Instrumental Variables

Eric French

University of Cambridge and Institute for Fiscal Studies

Outline: Instrumental Variables (IV)

Part I: Refresher: The basics of IV

- Motivation and basic idea
- IV assumptions and estimator

Outline: Instrumental Variables (IV)

Part I: Refresher: The basics of IV

- Motivation and basic idea
- IV assumptions and estimator

Part II: Issues in IV estimation

- Inference and weak instruments
- 2SLS
    - Multiple explanatory variables
    - Multiple instruments
- Testing for endogeneity
- Overidentification

Outline: Instrumental Variables (IV)

Part I: Refresher: The basics of IV

- Motivation and basic idea
- IV assumptions and estimator

Part II: Issues in IV estimation

- Inference and weak instruments
- 2SLS
  - Multiple explanatory variables
  - Multiple instruments
- Testing for endogeneity
- Overidentification

Part III: Parameter Heterogeneity

- What does IV recover when there is parameter heterogeneity?
- Local Average Treatment Effects
- Marginal Treatment Effects

Introduction

Remember the classical linear regression model:

$$y = \beta_0 + \beta_1 x + u$$

For the OLS estimator to be consistent (i.e. to get "close" to β_1 as the sample increases) we assume

$$\mathrm{cov}(x, u) = 0.$$

Introduction

Remember the classical linear regression model:

$$y = \beta_0 + \beta_1 x + u$$

For the OLS estimator to be consistent (i.e. to get “close” to β_1 as the sample increases) we assume

$$\text{cov}(x, u) = 0.$$

If $\text{cov}(x, u) \neq 0$ we say that x is (econometrically) endogenous.

When could $\text{cov}(x, u) \neq 0$ arise?

Common sources of endogeneity are:

- Omitted variables
  - from a variable that is correlated with x but is unobserved, so cannot be included in the regression
- Measurement error (i.e. errors-in-variables)
  - x is measured with error
- Simultaneity
  - x causes y, y causes x
  - Both x and y are determined within a system, and are thus endogeneous to the system
  - Since simultaneity was the original problem IV addressed, we use the term endogeneity anytime $\text{cov}(x, u) \neq 0$

When could $\text{cov}(x, u) \neq 0$ arise?

Common sources of endogeneity are:

- Omitted variables
  - from a variable that is correlated with x but is unobserved, so cannot be included in the regression
- Measurement error (i.e. errors-in-variables)
  - x is measured with error
- Simultaneity
  - x causes y, y causes x
  - Both x and y are determined within a system, and are thus endogeneous to the system
  - Since simultaneity was the original problem IV addressed, we use the term endogeneity anytime $\text{cov}(x, u) \neq 0$

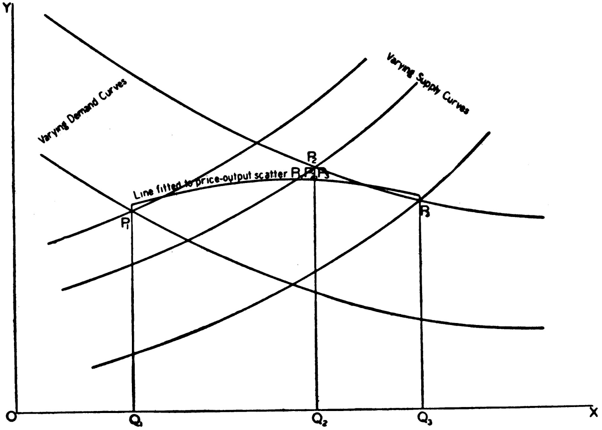
Instrumental variables regression can (potentially) eliminate bias from these three sources

History and Example

- How much money would be raised by an import tariff on animal and vegetable oils (butter, flaxseed oil, soy oil, etc.)?
- To do this calculation you need to know the elasticities of supply and demand, both domestic and foreign
- This problem was first solved using IV in Appendix B of Wright (1928), *The Tariff on Animal and Vegetable Oils*.

Figure 4, p. 296, from Appendix B (1928):

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



Who wrote Appendix B of Philip Wright (1928)?

…this appendix is thought to have been written with or by his son, Sewall Wright, an important statistician.

	
Philip Wright (1861-1934) *obscure economist and poet*	**Sewall Wright (1889-1988)** *famous genetic statistician*

History and Example: Butter

- IV regression was originally developed to estimate demand elasticities for agricultural goods, for example butter:

$$\ln(Q_i^{butter}) = \beta_0 + \beta_1 \ln(P_i^{butter}) + u_i$$

- β_1 = price elasticity of butter

History and Example: Butter

- IV regression was originally developed to estimate demand elasticities for agricultural goods, for example butter:

$$\ln(Q_i^{butter}) = \beta_0 + \beta_1 \ln(P_i^{butter}) + u_i$$

- β_1 = price elasticity of butter
- Data: observations on price and quantity of butter for different years
- The OLS regression of $\ln(Q_i^{butter})$ on $\ln(P_i^{butter})$ suffers from simultaneity bias
  - Price and quantity are determined by the interaction of demand and supply $\Rightarrow$ increase in u_i (e.g., preference shock) will impact prices

History and Example: Butter (continued)

- IV (or more generally TSLS) estimates the demand curve by isolating shifts in the supply curve.
- z is a variable that shifts supply curve but not demand.
- IV in the supply-demand example:

History and Example: Butter (continued)

- IV (or more generally TSLS) estimates the demand curve by isolating shifts in the supply curve.
- z is a variable that shifts supply curve but not demand.
- IV in the supply-demand example:

$$\ln(Q_i^{butter}) = \beta_0 + \beta_1 \ln(P_i^{butter}) + u_i$$

- Let $z_i = rain_i$ = rainfall in dairy-producing regions.
- Two key assumptions: exogeneity and relevance

History and Example: Butter (continued)

- IV (or more generally TSLS) estimates the demand curve by isolating shifts in the supply curve.
- z is a variable that shifts supply curve but not demand.
- IV in the supply-demand example:

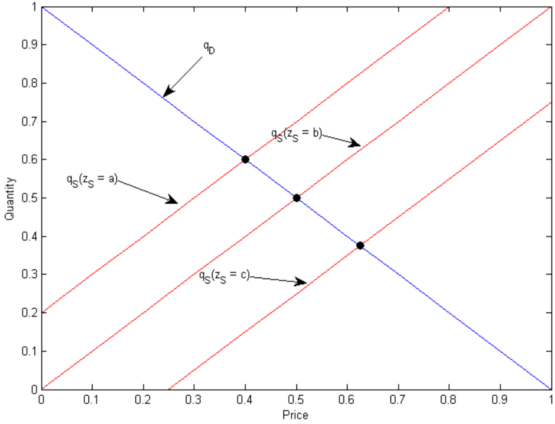
$$\ln(Q_i^{butter}) = \beta_0 + \beta_1 \ln(P_i^{butter}) + u_i$$

- Let $z_i = rain_i$ = rainfall in dairy-producing regions.
- Two key assumptions: exogeneity and relevance
  1. Exogenous? $corr(rain_i, u_i) = 0$ Plausible: whether it rains in dairy-producing regions shouldn't affect demand
  2. Relevant? $corr(rain_i, \ln(P_i^{butter})) \neq 0$ Plausible: insufficient rainfall means less grazing means less butter

How do instruments help?

