Instrumental Variables

Introduction

MIXTAPE SESSIONS

Roadmap

- Introductions
  - Who Am I?
  - Who Am I?
  - What is This Course?
- Regression Review
  - Models vs. Estimands vs. Estimators
  - Regression Identification and Endogeneity
- Intro to IV
  - Instrument Validity and Relevance
  - The 2SLS Estimator

Who Am I?

- Groos Family Assistant Professor of Economics, Brown University

Who Am I?

- Groos Family Assistant Professor of Economics, Brown University
- A big fan of instrumental variable (IV) methods

Who Am I?

- Groos Family Assistant Professor of Economics, Brown University
- A big fan of instrumental variable (IV) methods
  - → Lottery- and non-lottery IVs in studies of educational quality
    (Angrist et al. 2016, 2017, 2021, 2022; Abdulkadiroğlu et al. 2016)
  - → Quasi-experimental evaluations of healthcare quality
    (Hull 2020; Abaluck et al. 2021, 2022)
  - → IV-based analyses of discrimination and bias
    (Arnold et al. 2020, 2021, 2022; Hull 2021; Bohren et al. 2022)
  - → Shift-share instruments and related designs
    (Borusyak et al. 2022; Borusyak and Hull 2021, 2022; Goldsmith-Pinkham et al. 2022)

Who Am I?

- Groos Family Assistant Professor of Economics, Brown University
- A big fan of instrumental variable (IV) methods
  - → Lottery- and non-lottery IVs in studies of educational quality
    (Angrist et al. 2016, 2017, 2021, 2022; Abdulkadiroğlu et al. 2016)
  - → Quasi-experimental evaluations of healthcare quality
    (Hull 2020; Abaluck et al. 2021, 2022)
  - → IV-based analyses of discrimination and bias
    (Arnold et al. 2020, 2021, 2022; Hull 2021; Bohren et al. 2022)
  - → Shift-share instruments and related designs
    (Borusyak et al. 2022; Borusyak and Hull 2021, 2022; Goldsmith-Pinkham et al. 2022)
- A constant student of IV (and econometrics more generally)

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances
  - → Far from comprehensive; stay tuned for more “mixtape tracks” that take deeper dives on particular topics (judge IV, etc)
  - → Emphasis on *practical*: IV is meant to be used, not just studied!

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances
  - → Far from comprehensive; stay tuned for more "mixtape tracks" that take deeper dives on particular topics (judge IV, etc)
  - → Emphasis on *practical*: IV is meant to be used, not just studied!
- Four one-hour lectures: from IV basics to recent topics

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances
  - → Far from comprehensive; stay tuned for more “mixtape tracks” that take deeper dives on particular topics (judge IV, etc)
  - → Emphasis on *practical*: IV is meant to be used, not just studied!
- Four one-hour lectures: from IV basics to recent topics
  - → Please ask questions in the Discord chat!
  - → I will try to stick to the schedule but may improvise slightly

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances
  - → Far from comprehensive; stay tuned for more “mixtape tracks” that take deeper dives on particular topics (judge IV, etc)
  - → Emphasis on *practical*: IV is meant to be used, not just studied!
- Four one-hour lectures: from IV basics to recent topics
  - → Please ask questions in the Discord chat!
  - → I will try to stick to the schedule but may improvise slightly
- Two 75-minute coding labs, applying what we’ve learned

What is This Course?

- A one-day intensive on IV, with focus on recent practical advances
  - → Far from comprehensive; stay tuned for more "mixtape tracks" that take deeper dives on particular topics (judge IV, etc)
  - → Emphasis on *practical*: IV is meant to be used, not just studied!
- Four one-hour lectures: from IV basics to recent topics
  - → Please ask questions in the Discord chat!
  - → I will try to stick to the schedule but may improvise slightly
- Two 75-minute coding labs, applying what we've learned
  - → I will be live-coding in Stata, but R code will also be available
  - → Goal: demonstrate both methods & how I think about applying them

