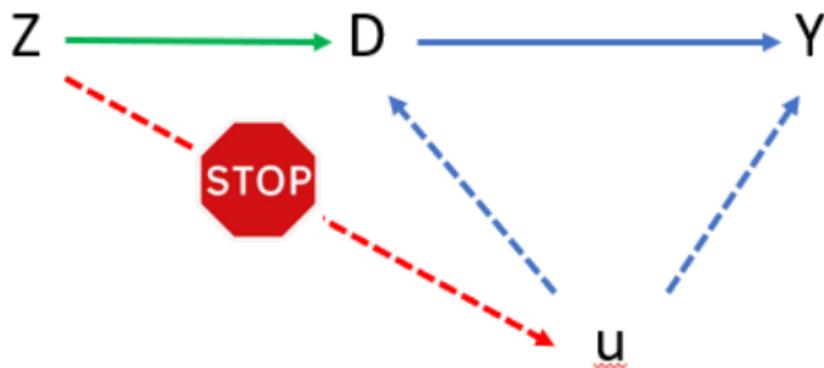
Instrumental Variables

One way to think about an IV:

- ▶ **people/firms make optimal choices** that affect their **treatment status**
- ▶ Z is a **shock that changes the behavior** of at least some people/firms
- ▶ Z has to be **unrelated to people's characteristics**
- ▶ i.e. it should be assigned as good as randomly

And another:

- ▶ The instrument Z is a **treatment/incentive that is offered** to units/people
- ▶ D measures if the unit **actually takes up the treatment**
- ▶ The instrument Z should be **as good as randomly assigned**
- ▶ Example: randomly assigned **school vouchers**

Instrumental Variables

And another:

- ▶ OLS uses all the variation in D to explain y
- ▶ IV uses **only the variation in D that is related to Z**
- ▶ So this means less variation is used, but at least Z is unrelated to u

Instrumental Variables Lingo
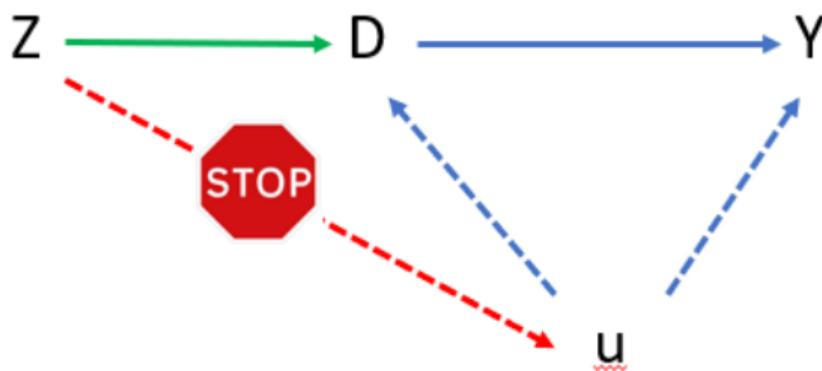
IV requires three ingredients:

- **First stage:** $cov(Z, D) \neq 0$
- **(Conditional) independence**: $cov(Z, u) = 0$
- **Exclusion restriction:** affects Y only through D and no other channel

First Stage and Exclusion Restriction

The **first-stage relationship is testable**

- ▶ we can run a regression of D on Z
- ▶ it is also possible to include covariates

The **exclusion restriction is not testable**

- ▶ it is an **identification assumption**
- ▶ we **need to make a convincing argument** in favor of it
- ▶ this is difficult and the reason for heated debates in seminars

Some say: **friends tell their friends not to use IV...**

IV Equations: Two-Stage Least Quares (2SLS)

Relationship of interest

$$y_i = \alpha + \beta D_i + X_i'\gamma + u_i$$

First stage

$$D_i = \delta_0 + \delta_1 Z_i + X_i'\rho + e_i$$

Second stage $\left(\widehat{D_i}\ \text{from first stage}\right)$

$$y_i = \tilde{\alpha} + \tilde{\beta}\widehat{D_i} + X_i'\kappa + \varepsilon_i$$

Reduced form

$$y_i = \lambda_0 + \lambda_1 Z_i + X_i'\sigma + \eta_i$$

IV in Theory

It can be shown that

$$\widehat{\beta^{IV}} = \frac{cov(Y, Z)}{cov(D, Z)} = \frac{\widehat{\lambda_1}}{\widehat{\delta_1}}$$

is a **consistent estimator of** β under the **exclusion restriction** $cov(Z, u) = 0$

This estimator is nothing but the **reduced-form coefficient** $\widehat{\lambda_1} = \frac{\widehat{cov(y, Z)}}{\widehat{var(Z)}}$

divided by the first stage $\widehat{\delta_1} = \frac{\widehat{cov(D, Z)}}{\widehat{var(Z)}}$

Later we will see that this interpretation is useful

IV Illustration

First stage: we predict the treatment X based on the instrument Z

(a) Raw Data — Instrument Values: ○ 0, ● 1

(b) Predict X with Z — X mean for Z = 1, X mean for Z = 0
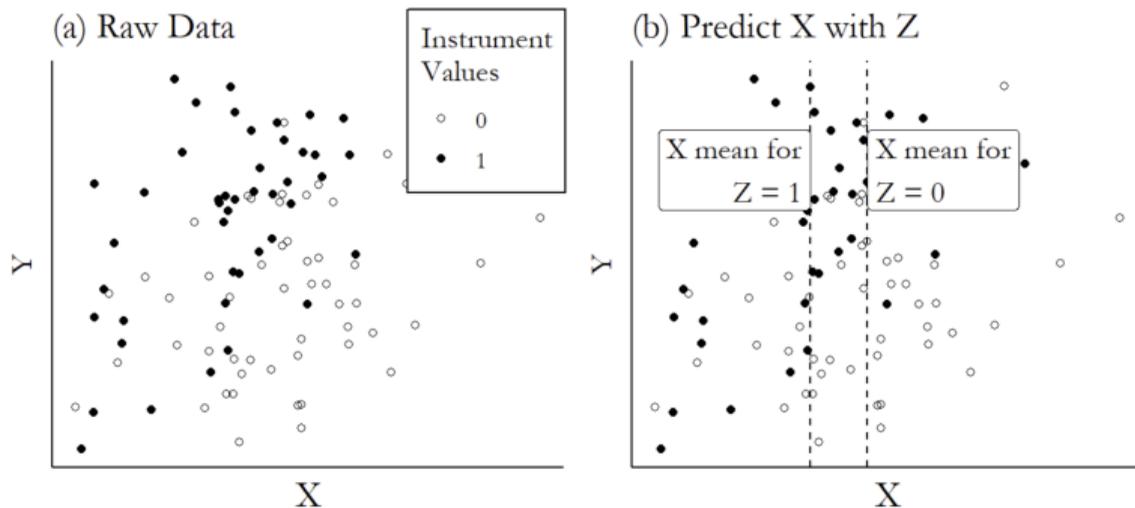
Credit: Huntington-Klein, The Effect, Ch. 19

IV Illustration

Relate the outcome Y to the **predicted X from the first stage**, and calculate the difference in outcomes for different levels of Z