- Heuristic: When z_S changes:
  - supply changes
  - demand remains the same (on average)

Movements in the supply curve induced by changing *z* trace out the demand curve



Estimation approach, in two stages

$$\ln(Q_i^{butter}) = \beta_0 + \beta_1 \ln(P_i^{butter}) + u_i$$

$z_i = rain_i$ = rainfall in dairy-producing regions.

- Stage 1: regress $\ln(P_i^{butter})$ on rain, get $\ln(\widehat{P_i^{butter}})$
  $\ln(\widehat{P_i^{butter}})$ isolates changes in log price that arise from shift in supply curve
- Stage 2: regress $\ln(Q_i^{butter})$ on $\ln(\widehat{P_i^{butter}})$)
  Use shifts in the supply curve to trace out the demand curve.

Because this approach is in two stages, the general approach is often referred to as Two Stage Least Squares (2SLS)

IV: The Basic Idea (arrows here means "affects")

OLS with $y = \beta_0 + \beta_1 x + u$: Exogenous regressor

$$\begin{array}{ccc} x & \rightarrow & y \\ & \nearrow & \\ u & & \end{array}$$

OLS: Endogenous regressor

$$\begin{array}{ccc} x & \rightarrow & y \\ \updownarrow & \nearrow & \\ u & & \end{array}$$

IV: Suppose there exists a variable z such that

$$\begin{array}{ccccc} z & \rightarrow & x & \rightarrow & y \\ & & \updownarrow & \nearrow & \\ & & u & & \end{array}$$

IV Assumptions

Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + u_i$ where $cov(u, x) \neq 0$

Assume now that we have an *instrumental variable* z_i that satisfies the following two conditions

- $cov(z_i, u_i) = 0$ $(+E(u_i) = 0)$: exogeneity or “validity”
- $cov(z_i, x_i) \neq 0$ (relevance)

IV Assumptions

The exogeneity assumption requires that:

- z_i affects y_i only through x_i
- z_i is unrelated to u_i

and is not testable (since it involves u_i).

IV Assumptions

The exogeneity assumption requires that:

- z_i affects y_i only through x_i
- z_i is unrelated to u_i

and is not testable (since it involves u_i).
Instrument relevance requires that z_i affect x_i

- It can be verified by estimating the following *first stage* regression

$$x_i = \pi_0 + \pi_1 z_i + v_i$$

Since $\pi_1 = cov(z_i, x_i)/var(z_i)$, we can (and must!) test relevance:

- $H_0 : \pi_1 = 0$: instrument irrelevant
- $H_1 : \pi_1 \neq 0$: instrument relevant
    - As usual, perform t-test

Questions?

- Where do instruments come from?
  –Intuition, subject matter knowledge, randomization
- Are there things I should look out for?
  –Weak instruments. Many instruments. Reasons to doubt the exclusion restriction. What exactly is estimated when treatment effects are heterogeneous.

IV Estimator: identifying β_1

(i) Note that

$$cov(z_i, y_i) = cov(z_i, \beta_0) + cov(z_i, \beta_1 x_i) + cov(z_i, u_i) = \beta_1 cov(z_i, x_i)$$

(ii) Use the IV assumptions to write

$$\beta_1 = \frac{cov(z_i, y_i)}{cov(z_i, x_i)} = \frac{cov(z_i, y_i)/var(z_i)}{cov(z_i, x_i)/var(z_i)}$$

which is the slope coefficient estimator from the *reduced form* divided by the slope coefficient estimator from the *first stage*.

IV Estimator: identifying β_1

(i) Note that

$$cov(z_i, y_i) = cov(z_i, \beta_0) + cov(z_i, \beta_1 x_i) + cov(z_i, u_i) = \beta_1 cov(z_i, x_i)$$

(ii) Use the IV assumptions to write

$$\beta_1 = \frac{cov(z_i, y_i)}{cov(z_i, x_i)} = \frac{cov(z_i, y_i)/var(z_i)}{cov(z_i, x_i)/var(z_i)}$$

which is the slope coefficient estimator from the *reduced form* divided by the slope coefficient estimator from the *first stage*.

Reduced form: $cov(z_i, y_i)/var(z_i)$

- slope coefficient from a regression of y_i on z_i and an intercept

First stage: $cov(z_i, x_i)/var(z_i)$

- slope coefficient from a regression of x_i on z_i and an intercept

IV Estimator

The sample analog is the *instrumental variable estimator* of β_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}$$

The IV estimator of β_0 is:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Note that $\hat{\beta}_1$ is the OLS estimator of β_1 when $z_i = x_i$.

Above assumptions + Random Sampling $\Rightarrow$ IV is consistent.

Special Case of IV: Wald Estimator

A common (and simple) example of IV is one where the instrument is binary

$$z_i \in \{0, 1\}$$

Note that

$$\begin{aligned} E[y_i|z_i = 1] &= \beta_0 + \beta_1 E[x_i|z_i = 1] \\ E[y_i|z_i = 0] &= \beta_0 + \beta_1 E[x_i|z_i = 0] \end{aligned}$$

so that

$$E[y_i|z_i = 1] - E[y_i|z_i = 0] = \beta_1 (E[x_i|z_i = 1] - E[x_i|z_i = 0])$$

Special Case of IV: Wald Estimator

A common (and simple) example of IV is one where the instrument is binary

$$z_i \in \{0, 1\}$$

Note that

$$\begin{aligned} E[y_i|z_i = 1] &= \beta_0 + \beta_1 E[x_i|z_i = 1] \\ E[y_i|z_i = 0] &= \beta_0 + \beta_1 E[x_i|z_i = 0] \end{aligned}$$

so that

$$E[y_i|z_i = 1] - E[y_i|z_i = 0] = \beta_1(E[x_i|z_i = 1] - E[x_i|z_i = 0])$$

rearranging and taking sample analogues gives the Wald estimator:

$$\hat{\beta}_1 = \frac{\overline{y}_{(z=1)} - \overline{y}_{(z=0)}}{\overline{x}_{(z=1)} - \overline{x}_{(z=0)}}$$

where $\overline{x}_{(z=1)} \equiv$ sample mean of x for cases where $z_i = 1$, for example.

Education example

Consider the impact of education on wages:

$$\ln(wage) = \beta_0 + \beta_1 educ + u.$$

We need to consider what factors are captured by u.

- u may include variables such as ability, and more educated individuals are potentially more "able".
- If $\text{cov}(educ, u) > 0$, OLS estimate of β_1 biased upwards.
- Although measurement error in *educ* would lead to downward bias.

Education example

Angrist and Krueger (1991) used quarter of birth as an instrument for education.