Schedule

9:00-10:00am	Lecture 1: Regression Review; Regression Endogeneity; Introduction to IV
10:00-10:10am	*Break*
10:10-11:10am	Lecture 2: Understanding Instrument Validity; 2SLS Mechanics; Applications
11:10-11:15am	*Break*
11:15am-12:30pm	Coding Lab 1: Angrist and Krueger (1991)
12:30-1:30pm	*Lunch*
1:30-2:30pm	Lecture 3: Heterogeneous Treatment Effects; Characterizing Compliers; MTEs
2:30-2:40pm	*Break*
2:40pm-3:40pm	Lecture 4: Judge Leniency Designs; Shift-Share IV; New IV Frontiers
3:40-3:45pm	*Break*
3:45-5:00pm	Coding Lab 2: Stevenson (2018)
5:00-5:15pm	Closing Remarks

Roadmap

Models vs. Estimands vs. Estimators

- Three distinct objects (though not always clearly distinguished)

Models vs. Estimands vs. Estimators

- Three distinct objects (though not always clearly distinguished)
- *Parameters* come from models of how observed data are generated
  - → E.g. a structural supply/demand model or potential outcomes
  - → They set the target for an empirical analysis: what we want to know

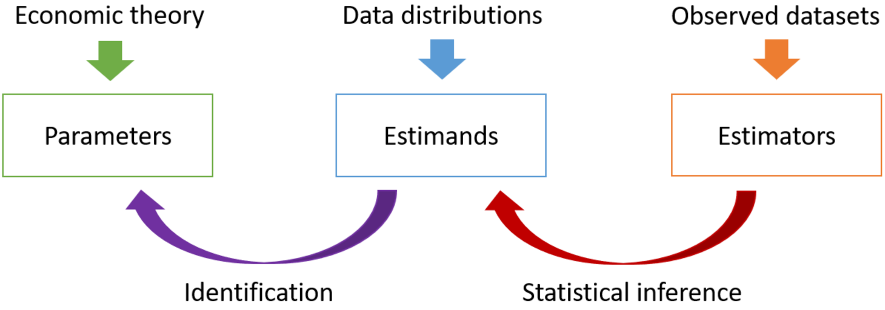
Models vs. Estimands vs. Estimators

- Three distinct objects (though not always clearly distinguished)
- *Parameters* come from models of how observed data are generated
  - → E.g. a structural supply/demand model or potential outcomes
  - → They set the target for an empirical analysis: what we want to know
- *Estimands* are functions of the distribution of observable data
  - → E.g. a difference in means or ratio of population regression coef's
  - → Make assumptions to link parameters & estimands ("identification")

Models vs. Estimands vs. Estimators

- Three distinct objects (though not always clearly distinguished)
- *Parameters* come from models of how observed data are generated
  - → E.g. a structural supply/demand model or potential outcomes
  - → They set the target for an empirical analysis: what we want to know
- *Estimands* are functions of the distribution of observable data
  - → E.g. a difference in means or ratio of population regression coef's
  - → Make assumptions to link parameters & estimands ("identification")
- *Estimators* are functions of the observed data itself (the "sample")
  - → E.g. a difference in sample means or ratio of OLS coefficients
  - → Since data are random, so are estimators. Each has a distribution
  - → Use knowledge of estimator distributions to make learn about estimands ("inference") and—hopefully—identified parameters

Identification vs. Estimation



This course will mostly focus on identification, but we'll cover some IV estimation / inference issues as well

Let’s Get Concrete

- Human capital theory (e.g. Becker, 1957) tells us that taking one-day IV intensives are likely to boost later-life productivity

Let's Get Concrete

- Human capital theory (e.g. Becker, 1957) tells us that taking one-day IV intensives are likely to boost later-life productivity
  - → Parameter: returns to taking this class β, measured in some outcome Y_i (e.g. lifetime top-5 pubs / earnings / twitter followers)
  - → Simple causal/structural model: $Y_i = \alpha + \beta D_i + \varepsilon_i$, where $D_i \in \{0, 1\}$ indicates taking this class

Let's Get Concrete

- Human capital theory (e.g. Becker, 1957) tells us that taking one-day IV intensives are likely to boost later-life productivity
  - → Parameter: returns to taking this class β, measured in some outcome Y_i (e.g. lifetime top-5 pubs / earnings / twitter followers)
  - → Simple causal/structural model: $Y_i = \alpha + \beta D_i + \varepsilon_i$, where $D_i \in \{0, 1\}$ indicates taking this class
- We see a sample of Y_i, D_i, and some other covariates $W_{1i}, \ldots, W_{Ki}$
  - → We fire up Stata and *reg y d w1-wk, r*. How do we interpret the output?

