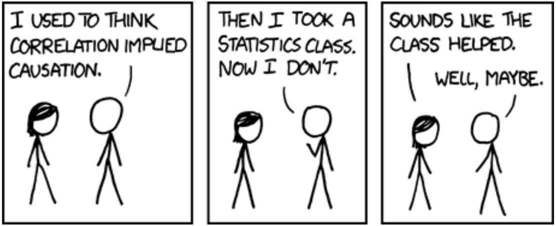
Advanced Applied Econometrics

Teacher: Felix Weinhardt



Slides heavily borrow from Scott Cunningham and Paul Goldsmith-Pinkham

Introduction to instrumental variables

- Pearl argues that there are three ways to estimate causal effects: the backdoor criterion ("selection on observables"), instrumental variables, and the front door criterion.
- Now we move into the instrumental variables research design
- Illustrious history – discovered by Philip Wright and published as an appendix in his 1928 book
- One of the most powerful methods for identifying causal effects in the social sciences

- Sometimes, the researcher *can* find the flag. That is, the researcher knows of a variable (Z) that actually *is* randomly assigned and that affects fertility decisions. Such a variable is called an "instrument".
- Example: Angrist and Evans (1998), "Children and their parents' labor supply" *American Economic Review*,
  - Z is a dummy variable indicating whether the first two children born were of the same gender
  - Many parents have a preference for having at least one child of each gender
    - Consider a couple whose first two kids were both boys; they will often have a third, hoping to have a girl
    - Consider a couple whose first two kids were girls; they will often have a third, hoping for a boy
    - Consider a couple with one boy and one girls; they will often not have a third kid
  - The gender of your kids is arguably randomly assigned (maybe not exactly, but close enough)

- “No causation without manipulation” (Holland, 1985). If you want to use IV, then ask:

  What moves around the covariate of interest that might be plausibly viewed as random?

- “Focusing on the selection process”
  - How was the covariate of interest selected?
  - Do researchers select randomly, as in drawing balls from an urn?
  - Or do subjects choose the covariate partially based on random factors and partially based on non-random factors?
  - If the random factors underlying the decision can be *observed by the researcher*, then you can use IV to estimate the effect of the covariate of interest on the outcome
- Angrist and Evans (1988) example:
  - Families with at least two kids are the subject population
  - The covariate of interest is the number of kids (`numkids`).
  - There are several outcomes of interest measuring labor market decisions of the couple, such as whether the mother worked for pay in the last year (`workforpay`).
- Once you have identified such a variable, begin to think about what data sets might have information on an outcome of interest, the covariate of interest, and the instrument you have put your finger on.

- In a pinch, you can even get by with two different data sets, one of which has information on the outcome and the instrument, and the other of which has information on the covariate of interest and the instrument.
- This is known as "Two sample IV" because there are two *samples* involved, rather than the traditional one sample.
- Once we define what IV is measuring carefully, you will see why this works.

- Instrumental variables strategies formalize *untainted inference*, which is the inference drawn by an intelligent layperson with no particular training or background in statistics.
- Example of what I mean by "untainted inference":
  - The researchers tell a layperson that the gender of a woman's first two children is predictive of her labor force attachment. In particular, women whose first two children are of the same gender work substantially less than women whose first two children are of different genders
  - On its face, this is a puzzling fact – without further information, it is hard to see why the gender of your children would be so predictive of labor market participation
  - The researchers additionally point out that women whose first two children are of the same gender are more likely to have additional children than women whose first two children are of different genders
  - The layperson then wonders whether the labor market differences are due *solely* to the differences in the number of kids the woman has

Traditional and Contemporary IV Pedagogy

We want to learn the IV framework in two iterations:

1. Constant treatment effects (i.e., β is constant across all individual units)
   - Constant treatment effects is the traditional econometric pedagogy when first learning instrumental variables, and is not based explicitly on the potential outcomes model or notation
   - Constant treatment effects is identical to assuming that ATE=ATT=ATU because constant treatment effects assumes $\beta_i = \beta_{-i} = \beta$ for all units
2. Heterogeneous treatment effects (i.e., β_i varies across individual units)
   - This is the "modern IV pedagogy", and you may not have learned it in econometrics if only because potential outcomes is not ordinarily the basis of most first year econometrics sequences
   - Heterogeneous treatment effects means that the $ATE \neq ATT \neq ATU$ because β_i differs across the population
   - This is equivalent to assuming the coefficient, β_i, is a random variable that varies across the population
   - Heterogenous treatment effects is based on work by Angrist, Imbens and Rubin (1996) and Imbens and Angrist (1994) which introduced the "local average treatment effect" (LATE) concept

When should you think of using instrumental variables?