Arguments for instrument:

- During sample period in US, education was compulsory by law until your 16th birthday
- School start in the year you turn 6:
  - children born early in the year begin school at an older age
  - and may therefore leave school with somewhat less education

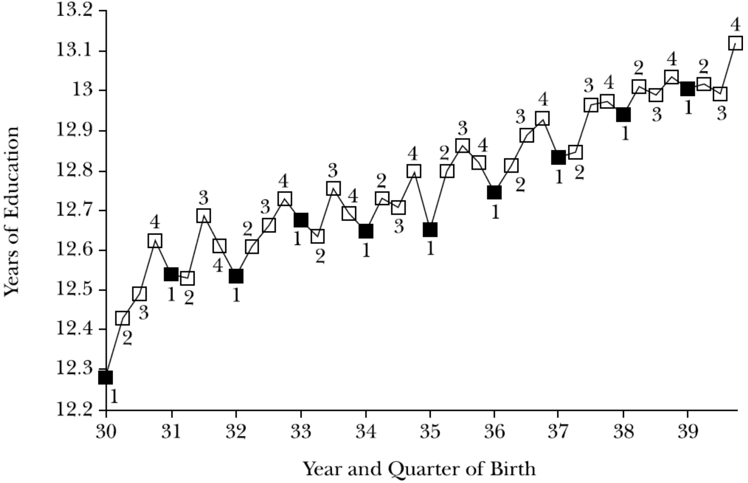
Education example

Angrist and Krueger (1991) used quarter of birth as an instrument for education.

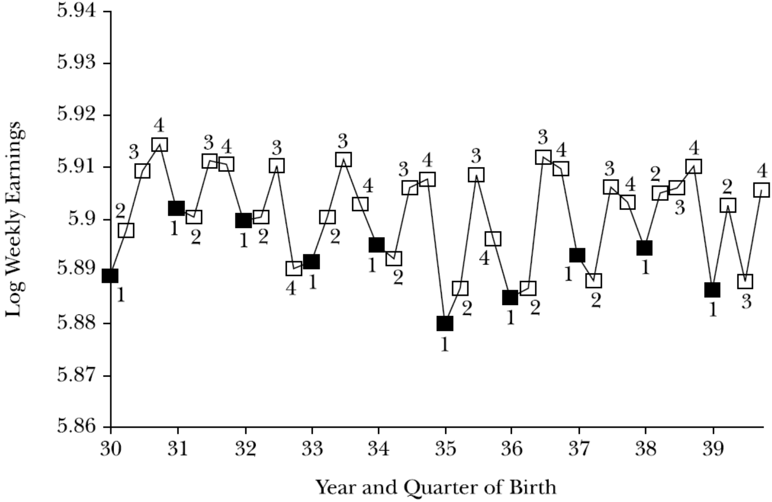
Arguments for instrument:

- During sample period in US, education was compulsory by law until your 16th birthday
- School start in the year you turn 6:
  - children born early in the year begin school at an older age
  - and may therefore leave school with somewhat less education
- Quarter of birth is arguably uncorrelated with unobservables affecting wages (though see Bound, Jaeger and Baker (1995), Buckles and Hungerman (2013) for arguments otherwise).

Mean Years of Completed Education, by Quarter of Birth



Mean Log Weekly Earnings, by Quarter of Birth



OLS and IV estimates

	Quarter of birth 1st (1)	4th (2)	Difference (2)-(1) (3)
ln(weekly wage)	5.892	5.905	0.0135
			(0.0034)
Years of education	12.688	12.839	0.151
			(0.016)
Wald estimate of return to education			0.089
			(0.021)
OLS estimate of return to education			0.070
			(0.0005)

And adding multiple instruments, controls

Structural Equation:

$$ln(wage_i) = \beta_0 + \beta_1 educ_i + \gamma' x_i + u_i$$

First-Stage Equation:

$$educ_i = \pi_{1,1} Q1_i + \pi_{1,2} Q2_i + \pi_{1,3} Q3_i + \pi'_{1,x} x_i + v_{1,i}$$

Reduced Form Equation

$$ln(wage_i) = \pi_{2,1} Q1_i + \pi_{2,2} Q2_i + \pi_{2,3} Q3_i + \pi'_{2,x} x_i + v_{2,i}$$

where

- $Q1 - Q3$: dummies for quarter of birth (our instruments)
- x: 59 control variables. Dummies for state of birth and year of birth
- Data from 1980 Census for men aged 40-49 in 1980

Example: Returns to Schooling

- OLS Results (from Stata):

```
xi: reg lwage educ i.yob i.sob , robust
i.yob             _Iyob_30-39         (naturally coded; _Iyob_30 omitted)
i.sob             _Isob_1-56          (naturally coded; _Isob_1 omitted)

Linear regression                                      Number of obs =  329509
                                                       F( 60,329448) =  649.29
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1288
                                                       Root MSE      =  .63366

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .067339   .0003883   173.40   0.000     .0665778    .0681001
.
.
.
```

If intuition about source of endogeneity is correct, this should be an over-estimate of the effect of schooling.

Example: Returns to Schooling

- First-Stage Results (from Stata):

```
 xi: regress educ i.qob i.sob i.yob , robust
Linear regression                                      Number of obs =   329509
                                                       F( 62,329446) =   292.87
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.0572
                                                       Root MSE      =   3.1863

------------------------------------------------------------------------------
             |               Robust
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iqob_2 |   .0455652    .015977     2.85   0.004     .0142508    .0768797
     _Iqob_3 |   .1060082   .0155308     6.83   0.000     .0755683     .136448
     _Iqob_4 |   .1525798   .0157993     9.66   0.000     .1216137    .1835459

.
.
.
 testparm _Iqob*

 ( 1)  _Iqob_2 = 0
 ( 2)  _Iqob_3 = 0
 ( 3)  _Iqob_4 = 0

       F(  3,329446) =   36.06
            Prob > F =    0.0000
```

First-stage F-statistic.

Example: Returns to Schooling

- Reduced-Form Results (from Stata):

```
xi: regress lwage i.qob i.sob i.yob , robust

Linear regression                                      Number of obs =  329509
                                                       F( 62,329446) =  147.83
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0290
                                                       Root MSE      =  .66899

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iqob_2 |   .0028362   .0033445     0.85   0.396    -.0037188    .0093912
     _Iqob_3 |   .0141472   .0032519     4.35   0.000     .0077736    .0205207
     _Iqob_4 |   .0144615   .0033236     4.35   0.000     .0079472    .0209757
.
.

testparm _Iqob*

 ( 1)  _Iqob_2 = 0
 ( 2)  _Iqob_3 = 0
 ( 3)  _Iqob_4 = 0

       F(  3,329446) =   10.43
            Prob > F =    0.0000
```

Example: Returns to Schooling

- 2SLS Results (from Stata):

```
xi: ivregress 2sls lwage (educ = i.qob) i.yob i.sob , robust

Instrumental variables (2SLS) regression               Number of obs =  329509
                                                       Wald chi2(60) = 9996.12
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.0929
                                                       Root MSE      =  .64652

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1076937   .0195571     5.51   0.000     .0693624     .146025
.
.
.
```

Bigger than OLS?

