

# Instrumental Variables with Heterogeneous Effects

Magne Mogstad

## Linear IV with heterogeneous effects

When estimating the effect of  $D$  on  $Y$  with IV  $Z$  the standard textbook case presents the outcome equation with homogenous effects

$$Y = \alpha + \beta D + U$$

But we can link observed outcome  $Y$  to potential outcomes ( $Y_0, Y_1$ )

$$\begin{aligned} Y_i &= \underbrace{E[Y_0]}_{\alpha} + \underbrace{(Y_1 - Y_0)}_{\beta_i} D + \underbrace{Y_0 - E[Y_0]}_{U_i} \\ &\equiv \alpha + \beta D + U \end{aligned}$$

**What does linear IV identify when treatment effects are heterogeneous?**

This question is the focus of much of applied micro.

Arguably reverse engineering. Like playing Jeopardy

Later we start with a question (target parameter), and then ask how to answer it (identify and estimate target parameter)]

## Heterogeneous potential outcome set-up

Instrument initiates a causal chain, whereby  $Z$  affects the variable of interest  $D$  which in turn affects  $Y$

Keeping this in mind we can adopt the potential outcome set-up:

- ▶  $D_z$  is treatment status at instrument value  $Z = z$
- ▶  $Y_{d,z}$  is outcome of individual  $i$  if he receives treatment  $D = d$  and instrument value  $Z = z$

We can now define various causal effects:

- ▶  $Y_{1,z} - Y_{0,z}$
- ▶  $Y_{D,1} - Y_{D,0}$
- ▶  $Y_{1,z} - Y_{0,z}$
- ▶  $Y_{d,1} - Y_{d,0}$
- ▶  $D_1 - D_0$

# Heterogeneous potential outcome set-up

The first assumption in the heterogeneous effects set-up is random assignment

## Random assignment

$$Y_{d,z}, D_z \perp Z \forall d, z$$

This is sufficient to identify average causal effect of  $Z$  on  $Y$  (and of  $Z$  on  $D$ ):

$$\begin{aligned} E[Y|Z=1] - E[Y|Z=0] \\ &= E[Y_{D_1,1}|Z=1] - E[Y_{D_0,0}|Z=0] \\ &= E[Y_{D_1,1} - Y_{D_0,0}] \end{aligned}$$

# Heterogeneous potential outcome set-up

The second assumption in the heterogeneous effects set-up is the exclusion restriction

## Exclusion restriction

$$Y_{d,1} = Y_{d,0}$$

This states that any effect of  $Z$  on  $Y$  must be via an effect of  $Z$  on  $D$

The **exclusion restriction** is often expressed by omitting  $Z$  in equation of interest:  $Y = \alpha + \beta \cdot D + U$

*Random assignment + exclusion restriction = instrument exogeneity*

Conceptually distinct problems – argue one at the time!

# Heterogeneous potential outcome set-up

The third assumption in the heterogeneous effects set up is the existence of a first stage

## First stage

$$E[D_1 - D_0] \neq 0$$

Which requires the instrument  $Z$  to have some effect on the average probability of treatment

Note: For the (usual) statistical inference (which relies on the standard first-order asymptotic approximation invoked in large-sample theory), the first stage should not be too close to zero (more on that later)

# Heterogeneous potential outcome set-up

The fourth assumption in the heterogeneous effects set-up is monotonicity

## Monotonicity

$$D_1 \geq D_0 \quad \forall i \text{ (or vice versa)}$$

Which says that all those affected by the instrument are affected in the same direction

Note: Uniformity would be a better terminology.

Monotonicity assumption does not imply that treatment is a monotonic function of the instrument (which becomes relevant with multiple instruments or when instrument takes multiple values).

## Local Average Treatment Effect (LATE)

A variable  $Z$  is an instrumental variable for the causal effect of  $D$  on  $Y$  if the following assumptions hold:

1. Random assignment:  $Y_{d,z}, D_z \perp Z \forall d, z$
2. Exclusion Restriction:  $Y_{d,1} = Y_{d,0} = Y_d$
3. Monotonicity:  $D_1 \geq D_0$  , or vice versa
4. First-Stage:  $E[D_1 - D_0] \neq 0$

The Wald estimand then gives the *Local Average Treatment Effect*:

$$\beta_{IV} = E[\beta | D_1 = 1, D_0 = 0]$$

the average treatment effect for those affected by the instrument



## Local Average Treatment Effect (LATE)

A variable  $Z$  is an instrumental variable for the causal effect of  $D$  on  $Y$  if the following assumptions hold:

1. Random assignment:  $Y_{d,z}, D_z \perp Z \forall d, z$ 
  - ▶ gives the causal effect of  $Z$  on  $D$  (1st stage) and  $Y$  (reduced form)
2. Exclusion Restriction:  $Y_{d,1} = Y_{d,0} = Y_d$
3. Monotonicity:  $D_1 \geq D_0$ , or vice versa
4. First-Stage:  $E[D_1 - D_0] \neq 0$

The Wald estimand then gives the *Local Average Treatment Effect*:

$$\beta_{IV} = E[\beta | D_1 = 1, D_0 = 0]$$

the average treatment effect for those affected by the instrument

## Local Average Treatment Effect (LATE)

A variable  $Z$  is an instrumental variable for the causal effect of  $D$  on  $Y$  if the following assumptions hold:

1. Random assignment:  $Y_{d,z}, D_z \perp Z \forall d, z$ 
  - ▶ gives the causal effect of  $Z$  on  $D$  (1st stage) and  $Y$  (reduced form)
2. Exclusion Restriction:  $Y_{d,1} = Y_{d,0} = Y_d$ 
  - ▶ so that the causal effect of  $Z$  on  $Y$  is only due to the effect of  $Z$  on  $D$
3. Monotonicity:  $D_1 \geq D_0$ , or vice versa
4. First-Stage:  $E[D_1 - D_0] \neq 0$