- Instrumental variables methods are typically used to address the following kinds of problems encountered in OLS regressions:
  1. Omitted variable bias
  2. Measurement error
  3. Simultaneity bias, or "reverse causality"

Omitted Variable Bias (Angrist and Pischke, 2009)

- Labor economists have been studying the returns to schooling a long time – typically some version of a "Mincer regression":

$$Y_i = \alpha + \rho S_i + \gamma A_i + \nu_i$$

$$\begin{aligned} Y_i &= \text{log of earnings} \\ S_i &= \text{schooling measured in years} \\ A_i &= \text{individual ability} \end{aligned}$$

- Typically the econometrician cannot observe A_i; for instance, the CPS tells us nothing about adult respondents' family background, intelligence, or motivation.
- What are the consequences of leaving ability out of the regression? Suppose you estimated this short regression instead:

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

where $\eta_i = \gamma A_i + \nu_i$; α, ρ, and γ are population regression coefficients; S_i is correlated with η_i through A_i only; and v_i is a regression residual uncorrelated with all regressors by definition.

Derivation of Ability Bias

- Suppressing the i subscripts, the OLS estimator for ρ is:

$$\widehat{\rho} = \frac{Cov(Y,S)}{Var(S)} = \frac{E[YS] - E[Y]E[S]}{Var(S)}$$

- Plugging in the true model for Y, we get:

$$\begin{aligned}
\widehat{\rho} &= \frac{Cov[(\alpha + \rho S + \gamma A + \nu), S]}{Var(S)} \\
&= \frac{E[(\alpha S + S^2\rho + \gamma SA + \nu S)] - E(S)E[\alpha + \rho S + A\gamma + \nu]}{Var(S)} \\
&= \frac{\rho E(S^2) - \rho E(S)^2 + \gamma E(AS) - \gamma E(S)E(A) + E(\nu S) - E(S)E(\nu)}{Var(S)} \\
&= \rho + \gamma\frac{Cov(AS)}{Var(S)}
\end{aligned}$$

- If $\gamma > 0$ and Cov$(A,S) > 0$ the coefficient on schooling in the shortened regression (without controlling for A) would be upward biased

How can IV be used to obtain unbiased estimates?

- Suppose there exists a variable, Z_i, that is correlated with S_i .
- We can estimate ρ with this variable, Z:

$$
\begin{aligned}
Cov(Y,Z) &= Cov(\alpha+\rho S+\gamma A+\nu, Z) \\
&= E[(\alpha+\rho S+\gamma A+\nu),Z] - E[\alpha+\rho S+\gamma A+\nu]E[Z] \\
&= \{\alpha E(Z)-\alpha E(Z)\}+\rho\{E(SZ)-E(S)E(Z)\} \\
&\quad +\gamma\{E(AZ)-E(A)E(Z)\}+E(\nu Z)-E(\nu)E(Z) \\
Cov(Y,Z) &= \rho Cov(S,Z)+\gamma Cov(A,Z)+Cov(\nu,Z)
\end{aligned}
$$

- What conditions must hold for a valid instrumental variable?
  - $Cov(S,Z)\neq 0$ – "first stage" exists. S and Z are correlated
  - $Cov(A,Z)=Cov(\nu,Z)=0$ – "exclusion restriction". Z is orthogonal to the factors in η, such as unobserved ability or the structural disturbance term, η
- Assuming the first stage exists and that the exclusion restriction holds, then we can estimate ρ with ρ_{IV}:

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

IV is Consistent if IV Assumptions are Satisfied

- The IV estimator is consistent if the IV assumptions are satisfied. Substitute true model for Y:

$$
\begin{aligned}
\rho_{IV} &= \frac{Cov([\alpha + \rho S + \gamma A + \nu], Z)}{Cov(S, Z)} \\
&= \rho \frac{Cov([S], Z)}{Cov(S, Z)} + \gamma \frac{Cov([A], Z)}{Cov(S, Z)} + \frac{Cov([\nu], Z)}{Cov(S, Z)} \\
&= \rho + \gamma \frac{Cov(\eta, Z)}{Cov(S, Z)}
\end{aligned}
$$

- Taking the plim:

$$\text{plim } \widehat{\rho}_{IV} = \rho$$

because $Cov([A], Z) = 0$ and $Cov([\nu], Z) = 0$ due to the exclusion restriction, and $Cov(S, Z) \neq 0$ (due to the first stage)

- But, if Z is *not* independent of η (either correlated with A or ν), *and* if the correlation between S and Z is "weak", then the second term blows up. We will return to this later when we discuss the problem of "weak instruments".

In which DAG is Z a valid instrument for D?

$$
\begin{tikzcd}
Z \arrow[dr] & & e \arrow[dl, dashed, leftrightarrow] \arrow[dr, dashed, leftrightarrow] & \\
& D \arrow[rr] & & Y
\end{tikzcd}
$$

(a)

$$
\begin{tikzcd}
Z \arrow[dr] \arrow[rr, dashed, leftrightarrow] & & e \arrow[dl, dashed, leftrightarrow] \arrow[dr, dashed, leftrightarrow] & \\
& D \arrow[rr] & & Y
\end{tikzcd}
$$

(b)

Reviewing some of the IV Jargon

- Causal model. Sometimes called the structural model:

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

- First-stage regression. Gets the name because of two-stage least squares:

$$S_i = \alpha + \rho Z_i + \zeta_i$$

- Second-stage regression. Notice the fitted values, $\widehat{S}$:

$$Y_i = \alpha + \rho \widehat{S}_i + \nu_i$$

- Reduced form. Notice this is just a regression of Y onto the instrument:

$$Y_i = \alpha + \pi Z_i + \varepsilon_i$$

Two-stages least squares

- Suppose you have a sample of data on Y, X, and Z. For each observation i we assume the data are generated according to

$$
\begin{aligned}
Y_i &= \alpha + \beta X_i + \varepsilon_i \\
X_i &= \gamma + \delta Z_i + \nu_i
\end{aligned}
$$

  where $Cov(Z, \varepsilon) = 0$ and $\delta \neq 0$.
- Plug in covariance, and using the result that $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, write out the IV estimator:

$$
\widehat{\beta_{IV}} = \frac{Cov(Z,Y)}{Cov(Z,X)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(Y_i - \overline{Y})}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})Y_i}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})X_i}
$$

- Substitute the causal model definition of Y to get:

$$
\begin{aligned}
\widehat{\beta_{IV}} &= \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})\{\alpha + \beta X + \varepsilon\}}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})X_i} \\
&= \beta + \frac{\frac{1}{n}(Z_i - \overline{Z})\varepsilon_i}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z})X_i} \\
&= \beta + \text{"small if } n \text{ is large"}
\end{aligned}
$$

Two-stages least squares

- Note β_{IV} is ratio of "reduced form" (π) to "first stage" coefficient (δ):

$$\widehat{\beta}_{IV} = \frac{Cov(Z,Y)}{Cov(Z,X)} = \frac{\frac{Cov(Z,Y)}{Var(Z)}}{\frac{Cov(Z,X)}{Var(Z)}} = \frac{\widehat{\pi}}{\widehat{\delta}}$$

- Rewrite $\widehat{\delta}$ as

$$\widehat{\delta} = \frac{Cov(Z,X)}{Var(Z)} \Leftrightarrow Cov(Z,X) = \widehat{\delta}Var(Z) \tag{2}$$

- Then rewrite β_{IV}

$$\begin{aligned} \widehat{\beta_{IV}} &= \frac{Cov(Z,Y)}{Cov(Z,X)} = \frac{\widehat{\delta}Cov(Z,Y)}{\widehat{\delta}Cov(Z,X)} = \frac{\widehat{\delta}Cov(Z,Y)}{\widehat{\delta}^2 Var(Z)} \\ &= \frac{Cov(\widehat{\delta}Z,Y)}{Var(\widehat{\delta}Z)} \end{aligned} \tag{3}$$

Two-stage least squares

- Recall $X = \gamma + \delta Z + \nu$; $\widehat{\beta_{IV}} = \frac{Cov(\widehat{\delta}Z,Y)}{Var(\widehat{\delta}Z)}$ and let $\widehat{X} = \widehat{\gamma} + \widehat{\delta}Z$.
- Then the two-stage least squares (2SLS) estimator is

$$\widehat{\beta_{IV}} = \frac{Cov(\widehat{\delta}Z, Y)}{Var(\widehat{\delta}Z)} = \frac{Cov(\widehat{X}, Y)}{Var(\widehat{X})}$$

Proof.

We will show that $\widehat{\delta}Cov(Y,Z) = Cov(\widehat{X}, Y)$. I will leave it to you to show that $Var(\widehat{\delta}Z) = Var(\widehat{X})$

$$\begin{aligned}
Cov(\widehat{X}, Y) &= E[\widehat{X}Y] - E[\widehat{X}]E[Y] \\
&= E(Y[\widehat{\gamma} + \widehat{\delta}Z]) - E(Y)E(\widehat{\gamma} + \widehat{\delta}Z) \\
&= \widehat{\gamma}E(Y) + \widehat{\delta}E(YZ) - \widehat{\gamma}E(Y) - \widehat{\delta}E(Y)E(Z) \\
&= \widehat{\delta}[E(YZ) - E(Y)E(Z)] \\
Cov(\widehat{X}, Y) &= \widehat{\delta}Cov(Y,Z)
\end{aligned}$$

□

- The 2SLS estimator replaces X with the fitted values of X (i.e., $\widehat{X}$) from the first stage regression of X onto Z and all other covariates.

Intuition of 2SLS

- I've said that learning about instrumental variables through the "intuition" of two-stage least squares is valuable, but what do I mean exactly?
- 2SLS emphasizes the use of the fitted values of the endogenous regressor – what has that transformation done?
- By using the fitted values of the endogenous regressor from the first stage regression, our regression now uses *only* the exogenous variation in the regressor due to the instrumental variable itself
- We recover exogenous variation in other words
- ... but think about it – that variation was there before, but was just a subset of all the variation in the regressor
- Instrumental variables therefore reduces the variation in the data, but that variation which is left is *exogenous*

Implementing 2SLS in STATA

- In a sample of data, you could get the reduced form and first stage coefficients manually by the following two regression commands in STATA:

```
. reg Y Z
```

```
. reg X Z
```

- While it is always a good idea to run these two regressions, don't compute your IV estimate this way
  - Example: It is often the case that a pattern of missing data will differ between Y and X; in such a case, the usual procedure of "casewise deletion" is to keep the subsample with non-missing data on Y, X, and Z.
  - But the reduced form and first stage regressions would be estimated off of different sub-samples if you used the two step method above
  - The standard errors from the second stage regression are also wrong
- Best practice is to use your built-in procedure (which also gives standard errors):

```
. ivregress Y (X=Z)
```

