Non-Experimental Data: IV Advanced Topics

Rajeev Dehejia

Local Average Treatment Effect

- Let T_{1i} and T_{0i} denote potential outcomes for T_i depending on whether the instrument is assigned or not (that is, the opportunity for treatment T is assigned, or not). We can express the observed outcomes as:

$$T_i = Z_i T_{1i} + (1 - Z_i) T_{0i}$$

- There are four possibilities for each individual:

$T_{0i} = 0$, $T_{1i} = 0$ **"Never-Takers"** Will not take up treatment if regardless of assignment status	$T_{0i} = 1$, $T_{1i} = 1$ **"Always-Takers"** Will take up treatment regardless of assignment status
$T_{0i} = 1$, $T_{1i} = 0$ **"Defiers"** Will not take up treatment if assignment to treatment (and vice-versa)	$T_{0i} = 0$, $T_{1i} = 1$ **"Compliers"** Will take up treatment if assigned, will not take up treatment if not assigned

Local Average Treatment Effect (cont'd)

$$T_i = Z_i T_{1i} + (1 - Z_i) T_{0i}$$

- Assume:
  - **Independence:** differences in Z are not correlated with differences in T_{1i} , T_{0i} , Y_{1i}, and Y_{0i}.
  - **T and Z are correlated** (first stage).
  - **Monotonicity:** either $T_{1i} > T_{0i}$ or vice versa. Monotonicity narrows the set of possible outcomes for T_i from 4 to 3. "Defiers" are ruled out.

LATE

$$E(Y_i \mid Z_i = 1) = E[(1-T_i)Y_{0i} + T_i Y_{1i} \mid Z_i = 1]$$
$$= E[Y_{0i} + (Y_{1i} - Y_{0i})T_i \mid Z_i = 1]$$
$$= E[Y_{0i} + (Y_{1i} - Y_{0i})T_{1i}]$$

(by random assignment | Z not needed)

$$E(Y_i \mid Z_i = 0) = E[Y_{0i} + (Y_{1i} - Y_{0i})T_{0i}]$$

$$E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)$$

$$= E[Y_{0i} + (Y_{1i} - Y_{0i})T_{1i}] - E[Y_{0i} + (Y_{1i} - Y_{0i})T_{0i}]$$

$$= E[(Y_{1i} - Y_{0i})(T_{1i} - T_{0i})]$$

But note $T_1 - T_0$ takes on only 2 values

(0 and1, with -1 ruled out by monotonicity)

$$= (1)\Pr(T_{1i} - T_{0i} = 1)E(Y_{1i} - Y_{0i} \mid T_{1i} - T_{0i} = 1)$$
$$+ (0)\Pr(T_{1i} - T_{0i} = 0)E(Y_{1i} - Y_{0i} \mid T_{1i} - T_{0i} = 0)$$

LATE

$$E(Y_{1i} - Y_{0i} \mid T_{1i} - T_{0i} = 1) = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{\Pr(T_{1i} - T_{0i} = 1)}$$

$$E(Y_{1i} - Y_{0i} \mid T_{1i} - T_{0i} = 1) = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(T_i \mid Z_i = 1) - E(T_i \mid Z_i = 0)}$$

IV with Heterogeneous Treatment Effects and Continuous Instrument

- If treatment effect same for everyone then TOT recovers this (obvious)
- But what if treatment effect heterogeneous?
- No simple answer to this question
- Suppose model for treatment effect is:

$$y_i = \beta_0 + \beta_{1i} T_i + \varepsilon_i$$

Proposition

Proposition

The IV estimate for the heterogeneous treatment case is a consistent estimate of:

$$p\lim \hat{\beta}_{1,IV} = \frac{E\left(\pi_i \beta_{1i}\right)}{E\left(\pi_i\right)}$$

where:

$$\pi_i = \Pr\left(T_i = 1 \middle| Z_i = 1\right) - \Pr\left(T_i = 1 \middle| Z_i = 0\right)$$

the difference in the probability of treatment for individual *i* when in treatment and control group

Sketch of the idea

- Model for effect of intention to treat on being treated:

$$p_{1i} = \Pr\left(T_i = 1 \middle| Z_i = 1\right), \qquad p_{0i} = \Pr\left(T_i = 1 \middle| Z_i = 0\right),$$

$$\pi_i = p_{1i} - p_{0i}$$

$$T_i = p_{0i} + \pi_i Z_i + v_i$$

$$E\left(v_i \middle| Z_i\right) = 0$$

Idea (continued)

- Can write ‘reduced-form’ as:

$$y_i = \beta_0 + \beta_{1i}\left(p_{0i} + \pi_i Z_i\right) + \varepsilon_i + \beta_{1i}\nu_i$$

- Wald estimator then becomes:

$$\text{plim}\hat{\beta}_{1IV} = \frac{E\left(y_i \middle| Z_i = 1\right) - E\left(y_i \middle| Z_i = 0\right)}{E\left(T_i \middle| Z_i = 1\right) - E\left(T_i \middle| Z_i = 0\right)} = \frac{E\left(\pi_i \beta_{1i}\right)}{E\left(\pi_i\right)}$$