The Wald estimand then gives the *Local Average Treatment Effect*:

$$\beta_{IV} = E[\beta | D_1 = 1, D_0 = 0]$$

the average treatment effect for those affected by the instrument

## Local Average Treatment Effect (LATE)

A variable  $Z$  is an instrumental variable for the causal effect of  $D$  on  $Y$  if the following assumptions hold:

1. Random assignment:  $Y_{d,z}, D_z \perp Z \forall d, z$ 
  - ▶ gives the causal effect of  $Z$  on  $D$  (1st stage) and  $Y$  (reduced form)
2. Exclusion Restriction:  $Y_{d,1} = Y_{d,0} = Y_d$ 
  - ▶ so that the causal effect of  $Z$  on  $Y$  is only due to the effect of  $Z$  on  $D$
3. Monotonicity:  $D_1 \geq D_0$ , or vice versa
  - ▶ to avoid offsetting effects
4. First-Stage:  $E[D_1 - D_0] \neq 0$

The Wald estimand then gives the *Local Average Treatment Effect*:

$$\beta_{IV} = E[\beta | D_1 = 1, D_0 = 0]$$

the average treatment effect for those affected by the instrument

## Local Average Treatment Effect (LATE)

A variable  $Z$  is an instrumental variable for the causal effect of  $D$  on  $Y$  if the following assumptions hold:

1. Random assignment:  $Y_{d,z}, D_z \perp Z \forall d, z$ 
  - ▶ gives the causal effect of  $Z$  on  $D$  (1st stage) and  $Y$  (reduced form)
2. Exclusion Restriction:  $Y_{d,1} = Y_{d,0} = Y_d$ 
  - ▶ so that the causal effect of  $Z$  on  $Y$  is only due to the effect of  $Z$  on  $D$
3. Monotonicity:  $D_1 \geq D_0$ , or vice versa
  - ▶ to avoid offsetting effects
4. First-Stage:  $E[D_1 - D_0] \neq 0$ 
  - ▶ because we need treatment variation in the sample

The Wald estimand then gives the *Local Average Treatment Effect*:

$$\beta_{IV} = E[\beta | D_1 = 1, D_0 = 0]$$

the average treatment effect for those affected by the instrument

## Local Average Treatment Effect (LATE)

Wald estimand can be interpreted as effect of treatment on outcomes for individuals who were treated because  $Z = 1$ , but who would not have been treated otherwise

To see why this is so, we can divide the population into four groups:

1. **Compliers:**  $D_1 = 1$  and  $D_0 = 0$ ;
2. **Always-takers:**  $D_1 = 1$  and  $D_0 = 1$ ;
3. **Never-takers:**  $D_1 = 0$  and  $D_0 = 0$ ;
4. **Defiers:**  $D_1 = 0$  and  $D_0 = 1$ ;

Note: The terminology is much used but a bit confusing (at least to me).

Always-takers are not always taking treatment. Never-takers are not never taking treatment. Everything is specific to the instrument at hand.

With other instruments, always-taker, never-taker and complier status may change

## Local Average Treatment Effect: Proof

We saw that (by independence)

$$E[Y|Z = 1] - E[Y|Z = 0] = E[Y_{D_1} - Y_{D_0}]$$

The average causal effect of  $Z$  on  $Y$  can be written as weighted average of the causal effects of the four sub-populations:

$$\begin{aligned} E[Y_{D_1} - Y_{D_0}] = & \\ & E[Y_{D_1} - Y_{D_0} | \text{Complier}] \times P(D_1 = 1, D_0 = 0) \\ & + E[Y_{D_1} - Y_{D_0} | \text{Never taker}] \times P(D_1 = 0, D_0 = 0) \\ & + E[Y_{D_1} - Y_{D_0} | \text{Always taker}] \times P(D_1 = 1, D_0 = 1) \\ & + E[Y_{D_1} - Y_{D_0} | \text{Defier}] \times P(D_1 = 0, D_0 = 1) \end{aligned}$$

## Local Average Treatment Effect: Proof

We saw that (by independence)

$$E[Y|Z = 1] - E[Y|Z = 0] = E[Y_{D_1} - Y_{D_0}]$$

The average causal effect of  $Z$  on  $Y$  can be written as weighted average of the causal effects of the four sub-populations:

$$\begin{aligned} E[Y_{D_1} - Y_{D_0}] = & E[Y_{D_1} - Y_{D_0} | \text{Complier}] \times P(D_1 = 1, D_0 = 0) \\ & + \underbrace{E[Y_{D_1} - Y_{D_0} | \text{Never taker}]}_{= Y_0 - Y_0 = 0} \times P(D_1 = 0, D_0 = 0) \\ & + \underbrace{E[Y_{D_1} - Y_{D_0} | \text{Always taker}]}_{= Y_1 - Y_1 = 0} \times P(D_1 = 1, D_0 = 1) \\ & + E[Y_{D_1} - Y_{D_0} | \text{Defier}] \times \underbrace{P(D_1 = 0, D_i(0) = 1)}_{= 0} \end{aligned}$$

## Local Average Treatment Effect: Proof

By monotonicity  $D_1 \geq D_0$ , which implies that there are no defiers.

$$\begin{aligned} E[Y|Z = 1] - E[Y|Z = 0] \\ = E[Y_1 - Y_0 | \text{Complier}] \times P(D_1 = 1, D_0 = 0) \end{aligned}$$

and by independence and monotonicity we can show that

$$E[D|Z = 1] - E[D|Z = 0] = E[D_1 - D_0] = P(D_1 = 1, D_0 = 0)$$

From this it follows that the Wald estimand is equal to the average treatment effect on the compliers

$$\begin{aligned} \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} \\ = \frac{E[Y_1 - Y_0 | \text{Complier}] \times P(D_1 = 1, D_0 = 0)}{P(D_1 = 1, D_0 = 0)} \\ = E[Y_1 - Y_0 | \text{Complier}] \end{aligned}$$



## LATE: Interpretation and relevance

With heterogeneous effects IV estimates the average causal effect for compliers

Different valid instruments for same causal relation therefore estimate different things (different groups of compliers)

- ▶ Overidentifying restrictions test (Sargan test) might reject even if all instruments are valid.
- ▶ Policy-relevance of IV estimate depends on policy relevance of instrument