Let's Get Concrete

- Human capital theory (e.g. Becker, 1957) tells us that taking one-day IV intensives are likely to boost later-life productivity
  - $\rightarrow$ Parameter: returns to taking this class β, measured in some outcome Y_i (e.g. lifetime top-5 pubs / earnings / twitter followers)
  - $\rightarrow$ Simple causal/structural model: $Y_i = \alpha + \beta D_i + \varepsilon_i$, where $D_i \in \{0, 1\}$ indicates taking this class
- We see a sample of Y_i, D_i, and some other covariates $W_{1i}, \dots, W_{Ki}$
  - $\rightarrow$ We fire up Stata and *reg y d w1-wk, r*. How do we interpret the output?
- The OLS *estimator* $\widehat{\beta}^{OLS}$ consistently estimates the regression *estimand* β^{OLS} under relatively weak conditions (e.g. *i.i.d.* data)
  - $\rightarrow$ Stata tells us $\widehat{\beta}^{OLS}$ and what we can infer about β^{OLS} from it
  - $\rightarrow$ It *doesn't* directly tell us about the relationship between β^{OLS} and β

Population Regression

- *Def.:* the population regression of Y_i on $\mathbf{X}_i = [1, D_i, W_{1i}, \ldots, W_{Ki}]'$ is given by $Y_i = \mathbf{X}_i' \boldsymbol{\beta}^{OLS} + U_i$ where $E[\mathbf{X}_i U_i] = 0$

Population Regression

- *Def.:* the population regression of Y_i on $\mathbf{X}_i = [1, D_i, W_{1i}, \ldots, W_{Ki}]'$ is given by $Y_i = \mathbf{X}_i'\boldsymbol{\beta}^{OLS} + U_i$ where $E[\mathbf{X}_i U_i] = 0$
  - $\rightarrow$ Equivalently, $\boldsymbol{\beta}^{OLS} = E[\mathbf{X}_i\mathbf{X}_i']^{-1}E[\mathbf{X}_i Y_i]$ and $U_i = Y_i - \mathbf{X}_i'\boldsymbol{\beta}^{OLS}$
  - $\rightarrow$ $\boldsymbol{\beta}^{OLS}$ contains regression *coefficients*; U_i is the regression *residual*

Population Regression

- *Def.:* the population regression of Y_i on $\mathbf{X}_i = [1, D_i, W_{1i}, \ldots, W_{Ki}]'$ is given by $Y_i = \mathbf{X}_i'\boldsymbol{\beta}^{OLS} + U_i$ where $E[\mathbf{X}_i U_i] = 0$
  - → Equivalently, $\boldsymbol{\beta}^{OLS} = E[\mathbf{X}_i\mathbf{X}_i']^{-1}E[\mathbf{X}_i Y_i]$ and $U_i = Y_i - \mathbf{X}_i'\boldsymbol{\beta}^{OLS}$
  - → $\boldsymbol{\beta}^{OLS}$ contains regression *coefficients*; U_i is the regression *residual*
- Key point: we can always define $\boldsymbol{\beta}^{OLS}$ for any Y_i and $\mathbf{X}_i$ (assuming no perfect collinearity); this is what Stata estimates
  - → Specifically it computes $\widehat{\boldsymbol{\beta}}^{OLS} = (\frac{1}{N}\sum_i \mathbf{X}_i\mathbf{X}_i')^{-1}(\frac{1}{N}\sum_i \mathbf{X}_i Y_i)$ and uses large-sample asymptotics (LLN/CLT) to get a standard error