- As:

$$E\left(y_i \middle| Z_i = 1\right) = \beta_0 + E\left(\beta_{1i} p_{0i}\right) + E\left(\beta_{1i}\pi_i\right)$$

$$E\left(y_i \middle| Z_i = 0\right) = \beta_0 + E\left(\beta_{1i} p_{0i}\right)$$

$$E\left(T_i \middle| Z_i = 1\right) = E\left(p_{0i}\right) + E\left(\pi_i\right)$$

$$E\left(T_i \middle| Z_i = 0\right) = E\left(p_{0i}\right)$$

Hence Wald estimator can be thought of as estimator as:

$$\hat{\beta}_{1IV} = \frac{E(y_i|Z_i = 1) - E(y_i|Z_i = 0)}{E(T_i|Z_i = 1) - E(T_i|Z_i = 0)} = \frac{E(\pi_i \beta_{1i})}{E(\pi_i)}$$

- This is a weighted average of treatment effects
- ‘Weights’ will vary with instrument
- Some cases in which can interpret IV estimate as ATE

The IV estimate is ATE if
(a) no heterogeneity in treatment effect
(b) β_{1i} uncorrelated with π_i

Idea:

A. This is immediate as:

$$\hat{\beta}_{1,IV} = \frac{E\left(\pi_i \beta_{1i}\right)}{E\left(\pi_i\right)} = \frac{E\left(\pi_i \beta_1\right)}{E\left(\pi_i\right)} = \frac{\beta_1 E\left(\pi_i\right)}{E\left(\pi_i\right)} = \beta_1$$

B. Can write as:

$$\hat{\beta}_{1,IV} = \frac{E\left(\pi_i \beta_{1i}\right)}{E\left(\pi_i\right)} = \frac{Cov\left(\pi_i, \beta_{1i}\right) + E\left(\pi_i\right) E\left(\beta_{1i}\right)}{E\left(\pi_i\right)} = E\left(\beta_{1i}\right) + \frac{Cov\left(\pi_i, \beta_{1i}\right)}{E\left(\pi_i\right)}$$

How will IV estimate differ from ATE

- Previous formula says depends on covariance of β_{1i} and π_i
- In some situations can sign – but not always
- Example 1: no-one gets treatment in the absence of the program so

$$\pi_i = p_{1i}$$

- If those who get treatment when in the treatment group are those with the highest returns then:

$$Cov\left(\pi_i, \beta_{1i}\right) = Cov\left(p_{1i}, \beta_{1i}\right) > 0$$

- IV>ATE

How will IV differs from ATE: Example 2

- Treatment is voluntary for those in the control group but compulsory for those in the treatment group
- This implies

$$\pi_i = 1 - p_{0i}$$

- If those who get treatment in control are those with highest returns then:

$$Cov\left(\pi_i, \beta_{1i}\right) = -Cov\left(p_{0i}, \beta_{1i}\right) < 0$$

- IV<ATE

Identifying compliers

the kappa function

$$E(x|D_1^Z - D_0^Z > 0) = \frac{E(\kappa^Z(x)x)}{E(\kappa^Z(x))}$$

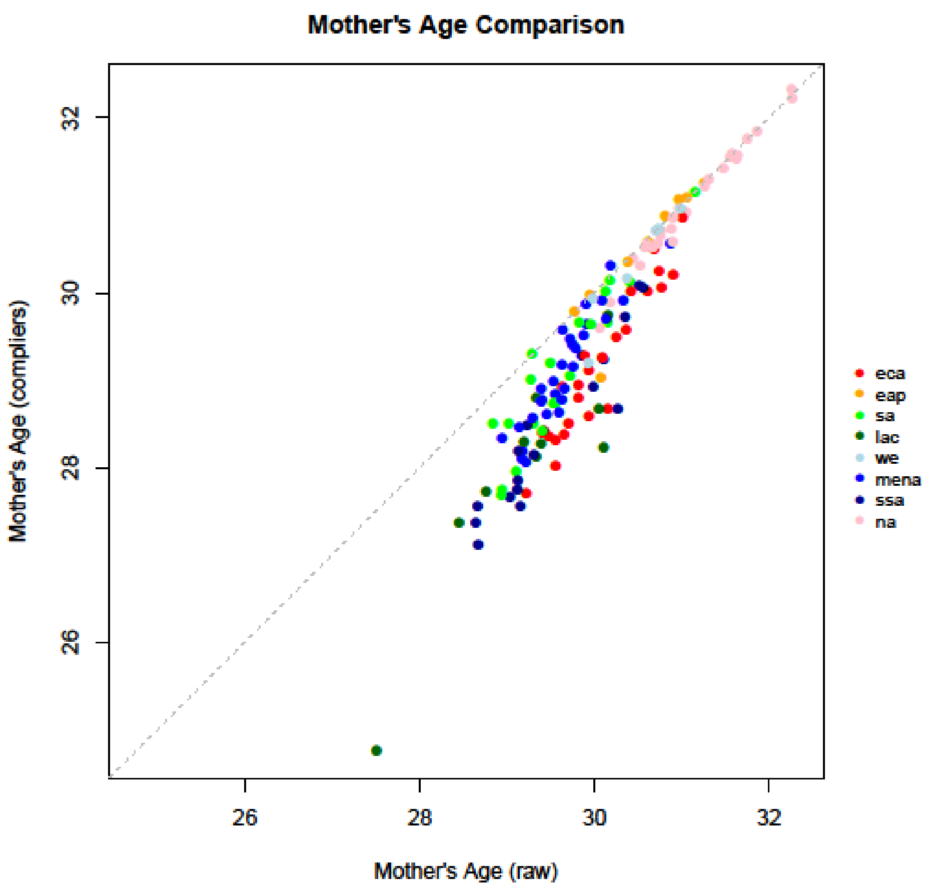
$$\kappa^Z(x) = 1 - \frac{D_i(1 - Z_i)}{1 - \Pr(Z_i = 1|X_i)} - \frac{(1 - D_i)Z_i}{\Pr(Z_i = 1|X_i)}$$

The kappa function “finds” compliers.