(c) Only Use Predicted Variation

X mean for Z = 1

X mean for Z = 0

(d) Predict Y with Z

Y mean for Z = 1

Y mean for Z = 0

Credit: Huntington-Klein, The Effect, Ch. 19
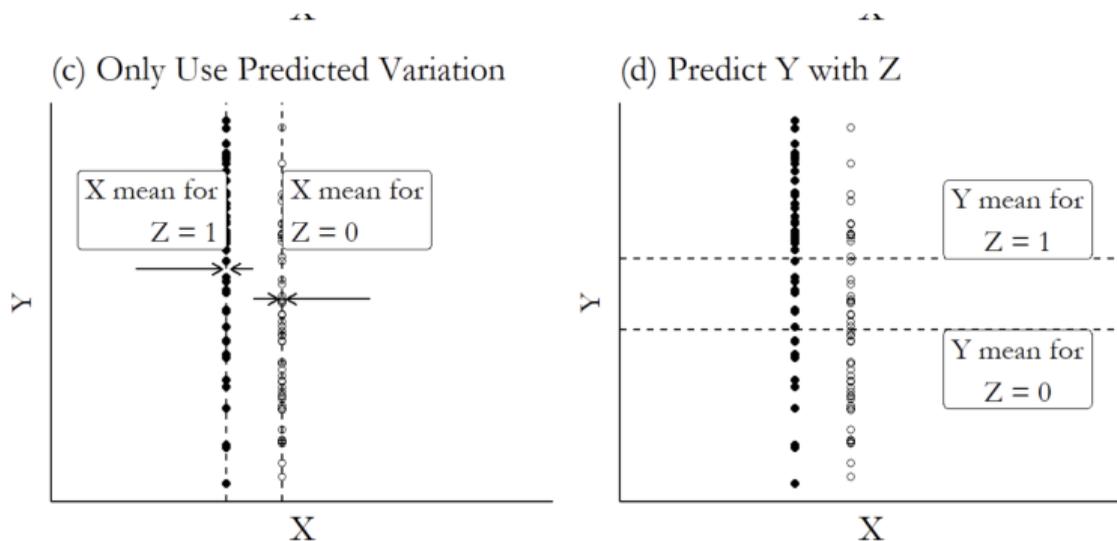
IV Illustration

Putting it all together: a change in the predicted X leads to a different Y

(e) Only Use Predicted Variation

(f) Relate Predicted Y to Predicted X
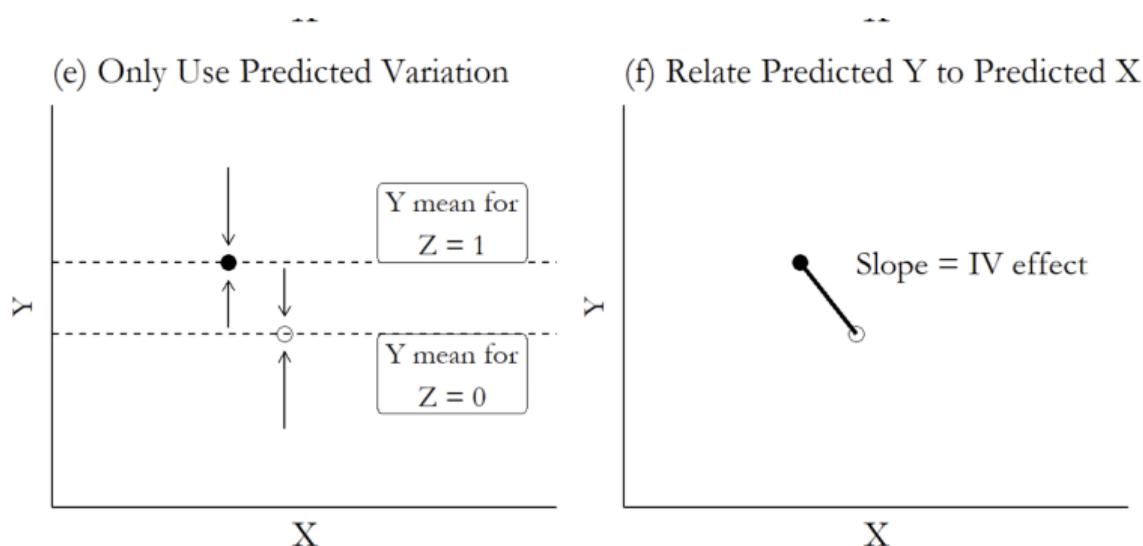
Credit: Huntington-Klein, The Effect, Ch. 19

Classic IV Example: Moving to Opportunity

Research question: does **moving to a better neighborhood** affect adults and children?

The **Moving to Opportunity Program (MTO)**

▶ **Large-scale experiment** with people in public housing in several US cities in 1996
▶ **Treatment group 1**: voucher for private rental housing in low-poverty neighborhood
▶ **Treatment group 2**: voucher for private rental housing (no strings attached)
▶ **Control group**: no voucher

This experiment has been evaluated by Kling *et al.* (2007).

Classic IV Example: Moving to Opportunity

50% of voucher recipients actually moved; most to better neighborhoods

DESCRIPTIVE STATISTICS OF NEIGHBORHOOD CHARACTERISTICS

	Experimental (i)	Section 8 (ii)	Control (iii)
Average census tract poverty rate	.33	.35	.45
Average census tract poverty rate above 30%	.52	.62	.87
Respondent saw illicit drugs being sold or used in neighborhood during past 30 days	.33	.34	.46
Streets are safe or very safe at night	.70	.65	.56
Member of household victimized by crime during past 6 months	.17	.16	.21
Average census tract share on public assistance	.16	.17	.23
Average census tract share of adults employed	.83	.83	.78
Average census tract share workers in professional and managerial occupations	.26	.23	.21
Average census tract share minority	.82	.87	.90

Moving to Opportunity: Empirical Challenge

MTO was a **randomized experiment**

- $Z \in \{0, 1\}$ is the instrument, $D \in \{0, 1\}$ is the treatment
- but **not everyone** who received a voucher **actually moved**

We can estimate an **Intention-to-Treat (ITT)** effect by using the **reduced form**

$$y_i = \gamma_0 + \gamma_1 Z_i + \varepsilon_i$$

ITT is useful for policy evaluation

- But it does not tell us much about the **causal effect of moving**

Moving to Opportunity

Suppose we are interested in the **treatment effect on the treated**, in this case the **causal effect of moving**

▶ but we cannot force voucher recipients to move...

IV allows us to estimate this treatment effect under three conditions

1. Z is as good as **randomly assigned**
2. Z has **no direct effect** on the outcome
3. Z has a **sufficiently strong effect** on D

Moving to Opportunity: The Wald Estimator

We can estimate **three causal effects**

1. **First stage:** the causal effect of Z on D: $P(D = 1|Z = 1) - P(D = 1|Z = 0)$
2. **Reduced form (ITT):** the causal effect of Z on Y: $E(Y|Z = 1) - E(Y|Z = 0)$
3. **Treatment effect of interest:** the causal effect of D on Y:
   $Y(1) - Y(0) = E(Y|D = 1) - E(Y|D = 0)$

The **Wald Estimator** relates all three effects

$$E(Y|D = 1) - E(Y|D = 0) = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{P(D = 1|Z = 1) - P(D = 1|Z = 0)} \tag{1}$$

Moving to Opportunity: The Wald Estimator

$$\widehat{\beta}^{IV} = E(Y|D=1) - E(Y|D=0) = \frac{E(Y|Z=1) - E(Y|Z=0)}{P(D=1|Z=1) - P(D=1|Z=0)}$$

▶ **difference in outcomes** by groups intended and not intended for treatment
▶ divided by **difference in the actual treatment**

Interpretation of the Wald Estimator

What we want to know: the impact of moving: $\Delta D = P(D = 1) - P(D = 0) = 1$

\brf{What we do know{:

- the impact of the instrument on moving:
  $\Delta D(Z) = P(D = 1 | Z = 1) - P(D = 1 | Z = 0) = 0.5$
- suppose the difference in outcomes $E(Y|Z = 1) - E(Y|Z = 0)$ is 10
- so the fact that **50% moved gives us an average difference in outcomes of 10**

If 0.5 movers gives us 10 then what would 1 mover give us?