Implementing 2SLS in STATA

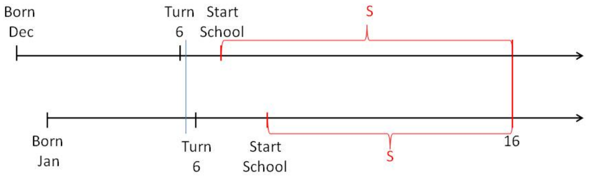
- You can also estimate 2SLS using the auxiliary regression approach we just covered:

```
. reg X Z
. predict Xhat
. reg Y Xhat
```

- For the same reasons that you shouldn't actually implement 2SLS manually using the ratio of the reduced form and first stage coefficients, you shouldn't manually use the auxiliary regression approach because, again, the standard errors are incorrect, and any complex missing patterns may leave you with different samples
- This "two stage least squares" interpretation of IV – called an *interpretation*, because it is not the actual suggested procedure – is useful for understanding what IV does, but stick with `ivregress`

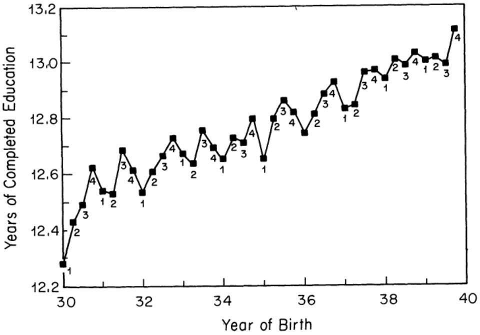
Instrument for Education using Compulsory Schooling Laws

- In practice, it is often difficult to find convincing instruments – usually because potential instruments don't satisfy the exclusion restriction
- In the returns to education literature, Angrist and Krueger (1991) had a very influential study where they used quarter of birth as an instrumental variable for schooling
- In the US, you could drop out of school once you turned 16
- "School districts typically require a student to have turned age six by January 1 of the year in which he or she enters school" (Angrist and Krueger 1991, p. 980)
- Children have different ages when they start school, though, and this creates different lengths of schooling at the time they turn 16 (potential drop out age):



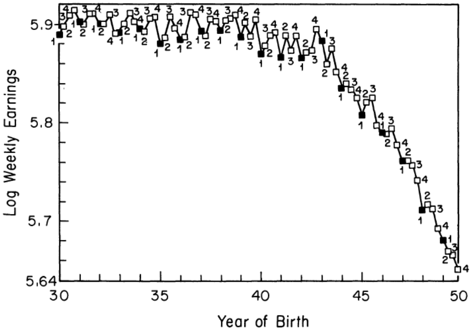
First Stage

- Men born earlier in the year have lower schooling. This indicates that there is a first stage.



Reduced Form

- Do differences in schooling due to different quarter of birth translate into different earnings?



Two Stage Least Squares model

- The first stage regression is:

$$S_i = X\pi_{10} + \pi_{11} Z_i + \eta_{1i}$$

- The reduced form regression is:

$$Y_i = X\pi_{20} + \pi_{21} Z_i + \eta_{2i}$$

- The covariate adjusted IV estimator is the sample analog of the ratio, $\frac{\pi_{21}}{\pi_{11}}$
- Again, how was this estimator calculated?
    1. Obtain the first stage fitted values:

    $$\widehat{S}_i = X\widehat{\pi}_{10} + \widehat{\pi}_{11} Z_i$$

    where $\widehat{\pi}_{1j}$ for $j = 1, 2$ are OLS estimates of the first stage regression
    2. Plug the first stage fitted values into the "second-stage equation" to then estimate

    $$Y_i = \alpha X + \widehat{S}\rho + \text{error}$$

Two Stage Least Squares

- But as we note, they don't actually manually do this – I remind you of this because the 2SLS intuition is very useful. 2SLS only retains the variation in S generated by the quasi-experimental variation, which we hope is exogenous
- Angrist and Krueger use more than one instrumental variable to instrument for schooling: they include a dummy for each quarter of birth. Their estimated first-stage regression is therefore:

$$S_i = X\pi_{10} + Z_{1i}\pi_{11} + Z_{2i}\pi_{12} + Z_{3i}\pi_{13} + \eta_1$$

- The second stage is the same as before, but the fitted values are from the new first stage

First stage regression results

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a] I	II	III	F-test[b] [P-value]
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]

First stage regression results

- Quarter of birth is a strong predictor of total years of education

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

First stage regression results: Placebos

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]
Completed doctoral degree	1930–1939	0.03	0.002	0.003	0.000	2.9
			(0.001)	(0.001)	(0.001)	[0.0332]
	1940–1949	0.04	−0.002	0.001	−0.001	4.3
			(0.001)	(0.001)	(0.001)	[0.0050]

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. F-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

IV Estimates Birth Cohorts 20-29, 1980 Census

Independent variable	(1) OLS	(2) TSLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)
Race (1 = black)	—	—
SMSA (1 = center city)	—	—
Married (1 = married)	—	—
9 Year-of-birth dummies	Yes	Yes
8 Region-of-residence dummies	No	No
Age	—	—
Age-squared	—	—
χ^2 [dof]	—	25.4 [29]