Intuition: by monotonicity, *Z*=0 & *D*=1 are always-takers and *Z*=1 & *D*=0 are never takers. So by subtracting out these two groups you are left with compliers.

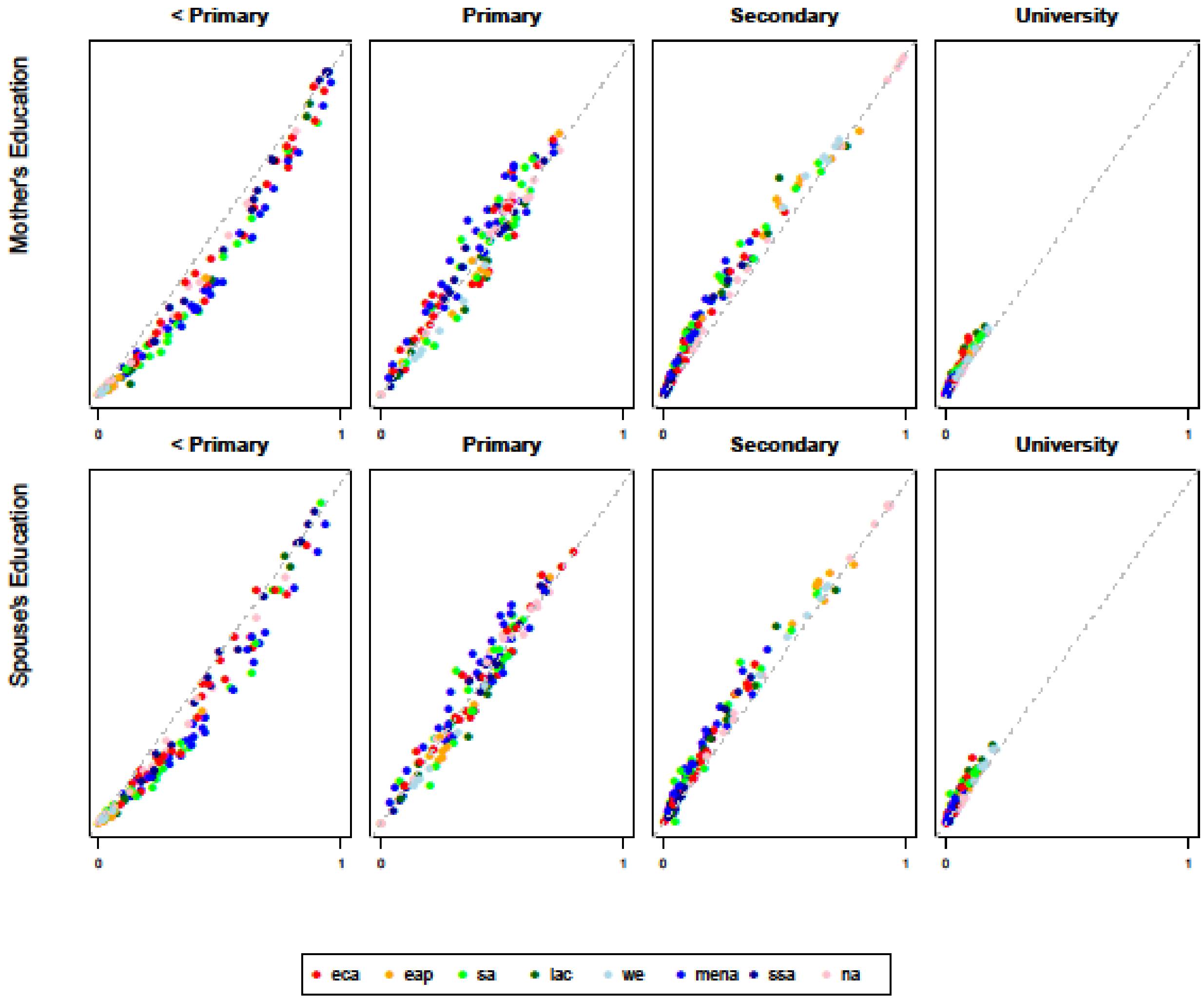
Example: a world of LATES

first stage compliers for same-sex on more kids



Compliers are younger than the full population.

Example: a world of LATES
first stage compliers for same-sex on more kids



Compliers and their spouses have less primary education, more secondary and tertiary.

The problem of under-identification

- Recall key assumption: instrument is correlated with the endogenous variable.
- But this assumption can actually be tested.
- In simplest case look at t or F-test statistic of the instrument in the first stage.
  - Generally, F-stat for the joint significance of the instrument(s) in the first stage should be >10
- There are specialized statistics for more advanced cases (e.g., multiple instruments), but same intuition.

The problem of weak instruments

- Say that instruments are ‘weak’ if correlation between *T* and *Z* low (after inclusion of other exogenous variables)
- Here *T* refers generically to any endogenous variable – could be a treatment or could be other variables as well

Why do weak instruments matter?

- A whole range of problems tend to arise if instruments are weak
- Asymptotic problems:
  - High asymptotic variance
  - Small departures from instrument exogeneity lead to big inconsistencies
- Finite-Sample Problems:
  - Small-sample distribution may be very different from asymptotic one
    - May be large bias
    - Computed variance may be wrong
    - Distribution may be very different from normal

Example

Small departures from instrument exclusion lead to big inconsistencies

- Suppose true causal model is

$$y = X\beta + Z\gamma + \varepsilon$$

So there is possibly a direct effect of *Z* on *y*.