Population Regression

- *Def.:* the population regression of Y_i on $\mathbf{X}_i = [1, D_i, W_{1i}, \dots, W_{Ki}]'$ is given by $Y_i = \mathbf{X}_i'\boldsymbol{\beta}^{OLS} + U_i$ where $E[\mathbf{X}_i U_i] = 0$
  - $\rightarrow$ Equivalently, $\boldsymbol{\beta}^{OLS} = E[\mathbf{X}_i\mathbf{X}_i']^{-1}E[\mathbf{X}_i Y_i]$ and $U_i = Y_i - \mathbf{X}_i'\boldsymbol{\beta}^{OLS}$
  - $\rightarrow$ $\boldsymbol{\beta}^{OLS}$ contains regression *coefficients*; U_i is the regression *residual*
- Key point: we can always define $\boldsymbol{\beta}^{OLS}$ for any Y_i and $\mathbf{X}_i$ (assuming no perfect collinearity); this is what Stata estimates
  - $\rightarrow$ Specifically it computes $\widehat{\boldsymbol{\beta}}^{OLS} = (\frac{1}{N}\sum_i \mathbf{X}_i\mathbf{X}_i')^{-1}(\frac{1}{N}\sum_i \mathbf{X}_i Y_i)$ and uses large-sample asymptotics (LLN/CLT) to get a standard error
- But what if this *estimand* is not what we want?

Population Regression

- *Def.:* the population regression of Y_i on $\mathbf{X}_i = [1, D_i, W_{1i}, \ldots, W_{Ki}]'$ is given by $Y_i = \mathbf{X}_i'\boldsymbol{\beta}^{OLS} + U_i$ where $E[\mathbf{X}_i U_i] = 0$
  - $\rightarrow$ Equivalently, $\boldsymbol{\beta}^{OLS} = E[\mathbf{X}_i\mathbf{X}_i']^{-1}E[\mathbf{X}_i Y_i]$ and $U_i = Y_i - \mathbf{X}_i'\boldsymbol{\beta}^{OLS}$
  - $\rightarrow$ $\boldsymbol{\beta}^{OLS}$ contains regression *coefficients*; U_i is the regression *residual*

- Key point: we can always define $\boldsymbol{\beta}^{OLS}$ for any Y_i and $\mathbf{X}_i$ (assuming no perfect collinearity); this is what Stata estimates
  - $\rightarrow$ Specifically it computes $\widehat{\boldsymbol{\beta}}^{OLS} = (\frac{1}{N}\sum_i \mathbf{X}_i\mathbf{X}_i')^{-1}(\frac{1}{N}\sum_i \mathbf{X}_i Y_i)$ and uses large-sample asymptotics (LLN/CLT) to get a standard error

- But what if this *estimand* is not what we want?
  - $\rightarrow$ What if $\boldsymbol{\beta}^{OLS}$ fails to coincide with our economic parameter of interest (e.g. returns to mixtape workshops)?

You Can't Always Get What you Want...

- The model parameter in $Y_i = \alpha + \beta D_i + \varepsilon_i$ need not coincide with the regression coefficient in $Y_i = \alpha^{OLS} + \beta^{OLS} D_i + U_i$
  - $\rightarrow$ I.e. we may not have $Cov(D_i, \varepsilon_i) = 0$ (always have $Cov(D_i, U_i) = 0$)

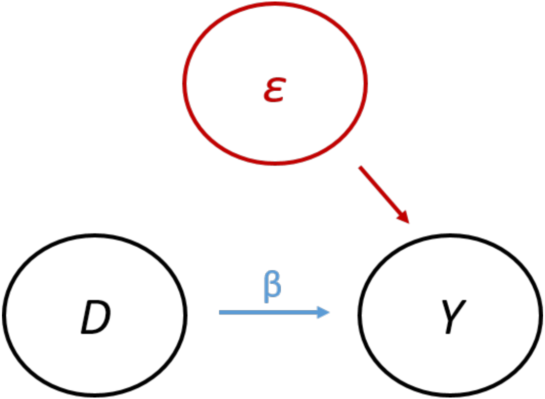
You Can't Always Get What you Want...