IV Estimates - including some covariates

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760
	(0.0003)	(0.0161)	(0.0003)	(0.0290)
Race (1 = black)	—	—	—	—
SMSA (1 = center city)	—	—	—	—
Married (1 = married)	—	—	—	—
9 Year-of-birth dummies	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No
Age	—	—	−0.0772	−0.0801
			(0.0621)	(0.0645)
Age-squared	—	—	0.0008	0.0008
			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]

Wald estimator

- They also present an alternative to 2SLS called the Wald estimator – both are versions of instrumental variables
- Recall that 2SLS uses the predicted values from a first stage regression – but we showed that the 2SLS method was equivalent to $\frac{Cov(Y,Z)}{Cov(X,Z)}$
- The Wald estimator simply calculates the return to education as the ratio of the difference in earnings by quarter of birth to the difference in years of education by quarter of birth – it's a version of the above
- Formally, $IV_{Wald} = \frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)}$

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)

IV Estimates - more covariates and interacting quarter of birth

- They also include specifications where they use 30 (quarter of birth $\times$ year) dummy variables and 150 (quarter of birth $\times$ state) dummies as instrumental variables
  - What's the intuition here? The effect of quarter of birth may vary by birth year or by state
- It reduced the standard errors, but that comes at a cost of potentially having a weak instruments problem

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS
Years of education	0.0673 (0.0003)	0.0928 (0.0093)	0.0673 (0.0003)	0.0907 (0.0107)
Race (1 = black)	—	—	—	—
SMSA (1 = center city)	—	—	—	—
Married (1 = married)	—	—	—	—
9 Year-of-birth dummies	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No
50 State-of-birth dummies	Yes	Yes	Yes	Yes
Age	—	—	−0.0757 (0.0617)	−0.0880 (0.0624)
Age-squared	—	—	0.0008 (0.0007)	0.0009 (0.0007)

Mechanism

- In addition to log weekly wage, they examined the impact of compulsory schooling on log annual salary and weeks worked
- The main impact of compulsory schooling is on the log weekly wage – not on weeks worked

Weak Instruments

- As we mentioned earlier, IV is consistent but biased
- For a long time, researchers were not attentive to this subtle difference and didn't care much about the small sample bias of IV
- But in the early 1990s, a number of papers highlighted that IV can be *severely* biased – in particular, when instruments have only a weak correlation with the endogenous variable of interest and when many instruments are used to instrument for one endogenous variable (i.e., there are many overidentifying restrictions).
- In the worst case, if the instruments are so weak that there is no first stage, then the 2SLS sampling distribution is centered on the probability limit of OLS

Weak instruments and bias towards OLS

- Let's consider a model with a single endogenous regressor and a simple constant treatment effect
- The causal model of interest is:

$$y = \beta x + \nu$$

- The matrix of instrumental variables is Z with the first stage equation:

$$x = \mathbf{Z'}\pi + \eta$$

- If ν_i and η_i are correlated, estimating the first equation by OLS would lead to biased results, wherein the OLS bias is:

$$E[\beta_{OLS} - \beta] = \frac{Cov(\nu, x)}{Var(x)}$$

- If ν_i and η_i are correlated the OLS bias is therefore: $\frac{\sigma_{\nu\eta}}{\sigma_x^2}$

Weak instruments and bias towards OLS

- It can be shown that the bias of 2SLS is approximately:

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

  where F is the population analogue of the F-statistic for the joint significance of the instruments in the first stage regression. See Angrist and Pischke pp. 206-208 for a derivation.
- If the first stage is weak (i.e, $F \to 0$), then the bias of 2SLS approaches $\frac{\sigma_{\nu\eta}}{\sigma_\eta^2}$
- This is the same as the OLS bias as for $\pi = 0$ in the second equation on the earlier slide (i.e., there is no first stage relationship between Z and D) $\sigma_x^2 = \sigma_\eta^2$ and therefore the OLS bias $\frac{\sigma_{\nu\eta}}{\sigma_x^2}$ becomes $\frac{\sigma_{\nu\eta}}{\sigma_x^2}$.
- But if the first stage is very strong ($F \to \infty$) then the IV bias goes to 0.

Weak Instruments - Adding More Instruments

- Adding more weak instruments will increase the bias of 2SLS
  - By adding further instruments without predictive power, the first stage F-statistic goes toward zero and the bias increases
- If the model is "just identified" – mean the same number of instrumental variables as there are endogenous covariates – weak instrument bias is less of a problem
  - See Angrist and Pischke, p. 209 where they write that IV is "approximately biased" – this is only true if the first stage is not zero
  - See `http://econ.lse.ac.uk/staff/spischke/mhe/josh/solon_justid_April14.pdf`
- Bound, Jaeger and Baker (1995) highlighted this problem for the Angrist and Krueger study. AK present findings from using different sets of instruments
  1. Quarter of birth dummies $\rightarrow$ 3 instruments
  2. Quarter of birth dummies + (quarter of birth) $\times$ (year of birth) + (quarter of birth) $\times$ (state of birth) $\rightarrow$ 180 instruments