- Instrument exclusion is $\gamma = 0$.
- Obviously want this to be zero but might hope that no big problem if 'close to zero' – a small deviation from the exclusion restriction
- But this isn't true.

An example:
The return to education

- Economists long-interested in whether investment in human capital a ‘good’ investment
- Some theory shows that coefficient on *s* in regression:

$$y = \beta_0 + \beta_1 s + \gamma x + \varepsilon$$

  is measure of rate of return to education
- OLS estimates around 8% - suggests very good investment
- Might be liquidity constraints
- Might be bias

Potential sources of bias

- Most commonly mentioned is ‘ability bias’
- Ability correlated with earnings independent of education
- Ability correlated with education
- If ability omitted from ‘X’ variables then usual formula for omitted variables bias suggests upward bias in OLS estimate

Potential solution

- Find an instrument correlated with education but uncorrelated with ‘ability’ (or other excluded variables)
- Angrist-Krueger “Does Compulsory Schooling Attendance Affect Schooling and Earnings”, QJE 1991, suggest using quarter of birth
- Argue correlated with education because of school start age policies and school leaving laws (instrument relevance)
- Don’t have to accept this – can test it

A graphical version of first-stage (correlation between education and Z)

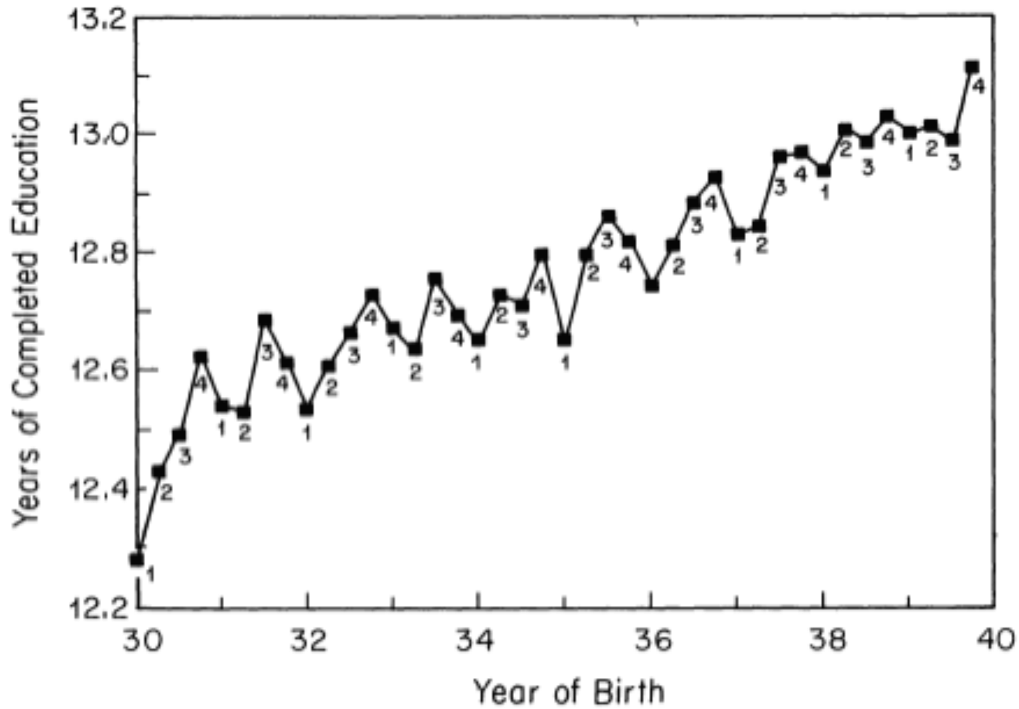


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

In this case…

- Their instrument is binary so IV estimator can be written in Wald form
- And this leads to following expression for potential inconsistency:

$$p\lim \hat{\beta}^{IV} = \frac{E(y|Z=1)-E(y|Z=0)}{E(T|Z=1)-E(T|Z=0)} = \beta + \frac{\gamma}{E(T|Z=1)-E(T|Z=0)}$$

- Note denominator is difference in schooling for those born in first- and other quarters
- Instrument will be ‘weak’ if this difference is small

Their results

TABLE III

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)

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(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)	5.8916	5.9027	−0.01110 (0.00274)
Education	12.6881	12.7969	−0.1088 (0.0132)
Wald est. of return to education			0.1020 (0.0239)
OLS return to education			0.0709 (0.0003)