- The model parameter in $Y_i = \alpha + \beta D_i + \varepsilon_i$ need not coincide with the regression coefficient in $Y_i = \alpha^{OLS} + \beta^{OLS} D_i + U_i$
  - $\rightarrow$ I.e. we may not have $Cov(D_i, \varepsilon_i) = 0$ (always have $Cov(D_i, U_i) = 0$)
- Selection bias (a.k.a. omitted variables bias): students with higher latent earnings potential ε_i are more likely to take this class D_i
  - $\rightarrow$ $Cov(D_i, \varepsilon_i) > 0$ means $\beta^{OLS} > \beta$: overstate the returns-to-mixtape

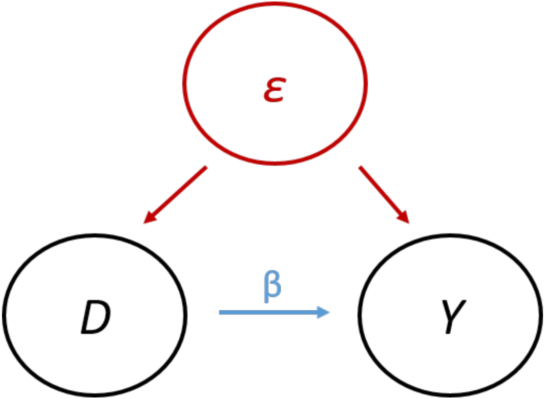
You Can't Always Get What you Want...

- The model parameter in $Y_i = \alpha + \beta D_i + \varepsilon_i$ need not coincide with the regression coefficient in $Y_i = \alpha^{OLS} + \beta^{OLS} D_i + U_i$
  - → I.e. we may not have $Cov(D_i, \varepsilon_i) = 0$ (always have $Cov(D_i, U_i) = 0$)
- Selection bias (a.k.a. omitted variables bias): students with higher latent earnings potential ε_i are more likely to take this class D_i
  - → $Cov(D_i, \varepsilon_i) > 0$ means $\beta^{OLS} > \beta$: overstate the returns-to-mixtape
- Adding more controls (e.g. demographics) may or may not help
  - → Projecting ε_i on X_i, we get $Y_i = \alpha + \beta D_i + \gamma X_i + \tilde{\varepsilon}_i$, $Cov(X_i, \tilde{\varepsilon}_i) = 0$
  - → Whether or not $Cov(D_i, \tilde{\varepsilon}_i) = 0$ depends on whether X_i sufficiently accounts for the confounding relationship $Cov(D_i, \varepsilon_i) \neq 0$

Regression "Exogeneity"



Regression "Endogeneity"



...But Sometimes, You Get What you Need

- Imagine this course was "oversubscribed," and admission was determined by lottery
  - → Among those interested in taking the course, a random sample denoted by $Z_i = 1$ was given access
  - → The rest, with $Z_i = 0$ not initially given access (maybe got in later)

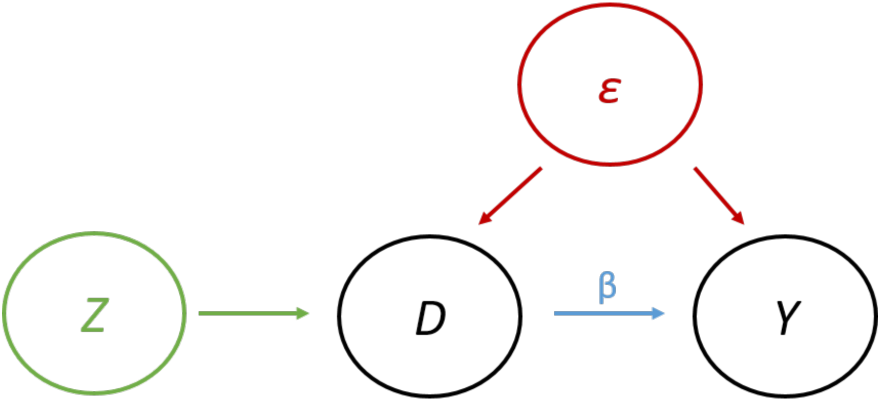
...But Sometimes, You Get What you Need

- Imagine this course was "oversubscribed," and admission was determined by lottery
  - → Among those interested in taking the course, a random sample denoted by $Z_i = 1$ was given access
  - → The rest, with $Z_i = 0$ not initially given access (maybe got in later)
- Intuitively, this external shock Z_i should be helpful for identifying β
  - → Affects D_i, so relevant to the "treatment" of interest
  - → Randomly assigned, so unconfounded by selection (unlike D_i)