Example: Returns to Schooling

- GMM Results (from Stata, efficient under heteroskedasticity):

```
xi: ivregress gmm lwage (educ = i.qob) i.yob i.sob , robust

Instrumental variables (GMM) regression                Number of obs =   329509
                                                       Wald chi2(60) = 9992.90
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.0927
GMM weight matrix: Robust                              Root MSE      =  .64658

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1077817   .0195588     5.51   0.000     .0694472    .1461163
.
.
.

 estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(2) = 3.10009 (p = 0.2122)
```

Bigger than OLS?

Fail to reject over-id test

Inference

Assuming homoscedasticity $E[u_i^2|z_i] = \sigma^2$, it can be shown that the estimated (asymptotic) variance of the IV estimator is:

$$\widehat{var}(\hat{\beta}_{IV}) = \frac{\hat{\sigma}^2}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right) R_{x,z}^2}$$

- $R_{x,z}^2$: the R-squared from a regression of x_i on z_i and an intercept
- $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$ where $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

Inference

Assuming homoscedasticity $E[u_i^2|z_i] = \sigma^2$, it can be shown that the estimated (asymptotic) variance of the IV estimator is:

$$\widehat{var}(\hat{\beta}_{IV}) = \frac{\hat{\sigma}^2}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right) R^2_{x,z}}$$

- $R^2_{x,z}$: the R-squared from a regression of x_i on z_i and an intercept
- $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$ where $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$
- Recall: $\widehat{var}(\hat{\beta}_{OLS}) = \frac{\hat{\sigma}^2}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)}$, $R^2_{x,z} \leq 1$
- $\Rightarrow \widehat{var}(\hat{\beta}_{IV}) \geq \widehat{var}(\hat{\beta}_{OLS})$: variance of the IV estimator is greater than varaince of OLS estimator

Punchline: IV can be data hungry, especially when $R^2_{x,z}$ is low

Unbiasedness $\neq$ consistency!

- Unbiasedness: small sample; Consistency: large sample.
- OLS: consistent and **unbiased**
- IV: consistent but **biased**

Weak Instruments and Bias

Weak instrument $\Leftrightarrow$ z and x are only weakly correlated

- leads to imprecise IV estimates,
- but can also give large bias.

Recall that:

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(u_i - \overline{u})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}$$

Weak Instruments and Bias

Weak instrument $\Leftrightarrow$ z and x are only weakly correlated

- leads to imprecise IV estimates,
- but can also give large bias.

Recall that:

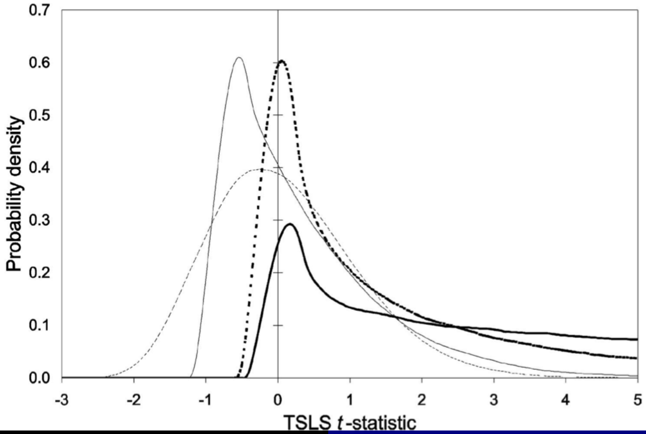
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(u_i - \overline{u})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}$$

Even if $cov(z,u) = 0, cov(z,x) \neq 0$

- $\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})$ a random variable that can be close to 0 in a small sample $\Rightarrow \hat{\beta}_1$ becomes very unstable
- Random fluctuations in u are also captured in x (if $cov(x,u) \neq 0$) $\Rightarrow$ **IV estimate biased toward OLS estimate**

- Figure shows distribution of t-statistic
- Dark line = irrelevant instruments
- Dashed light line = strong instruments

Weak Instruments and Bias



How to test for weak instruments?

- Compute the first stage F-statistic.
- Stock and Yogo (2005) showed that

$$E(\beta_{2SLS}) - \beta \approx \frac{E(\beta_{OLS} - \beta)}{E(F) - 1}$$

- $E(F)$ is the expectation of the first stage $F-$statistic.
  - If $E(F) = 10$, the bias of 2SLS, relative to the bias of OLS, is approximately $\frac{1}{9}$, which is small enough to be acceptable.

How to test for weak instruments?

- Compute the first stage F-statistic.
- Stock and Yogo (2005) showed that

$$E(\beta_{2SLS}) - \beta \approx \frac{E(\beta_{OLS} - \beta)}{E(F) - 1}$$

- $E(F)$ is the expectation of the first stage $F-$statistic.
  - If $E(F) = 10$, the bias of 2SLS, relative to the bias of OLS, is approximately $\frac{1}{9}$, which is small enough to be acceptable.
- A Rule of Thumb: if F-statistic exceeds 10 (same as t-statistic above $\sqrt{10}$ with only one r.h.s. variable),then don't worry too much.
- Lee, McCrary, Moreira, Porter: even when F-stat is 10, standard method of calculating s.e.s creates confidence intervals that are much too small

Weak Instruments and Bias: Punchlines

- **If the IVs are weak,** the sampling distribution of the TSLS estimator (and its *t*-statistic) **is not well approximated** by its large *n* **normal approximation.**
- IV estimation thus requires fairly strong instruments

Weak Instruments and Bias: Punchlines

- **If the IVs are weak,** the sampling distribution of the TSLS estimator (and its t-statistic) **is not well approximated** by its large n **normal approximation.**
- IV estimation thus requires fairly strong instruments
- Rule of thumb: F-statistic above 10 (same as t-statistic above $\sqrt{10}$ with only one r.h.s. variable) for the instrument in the first stage

IV in the Multiple Linear Regression model

We can add an additional explanatory variable x_2 to the model :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Assume that x_2 is uncorrelated with u, while x_1 is correlated with u

- x_{i2}: **exogenous** explanatory variable
- x_{i1}: **endogenous** explanatory variable

IV in the Multiple Linear Regression model

We can add an additional explanatory variable x_2 to the model :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Assume that x_2 is uncorrelated with u, while x_1 is correlated with u

- x_{i2}: **exogenous** explanatory variable
- x_{i1}: **endogenous** explanatory variable

Adding multiple exogenous and endogenous explanatory variables is straightforward, but we will not cover this here

IV in the Multiple Linear Regression model

Instrument relevance implies z_1 is correlated with x_1, but now over and above x_2. We can test this by estimating the following regression

$$x_{i1} = \pi_0 + \pi_1 z_{i1} + \pi_2 x_{i2} + v_i$$

Instrument relevance is tested as:

$$H_0 : \pi_1 = 0 \text{ vs } H_1 : \pi_1 \neq 0$$

IV in the multiple linear regression model is just as IV in the simple linear regression model but the exogeneity/validity assumption is now:

$$cov(z_i, u_i | \, x_{i2}) = 0$$

(Note that z_i and x_{i2} can be correlated!)