- The answer is $\widehat{\beta}^{IV} = \frac{10}{0.5} = 20$

Moving to Opportunity

	E/S (i)	CM (ii)	ITT (iii)	TOT (iv)	CCM (v)
A. Adult outcomes					
Obese, BMI ≥ 30	E − C	0.468	−0.048 (0.022)	−0.103 (0.047)	0.502
Calm and peaceful	E − C	0.466	0.061 (0.022)	0.131 (0.047)	0.443
Psychological distress, K6 z-score	E − C	0.050	−0.092 (0.046)	−0.196 (0.099)	0.150
B. Youth (female and male) outcomes					
Ever had generalized anxiety symptoms	E − C	0.089	−0.044 (0.019)	−0.099 (0.042)	0.164
	S − C	0.089	−0.063 (0.019)	−0.114 (0.035)	0.147
Ever had depression symptoms	S − C	0.121	−0.039 (0.019)	−0.069 (0.035)	0.134
C. Female youth outcomes					
Psychological distress, K6 scale z-score	E − C	0.268	−0.289 (0.094)	−0.586 (0.197)	0.634
Ever had generalized anxiety symptoms	E − C	0.121	−0.069 (0.027)	−0.138 (0.055)	0.207
	S − C	0.121	−0.075 (0.029)	−0.131 (0.051)	0.168
Used marijuana in the past 30 days	E − C	0.131	−0.065 (0.029)	−0.130 (0.059)	0.202
	S − C	0.131	−0.072 (0.032)	−0.124 (0.056)	0.209
Used alcohol in past 30 days	S − C	0.206	−0.091 (0.038)	−0.155 (0.056)	0.306

Wald estimator: TOT; denominator: CM

Classic IV Example: Angrist & Evans (1998)

Angrist & Evans (1998) study the effect of **children on female labor supply**

Their **most basic regression** is

$$hours_i = \alpha + \beta kids_i + u_i$$

The **number of children** is almost certainly **endogenous**:

- ▶ fertility is a choice, and so is labor supply
- ▶ richer families can afford more children and lower labor supply
- ▶ couples differ in their preferences over fertility and labor supply

Classic IV Example: Angrist & Evans (1998)

Ideal experiment: randomly assign children to families

IV in Angrist & Evans (1998): **sex of the first two children**

- ▶ the sex of a child is as good as random
- ▶ couples tend to have a preferences for mixed-sex offspring
- ▶ couples with two boys or two girls are more likely to have a third child

Analysis is purely based on families with two or more children

Classic IV Example: Angrist & Evans (1998)

The **components of the IV estimator**

First stage: effect of same-sex children on the likelihood of having a third child

$$kids_i = \delta_0 + \delta_1 samesex_i + e_i$$

Reduced form:

$$hours_i = \lambda_0 + \lambda_1 samesex_i + \eta_i$$

Exclusion restriction: same-sex children unrelated with personal characteristics
$\Rightarrow cov(samesex_i, u_i) = 0$

Classic IV Example: Angrist & Evans (1998)

The following analysis is based on a small sub-sample of Angrist & Evans (1998)

```
. sum hours kids samesex

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       hours |     31857    21.22011    19.49892         0         99
        kids |     31857    2.752237    .9771916         2         12
     samesex |     31857     .502778    .5000001         0          1
```

Descriptive statistics indicate that in **50% of all families** the **first two children had the same sex**

This is **what we would expect**. Any different result would be a red flag

Classic IV Example: Angrist & Evans (1998)

Now let's look at the **simple OLS regression**

```
. reg hours kids, robust

Linear regression                              Number of obs =   31857
                                               F(  1, 31855) =  585.25
                                               Prob > F      =  0.0000
                                               R-squared     =  0.0178
                                               Root MSE      =  19.325

------------------------------------------------------------------------
             |               Robust
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------
        kids |  -2.664309   .1101318   -24.19   0.000    -2.880171   -2.448446
       _cons |   28.55292   .3200455    89.22   0.000     27.92562    29.18022
------------------------------------------------------------------------
```

Each additional child (above two) decreases weekly work hours on average by 2.66

Classic IV Example: Angrist & Evans (1998)

The first stage: is the **instrument relevant to explain the number of kids?**

```
. reg kids samesex, robust
Linear regression                                  Number of obs =    31857
                                                   F(  1, 31855) =    40.90
                                                   Prob > F      =   0.0000
                                                   R-squared     =   0.0013
                                                   Root MSE      =  .97658

------------------------------------------------------------------------------
             |               Robust
        kids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     samesex |   .0699933   .0109439     6.40   0.000     .0485429    .0914437
       _cons |   2.717045    .007806   348.07   0.000     2.701745    2.732346
------------------------------------------------------------------------------
```

Important things to discuss in an IV paper

- ▶ Does the first-stage coefficient make sense (sign, magnitude)?
- ▶ Is the first-stage correlation strong enough (is the F-Statistic of the instrument >10)

Classic IV Example: Angrist & Evans (1998)

```
. reg kids samesex, robust
Linear regression                              Number of obs =   31857
                                               F(  1, 31855) =   40.90
                                               Prob > F      =  0.0000
                                               R-squared     =  0.0013
                                               Root MSE      =  .97658

------------------------------------------------------------------------
             |              Robust
        kids |    Coef.   Std. Err.      t     P>|t|    [95% Conf. Interval]
-------------+----------------------------------------------------------
     samesex |  .0699933   .0109439    6.40   0.000     .0485429   .0914437
       _cons |  2.717045    .007806  348.07   0.000     2.701745   2.732346
------------------------------------------------------------------------
```

In this case...

▶ families with same-sex children have more children
▶ the coefficient is small: out of 14 families with same-sex children, one has an additional child
▶ the t-statistic of the instrument is strong enough (implied F-Statistic: $F = 40.96$)

Classic IV Example: Angrist & Evans (1998)
2SLS estimate

```
. ivreg hours (kids = samesex), robust

Instrumental variables (2SLS) regression        Number of obs =   31857
                                                 F(  1, 31855) =    3.19
                                                 Prob > F      = 0.0743
                                                 R-squared     =       .
                                                 Root MSE      = 19.534

------------------------------------------------------------------------
            |               Robust
      hours |     Coef.   Std. Err.      t    P>|t|    [95% Conf. Interval]
------------+-----------------------------------------------------------
       kids |  -5.58186   3.127136    -1.78   0.074   -11.71117   .5474471
      _cons |  36.58271   8.606509     4.25   0.000    19.71362   53.45179
------------------------------------------------------------------------
Instrumented:  kids
Instruments:   samesex
```

This table reports the **second-stage estimates**

- ▶ the regressor is the **number of children predicted by the same-sex instrument**
- ▶ the effect is stronger than the OLS estimate (-2.66)
- ▶ it is statistically significant at the 10%-level

Classic IV Example: Angrist & Evans (1998)

To **develop a better intuition of how IV works**, it is useful to look at the reduced form and first stage

The IV estimator is the **reduced-form divided by the first stage**

$$\widehat{\beta^{IV}} = \frac{\widehat{\lambda_1}}{\widehat{\delta_1}}$$

```
. reg hours samesex

      Source |       SS           df       MS        Number of obs   =    31,857
-------------+----------------------------------      F(1, 31855)     =       3.20
       Model |  1215.63289         1  1215.63289      Prob > F        =     0.0738
    Residual |    12110681    31,855  380.181477      R-squared       =     0.0001
-------------+----------------------------------      Adj R-squared   =     0.0001
       Total |  12111896.6    31,856  380.207703      Root MSE        =     19.498


-------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
     samesex |  -.3906929   .2184891    -1.79   0.074    -.8189399    .0375541
       _cons |   21.41654   .1549237   138.24   0.000     21.11288     21.7202
-------------------------------------------------------------------------------
```

Intuition behind the IV

What we want to know: the **impact of having one more child**

Consider the **first stage** and **reduced form**:

- ▶ having same-sex children increases the number of children by 0.07
- ▶ having same-sex children decreases weekly work hours by 0.39

So, 0.07 additional children lead to 0.39 fewer work hours

What reduction in work hours would we expect from **one additional child**?