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

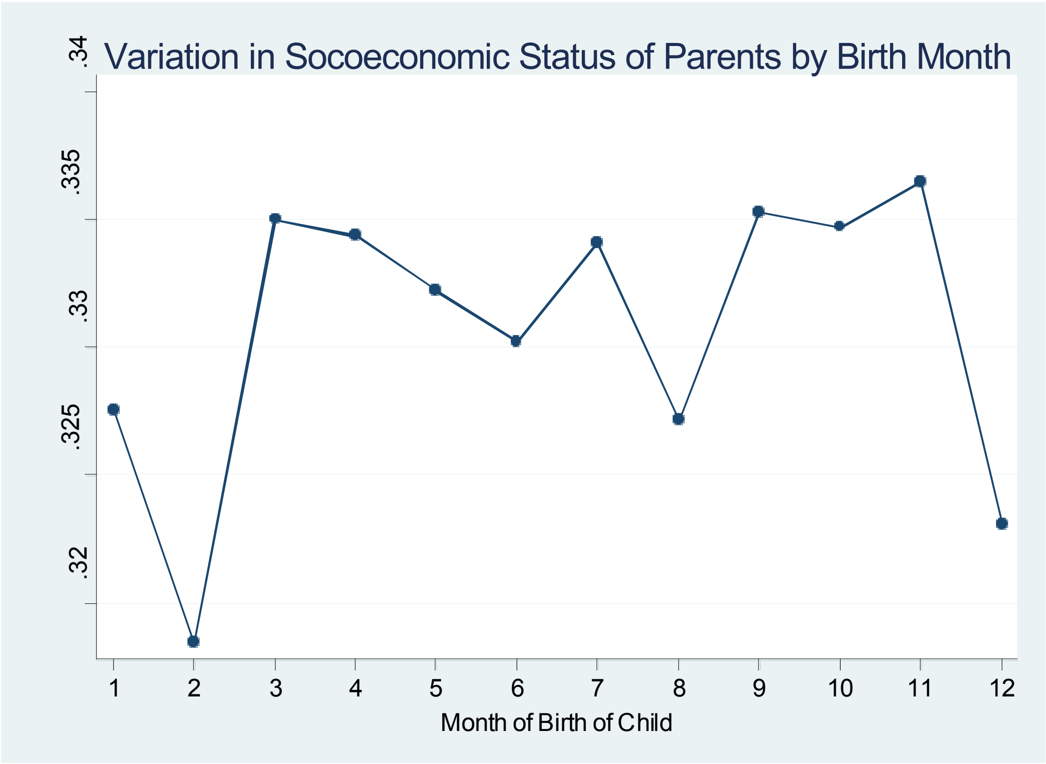
Interpretation (and potential criticism)

- IV estimates not much below OLS estimates (higher in one case)
- Suggests 'ability bias' no big deal
- But instrument is weak
- Being born in 1^{st} quarter reduces education by 0.1 years
- Means γ will be multiplied by 10

But why should we have $\gamma \neq 0$?

- Remember this would imply a direct effect of quarter of birth on earnings, not just one that works through the effect on education
- Bound, Jaeger and Baker argued that evidence that quarter of birth correlated with:
  - Mental and physical health
  - Socioeconomic status of parents
- Unlikely that any effects are large but don't have to be when instruments are weak

An example: UK data



Fraction of children <10 where head of HH is manager or professional. Effect is small but significantly different from zero

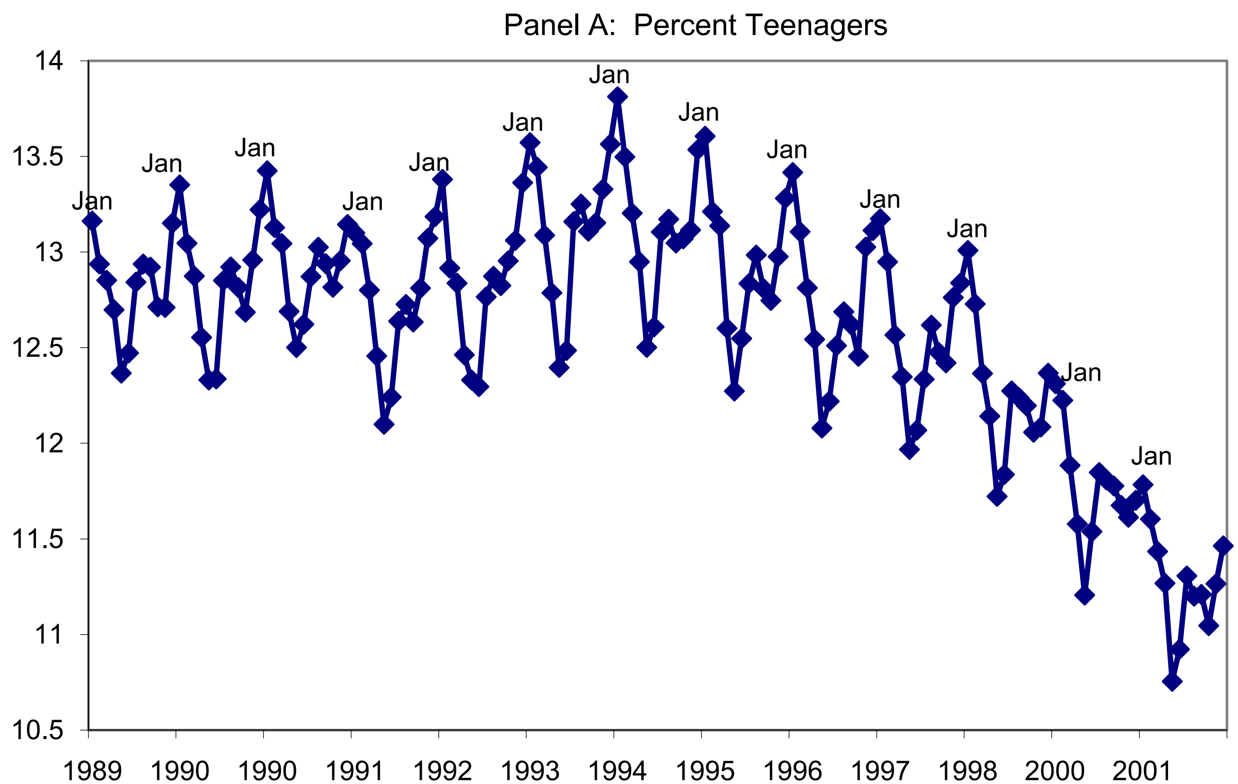
A back-of-the-envelope calculation

- Being born in first quarter means 0.01 less likely to have a managerial/professional parent
- Being a manager/professional raises earnings by 64%
- Correlation between earnings of children and parents 0.4
- Effect on earnings through this route 0.01*0.64*0.4=0.00256 i.e. ¼ of 1 per cent
- Small but weak instrument causes effect on inconsistency of IV estimate to be multiplied by 10 – 0.0256
- Now large relative to OLS estimate of 0.08