Two-stage least squares (2SLS)

We still consider the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

but now with **multiple instruments** $M > 1$

$$cov(u_i, z_{im} | \, x_{i2}) = 0 \quad m = 1, \ldots, M$$

The **first stage regression** is:

$$x_{i1} = \pi_0 + \pi_1 z_{i1} + \ldots + \pi_M z_{iM} + \pi_{M+1} x_{i2} + v_i$$

Instrument relevance is tested using an F statistic for

$$H_0 : \pi_1 = \ldots = \pi_M = 0$$

2SLS: Step-by-step

1. Estimate the *first-stage* regression:

$$x_{i1} = \pi_0 + \pi_1 z_{i1} + \ldots + \pi_M z_{iM} + \pi_{M+1} x_{i2} + v_i$$

   - regress the endogenous explanatory variable on all instruments and *all* other exogenous explanatory variables

2. Compute the predicted value of x_1:

$$\hat{x}_{i1} = \hat{\pi}_0 + \hat{\pi}_1 z_{i1} + \ldots + \hat{\pi}_M z_{iM} + \hat{\pi}_{M+1} x_{i2}$$

3. Estimate the *second-stage* regression:

$$y_i = \beta_0 + \beta_1 \hat{x}_{i1} + \beta_2 x_{i2} + u_i$$

   - regress the outcome variable on $\hat{x}_{i1}$ and *all* the other exogenous explanatory variables

Note that using more than one instrument is not necessary, but it can give you more efficient IV estimates.

Weak IV Revisited

- Weak IV are also a problem with **many instruments.**

- **Adding instruments with low predictive power** in the first stage lowers the F-statistic and **exacerbates the bias in the 2SLS estimator.**

- Bound, Jaeger, and Baker (1995) illustrate this using the Angrist and Krueger (1991). Angrist and Krueger present results using different sets of IVs (plus other covariates):
  - quarter of birth dummies: $M = 3$ instruments.
  - quarter of birth + (quarter of birth) x (year of birth) dummies: $M = 30$ instruments.
  - quarter of birth + (quarter of birth) x (year of birth) + (quarter of birth) x (state of birth): $M = 180$ instruments.

Weak IV Revisited

Table 1. *Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings (standard errors of coefficients in parentheses)*

	(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

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. F (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basmann (1960).

Testing Overidentification Restrictions

- When there are as many IVs as endogenous variables, exogeneity of the instruments is not testable.
- If there are more IVs than endogenous variables, we can test whether some of them are correlated with the u.
- **Logic:** With two IVs and one endogenous variable we can compute alternative 2SLS estimates using each of the IVs. If the IVs are both exogenous, the 2SLS will converge to the same parameter and they will differ only by sampling error.
- Test is often called “Hansen’s Over-Identification Test”, “J-Test”, “Sargan test”, “S-Test”
- Based on GMM criterion function (distributed chi-square under the null)

Testing Overidentification Restrictions

We can only reject hypotheses, **we can never accept!**

- If the two estimates are statistically different, we can reject the hypothesis that both IVs are valid.
- But we would not be able to ascertain which one!
- Moreover, if they are similar it could because both IVs fail the exogeneity requirement.

Heterogeneous Treatment Effects: LATE

1. Previous discussion posits a constant treatment effect β = average treatment effect (ATE)
2. It is unlikely that "treatment effects" are constant across the population
   - You would likely benefit little from auto repair training
   - mechanic would likely benefit little from this class
3. So what does IV recover when there is parameter, or "treatment effect", heterogeneity β_i?
   - If "treatment effect" is independent of instrument, we can recover "average treatment effect"
   - If "treatment effect" is correlated with instrument, Angrist-Imbens (1994) show we can still estimate a causal effect among a subpopulation (given several assumptions)

If "Treatment effect" is independent of instrument $\Rightarrow$ IV recovers "Average Treatment Effect"

- Suppose β_i independent of x_i, z_i and $E[z_i u_i] = 0$. Then define ATE= $\beta = E[\beta_i]$

$$
\begin{aligned}
y_i &= \beta_0 + \beta_i x_i + u_i \\
&= \beta_0 + \beta x_i + u_i + (\beta_i - \beta)\, x_i \\
&= \beta_0 + \beta x_i + \omega_i,
\end{aligned}
$$

where

$$
\begin{aligned}
E[z_i \omega_i] &= E[z_i u_i] + E[z_i (\beta_i - \beta)\, x_i] \\
&= 0 + \; E[z_i x_i] E[(\beta_i - \beta)] = 0
\end{aligned}
$$

- **IV estimate recovers ATE**
- But exclusion restriction will not hold if heterogeneous effect is not independent of instrument or treatment!

LATE: set-up

Strip model down to binary treatment x and binary instrument z
Four subpopulations (Wald estimator):

1. Always-takers: $z_i = 1, x_i = 1$ and $z_i = 0, x_i = 1$
2. Never-takers: $z_i = 1, x_i = 0$ and $z_i = 0, x_i = 0$
3. Complier: $z_i = 1, x_i = 1$ and $z_i = 0, x_i = 0$
4. Defier: $z_i = 1, x_i = 0$ and $z_i = 0, x_i = 1$

Notes:

- We never observe individuals at both instrument states (z_i=0 and $z_i = 1$).
- We cannot determine an observation's subpopulation

Heterogeneous Treatment Effects: LATE

- ATE among subpopulation of compliers = Local Average Treatment Effect (LATE)
- Conditions (first two similar to constant Treatment Effect care):
  - **First-stage or relevance:** instrument predicts treatment
  - **Independence:** instrument independent of all unobservables affecting outcome and treatment/endogenous variable state
  - **Exclusion:** instrument only affects outcome though treatment receipt. (No direct effect of the instrument.)
  - **Monotonicity:** effect of instrument on probability of receiving treatment is ≥ 0 for everyone or ≤ 0 for everyone. (No defiers)

Heterogeneous Treatment Effects: LATE

LATE can be generalized (with varying degrees of difficulty) to
multi-valued/continous treatments or instruments,
over-identified models, models with controls

Heterogeneous Treatment Effects: LATE

LATE can be generalized (with varying degrees of difficulty) to multi-valued/continous treatments or instruments, over-identified models, models with controls

- 2SLS coefficient estimates (or approximates) a weighted average of different LATEs
- Each instrument/value of the instrument potentially affects a different set of compliers $\Rightarrow$ a different LATE
  - Rejection of over-id test does not mean instruments are invalid as all could be valid but affects different complier populations

Learning about compliers in LATE framework

- We can learn some things about compliers from data, e.g.
- Probability of being a complier (assuming no defiers)

$$\underbrace{\Pr[x_i = 1|z_i = 1]}_{\text{Probability complier or always-taker}} - \underbrace{\Pr[x_i = 1|z_i = 0]}_{\text{Probability always-taker}}$$