- ▶ answer: $\dfrac{\widehat{\lambda_1}}{\widehat{\delta_1}} = \dfrac{0.39}{0.07} = 5.57$ hours

Classic IV Example: Angrist & Evans (1998)

So we have that $\widehat{\beta^{IV}} < \widehat{\beta^{OLS}}$. Does this make sense?

Explanation 1: OLS estimator is upward biased (i.e. closer to zero)

- ▶ there could be an omitted variable (for example family wealth)
- ▶ both the correlation with kids and the direct effect on hours need to have the same sign
- ▶ e.g. $cov(wealth, kids) > 0$ and $cov(wealth, hours|kids) > 0$ or both negative

Explanation 2: IV effect measures the **effect for a specific population**

- ▶ only 1 in 14 families "respond'' to the instrument
- ▶ families who respond may not be the average family...

Local Average Treatment Effects (LATE)

So far, we implicitly assumed that the **potential outcomes are constant across units**. But what if potential outcomes are heterogeneous?

Consider a case with a binary instrument $Z_i \in \{0, 1\}$ the the treatment statuses

- $D_{1i} = $ i's treatment status when $Z_i = 1$
- $D_{0i} = $ i's treatment status when $Z_i = 0$

The **observed treatment status** is

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \delta_0 + \delta_{1i}Z_i + \eta_i$$

Note that the effect of the IV on treatment may differ between individuals

Local Average Treatment Effects (LATE)

We **divide the population into four groups** depending on their reaction to the instrument

1. **Compliers**: people who react to the instrument as expected, $D_{1i} = 1$ and $D_{0i} = 0$
2. **Always-takers**: people who always take the treatment regardles of Z, $D_{1i} = D_{0i} = 1$
3. **Never-takers**: people who never take the treatment regardless of Z, $D_{1i} = D_{0i} = 0$
4. **Defiers**: people who react to the instrument in the wrong direction, $D_{1i} = 0$ and $D_{0i} = 1$

From any dataset, it is impossible to see who belongs to what group

The Angrist-Imbens-Rubin Causal Model

Angrist *et al.* (1996) define the **minimum set of assumptions** for the **identification of a causal effect** for the relevant subgroup of the population

As an example, consider Angrist (1990): the impact of **being a Vietnam veteran on earnings**

The Vietnam Draft Lottery (Angrist, 1990)

Context:

- In the 1960s and 70s young men in the US were at **risk of being drafted for military service** in Vietnam.
- Fairness concerns led to the institution of a **draft lottery** in 1970 that was used to determine **priority for conscription**

In each year from 1970 to 1972, **random sequence numbers were randomly assigned** to each birth date in cohorts of 19-year-olds.

- Men with lottery numbers below a cutoff were eligible for the draft.
- Men with lottery numbers above the cutoff were not.

But **compliance was not perfect**

- Many eligible men were exempted from service for health or other reasons.
- Others, who were not eligible, nevertheless volunteered for service.

The Vietnam Draft Lottery (Angrist, 1990)

Idea: use **lottery outcome as an instrument** for veteran status

Is there a first stage? the lottery did not completely determine veteran status, but it certainly mattered

What about the exclusion restriction?

- ▶ the lottery was random
- ▶ it seems reasonable to assume that its only effect was on veteran status

The Vietnam Draft Lottery (Angrist, 1990)

The **instrument is thus defined** as follows:

- $Z_i = 1$ if lottery implied individual i would be draft eligible,
- $Z_i = 0$ if lottery implied individual i would not be draft eligible.

The instrument affects **treatment**, which in this application amounts to **entering military service**.

The econometrician observes **treatment status** as follows:

- $D_i = 1$ if individual i served in the Vietnam war (veteran),
- $D_i = 0$ if individual i did not serve in the Vietnam war (not veteran).

The Angrist-Imbens-Rubin Causal Model

In Angrist (1990), the **main research question** is whether veteran status has a causal effect on earnings

The **causal effect of veteran status**, conditional on draft eligibility status, is defined as

$$Y_i(1, Z_i) - Y_i(0, Z_i)$$

We are **unable to identify individual treatment effects**, because we **do not observe all potential outcomes**

The Angrist-Imbens-Rubin Causal Model: Assumptions

Assumption 1: Random Assignment (ignorability)

All units have the **same probability of assignment to treatment**%

$$Pr(Z_i = 1) = Pr(Z_j = 1).$$

Given random assignment we can **identify and estimate the two intention to treat causal effects**:

$$E(D_i|Z_i = 1) - E(D_i|Z_i = 0) = \frac{cov(D_i, Z_i)}{var(Z_i)}$$

%

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) = \frac{cov(Y_i, Z_i)}{var(Z_i)}.$$

The Angrist-Imbens-Rubin Causal Model: Assumptions

Assumption 2: Non-zero average causal effect of Z on D

The **probability of treatment must be different** in the two assignment groups:%

$$Pr(D_{i1} = 1) \neq Pr(D_{i0} = 1)$$

This is the equivalent of the **first stage in the conventional I**V approach.

The Angrist-Imbens-Rubin Causal Model: Assumptions

Assumption 3: Exclusion Restriction

The **instrument** affects the **outcome** **only** through the treatment

$$Y_i(D_i, 0) = Y_i(D_i, 1) = Y_i(D_i)$$

Given treatment, assignment does not affect the outcome. So we can define the causal effect of D_i on Y_i as%

$$Y_{i1} - Y_{i0}.$$

This difference is not observed in the data. We **need to assume that assumption 3 holds** and bring good arguments in favour of it.

The Angrist-Imbens-Rubin Causal Model: Assumptions

Assumption 4: Monotonicity

- ▶ The instrument affects the **treatment status of all units** in the **same direction**
- ▶ Binary case: **no one does the opposite** of his/her assignment
- ▶ I.e. there are **no defiers**

$$D_{i1} \geq D_{i0} \ \forall i$$

Assumptions 2 and 4 together give **Strong Monotonicity** and ensure that:

- ▶ there is **no defier** and
- ▶ there exists **at least one complier**

Compliance types

		D_{i0}	
		0	1
	0	never-taker	defier
D_{i1}	1	complier	always-taker

Compliance types by treatment status and instrument

		Z_i	
		0	1
	0	complier OR never-taker	never-taker OR defier
D_i	1	always-taker or defier	complier OR always-taker

Compliance types

Compliance types by treatment status and instrument given monotonicity

		Z_i	
		0	1
	0	complier OR never-taker	never-taker
D_i	1	always-taker	complier OR always-taker

Back to the example (Angrist, 1990)

A1: instrument is as good as **randomly assigned**

- ▶ draft eligibility was assigned by a lottery...