Adding instruments in Angrist and Krueger

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)
F (excluded instruments)		13.486		4.747
Partial R^2 (excluded instruments, ×100)		.012		.043
F (overidentification)		.932		.775
	Age Control Variables			
Age, Age^2	x	x		
9 Year of birth dummies			x	x
	Excluded Instruments			
Quarter of birth		x		x
Quarter of birth × year of birth				x
Number of excluded instruments		3		30

- Adding more weak instruments reduced the first stage F-statistic and moves the coefficient towards the OLS coefficient

Adding instruments in Angrist and Krueger

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

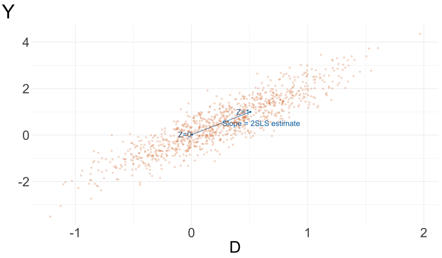
- Adding more weak instruments reduced the first stage F-statistic and moves the coefficient towards the OLS coefficient

What can you do if you have weak instruments?

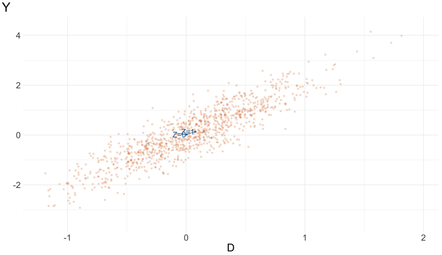
- WIth weak instruments, you have the following options:
  1. Use a just identified model with your strongest IV
  2. Proposition from the Mixed Tape textbook: use a limited information maximum likelihood estimator (LIML). This is approximately median unbiased for over identified constant effects models. It provides the same asymptotic distribution as 2SLS (under constant effects) but provides a finite-sample bias reduction. (LIML is programmed for STATA in the `ivregress` command.)
  3. Find stronger instruments.
  4. Currently there is an active econometric literature that looks deeper into the issue of weak instruments. We will now touch on this a little bit.

Lets look at some simulated data for the intuition before looking at possible solutions

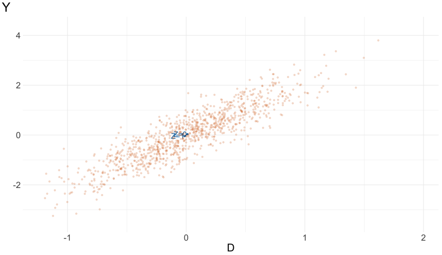
- Simple 2SLS simulation, with binary instrument
  - First stage coef = 0.5, true beta = 2
- Note that the estimation on the x-axis comes from variation in the first stage
- The larger this is, the stronger the first stage
- However, if the first stage is weak, this interval is quite short, even if the variation in D stays the same



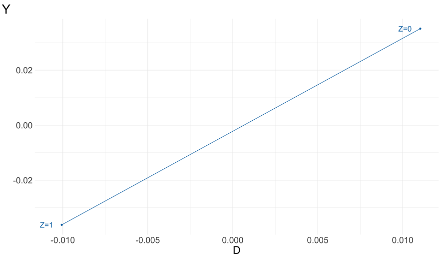
- With a first stage coefficient of 0.1, it becomes hard to distinguish the points
    - Note: I hold fixed the overall variance of D here to keep the correct comparison!
- Given that the model is correctly specified, with enough data it should converge to the right β
- But small shifts in the x-axis will massively swing the estimate!



- With a first stage coefficient of 0.01, the problem is even worse
- We see that the relevant variation being exploited is tiny
- A small change in the x-axis points would even flip the sign!
- What does that do to our estimation procedure?



- With a first stage coefficient of 0.01, the problem is even worse
- We see that the relevant variation being exploited is tiny
- A small change in the x-axis points would even flip the sign!
- What does that do to our estimation procedure?



- This graphical intuition should guide your understanding of the statistical problem
- For simplicity, assume the following: variables are demeaned (mean zero) and there are no additional controls (e.g. no constant). Hence,

$$
\begin{aligned}
Y_i &= D_i\beta + \epsilon_i \\
D_i &= Z_i\pi + u_i \\
\rightarrow Y_i &= Z_i \underbrace{\pi\beta}_{\delta} + u_i\beta + \epsilon_i
\end{aligned}
$$

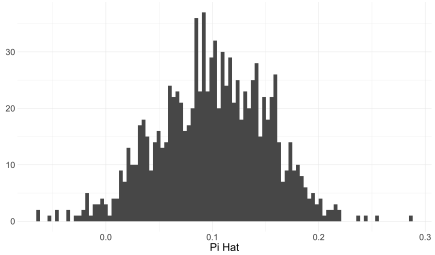
- The 2SLS estimator (for single endog. variable) can then be written as:

$$\hat{\beta}_{2SLS} = \frac{D'Z(Z'Z)^{-1}Z'Y}{D'Z(Z'Z)^{-1}Z'D} = \frac{D'P_Z P_Z' Y}{D'P_Z P_Z D} = \frac{\hat{D}'\hat{Y}}{\hat{D}'\hat{D}} = \frac{\hat{\pi}'\hat{Q}\hat{\delta}}{\hat{\pi}'\hat{Q}\hat{\pi}} = \underbrace{\frac{\hat{\delta}}{\hat{\pi}}}_{\text{Single Instrum}},$$