Note: We **cannot** identify the compliers because we can never observe both  $D_0$  and  $D_1$  (thus, we don't know who the compliers are)

- ▶ those with  $Z = 1$  and  $D = 1$  can be compliers or always-takers
- ▶ those with  $Z = 0$  and  $D = 0$  can be compliers or never-takers

## Compliers: How many and what do they look like

The size of the complier group is the Wald 1st-stage:

$$P(D_1 = 1, D_0 = 0) = E[D|Z = 1] - E[D|Z = 0]$$

Or among the treated

$$\begin{aligned} P(D_1 - D_0 = 1 | D = 1) &= \frac{P(D = 1 | D_1 > D_0) P(D_1 > D_0)}{P(D = 1)} \\ &= \frac{P(Z = 1)(E[D|Z = 1] - E[D|Z = 0])}{P(D = 1)} \end{aligned}$$

We cannot identify compliers, but we can describe them

$$\begin{aligned} \frac{P(X = x | D_1 > D_0)}{P(X = x)} &= \frac{P(D_1 > D_0 | X = x)}{P(D_1 > D_0)} \\ &= \frac{E[D|Z = 1, X = x] - E[D|Z = 0, X = x]}{E[D|Z = 1] - E[D|Z = 0]} \end{aligned}$$

## LATE extensions

Until now we considered the IV model with heterogeneity in the simple case of

- ▶ average effects (for compliers)
- ▶ binary treatment, binary instrument
- ▶ no covariates

What happens when we relax these assumptions?

Angrist and Pischke (2009, p. 173) write that “The econometric tool remains 2SLS and the interpretation remains fundamentally similar to the basic LATE result, with a few bells and whistles.”

Is this really true? (spoiler: no, it's not!)

But first, let's see that even in the simple case, linear IV is not revealing all the information about potential outcomes available in the data

## Extension I: Counterfactual distributions

## Counterfactual distributions

Imbens & Rubin (1997) show that we can estimate more than average causal effects for compliers

They show how to recover the complete marginal distributions of the outcome

- ▶ under different treatments for the compliers
- ▶ under the treatment for the always-takers
- ▶ without the treatment for the never-takers

These results allow us to draw inference about effect on the outcome distribution of compliers (QTE of compliers)

Can also be used to test instrument exogeneity & monotonicity

Even exactly identified models can have testable implications (unlike what is claimed in MHE).

# Counterfactual distributions

First introduce some shorthand notation

$$C_i = n \iff D_1 = D_0 = 0$$

$$C_i = a \iff D_1 = D_0 = 1$$

$$C_i = c \iff D_1 = 1, D_0 = 0$$

$$C_i = d \iff D_1 = 0, D_0 = 1$$

For the different combinations of  $Z$  and  $D$ , we know the following:

		D	
		0	1
Z	0	n, c	a
	1	n	a, c

# Counterfactual distributions

## Distribution of types

Since  $Z$  is random we know that the distribution of types  $a$ ,  $n$ ,  $c$  is the same for each value of  $Z$  and in the population as a whole

Therefore, this...

		D	
		0	1
Z	0	n, c	a
	1	n	a, c

...implies the following:

$$p_a = \Pr(D = 1 | Z = 0)$$

$$p_n = \Pr(D = 0 | Z = 1)$$

$$p_c = 1 - p_a - p_n$$

# Counterfactual distributions

## Identifying distributions

Let's use the following notation for the observed marginal distribution of  $Y$  conditional on  $Z$  and  $D$ :

$$f_{zd}(y) \equiv f(y|Z = z, D = d)$$

Therefore, this...

		D	
		0	1
Z	0	n, c	a
	1	n	a, c

...implies the following:

$$f_{10}(y) = g_n(y)$$

$$f_{01}(y) = g_a(y)$$

$$f_{00}(y) = g_{c0}(y) \cdot (p_c / (p_c + p_n)) \\ + g_n(y) \cdot (p_n / (p_c + p_n))$$

$$f_{11}(y) = g_{c1}(y) \cdot (p_c / (p_c + p_a)) \\ + g_a(y) \cdot (p_a / (p_c + p_a))$$



# Counterfactual distributions

## Example

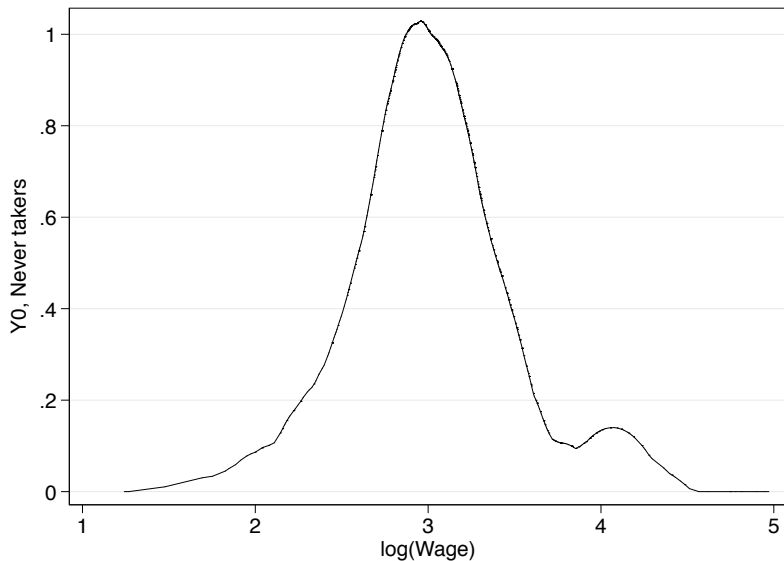
To illustrate the above, consider Dutch data (see Ketel et al., 2016, AEJ applied).