A2: can have **no direct effect on the outcome** variable (earnings)

- ▶ this is debatable. Angrist argues that it holds

A3: **instrument affects the treatment**

- ▶ this can be checked

A4: **monotonicity**: a man who serves if not draft eligible, would also serve if draft eligible

- ▶ this seems plausible

Local Average Treatment Effect (Angrist, 1990)

Under the assumptions A1-A4, the IV approach in Angrist (1990) identifies a **local average treatment effect (LATE)**

The **effect is "local"** because

- it identifies the **effect on the compliers**
- ... the **causal effect of the draft on earnings** for men whose treatment status is changed by the instrument
- i.e. on men who are **drafted if eligible** but who **wouldn't volunteer if not eligible**

The **LATE is different from the ATE** because it excludes men who

- would be exempt from the draft regardless of their eligibility (never-takers)
- would volunteer regardless of their eligibility (always-takers)

The LATE theorem

Given assumptions 1-4,

$$\frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)} = E(Y_{i1} - Y_{i0}|D_{i1} > D_{i0})$$
$$= E(Y_{i1} - Y_{i0}|\text{complier}).$$

It shows that the **Wald estimator** equals the **average treatment effect for compliers**

LATE: Summary

The IV approach identifies a **local average treatment effect (LATE)**

- ▶ the IV needs to be **as good as randomly assigned** and satisfy the exclusion restriction
- ▶ the LATE is the **average treatment effect on the compliers**

Is **LATE an interesting parameter?**

- ▶ It depends on the question and who the compliers are
- ▶ Problem: we **cannot easily pinpoint the compliers**
- ▶ Newer methods allow us to **extrapolate from LATE** to other populations, e.g. Mogstad & Torgovitsky (2018)

Weak Instruments

Identification of the LATE requires the **existence of a first stage**

Otherwise, the **numerator of the Wald estimator is zero**, and the estimator not defined

$$\frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

Problem: existence of a **first stage is not enough**. It needs to be **sufficiently strong**

Weak IV Example: Angrist & Krueger (1991)

Research question: what is the effect of **compulsory schooling** on **earnings**?

It is **difficult to randomise**

- ▶ whether someone is affected compulsory schooling laws
- ▶ or how long someone stays in school

Trick of Angrist & Krueger (1991): when in the year you are born affects when you have to leave school

Compulsory Schooling and School Leaving Age
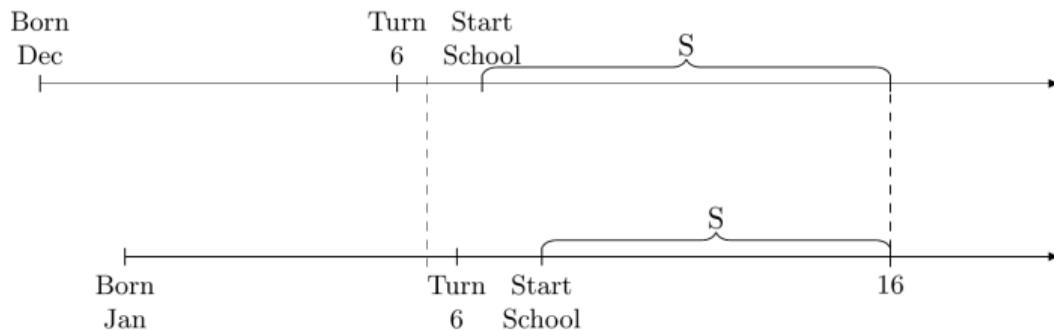
Quirk in the U.S. education system: **assignment to a cohort** is determined by **birth date**

▶ Children **born up until December 31** were assigned to **first grade**
▶ Children **born from January 1** were assigned to **kindergarten**

Compulsory Schooling and School Leaving Age
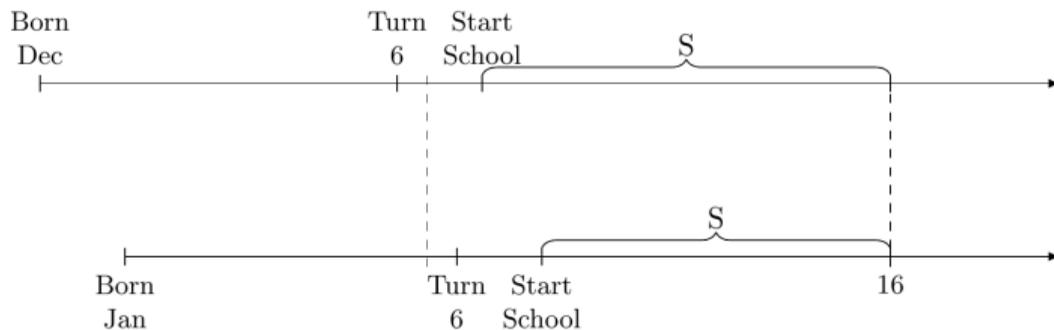
- Schooling **was compulsory until age 16**
- Children born in December had (exogenously) more education than children born in January

First Stage: Quarter of Birth and Years of Education
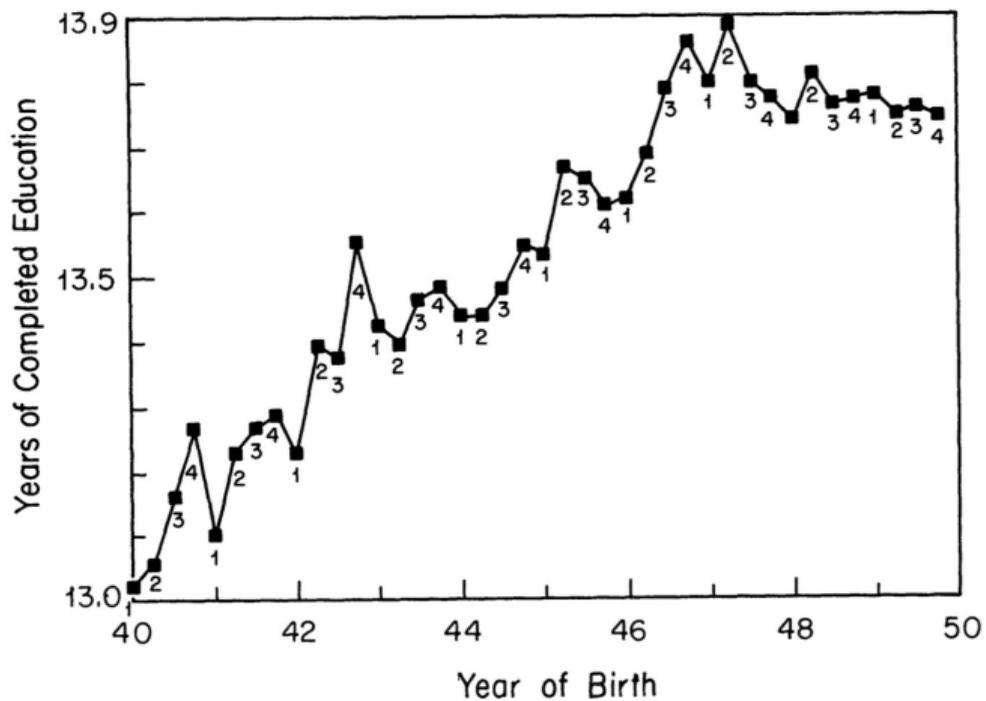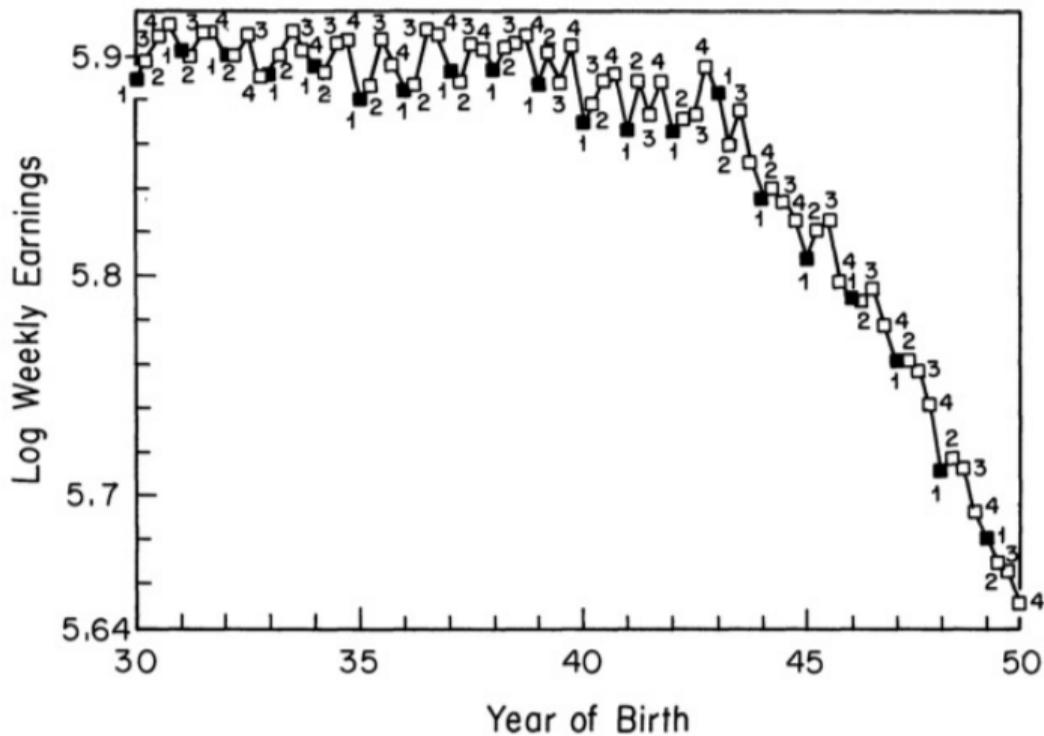
IV Relevance