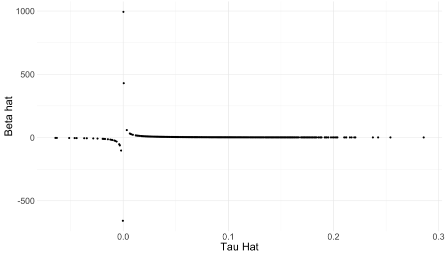
- Intuitively, the 2SLS estimate is just the ratio of the reduced form and the first stage
  - This ratio can be highly non-linear with the denominator
- Notice that under traditional asymptotic approximations, the small value for π is not a big deal.
  - Given a large enough sample, $\hat{\pi} \to \pi$, and you will consistently estimate β
- That's not really what we want to approximate though
  - In a finite sample, $\hat{\pi}$ is noisy, and if the s.e. of $\hat{\pi}$ is large relative to $\hat{\pi}$, that can cause very weird behavior in $\hat{\beta}$

$$\hat{\beta}_{2SLS} = \frac{\hat{\delta}}{\hat{\pi}}$$

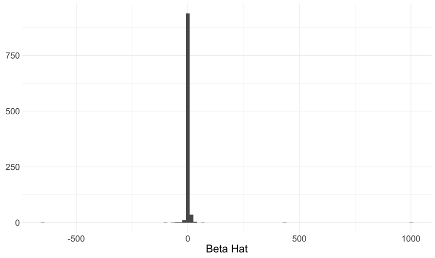
- Imagine a marginally significant first stage (se $= 0.05$, estimate $= 0.1$)
- This estimator is normal, and reasonably well-behaved



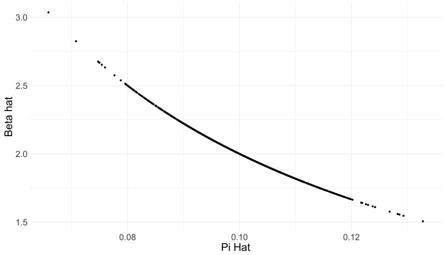
- Imagine a marginally significant first stage (se = 0.05, estimate = 0.1)
- This estimator is normal, and reasonably well-behaved
- However, the relationship between $\hat{\pi}$ and $\hat{\beta}$ is highly nonlinear near zero



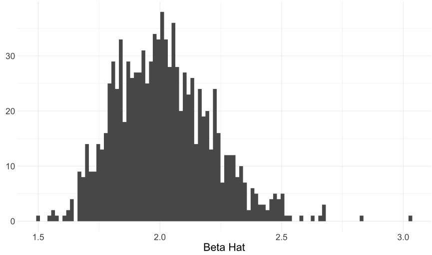
- Imagine a marginally significant first stage (se = 0.05, estimate = 0.1)
- This estimator is normal, and reasonably well-behaved
- However, the relationship between $\hat{\pi}$ and $\hat{\beta}$ is highly nonlinear near zero
- This makes the distribution for $\hat{\beta}$ very non-normal
  - Asymptotic normality is a bad approximation!



- Interestingly, this is not the case if π is sufficiently large!
- Then the relationship is quite linear



- This makes the distribution for $\hat{\beta}$ reasonably good
    - We have a problem about differentiating between these regimes



Solution 1: Pretesting

- A natural solution to this is to just check if the F-statistic is large enough that these highlighted problems are not an issue.
- This is the approach initially developed by Staiger and Stock (1997) and Stock and Yogo (2005).
    - Typical rule of thumb: first-stage F-statistic above 10 means that bias won't be larger than 10% with size of 5%. Very popular!
- Key assumption: homoskedastic. This is a strong assumption!

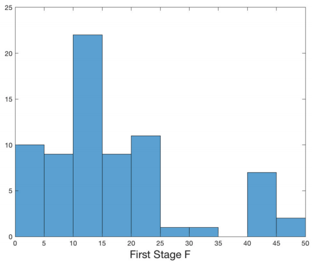
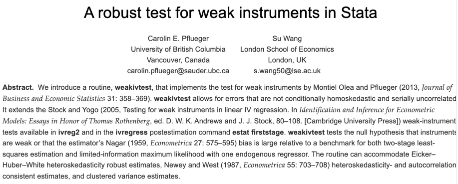


Figure 1: Distribution of reported first-stage F-statistics (and their non-homoskedastic generalizations) in 72 specifications with a single endogenous regressor and first-stage F smaller than 50. Total number of single endogenous regressor specifications reporting F-statistics is 108.