- ▶ Lottery outcome as instrument of medical school completion
  - ▶  $D = 1$  if completed medical school
  - ▶  $Z = 1$  if offered medical school after successful lottery

. ta z d				
	z	d		
		0	1	Total
	0	269	187	456
	1	71	949	1,020
Total		340	1,136	1,476

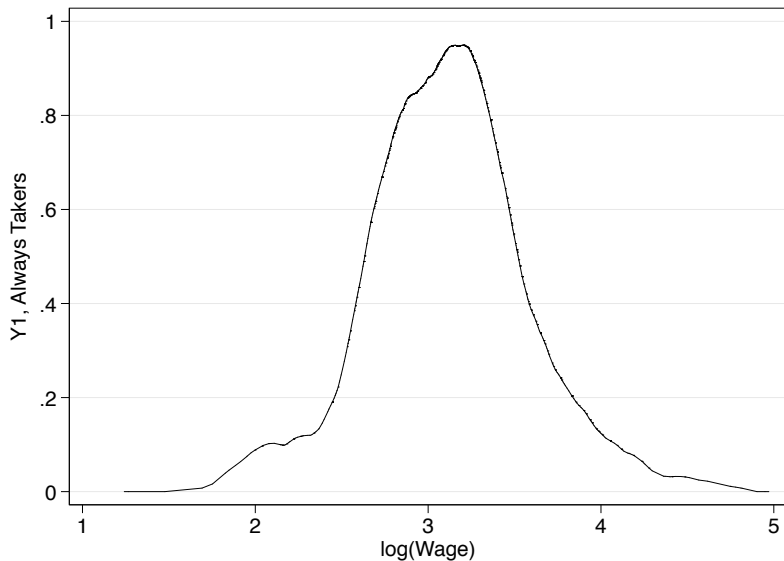
# Counterfactual distributions

$$f_{10}(y) = g_n(y)$$



# Counterfactual distributions

$$f_{01}(y) = g_a(y)$$



# Counterfactual distributions

We have seen that we can estimate  $p_a$ ,  $p_n$ ,  $p_c$  and also  $g_n(y)$  ( $=f_{10}(y)$ ) and  $g_a(y)$  ( $=f_{01}(y)$ )

By rearranging the following

$$f_{00}(y) = g_{c0}(y) \cdot (p_c / (p_c + p_n)) + g_n(y) \cdot (p_n / (p_c + p_n))$$

$$f_{11}(y) = g_{c1}(y) \cdot (p_c / (p_c + p_a)) + g_a(y) \cdot (p_a / (p_c + p_a))$$

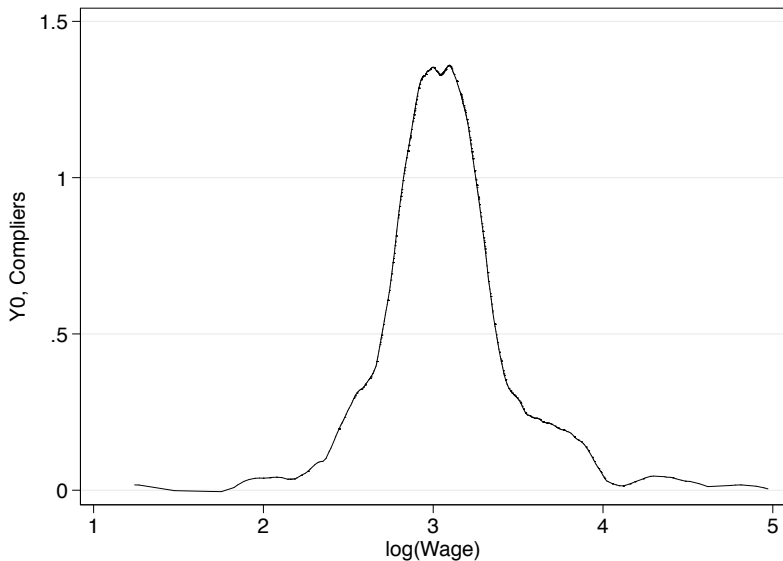
we can back out the counterfactual distributions for the compliers:

$$g_{c0}(y) = f_{00}(y) \cdot (p_c + p_n) / p_c - f_{10}(y) \cdot p_n / p_c$$

$$g_{c1}(y) = f_{11}(y) \cdot (p_c + p_a) / p_c - f_{01}(y) \cdot p_a / p_c$$

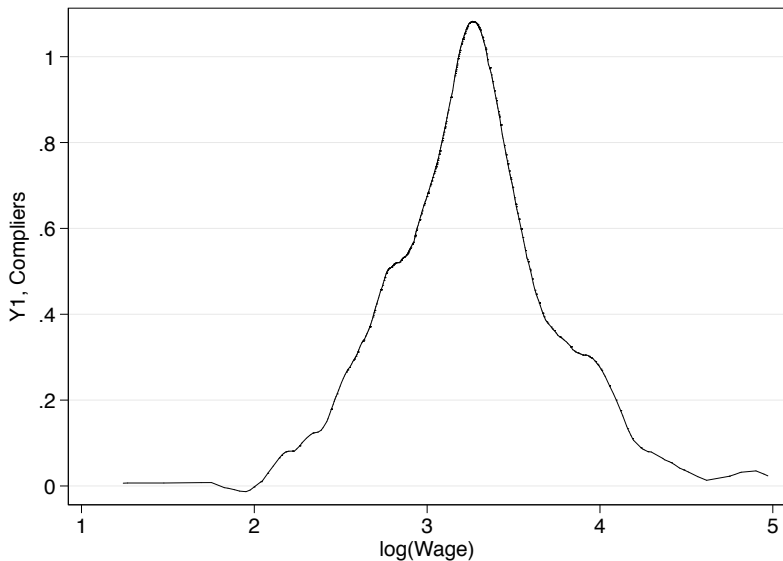
# Counterfactual distributions

$$g_{c0}(y) = f_{00}(y) \cdot (p_c + p_n)/p_c - f_{10}(y) \cdot p_n/p_c$$

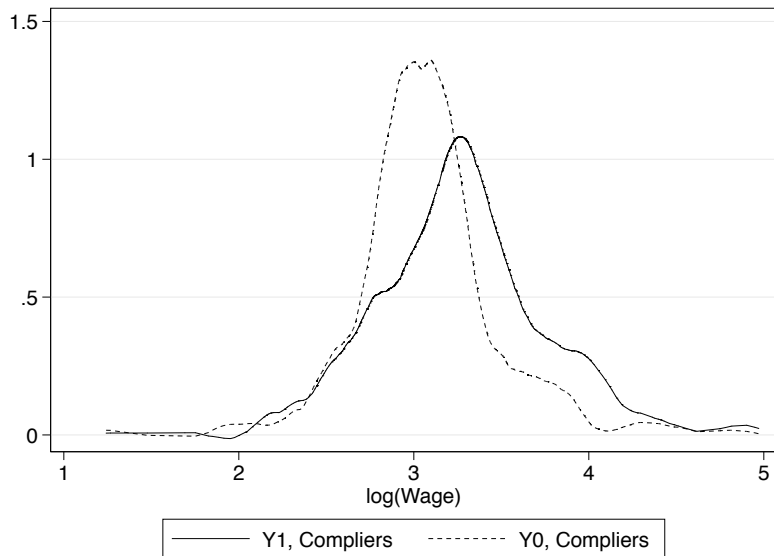


# Counterfactual distributions

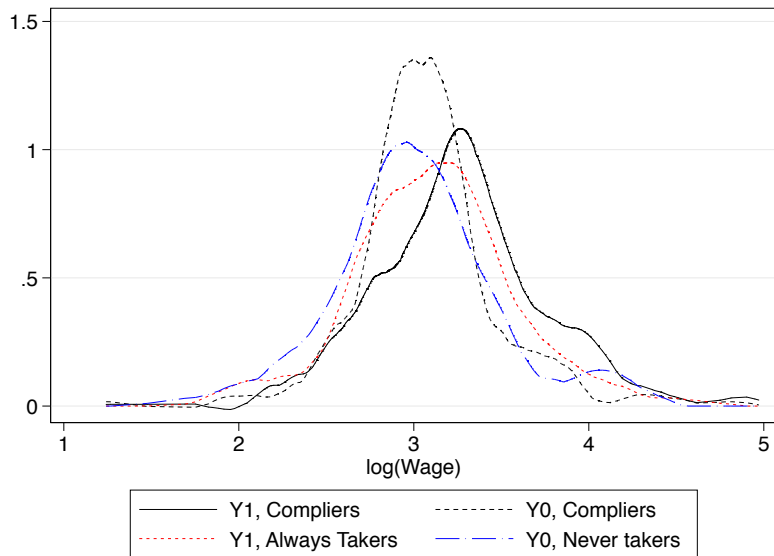
$$g_{c1}(y) = f_{11}(y) \cdot (p_c + p_a)/p_c - f_{01}(y) \cdot p_a/p_c$$



# Counterfactual distributions



# Counterfactual distributions





# Counterfactual distributions

We can also show that

$$E[Y_1|C = c] = \frac{E[Y \cdot D|Z = 1] - E[Y \cdot D|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]}$$

and

$$E[Y_0|C = c] = \frac{E[Y \cdot (1 - D)|Z = 1] - E[Y \cdot (1 - D)|Z = 0]}{E[1 - D|Z = 1] - E[1 - D|Z = 0]}$$

# Counterfactual distributions

```
. ivregress 2sls lnw (d = z), robust noheader
```

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lnw							
d		.1871175	.0485501	3.85	0.000	.0919609	.282274
_cons		3.010613	.0382073	78.80	0.000	2.935728	3.085498

```
. g y1 = lnw*d  
. ivregress 2sls y1 (d = z), robust noheader
```

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
y1							
d		3.264167	.0387887	84.15	0.000	3.188142	3.340191
_cons		-.0617161	.0275252	-2.24	0.025	-.1156644	-.0077678

```
. g y0 = lnw*(1-d)  
. g md = 1-d  
. ivregress 2sls y0 (md = z), robust noheader
```

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
y0							
md		3.077049	.0293153	104.96	0.000	3.019592	3.134506
_cons		-.0047203	.0047455	-0.99	0.320	-.0140213	.0045806

```
. di 3.264167 - 3.077049  
.187118
```

# Testing instrument validity

The above discussion points to a test for instrument validity (or, equivalently, a test for monotonicity given exogeneity)

Basic idea: Under the IV assumptions, the complier distribution should actually be a distribution

- ▶ By definition, probability can never be negative.
- ▶ Thus, density can never be negative
- ▶ For binary  $Y$ , it means that  $E(Y|C = c)$  needs to be between 0 and 1

Kitagawa (2015) develops a formal statistical test based on these implication

## Extension II: Multiple instruments

## LATE with multiple instruments

Assume we have 2 mutually exclusive (and for simplicity independent) binary instruments

(Without loss of generality: make two non-exclusive instruments mutually exclusive by working with  $Z_1(1-Z_2)$ ,  $Z_2(1-Z_1)$ ,  $Z_1Z_2$ )

We can then estimate two different LATEs:

$$\begin{aligned}\beta_{Z_j} &= \frac{\text{cov}(Y, Z_j)}{\text{cov}(D, Z_j)} \\ &= E[Y_1 - Y_0 | D_{Z_j=1} - D_{Z_j=0} = 1]\end{aligned}$$

In practice researchers often combine the instruments using 2SLS

The 2SLS estimator is

$$\beta_{2SLS} = \frac{\text{cov}(Y, \hat{D})}{\text{cov}(D, \hat{D})}$$

where  $\hat{D} = \pi_1 Z_1 + \pi_2 Z_2$

# LATE with multiple instruments

Expanding  $\beta_{2SLS}$  gives

$$\begin{aligned}\beta_{2SLS} &= \pi_1 \frac{\text{cov}(Y, Z_1)}{\text{cov}(D, \hat{D})} + \pi_2 \frac{\text{cov}(Y, Z_2)}{\text{cov}(D, \hat{D})} \\ &= \pi_1 \frac{\text{cov}(D, Z_1)}{\text{cov}(D, \hat{D})} \frac{\text{cov}(Y, Z_1)}{\text{cov}(D, Z_1)} + \pi_2 \frac{\text{cov}(D, Z_2)}{\text{cov}(D, \hat{D})} \frac{\text{cov}(Y, Z_2)}{\text{cov}(D, Z_2)} \\ &= \psi \beta_{Z_1} + (1 - \psi) \beta_{Z_2}\end{aligned}$$

where

$$\psi \equiv \frac{\pi_1 \text{cov}(D, Z_1)}{\pi_1 \text{cov}(D, Z_1) + \pi_2 \text{cov}(D, Z_2)}$$

is the relative strength of  $Z_1$  in the first stage

Under assumptions 1-4, the 2SLS estimate is an instrument-strength weighted average of the instrument specific LATEs

## Questions with multiple instruments?

- ▶ What question does the 2SLS weighted average of LATEs answer?
- ▶ Why not some other weighted average (e.g. use GMM or LIML)?
- ▶ Is monotonicity more restrictive with multiple instruments?
- ▶ Can one do without monotonicity?