Example from the US

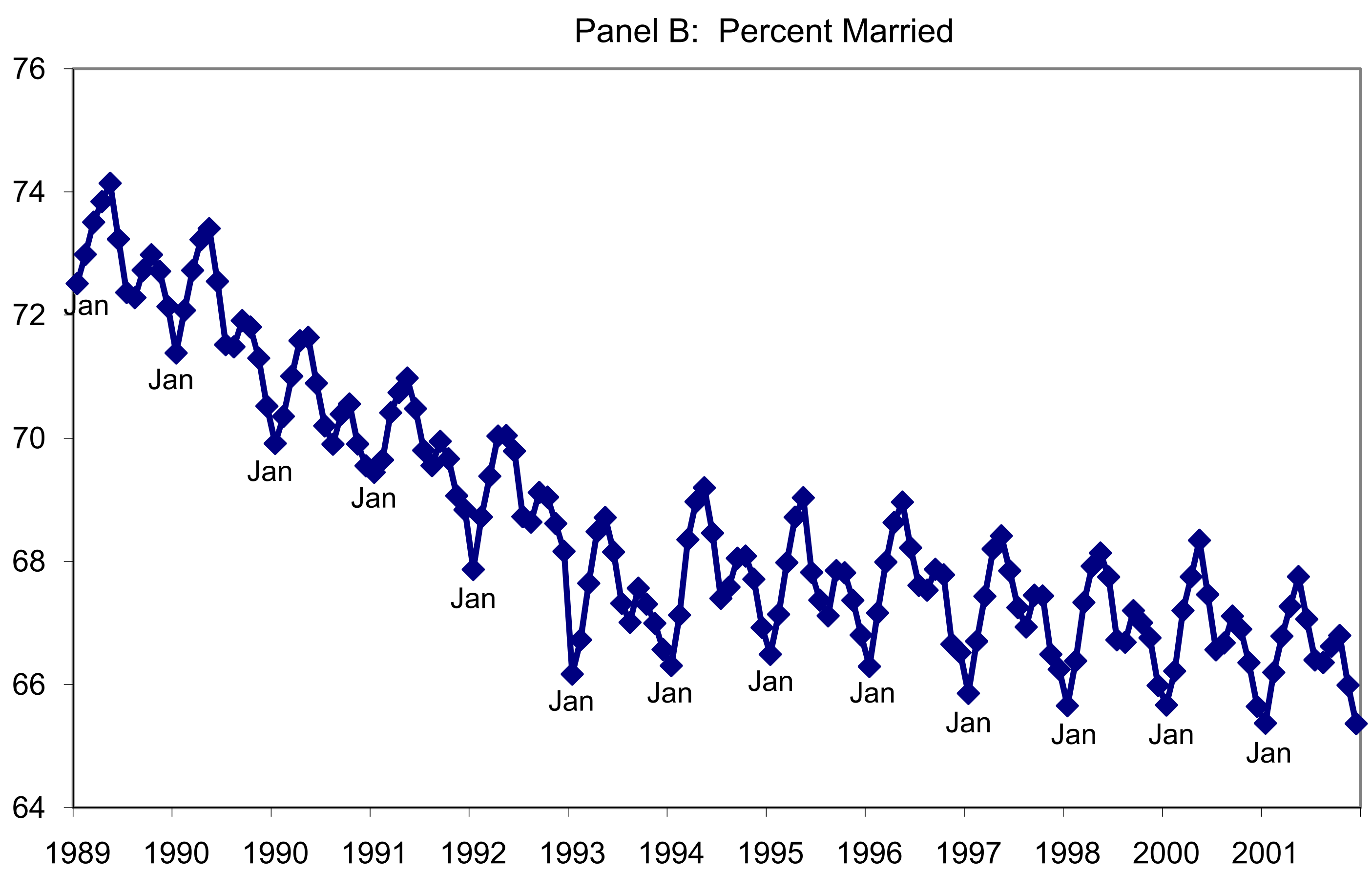
Buckles and Hungerman

FIGURE 1. MATERNAL CHARACTERISTICS BY MONTH, NATALITY FILES, 1989-2001



Example from the US

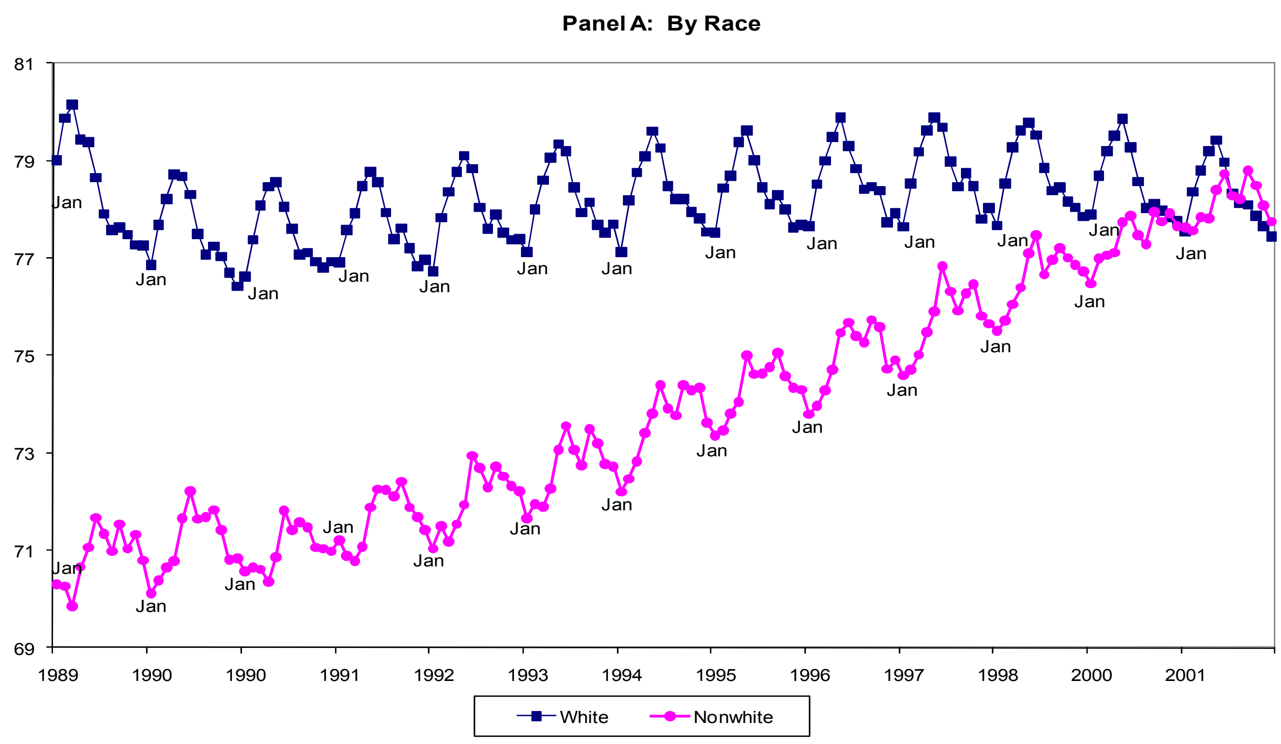
Buckles and Hungerman



Example from the US

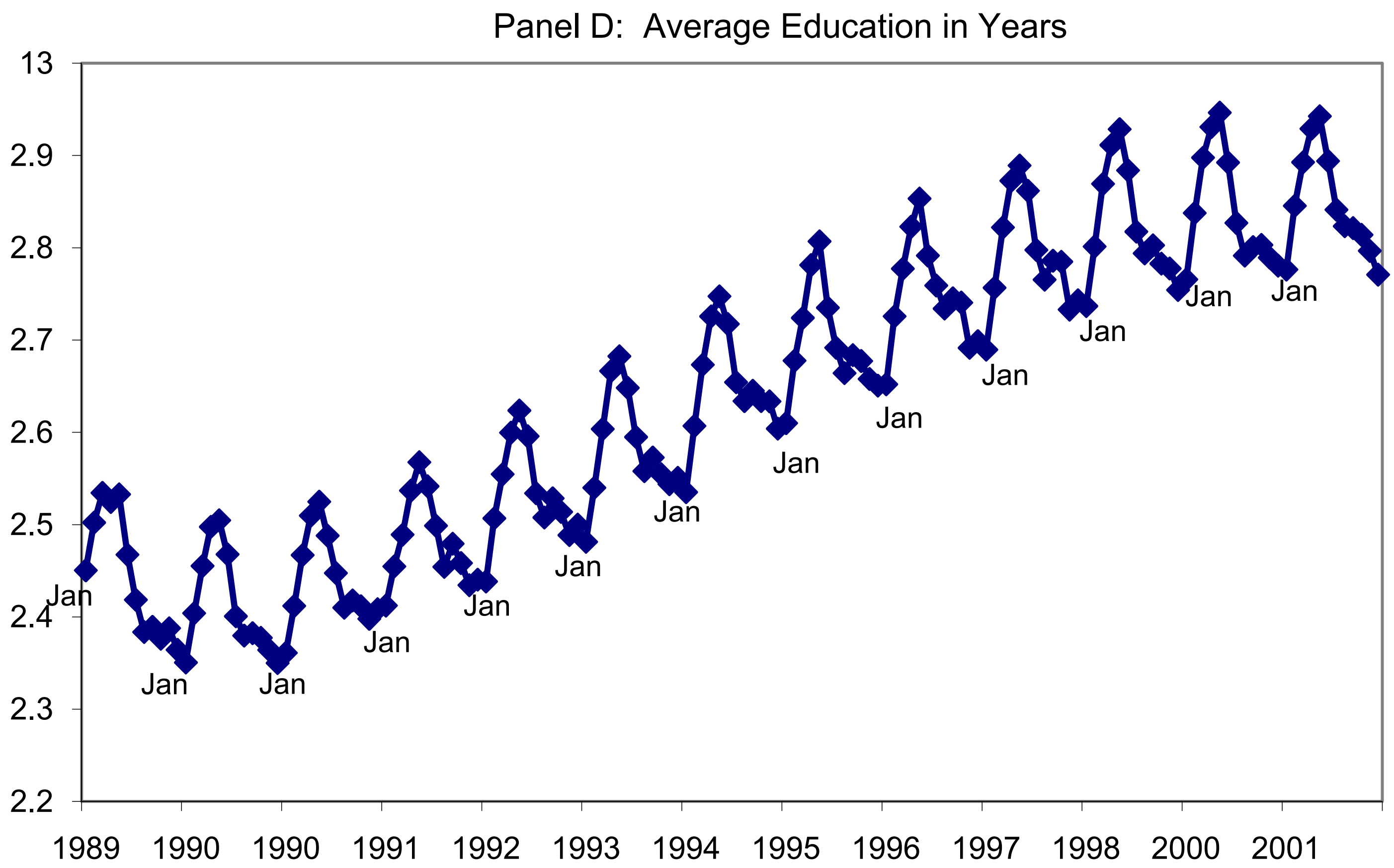
Buckles and Hungerman

FIGURE 2. PERCENT OF WOMEN GIVING BIRTH EACH MONTH WHO HAVE A HIGH SCHOOL DEGREE, NATALITY FILES, 1989-2001



Example from the US

Buckles and Hungerman



Summary

- Small deviations from instrument exogeneity lead to big inconsistencies in IV estimate if instruments are weak
- Suspect this is often of great practical importance
- Quite common to use ‘odd’ instrument – argue that ‘no reason to believe’ it is correlated with ε but show correlation with T

Review and reminder

- If ask STATA to estimate equation by IV
- Coefficients computed using formula given
- Standard errors computed using formula for asymptotic variance
- t-statistics, confidence intervals and p-values computed using assumption that estimator is unbiased with variance as computed and normally distributed
- All are asymptotic results

Difference between asymptotic and finite-sample distributions

- This is the normal case
- Only in special cases e.g. linear regression model with normally distributed errors are small-sample and asymptotic distributions the same.
- Difference likely to be bigger
  - The smaller the sample size
  - The weaker the instruments

Finite sample problems

- This is a very complicated topic
- Exact results for special cases, approximations for more general cases
- Hard to say anything that is definitely true but can give useful guidance
- Will divide problems into 3 areas
  - Bias
  - Incorrect measurement of variance
  - Non-normal distribution
- But really all different symptoms of same thing

Some intuition for why strength of instruments is important