...But Sometimes, You Get What you Need

- Imagine this course was "oversubscribed," and admission was determined by lottery
  - → Among those interested in taking the course, a random sample denoted by $Z_i = 1$ was given access
  - → The rest, with $Z_i = 0$ not initially given access (maybe got in later)
- Intuitively, this external shock Z_i should be helpful for identifying β
  - → Affects D_i, so relevant to the "treatment" of interest
  - → Randomly assigned, so unconfounded by selection (unlike D_i)
- Indeed, this leads us to IV estimands (and estimators)

The IV Solution



Roadmap

Instrument Validity and Relevance

- Causal/structural model $Y_i = \alpha + \beta D_i + \varepsilon_i$ and a candidate IV Z_i
  - $\rightarrow$ Single D_i and Z_i and no further controls, for now

Instrument Validity and Relevance

- Causal/structural model $Y_i = \alpha + \beta D_i + \varepsilon_i$ and a candidate IV Z_i
  - $\rightarrow$ Single D_i and Z_i and no further controls, for now
- Two key assumptions:
  - $\rightarrow$ Relevance: Z_i and D_i are correlated: $Cov(Z_i, D_i) \neq 0$
  - $\rightarrow$ Validity: Z_i and ε_i are *un*correlated: $Cov(Z_i, \varepsilon_i) = 0$

Instrument Validity and Relevance

- Causal/structural model $Y_i = \alpha + \beta D_i + \varepsilon_i$ and a candidate IV Z_i
  - $\rightarrow$ Single D_i and Z_i and no further controls, for now
- Two key assumptions:
  - $\rightarrow$ Relevance: Z_i and D_i are correlated: $Cov(Z_i, D_i) \neq 0$
  - $\rightarrow$ Validity: Z_i and ε_i are *un*correlated: $Cov(Z_i, \varepsilon_i) = 0$
- We then have identification:

$$Cov(Z_i, Y_i) = Cov(Z_i, \alpha + \beta D_i + \varepsilon_i) = \beta Cov(Z_i, D_i) + Cov(Z_i, \varepsilon_i)$$

$$= \beta Cov(Z_i, D_i), \text{ Implying } \beta = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)}$$

Instrument Validity and Relevance

- Causal/structural model $Y_i = \alpha + \beta D_i + \varepsilon_i$ and a candidate IV Z_i
  - $\rightarrow$ Single D_i and Z_i and no further controls, for now
- Two key assumptions:
  - $\rightarrow$ Relevance: Z_i and D_i are correlated: $Cov(Z_i, D_i) \neq 0$
  - $\rightarrow$ Validity: Z_i and ε_i are *un*correlated: $Cov(Z_i, \varepsilon_i) = 0$
- We then have identification:

$$Cov(Z_i, Y_i) = Cov(Z_i, \alpha + \beta D_i + \varepsilon_i) = \beta Cov(Z_i, D_i) + Cov(Z_i, \varepsilon_i)$$

$$= \beta Cov(Z_i, D_i), \text{ Implying } \beta = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)}$$

- This $\beta^{IV} = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,D_i)}$ is the (simple) IV *estimand*

Instrument Validity and Relevance

- Causal/structural model $Y_i = \alpha + \beta D_i + \varepsilon_i$ and a candidate IV Z_i
  - $\rightarrow$ Single D_i and Z_i and no further controls, for now
- Two key assumptions:
  - $\rightarrow$ Relevance: Z_i and D_i are correlated: $Cov(Z_i, D_i) \neq 0$
  - $\rightarrow$ Validity: Z_i and ε_i are *un*correlated: $Cov(Z_i, \varepsilon_i) = 0$
- We then have identification:

$$Cov(Z_i, Y_i) = Cov(Z_i, \alpha + \beta D_i + \varepsilon_i) = \beta Cov(Z_i, D_i) + Cov(Z_i, \varepsilon_i)$$

$$= \beta Cov(Z_i, D_i), \text{ Implying } \beta = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)}$$