- Proportion of treated who are compliers:

$$\frac{\Pr(z_i = 1)}{\Pr(x_i = 1)}(\Pr[x_i = 1|z_i = 1] - \Pr[x_i = 1|z_i = 0])$$

Example: Returns to education

- Returns to education complicated
  - Multi-valued treatment (education)
  - Multiple instruments
- IV estimand with one binary instrument = weighted average of effects of increasing education by one year across all education levels among compliers
- 2SLS estimand with multiple instruments= weighted average of individual IV estimands
- **Monotonicity condition** => change in instrument causes weakly increasing (decreasing) level of education for everyone

Example: Returns to Education

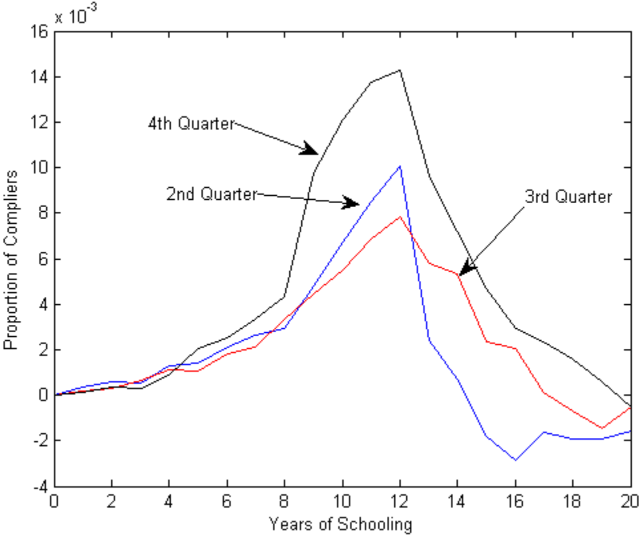
- Consider dummy for being born quarter j (*Qj*)
- Maintain exclusion restriction as before
- First-stage: Regress education on *Qj*
- Can estimate fraction of compliers at each education level as

$$P(educ_i < e|Qj_i = 0) - P(educ_i < e|Qj_i = 1)$$

  (Assuming monotonicity such that changing instrument from 0 to 1 increases education)
- Can estimate weight given to each education value as (fraction of compliers)/first-stage

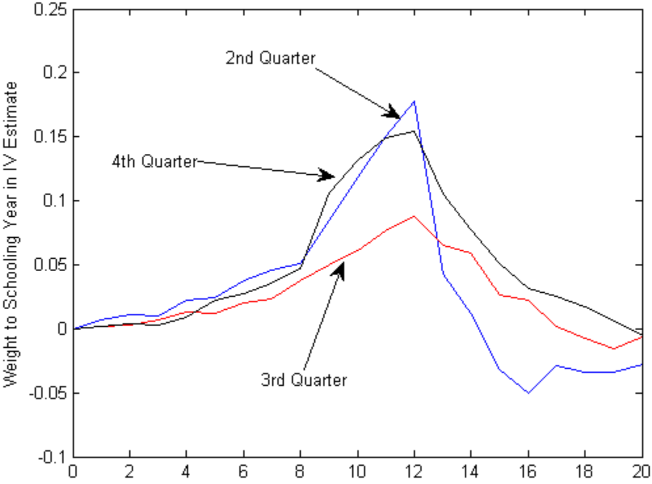
Example: Returns to Schooling

Fractions of compliers for different quarter of birth instruments.



Example: Returns to Schooling

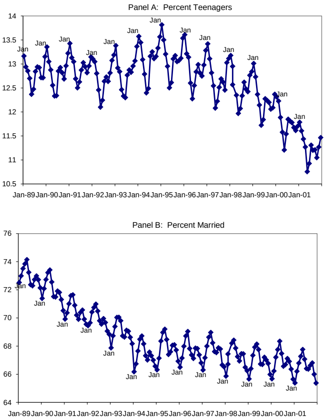
Weighting functions for different quarter of birth instruments. Note this is the fraction of compliers scaled by the first-stage coefficient.



Finally: nothing is perfect

Buckles and Hungerman (2013): Quarter of birth not random: correlated with many parental vars $\Rightarrow$ inconsistent estimates

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



27

Marginal Treatment Effects: MTEs

- So far we have considered heterogeneity in treatment effects for a discrete valued instrument
- **Question:** can we extend the logic of this to a continuous instrument?
- **Answer**: yes! These treatment effects are Marginal Treatment Effects (MTEs)
- MTE can be thought of as impact of changing x a small amount

Example: what is the effect of disability benefit receipt on labor supply and mortality?

- DI program is large
  - 4.7% of people ages 18-64 receive DI benefits
  - 6.4% of people ages 18-64 receive DI or SSI disability benefits
- Method: compare labor supply/mortality of those who applied for disability benefits and were Allowed versus Denied
  - Instrumental Variables: use assignment of cases to judges to predict allowance
  - judges randomly assigned to cases $\Rightarrow$ assigned judge independent of applicant's leisure preferences and productivity
- French and Song (2014): DI and employment
- Black, French McCauley, and Song (2024): DI and mortality

Key idea.

Compare labor supply of those who applied for Disability Insurance (DI) and were allowed vs. denied

Key issue: Those allowed different from those denied

Key idea.

Compare labor supply of those who applied for Disability Insurance (DI) and were allowed vs. denied

Key issue: Those allowed different from those denied

Solution: Use IVs: exploit random assignment of cases to Administrative Law Judges (ALJs)

The Judge Instrument:

- Idea: Some judges are stricter than others. Instrument receipt of DI with judge who (randomly) decided your case
- Instrument likely impacts marginal cases

Key idea.

Compare labor supply of those who applied for Disability Insurance (DI) and were allowed vs. denied

Key issue: Those allowed different from those denied

Solution: Use IVs: exploit random assignment of cases to Administrative Law Judges (ALJs)

The Judge Instrument:

- Idea: Some judges are stricter than others. Instrument receipt of DI with judge who (randomly) decided your case
- Instrument likely impacts marginal cases

Implementation idea (what is actually done is slightly fancier)

1. First stage: Regress allowance on a full set of judge dummies
2. Second stage: Regress labor supply on predicted allowance

Thousands of judges, millions of applicants

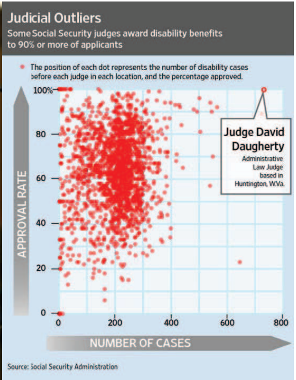
IV Procedure

- **Use judges as instruments**
  - “the Hearing Office Chief Administrative Law Judge generally assigns cases to ALJs from the master docket on a rotational basis, with the earliest (i.e., oldest) Request for Hearing receiving priority.” (Social Security Administration, 2009).