- Consider very strong instrument
- *Z* can explain a lot of variation in *T*
- *T* very close to *T*-hat
- Think of limiting case where correlation perfect – then *T*-hat=*T*
- IV estimator identical to OLS estimator
- Will have same distribution
- If errors normal then this is same as asymptotic distribution

Now consider case of weak instrument…

- Think of extreme case where true correlation between *T* and *Z* is useless
- First-stage tries to find some correlation so estimate of coefficients will not normally be zero and will have some variation in *T*-hat
- No reason to believe *T*-hat contains more ‘good’ variation than *T* itself
- So central tendency is OLS estimate
- But a lot more noise – so very big variance

Finite sample problems 2: Bias: exogenous but weak instruments

Even when assumptions are perfectly met, IV is *not unbiased* in small (finite) samples

Finite sample bias can be non-negligible (e.g., 20 - 30%), even when the sample size is over 100,000 if the instrument is weak (*Z* is only weakly correlated with *T*)

The relative bias of b_{IV} (versus b_{OLS}) is approximately 1/*F* where *F* is the *F*-statistic for testing the relation between the instrument (*Z*) and endogenous variable (*T*)

A small value of *F*, even if it is large enough to be statistically significant signals possible large bias in b_{IV}

Testing for weak instruments

It is *not* sufficient that the relation between *Z* and *T* is statistically significant

Need to test whether F-stat exceeds a special threshold (below which instruments are weak enough to imperil inference)

Two definitions of 'weak enough to imperil inference,' and both can be tested with first stage *F* for relation of *Z* and *T* (Stock & Yogo, 2005):

1. Squared bias of b_{IV} exceeds 10% of the squared bias of b_{OLS}

   E.g., requires $F > 10$