- This $\beta^{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)}$ is the (simple) IV *estimand*
  - $\rightarrow$ Compare to the OLS estimand: $\beta^{OLS} = \frac{Cov(D_i, Y_i)}{Var(D_i)}$

“Reduced Form” and “First Stage”

- Note we can write

$$\beta^{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, D_i)/Var(Z_i)} = \frac{\rho^{OLS}}{\pi^{OLS}}$$

where ρ^{OLS} and π^{OLS} are two OLS coefficients:

“Reduced Form” and “First Stage”

- Note we can write

$$\beta^{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, D_i)/Var(Z_i)} = \frac{\rho^{OLS}}{\pi^{OLS}}$$

where ρ^{OLS} and π^{OLS} are two OLS coefficients:

$$Y_i = \kappa^{OLS} + \rho^{OLS} Z_i + V_i \quad \text{“reduced form”}$$
$$D_i = \mu^{OLS} + \pi^{OLS} Z_i + W_i \quad \text{“first stage”}$$

“Reduced Form” and “First Stage”

- Note we can write

$$\beta^{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, D_i)/Var(Z_i)} = \frac{\rho^{OLS}}{\pi^{OLS}}$$

where ρ^{OLS} and π^{OLS} are two OLS coefficients:

$$Y_i = \kappa^{OLS} + \rho^{OLS} Z_i + V_i \quad \text{“reduced form”}$$
$$D_i = \mu^{OLS} + \pi^{OLS} Z_i + W_i \quad \text{“first stage”}$$

- Sometimes we refer to the IV estimand as the “second stage”:

$$Y_i = \alpha^{IV} + \beta^{IV} D_i + U_i$$

where now $Cov(Z_i, U_i) = 0$. Thus “IV=RF/FS” ($\beta^{IV} = \rho^{OLS}/\pi^{OLS}$)

The 2SLS Estimator

- As with OLS, we estimate IV by sample analog:

$$\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i, Y_i)}{\widehat{Cov}(Z_i, D_i)} = \frac{\widehat{\rho}^{OLS}}{\widehat{\pi}^{OLS}}$$

where $\widehat{Cov}(X_i, W_i) = \frac{1}{N}\sum_i X_i W_i - \left(\frac{1}{N}\sum_i X_i\right)\left(\frac{1}{N}\sum_i W_i\right)$,
$\hat{\rho}^{OLS} = \widehat{Cov}(Z_i, Y_i)/\widehat{Var}(Z_i)$ and $\hat{\pi}^{OLS} = \widehat{Cov}(Z_i, D_i)/\widehat{Var}(Z_i)$

The 2SLS Estimator

- As with OLS, we estimate IV by sample analog:

$$\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i, Y_i)}{\widehat{Cov}(Z_i, D_i)} = \frac{\widehat{\rho}^{OLS}}{\widehat{\pi}^{OLS}}$$

where $\widehat{Cov}(X_i, W_i) = \frac{1}{N}\sum_i X_i W_i - \left(\frac{1}{N}\sum_i X_i\right)\left(\frac{1}{N}\sum_i W_i\right)$, $\hat{\rho}^{OLS} = \widehat{Cov}(Z_i, Y_i)/\widehat{Var}(Z_i)$ and $\hat{\pi}^{OLS} = \widehat{Cov}(Z_i, D_i)/\widehat{Var}(Z_i)$

$\rightarrow$ This is what Stata does when you type "ivreg2 y (d=z), r"

$\rightarrow$ Standard errors come from the usual large-sample asymptotics

The 2SLS Estimator

- As with OLS, we estimate IV by sample analog:

$$\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i, Y_i)}{\widehat{Cov}(Z_i, D_i)} = \frac{\widehat{\rho}^{OLS}}{\widehat{\pi}^{OLS}}$$

  where $\widehat{Cov}(X_i, W_i) = \frac{1}{N}\sum_i X_i W_i - \left(\frac{1}{N}\sum_i X_i\right)\left(\frac{1}{N}\sum_i W_i\right)$, $\hat{\rho}^{OLS} = \widehat{Cov}(Z_i, Y_i)/\widehat{Var}(Z_i)$ and $\hat{\pi}^{OLS} = \widehat{Cov}(Z_i, D_i)/\widehat{Var}(Z_i)$