- We can do better, however. Montiel Olea and Pfluger(2013) have a heteroskedasticity-robust test, which proposes a more appropriate F statistic (allows for clustering, autocorrelation, etc.)
    - Cutoff is more like 23.1
    - An arms race in F-statistics!
- Stata package `weakivtest` here: `https://www.stata-journal.com/article.html?article=st0377`



A robust test for weak instruments in Stata

Carolin E. Pflueger
University of British Columbia
Vancouver, Canada
carolin.pflueger@sauder.ubc.ca

Su Wang
London School of Economics
London, UK
s.wang50@lse.ac.uk

Abstract. We introduce a routine, **weakivtest**, that implements the test for weak instruments by Montiel Olea and Pflueger (2013, *Journal of Business and Economic Statistics* 31: 358–369). **weakivtest** allows for errors that are not conditionally homoskedastic and serially uncorrelated. It extends the Stock and Yogo (2005, Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. J. Stock, 80–108. [Cambridge University Press]) weak-instrument tests available in **ivreg2** and in the **ivregress** postestimation command **estat firststage**. **weakivtest** tests the null hypothesis that instruments are weak or that the estimator's Nagar (1959, *Econometrica* 27: 575–595) bias is large relative to a benchmark for both two-stage least-squares estimation and limited-information maximum likelihood with one endogenous regressor. The routine can accommodate Eicker–Huber–White heteroskedasticity robust estimates, Newey and West (1987, *Econometrica* 55: 703–708) heteroskedasticity- and autocorrelation consistent estimates, and clustered variance estimates.

- The arms race continues. Lee et al. (2020) point out that current practice focuses on the β term, rather than on the t-statistic
  - which is how we claim statistical significance
- Need a much stronger first stage (F $= 104!$) for this
- Highlights the challenge of using pre-testing
  - Moreover, pre-testing for IV, much like pre-testing in dind trends, cause distort inference for your parameters

Valid t-ratio Inference for IV[1]

David S. Lee[2]
Princeton University and NBER

Marcelo J. Moreira[4]
FGV EPGE

Justin McCrary[3]
Columbia University and NBER

Jack Porter[5]
University of Wisconsin

October 15, 2020

Abstract

In the single IV model, current practice relies on the first-stage F exceeding some threshold (e.g., 10) as a criterion for trusting t-ratio inferences, even though this yields an anti-conservative test. We show that a true 5 percent test instead requires an F greater than 104.7. Maintaining 10 as a threshold requires replacing the critical value 1.96 with 3.43. We re-examine 57 AER papers and find that corrected inference causes half of the initially presumed statistically significant results to be insignificant. We introduce a more powerful test, the tF procedure, which provides F-dependent adjusted t-ratio critical values.

Keywords: Instrumental Variables, Weak Instruments, t-ratio, First-stage F statistic

- The arms race continues (again!). Angirst and Kolesar (2022) argue that weak instruments are generally *not* a concern in the just identified case
- Why? How to reconcile with Lee?
- Punchline: endogenity has to be *extremely* high for the bias to outweigh the increase in noise

One Instrument to Rule Them All:
The Bias and Coverage of Just-ID IV*

Joshua Angrist Michal Kolesár

December 2022

Abstract

We revisit the finite-sample behavior of single-variable just-identified instrumental variables (just-ID IV) estimators, arguing that in most microeconometric applications, the usual inference strategies are likely reliable. Three widely-cited applications are used to explain why this is so. We then consider pretesting strategies of the form $t_1 > c$, where t_1 is the first-stage t-statistic, and the first-stage sign is given. Although pervasive in empirical practice, pretesting on the first-stage F-statistic exacerbates bias and distorts inference. We show, however, that median bias is both minimized and roughly halved by setting $c = 0$, that is by screening on the sign of the *estimated* first stage. This bias reduction is a free lunch: conventional confidence interval coverage is unchanged by screening on the estimated first-stage sign. To the extent that IV analysts sign-screen already, these results strengthen the case for a sanguine view of the finite-sample behavior of just-ID IV.

Solution 2: Robust confidence intervals

- With a just-identified single endogenous regressor, Anderson-Rubin confidence intervals are valid, irrespective of the weakness of the first stage
- This is the easiest way to deal with this inference problem! These results are robust regardless of your first stage
  - Chernozhukov and Hansen (2008) "The reduced form: A simple approach to inference with weak instruments" discuss a very easy and simple way to implement these confidence intervals
  - Stata and R packages are also available

Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 100 (2008) 68 – 71
economics letters
www.elsevier.com/locate/econbase

The reduced form: A simple approach to inference with weak instruments

Victor Chernozhukov [a], Christian Hansen [b,*]

[a] MIT, United States
[b] University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, United States

Received 30 July 2007; received in revised form 19 October 2007; accepted 14 November 2007
Available online 28 November 2007

Abstract

In this paper, we show that conventional heteroskedasticity and autocorrelation robust inference procedures based on the reduced form provide tests and confidence intervals for structural parameters that are valid when instruments are strongly or weakly correlated to the endogenous variables.

Keywords: Heteroskedasticity; Autocorrelation; Weak identification

JEL classification: C12C30

Practical Tips for IV Papers

3. If you have many IVs, pick your best instrument and report the just identified model (weak instrument problem is much less problematic)
4. Look at the reduced form
   - The reduced form is estimated with OLS and is therefore unbiased
   - If you can't see the causal relationship of interest in the reduced form, it is probably not there
5. In case your F is not very high, think about alternative weak-instrument robust ways of estimating your statistical significance