WSJ, May 19, 2011

Disability-Claim Judge Has Trouble Saying ’No’

Near-Perfect Approval Record; Social-Security Program Strained



"Some of these judges act like it’s their own damn money we’re giving away"

Heterogeneity in responses

Labor supply (or mortality) response β_i potentially varies across people

With a few assumptions the procedure below allows us to **identify a (weighted) average** of responses

- **Rank/first stage/relevance**: judges affect the allowance rate
- **Independence:** judge assignment independent of characteristics (e.g., health) of their caseload
- **Exclusion restriction:** judges affect outcome only through their allowance decisions (and not other channels such as wait times)
- **Monotonicity:** a “high allowance” judge will allow any case a “low allowance” judge will allow

All of this looks looks like checking the LATE assumptions

MTEs and Selection Models

- To derive MTEs it is helpful to write down a model of selection into treatment (a “selection model”)
- Side benefit: the selection model allows us to draw links between the treatment effect literature and literature using choice models
  - e.g., Vytlacil (2002): selection models imply monotonicity
  - Furthermore, MTEs can also be thought of as a **unifying concept**, as ATEs and LATEs can be derived from MTEs (if we know a little about densities of some variables) : Heckman et al. 2006, French and Taber (2011)

Marginal Treatment Effects: MTEs

Define x_i as a 0-1 indicator =1 if individual i is allowed benefits (the choice), $\mathbf{X}_i$ controls, y_i is an outcome (e.g., earnings). The outcome equation is

$$y_i = x_i \beta_i + \mathbf{X}_i \delta_y + u_i. \tag{1}$$

Allowance (or "treatment") is determined by

$$\Pr(x_i = 1 | \mathbf{X}_i, z_i) = \Pr(g(\mathbf{X}_i, z_i) > V_i) \tag{2}$$

where

- $\Pr(g(\mathbf{X}_i, z_i) > V_i)$ is the probability $g(\mathbf{X}_i, z_i) > V_i$
  - $g(\mathbf{X}_i, z_i)$ is an arbitrary function of $(\mathbf{X}_i, z_i)$, typically estimated using a probit or linear probability model.
- z_i is our exclusion restriction (e.g., "judge leniency")
- Residual V_i can be interpreted as a measure of the health of individual i that is observed by the judge (and thus impacts the allowance decision) but not the econometrician.

Marginal Treatment Effects: MTEs

assume V_i is independent of Z_i and $\mathbf{X}_i$ (standard one for binary choice models) The Marginal Treatment Effect is

$$MTE(\mathbf{X}_i = \mathbf{X}, V_i = a) \equiv E[\beta_i.|\mathbf{X}_i = \mathbf{X}, V_i = a]$$

With a few more assumptions (largely those implied in the LATE framework), Heckman, others have shown that:

$$\frac{\partial E[y_i|\mathbf{X}_i = x, \Pr(x_i = 1|\mathbf{X}_i, z_i) = a]}{\partial a} = MTE(\mathbf{X}_i = x, V_i = a).$$

In practice, model y_i as a polynomial, local linear, local polynomial smoother in $\Pr(x_i = 1|\mathbf{X}_i, z_i)$

Empirical Methods

Empirical Methods

- **Instrumental variable:**
  - Judge leniency

$$x_i = \mathbf{j}_i\gamma + \mathbf{X}_i\delta_A + e_i.$$

where $\mathbf{j}_i$ a full set of judge indicators

- **Estimating equation:**

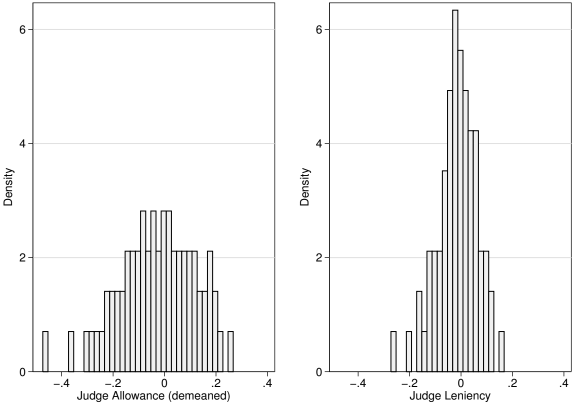
$$y_i = \hat{x}_i\beta + \mathbf{X}_i\delta_y + u_i.$$

  - $y_i =$ mortality
  - $x_i = 1$ if case i was allowed by case i ALJ , 0 otherwise
  - X_i = a full set of 0-1 hearing office-day indicators

- This procedure β should recover a weighted average of marginal treatment effects
- Add a polynomial in $\hat{x}_i$ to outcome equation and take derivatives: recover distribution of MTEs

Allowance Rate of ALJs–Judge Leniency

De-meaned, and de-meaned by hearing office and day



Lots of variation $\Rightarrow$ strong first stage

Predictors of Allowance and Judge Leniency

	Dependent variable: Allowed		Dependent variable: judge leniency	
	Coefficient	t-stat	Coefficient	t-stat
	(1)	(2)	(3)	(4)
LF particip yrs -11 to -2	0.0067	7.8	0.0004	1.0
Earnings/billion, yrs -11 to -2	0.0004	8.8	0.0000	1.0
Rep. by lawyer	0.0188	3.1	-0.0052	-2.2
Neoplasms/cancer	0.0323	11.5	0.0020	1.3
Mental disorders	0.0017	0.8	0.0001	0.1
Mental retardation	0.0183	3.3	-0.0005	-0.3
Nervous system	0.0154	6.9	0.0011	1.2
Circulatory/heart	0.0321	17.2	0.0020	1.3
Musculoskeletal	0.0278	16.2	0.0020	1.6
Respiratory	0.0192	8.6	0.0004	0.3
Injuries	0.0218	9.4	0.0008	0.6
Endocrine/diabetes	0.0275	12.5	0.0011	1.0
R^2	0.0126		0.0018	
Number of applicants = 599,353		Number of judges = 1,404		

Note: covariates not shown are: female, age, race, education, eligible for SSDI.
Omitted category: no lawyer, health condition is "other"

Independence implies we cannot predict our judge leniency instrument $\Rightarrow$ we cannot reject!

First Stage Regression: Allowance on Judge Leniency

	Observations	Allowance Rate ALJ stage	Coeff. on IV	SE	T-ratio
	(1)	(2)	(3)	(4)	(5)
All groups	599,353	0.841	0.940	(0.025)	38
Avg earnings<$10,000	277,336	0.785	1.032	(0.035)	30
Avg earnings>$10,000	322,017	0.889	0.833	(0.021)	40
Rep. by lawyer	378,807	0.853	0.900	(0.016)	56
Not represented	220,546	0.819	1.020	(0.044)	23
Neoplasms/cancer	18,647	0.868	0.931	(0.036)	26
Mental disorders	60,487	0.794	1.066	(0.035)	31
Mental retardation	3,109	0.812	0.964	(0.088)	11
Nervous system	33,826	0.828	0.928	(0.040)	23
Circulatory/heart	101,718	0.860	0.900	(0.030)	30
Musculoskeletal	229,085	0.856	0.911	(0.022)	41
Respiratory	29,258	0.844	0.942	(0.033)	29
Injuries	26,752	0.840	0.939	(0.046)	20
Endocrine/diabetes	38,405	0.840	0.951	(0.033)	29
All other	58,066	0.793	0.965	(0.040)	24

- Strong First Stage (t-ratios large)
- Lenient judges lenient across the board ... cannot reject monotonicity assumption. (Fransdsen et al.(2022) provide a more rigorous test)

Note: allowance is de-meaned by hearing office and day.