  → This is what Stata does when you type “ivreg2 y (d=z), r”

  → Standard errors come from the usual large-sample asymptotics

- We will soon consider extensions of all of this, with controls / multiple instruments / etc

Angrist (1990): The "Draft Lottery Paper"

- Angrist famously used Vietnam-era draft eligibility as an instrument to estimate the earnings effects of military service
  - → Let Z_i be an indicator for draft eligibility, D_i be an indicator for military service, and Y_i measure later-life earnings

Angrist (1990): The "Draft Lottery Paper"

- Angrist famously used Vietnam-era draft eligibility as an instrument to estimate the earnings effects of military service
  - → Let Z_i be an indicator for draft eligibility, D_i be an indicator for military service, and Y_i measure later-life earnings
- Here $\beta^{IV} = \frac{Cov(Z_i,Y_i)/Var(Z_i)}{Cov(Z_i,D_i)/Var(Z_i)} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ has a special name, because Z_i is binary: the *Wald estimand*
  - → First stage $E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]$: effect of eligibility on the *probability* of military service (b/c D_i is binary)
  - → Reduced form $E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]$: effect of eligibility on adult earnings (measured in 1971, 1981...)

Angrist (1990): The “Draft Lottery Paper”

- Angrist famously used Vietnam-era draft eligibility as an instrument to estimate the earnings effects of military service
  - → Let Z_i be an indicator for draft eligibility, D_i be an indicator for military service, and Y_i measure later-life earnings
- Here $\beta^{IV} = \frac{Cov(Z_i,Y_i)/Var(Z_i)}{Cov(Z_i,D_i)/Var(Z_i)} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ has a special name, because Z_i is binary: the *Wald estimand*
  - → First stage $E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]$: effect of eligibility on the *probability* of military service (b/c D_i is binary)
  - → Reduced form $E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]$: effect of eligibility on adult earnings (measured in 1971, 1981...)
- IV interprets the latter causal effect in terms of the former

Draft Lottery Reduced Form, First Stage, and IV

IV Estimates of the Effects of Military Service on the Earnings of White Men born in 1950

	Earnings		Veteran Status		Wald Estimate of Veteran Effect
Earnings year	Mean	Eligibility Effect	Mean	Eligibility Effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8 (210.5)	.267	.159 (.040)	-2,741 (1,324)
1971	3,338	-325.9 (46.6)			-2050 (293)
1969	2,299	-2.0 (34.5)			

Note: Adapted from Table 5 in Angrist and Krueger (1999) and author tabulations. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.

Draft Lottery Reduced Form and First Stage Visualized

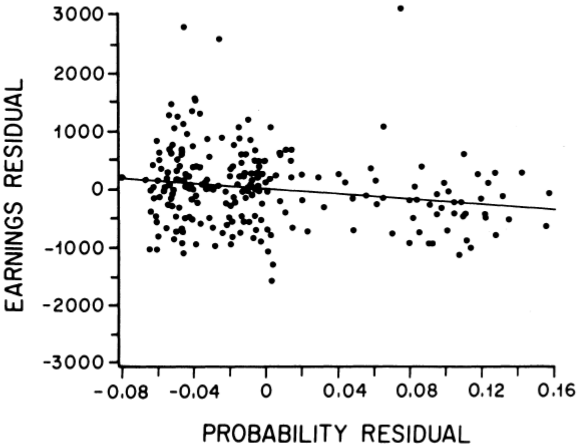


FIGURE 3. EARNINGS AND THE PROBABILITY OF VETERAN STATUS BY LOTTERY NUMBER

Notes: The figure plots mean W-2 compensation in 1981–4 against probabilities of veteran status by cohort and groups of five consecutive lottery numbers for white men born 1950–3. Plotted points consist of the average residuals (over four years of earnings) from regressions on period and cohort effects. The slope of the least-squares regression line drawn through the points is −2,384, with a standard error of 778, and is an estimate of α in the equation

$$\bar{y}_{ctj} = \beta_c + \delta_t + \hat{p}_{cj}\alpha + \bar{u}_{ctj}.$$