Visual inspection suggests that a **first stage exists**

- ▶ Children born in Q4 have more schooling than children born in Q1
- ▶ This is on top of a general trend in more schooling

A **reduced form appears to exist** as well

- ▶ Children born in Q4 seemingly have slightly higher earnings than children born in Q1
- ▶ Again, this is on top of an overall trend in earnings

IV validity

Conditional independence: is quarter of birth as good as randomly assigned?

- ▶ Yes, because children can't pick their birth date
- ▶ But: recent evidence suggest that parents characteristics differ by season of conception/birth (Buckles & Hungerman, 2013; Rietveld & Webbink, 2016; Fan *et al.*, 2017).

Exclusion restriction: does quarter of birth affect earnings only through education?

- ▶ presumably yes
- ▶ but it is possible that people enter the labour market in different seasons...

First Stage: Quarter of Birth and Years of Education

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a]			F-test[b] [P-value]
			I	II	III	
Total years of education	1930–1939	12.79	−0.124	−0.086	−0.015	24.9
			(0.017)	(0.017)	(0.016)	[0.0001]
	1940–1949	13.56	−0.085	−0.035	−0.017	18.6
			(0.012)	(0.012)	(0.011)	[0.0001]
High school graduate	1930–1939	0.77	−0.019	−0.020	−0.004	46.4
			(0.002)	(0.002)	(0.002)	[0.0001]
	1940–1949	0.86	−0.015	−0.012	−0.002	54.4
			(0.001)	(0.001)	(0.001)	[0.0001]
Years of educ. for high school graduates	1930–1939	13.99	−0.004	0.051	0.012	5.9
			(0.014)	(0.014)	(0.014)	[0.0006]
	1940–1949	14.28	0.005	0.043	−0.003	7.8
			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]
Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7
			(0.001)	(0.001)	(0.001)	[0.1599]
	1940–1949	0.11	0.000	0.004	0.001	3.9
			(0.001)	(0.001)	(0.001)	[0.0091]

First Stage: Quarter of Birth and Years of Education

Previous slide: first stage regression results

$$S_i = X\pi_{10} + Z_1\pi_{11} + Z_2\pi_{12} + Z_3\pi_{13} + \eta_1$$

Z_1, Z_2, Z_3 are **quarter of birth dummies**

There ^brf{appears to be a first stage}:

- ▶ children born in Q4 have more schooling than children born in Q1
- ▶ the IV does not affect college graduation (which it shouldn't)

Angrist & Krueger (1991): 2SLS Results

PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 1920–1929[a]

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.1484	5.1574	−0.00898 (0.00301)
Education	11.3996	11.5252	−0.1256 (0.0155)
Wald est. of return to education			0.0715 (0.0219)
OLS return to education[b]			0.0801 (0.0004)

$\widehat{\beta^{OLS}} > \widehat{\beta^{2SLS}}$ as one would expect (?)

Note the much larger standard error of $\widehat{\beta^{2SLS}}$

Angrist & Krueger (1991): Many Many IVs

In their analysis, Angrist & Krueger (1991) use specifications with

- ▶ 30 (quarter-of-birth × year) dummies to **account for cohort effects**
- ▶ 150 (quarter-of-birth × state) dummies to **account for differences across states**

This means that they use up to 150 instruments for education

- ▶ By controlling for state differences, they **reduce bias**
- ▶ But they also **reduce the amount of variation in education** that is used for identification

Low degree of identifying variation ⇒ **weak IV** problem

Bound *et al.* (1995): The Weak Instrument Problem

Causal model: $y = \beta s + \varepsilon$

First stage: $s = \pi z + \eta$

Suppose ε and η are correlated. Estimating β using OLS will be biased:

$$E\big[\widehat{\beta}_{OLS} - \beta\big] = \frac{C(\varepsilon, s)}{V(s)}$$

Bound *et al.* (1995): The Weak Instrument Problem

Bound *et al.* (1995) show that weak instruments bias the 2SLS estimator towards the OLS estimator

One way of expressing the |weak instrument bias} is

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

where F is the **first stage F-statistic** of the instruments in the first stage

▶ Strong instruments: $F \to \infty$, bias $\to 0$
▶ Weak instruments: $F \to 0$, bias $\to \dfrac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2}$

Weak IVs in Angrist & Krueger (1991)

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
F (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, $\times 100$)		.012		.043		.014
F (overidentification)		.932		.775		.725
Age Control Variables						
Age, Age^2	x	x			x	x
9 Year of birth dummies			x	x	x	x
Excluded Instruments						
Quarter of birth		x		x		x
Quarter of birth $\times$ year of birth				x		x
Number of excluded instruments		3		30		28

With **more IVs added** the first stage of the IV gets weaker

Weak IVs in Angrist & Krueger (1991)

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
F (excluded instruments)		2.428		1.869
Partial R^2 (excluded instruments, $\times 100$)		.133		.101
F (overidentification)		.919		.917
Age Control Variables				
Age, Age^2			X	X
9 Year of birth dummies	X	X	X	X
Excluded Instruments				
Quarter of birth		X		X
Quarter of birth $\times$ year of birth		X		X
Quarter of birth $\times$ state of birth		X		X
Number of excluded instruments		180		178

When 180 IVs are included, the first stage is very weak; the IV bias gets close to the OLS bis

Variance of the 2SLS estimator

It can be shown that the asymptotic variance of the 2SLS estimator is

$$\widehat{Avar}\left(\hat{\beta}^{2SLS}\right) = \hat{\sigma}^2 \frac{1}{N\rho_{xz}^2\sigma_x^2},$$

where $\rho_{xz} = \text{cov}\left(z_i, x_i\right) / \left(\sigma_z\sigma_x\right)$.