First Stage Regression: Allowance on Judge Leniency

	Observations	Allowance Rate ALJ stage	Coeff. on IV	SE	T-ratio
	(1)	(2)	(3)	(4)	(5)
All groups	599,353	0.841	0.940	(0.025)	38
Avg earnings<$10,000	277,336	0.785	1.032	(0.035)	30
Avg earnings>$10,000	322,017	0.889	0.833	(0.021)	40
Rep. by lawyer	378,807	0.853	0.900	(0.016)	56
Not represented	220,546	0.819	1.020	(0.044)	23
Neoplasms/cancer	18,647	0.868	0.931	(0.036)	26
Mental disorders	60,487	0.794	1.066	(0.035)	31
Mental retardation	3,109	0.812	0.964	(0.088)	11
Nervous system	33,826	0.828	0.928	(0.040)	23
Circulatory/heart	101,718	0.860	0.900	(0.030)	30
Musculoskeletal	229,085	0.856	0.911	(0.022)	41
Respiratory	29,258	0.844	0.942	(0.033)	29
Injuries	26,752	0.840	0.939	(0.046)	20
Endocrine/diabetes	38,405	0.840	0.951	(0.033)	29
All other	58,066	0.793	0.965	(0.040)	24

- Abadie (2003) shows that ratio of subsample to full sample IV estimate is informative of the relative likelihood that someone with a given characteristic is allowed given a small increase in judge leniency.

Note: allowance is de-meaned by hearing office and day.

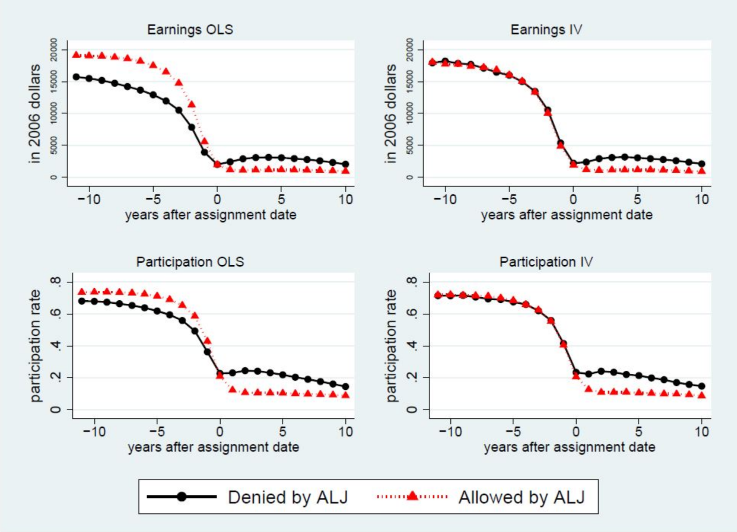
TABLE 4: ESTIMATED EFFECT OF DI RECIPIENCY ON PARTICIPATION, DISAGGREGATED

	OLS			IV	
	Allowed	Denied	Difference	Difference	Std. Error
All groups					
All groups	0.130	0.395	-0.265	-0.256	0.006
Education					
Less than high school	0.076	0.327	-0.251	-0.230	0.009
High school graduate, no college	0.148	0.425	-0.277	-0.279	0.009
Some college	0.210	0.479	-0.269	-0.261	0.019
College graduate	**0.254**	**0.472**	**-0.219**	**-0.179**	**0.031**
Health conditions (by diagnosis group)					
Neoplasms (e.g., cancer)	**0.128**	**0.365**	**-0.237**	**-0.211**	**0.015**
Mental disorders	**0.155**	**0.457**	**-0.302**	**-0.194**	**0.043**
Mental retardation	**0.146**	**0.383**	**-0.237**	**-0.202**	**0.016**
Nervous system	0.094	0.322	-0.227	-0.282	0.048
Circulatory system (e.g., heart disease)	0.140	0.392	-0.251	-0.237	0.027
Musculoskeletal disorders (e.g., back pain)	0.111	0.367	-0.256	-0.250	0.018
Respiratory system	0.136	0.419	-0.283	-0.285	0.009
Injuries	0.089	0.363	-0.274	-0.254	0.023
Endocrine system (e.g., diabetes)	0.147	0.468	-0.320	-0.367	0.022
All other	0.089	0.324	-0.235	-0.224	0.024

Notes: OLS estimates are in levels with no covariates
IV estimates use demeaned variables and the judge allowance differential as the instrument
Allowance and participation measured 3 years after assignment to an ALJ

Dynamics of Earnings and Participation,

Allowed Versus Denied by ALJ, age 50-54 (French and Song, 2014)



Recovering MTEs

- Most lenient judges $\Rightarrow$ marginal case less in need of the benefit
- Marginal case heard by more lenient judge likely healthier

$$\frac{\partial E[y_i|\mathbf{X}_i = x, \Pr(x_i = 1|\mathbf{X}_i, z_i) = a]}{\partial a} = MTE(\mathbf{X}_i = x, V_i = a).$$

We model y_i as a polynomial in $\Pr(x_i = 1|\mathbf{X}_i, z_i)$

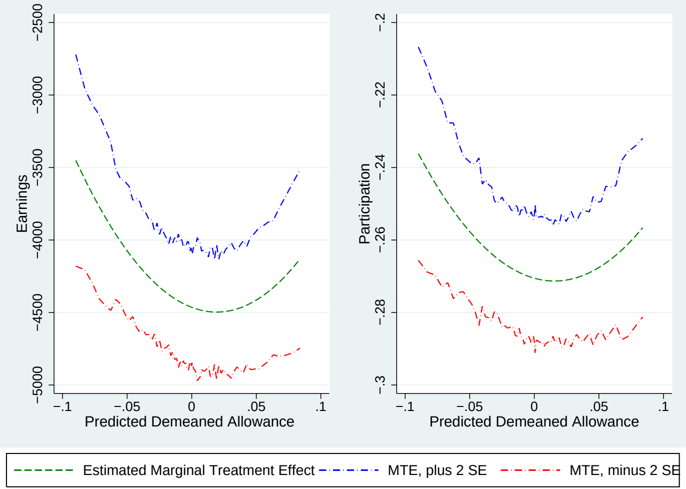


Figure 3: EARNINGS AND PARTICIPATION DECLINE FOR MARGINAL CASE WHEN ALLOWED.

MTE for mortality: 5 years after assignment (from Black French McCauley Song (2014))