Some papers do IV with heterogeneity without invoking monotonicity

See, for example, much of the work by Manski but also Heckman and Pinto (2018) and Mogstad, Walters and Torgovitsky (2019)

# Interpreting Monotonicity with Multiple Instruments

## Notation

- ▶ Binary treatment  $D \in \{0, 1\}$
- ▶ Potential treatments  $D_z$  for instrument values  $z \in Z$

## IA monotonicity condition (IAM)

For all  $z, z' \in Z$  either:

- ▶  $D_z \geq D_{z'}$  or
  - ▶  $D_z \leq D_{z'}$
- 
- ▶ IA Monotonicity is uniformity, not monotonicity
  - ▶ Pairwise instrument shifts push everyone to or from treatment



# Choice Behavior

- ▶ Random utility model

$V(d, z)$  is indirect utility from choosing  $d$  when instrument  $z$ :

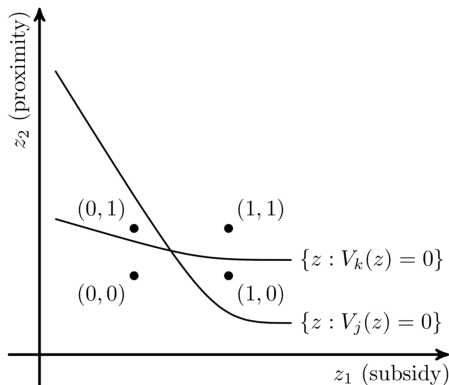
$$D_z = \arg \max_{d \in \{0,1\}} V(d, z) = \mathbb{1}[V_z \geq 0]$$

where  $V(z) \equiv V(1, z) - V(0, z)$  is net indirect utility

Illustrative example:

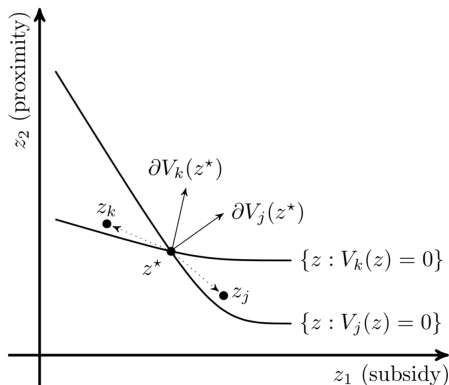
- ▶  $D_z \in \{0, 1\}$  is whether to attend college
- ▶  $Z_1$  is a tuition subsidy
- ▶  $Z_2$  is proximity to a college
- ▶  $D_z$  should be an increasing function of  $z$
- ▶ Neither implies nor is implied by IA monotonicity
- ▶ What is implied by IA monotonicity? Restrictions on  $V(z)$ ?

# Binary Instruments



- ▶ IA monotonicity does not permit individuals to differ in responses
- ▶ All individuals must find either tuition or distance more compelling

# Continuous Instruments



- ▶  $z^*$  is a point of indifference for  $j$  and  $k$
- ▶ IA monotonicity fails if marginal rates of substitution are different

# Homogenous Marginal Rates of Substitution

- ▶ Let  $z^*$  be a point at which  $V(z)$  is differentiable

- ▶ Let  $\mathcal{I}(z^*) = \{i \in \mathcal{I} : V(z^*) = 0\}$

- ▶ IA monotonicity implies that

$$\partial_1 V_j(z^*) \partial_2 V_k(z^*) = \partial_1 V_k(z^*) \partial_2 V_j(z^*), \forall j, k \in \mathcal{I}(z^*)$$

- ▶ Natural discrete choice specification:

$$V(z) = B_0 + B_1 Z_1 + 1 \times Z_2$$

- ▶ Where  $(B_0, B_1)$  are unobserved
  - ▶  $B_1$  controls variation in taste for tuition relative to proximity
  - ▶ IA monotonicity requires no variation over individuals:  $\text{Var}(B_1) = 0$

## Extension III: Variable treatment intensity

## Variable treatment intensity

Assume treatment is no longer binary but varies in its level

$$S \in \{0, 1, 2, \dots, J\}$$

such as for example years of schooling.

We can then define potential outcomes indexed by the *level* of treatment

$$Y_S$$

Potential treatments (schooling level) are as before indexed by the value of the instrument

$$S_Z$$

so that with a binary instrument the observed level of schooling is

$$S = ZS_1 + (1 - Z)S_0$$

## Variable treatment intensity

The observed outcome

$$Y = \sum_{s=0}^J Y_s \mathbb{1}[S = s] = Y_0 + \sum_{s=1}^J (Y_s - Y_{s-1}) \mathbb{1}[S \geq s]$$

The average effect of the  $s$ -th year of schooling is then

$$E[Y_s - Y_{s-1}]$$

and we have now  $J$  different treatment effects

Even so, researchers often estimate a linear-in-parameter model:

$$Y = \alpha + \beta S + u$$

One possibility is to take the linearity restriction literally

Another option is to reverse-engineer

(A third possibility is to start with a target parameter.....)