2. Actual level of 5% significance test is at most 15% (i.e., a 5% test falsely rejects no more than 15% of the time).

   E.g., requires $F > 8.96$

   * This is what is given in ivreg2

Exogenous, but weak instruments

- Exact (small sample) results are available, but *very* complex (almost to the point of being uninformative)

- In general, more instruments increases the relevance of the instrument set (increases the first stage F)

- But, too many instruments increases small sample bias (compared to few instruments)

- In general it is best to have as few instruments as possible, and for them to be strongly correlated with X (the endogenous variable)

Practical guidance

- Whenever you use IV you must worry about the problems of under-identification and weak instruments.
- Under-identification can be tested in the first stage using standard tests.
- Weak instruments can also be tested in the first stage, using the extended Stock-Yogo critical values.

Conclusion

- Natural experiments useful source of knowledge
- Often requires use of IV
- Instrument exogeneity and relevance need justification
- Weak instruments potentially serious
- Good practice to present first-stage regression
- Finding more robust alternative to IV an active research area

Additional Resources

- Kurt Shmidheiny (Unversität Basel)'s IV Chapter in "Short Guides to Microeconometrics" (2019); great for theory and stata application
- For stata demonstrations and more theory explanation, check out Econometrics Academy's page on IV
- The Urban Institute has a very simple one-page explainer on IV for very succinct overview
- #EconTwitter … Prof. Nick Huntington-Klein has an animated IV graphic of correlations between Z, X, and Y (in a few combos)

What we will cover in recitation

- ✓ Essential skills and questions regarding PS2
- ✓ Additional explanation of notation, theory, or validity tests for IV
- ✓ Discussion on papers if time allows