This equation offers **several important insights**:

▶ An increase in the sample size decreases the standard errors
▶ The standard error is higher the higher the variance of the residuals $\hat{\sigma}^2$ and the lower the variation in x_i
▶ The standard error decreases with the strength of the first stage
▶ Also: $\widehat{Avar}\left(\hat{\beta}^{2SLS}\right) > \widehat{Avar}\left(\hat{\beta}^{OLS}\right)$ because $\rho_{xx} = 1$

Note: we assumed here **homoskedasticity of the error terms**

Simulation: Strong vs. Weak IVs

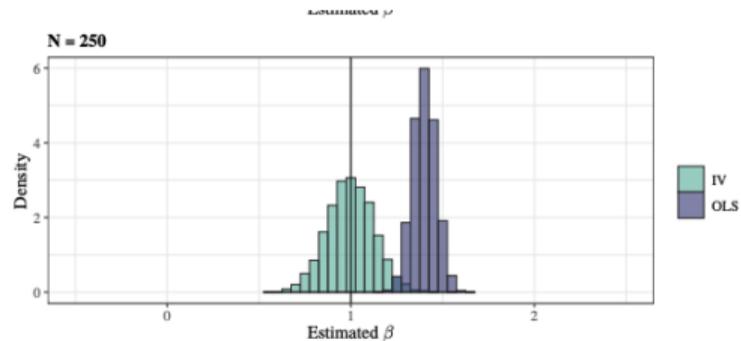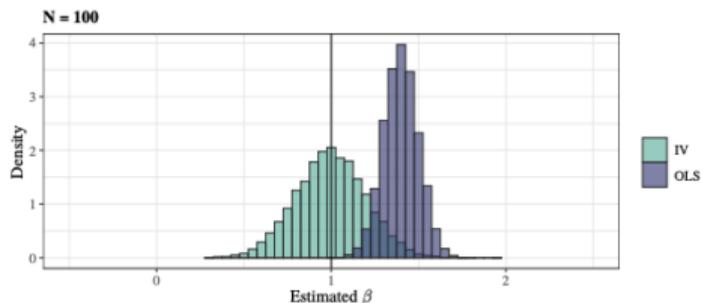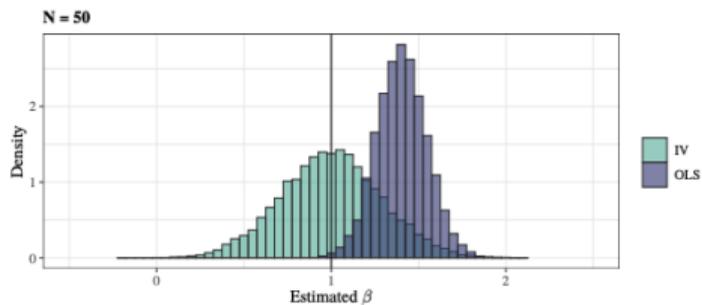We can illustrate the issues with weak IVs in a simulation

$$y = x + \varepsilon$$
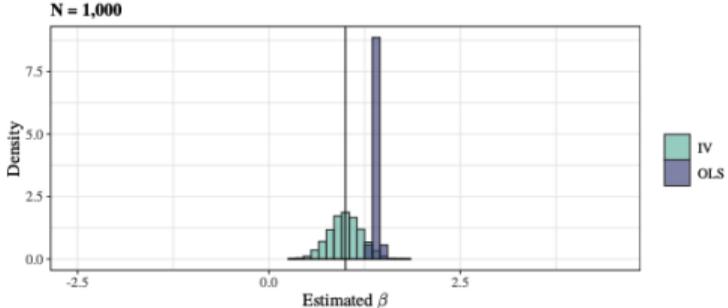$$x = \gamma_1 z + \nu$$
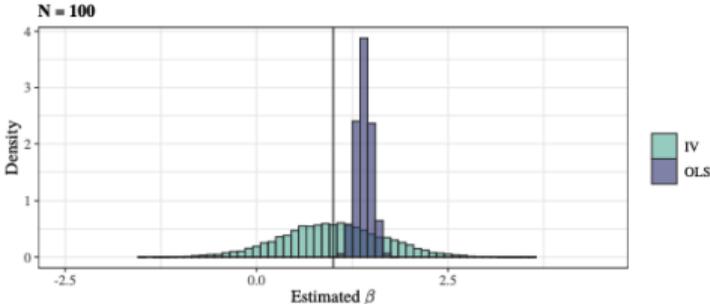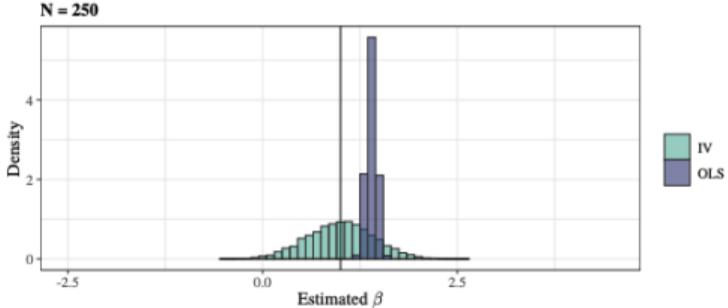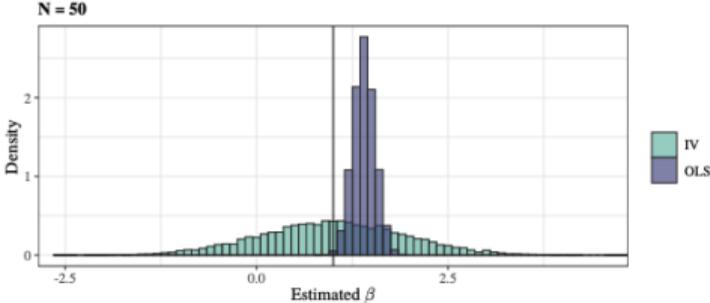$$\rho_{x,\varepsilon} = 0.4$$

- Strong IVs: $\rho_{x,z} = 0.5$
- Weak IVs: $\rho_{x,z} = 0.15$

Simulation: different sample sizes; 10,000 replications

Simulation of Strong IV
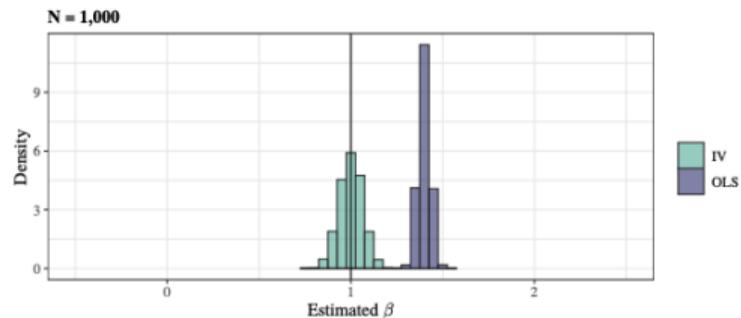
Simulation of Weak IV

Simulation Results

2SLS generally has a **wider sampling distribution** than OLS

If we want to **distinguish $\widehat{\beta}_{2SLS}$ from $\widehat{\beta}_{OLS}$**, we need

- ▶ large samples
- ▶ and a strong first stage

Otherwise we **cannot really distinguish between both estimates**; (biased) OLS estimator may be preferable

Weak Instruments - What to Do?

Show the **F-Statistic of the first stage**

- ▶ Stock *et al.* (2002) suggest that an F-Statistic > 10 indicates that the instruments are sufficiently strong
- ▶ But this is a rule of thumb, nothing more; nowadays, people say 10 is too small

Best solution: find a **better instrument**

Alternatives:

- ▶ use **LIML** (Limited Information Maximum Likelihood) instead of 2SLS
- ▶ report Anderson-Rubin confidence intervals that account for weak IVs

Where Do Good IVs Come from?