## Variable treatment intensity

As before we need to make an independence assumption

$$Y_{s,z}, S_z \perp Z \quad \forall s, z$$

and an exclusion restriction

$$Y_{s,z} = Y_s$$

We further need a monotonicity assumption

$$S_1 \geq S_0$$

and instrument relevance

$$E[S_1 - S_0] \neq 0$$



# Variable treatment intensity

Example with 3 levels

Monotonicity implies

$$\mathbb{1}[S_1 \geq s] - \mathbb{1}[S_0 \geq s] \in \{0, 1\}$$

so that

$$\Pr(\mathbb{1}[S_1 \geq s] > \mathbb{1}[S_0 \geq s]) = \Pr(S_1 \geq s > S_0)$$

if this probability is greater than 0, then the instrument affects the incidence of treatment level  $s$ .

$$\begin{aligned} E[S|Z = 1] - E[S|Z = 0] &\stackrel{(1)}{=} \\ &[\Pr(S_1 < 1|Z = 1) - \Pr(S_0 < 1|Z = 0)] \\ &+ [\Pr(S_1 < 2|Z = 1) - \Pr(S_0 < 2|Z = 0)] \\ &\stackrel{(2)}{=} \Pr(S_1 \geq 1 > S_0) + \Pr(S_1 \geq 2 > S_0) \end{aligned}$$

where (1) follows because the mean is the sum (or integral) of 1 minus the CDF, and (2) because of independence.

# Variable treatment intensity

## Example with 3 levels

With three treatment intensities  $S \in \{0, 1, 2\}$  we observe

$$Y = Y_0 + (Y_1 - Y_0)\mathbb{1}[S \geq 1] + (Y_2 - Y_1)\mathbb{1}[S \geq 2]$$

Using this we can expand the reduced form as follows

$$\begin{aligned} E[Y|Z = 1] - E[Y|Z = 0] &= E[(Y_1 - Y_0)(\mathbb{1}[S_1 \geq 1] - \mathbb{1}[S_0 \geq 1])] \\ &\quad + E[(Y_2 - Y_1)(\mathbb{1}[S_1 \geq 2] - \mathbb{1}[S_0 \geq 2])] \end{aligned}$$

# Variable treatment intensity

## Average Causal Response

We can now define

$$\omega_s = \frac{\Pr(S_1 \geq s > S_0)}{\sum_{j=1}^J \Pr(S_1 \geq j > S_0)}$$

and express the Wald estimate as follows

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[S|Z=1] - E[S|Z=0]} = \sum_{s=1}^J \omega_s E[Y_s - Y_{s-1} | S_1 \geq s > S_0]$$

which Angrist and Imbens call the **average causal response** (ACR).

# Variable treatment intensity

## Average Causal Response

We cannot estimate  $E[Y_s - Y_{s-1} | S_1 \geq s > S_0]$  for the different local complier groups

What we can do is estimate their weights in the ACR, since

$$\begin{aligned}\Pr(S_1 \geq s > S_0) &= \Pr(S_1 \geq s) - \Pr(S_0 \geq s) \\ &= \Pr(S_0 < s) - \Pr(S_1 < s) \\ &= \Pr(S < s | Z = 0) - \Pr(S < s | Z = 1)\end{aligned}$$

which allows us to estimate  $\omega_s$

Note: although ACR is a positive weighted average, it

- averages together components that are potentially overlapping
- cannot be expressed as a positive weighted average of causal effects across mutually exclusive subgroups (unlike the LATE)

# Variable treatment intensity

## Example

Angrist & Krueger (1991) use quarter of birth as an instrument for schooling

- ▶  $D = 1$  if education is at least high school
- ▶  $Z = 1$  if born in the 4th quarter,  $Z = 0$  if born in the 1st quarter

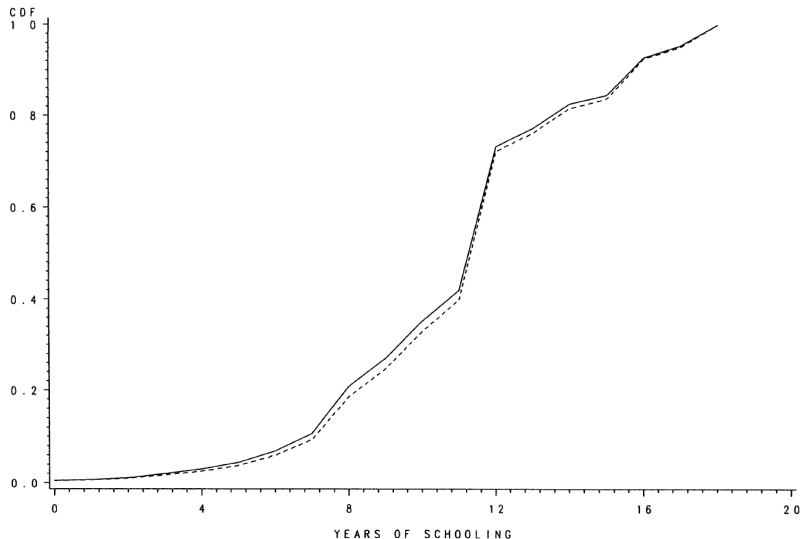
How does the Wald estimator weighs the average unit causal response

$$E[Y_s - Y_{s-1} | S_1 \geq s > S_0]$$

for the complier at the different points  $s$ ?