Theory combined with clever datas collection. Examples

- ▶ Distance from job training centers
- ▶ College openings

Variation in policies. This requires a **deep institutional knowledge**. Examples

- ▶ assignment to judges with different severity
- ▶ differences in budgets across job training centers
- ▶ ...

Nature. Examples

- ▶ different levels of pollution in different places
- ▶ sex of the first two children
- ▶ ...

IV: Cookbook

1) Explain your identification strategy very clearly

- ▶ start with the **ideal experiment**; why is your setting different? Why is your **regressor endogenous**?
- ▶ Explain theoretically **why there should be a first stage** and what coefficient we should expect
- ▶ Explain why the instrument is **as good as randomly assigned**
- ▶ Explain theoretically **why the exclusion restriction holds** in your setting

2) Show and discuss the first stage

- ▶ Best to start with a **raw correlation**
- ▶ Do the **sign and magnitude make sense**?
- ▶ Assess the **strength of the instrument** using state-of-the-art techniques

IV: Cookbook

3) Bring supportive evidence for instrument validity

- ▶ Show that the **instrument does not predict pre-treatment characteristics**
- ▶ Can you provide evidence in support of the exclusion restriction?
- ▶ Use auxiliary tests, for example Kitagawa (2015) and Huber & Mellace (2015)
- ▶ Consider using the *plausibly exogenous* bounding procedure by Conley *et al.* (2012)

4) Discuss the results in detail

- ▶ Show the **OLS and 2SLS results**, both with **varying sets of controls**
- ▶ Comment on the differences between both (bias, LATE, etc)
- ▶ Show the **reduced form**
- ▶ If the reduced form isn't there, the effect isn't there (MHE)

Instrumental Variables: Conclusion

IV is a **powerful approach to deal with endogeneity**

The **bar for finding a credible instrument is high**

- ▶ Exclusion restriction cannot be tested
- ▶ Defending an IV requires deep knowledge of institutions and context

For **canonical IV designs**, see the Mixtape, Section 7.8.

APPENDIX

How to do IV using R

Classic example: **Card (1995)'s study on returns to higher education**

▶ Uses distance |birthplace - nearest college| as an IV
▶ This is obviously questionable, but serves as a good example

There are **two main packages for IV in R**

▶ `AER` (Applied Econometrics with R) and the `ivreg` command
▶ `fixest` and the `feols` command; this is very useful for IV estimation with FE

How to do IV using R

Loading in packages and data; `haven` is for reading datasets in non-R format

```r
library(AER)
library(haven)
library(tidyverse)
library(modelsummary)

read_data <- function(df)
{
  full_path <- paste("https://github.com/scunning1975/mixtape/raw/master/"
                     df, sep = "")
  df <- read_dta(full_path)
  return(df)
}

card <- read_data("card.dta")
```

How to do IV using R

Prep data and run OLS

```r
attach(card)

Y1 <- lwage
Y2 <- educ
X1 <- cbind(exper, black, south, married, smsa)
X2 <- nearc4

#OLS
ols_reg <- lm(Y1 ~ Y2 + X1)
```

How to do IV using R

	(1)
(Intercept)	5.063
	(0.064)
Y2	0.071
	(0.003)
X1exper	0.034
	(0.002)
X1black	−0.166
	(0.018)
X1south	−0.132
	(0.015)
X1married	−0.036
	(0.003)
X1smsa	0.176
	(0.015)
Num.Obs.	3003
R2	0.305
R2 Adj.	0.304
AIC	2563.7

How to do IV using R

OLS would yield a return to education of 7%. Let's see what IV gives us

```
#2SLS
# Notice how we need to include all exogenous variables behind the "|"
iv_reg <- ivreg(Y1 ~ Y2 + X1 | X1 + X2)
```

How to do IV using R: First Stage

```
#2SLS
# Check the first stage
firststage <-  lm(Y2 ~ X1 + X2)
models <- list(ols_reg, firststage, iv_reg)
names(models) <- c("OLS", "First", "2SLS")
```

How to do IV using R: 2SLS estimates

	OLS	First	2SLS
(Intercept)	5.063***	16.831***	4.162***
	(0.064)	(0.131)	(0.850)
Y2	0.071***		0.124*
	(0.003)		(0.050)
X1exper	0.034***	−0.404***	0.056**
	(0.002)	(0.009)	(0.020)
X1black	−0.166***	−0.948***	−0.116*
	(0.018)	(0.091)	(0.051)
X1south	−0.132***	−0.297***	−0.113***
	(0.015)	(0.079)	(0.023)
X1married	−0.036***	−0.073***	−0.032***
	(0.003)	(0.018)	(0.005)
X1smsa	0.176***	0.421***	0.148***
	(0.015)	(0.085)	(0.031)
X2		0.327***	
		(0.082)	
Num.Obs.	3003	3003	3003
R2	0.305	0.477	0.251

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

References

Angrist, Joshua, & Evans, William. 1998. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, **88**(3), 450–77.

Angrist, Joshua D. 1990. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. **80**(3), 313–336.

Angrist, Joshua D., & Krueger, Alan B. 1991. Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, **106**(4), 979–1014.

Angrist, Joshua D., Imbens, Guido W., & Rublin, Donald B. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, **91**(434), 444–455.

Bound, John, Jaeger, David A., & Baker, Regina M. 1995. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association*, **90**(430), 443–450.

Buckles, Kasey S., & Hungerman, Daniel M. 2013. SEASON OF BIRTH AND LATER OUTCOMES: OLD QUESTIONS, NEW ANSWERS. *The Review of Economics and Statistics*, **95**(3), 711–724.

Card, David E. 1995. Using geographic variation in college proximity to estimate the return to schooling. *In:* Christofides, L., Grant, E. Kenneth, & Swindinsky, Robert (eds), *Aspects of Labour Economics: Essays in Honour of John Vanderkamp*. Toronto, Canada: University of Toronto Press.

Conley, Timothy G., Hansen, Christian B., & Rossi, Peter E. 2012. Plausibly Exogenous. **94**(1), 260–272.

Fan, Elliott, Liu, Jin-Tan, & Chen, Yen-Chien. 2017. Is the Quarter of Birth Endogenous? New Evidence from Taiwan, the US, and Indonesia. *Oxford Bulletin of Economics and Statistics*, **79**(6), 1087–1124.

Huber, Martin, & Mellace, Giovanni. 2015. Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints. *The Review of Economics and Statistics*, **97**(2), 398–411.

Kitagawa, Toru. 2015. A Test for Instrument Validity. *Econometrica*, **83**(5), 2043–2063.

Kling, Jeffrey R., Liebman, Jeffrey B., & Katz, Lawrence F. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica*, **75**(1), 83–119.

Mogstad, Magne, & Torgovitsky, Alexander. 2018. Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, **10**(1), 577–613.

Rietveld, Cornelius A., & Webbink, Dinand. 2016. On the genetic bias of the quarter of birth instrument. *Economics & Human Biology*, **21**, 137–146.

Stock, James H., Wright, Jonathan H., & Yogo, Motohiro. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, **20**(4), 518–529.

benjamin.elsner@ucd.ie

www.benjaminelsner.com

Sign up for office hours

YouTube Channel

@ben_elsner

LinkedIn

Contact

Prof. Benjamin Elsner
University College Dublin
School of Economics
Newman Building, Office G206
benjamin.elsner@ucd.ie

Office hours: book on Calendly