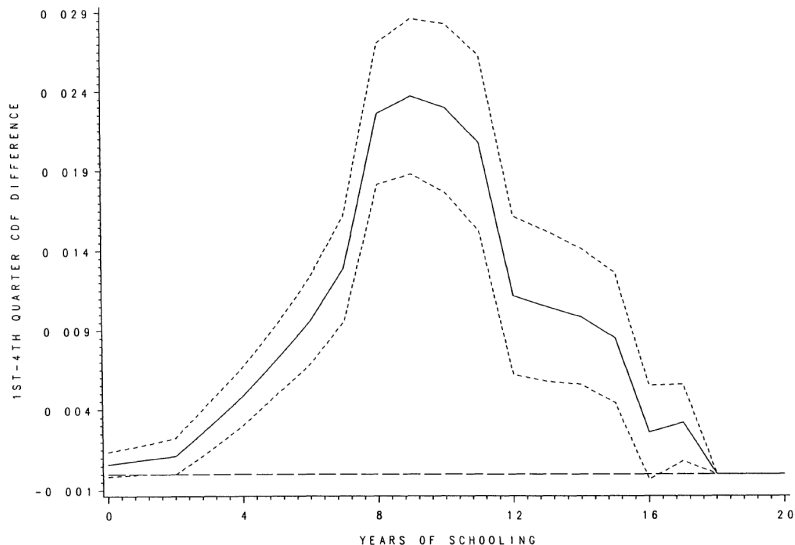
# Variable treatment intensity

Example, Schooling CDF by QoB (= 1, 4)



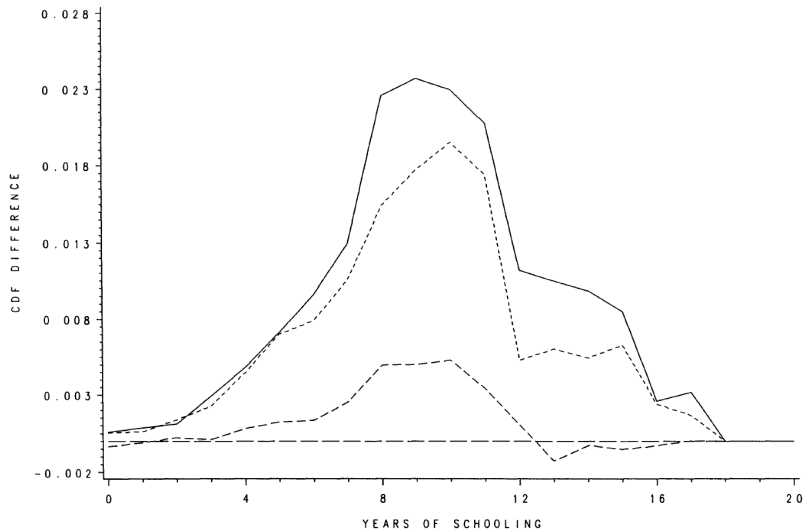
# Variable treatment intensity

Example, Differences in Schooling CDF by QoB (= 1, 4)



# Variable treatment intensity

Example, for different QoB's: 4vs1, 4vs2, 4vs3





# Can the weighing matter?

Loken et al. (2012) reports OLS, IV and family fixed effects estimates of how family income affects kid's outcomes

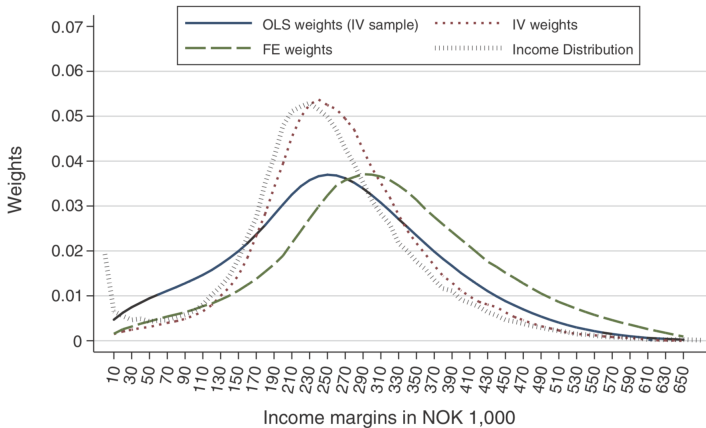


FIGURE 2. LINEAR OLS, IV, AND FE WEIGHTS

*Notes:* This figure reports the weights for the linear OLS and IV estimates (using the IV sample) shown in panels A and B, and the linear FE estimate (using the FE sample) shown in panel D of Table 3. To compute these weights, we use the decomposition in (9)–(11), and income margins of NOK 10,000. For comparison, this figure also graphs the family income distribution.

# Can the weighing matter?

	Decomposition		
	Absolute difference	Weights	Marginal effects
<i>Panel B. Estimates</i>			
$\beta(OLS) - \beta(Z)$	0.0207	7%	93%
$\beta(OLS) - \beta(FE)$	0.0410	11%	89%
$\beta(Z) - \beta(FE)$	0.0203	88%	12%

# Covariates

## Extensions to Covariates - Nonparametric

- ▶ Often, one wants covariates  $X$  to help justify the exogeneity of  $Z$
- ▶ And/or to reduce residual noise in  $Y$
- ▶ And/or to look at observed heterogeneity in treatment effects

### Adjust the assumptions to be conditional on $X$

- ▶ Exogeneity:  $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z | X$
- ▶ Relevance:  $\mathbb{P}[D = 1 | X, Z = 1] \neq \mathbb{P}[D = 1 | X, Z = 0]$  a.s.
- ▶ Monotonicity:  $\mathbb{P}[D_1 \geq D_0 | X] = 1$  a.s.
- ▶ Overlap:  $\mathbb{P}[Z = 1 | X] \in (0, 1)$  a.s.

# Non-parametric IV with Covariates

- Suppose we can estimate stratified LATEs

$$\begin{aligned}\beta(x) &= \frac{E[Y|Z = 1, X = x] - E[Y|Z = 0, X = x]}{E[D|Z = 1, X = x] - E[D|Z = 0, X = x]} \\ &= E[Y_1 - Y_0 | D_1 - D_0 = 1, X = x]\end{aligned}$$

- We want to go from here to some population averaged LATE
- Which one would we like to have? Complier weighted? Population weighted?

## 2SLS regression with Covariates

- ▶ What does a saturated 2SLS estimation gives us?

$$Y = \beta D + \alpha_x + e$$

$$D = \pi_x Z + \gamma_x + u$$

- ▶ i.e.  $x$ -dummies in both stages, and  $x$ -specific first-stage coefficients
- ▶ Angrist & Imbens (1995) show that

$$\beta = E[\beta(x)\omega(x)]$$

- ▶ where  $\beta(x)$  is the  $x$ -specific LATE, and

$$\omega(x) = \frac{\sigma_D^2(x)}{E[\sigma_D^2(x)]} = \frac{\pi_x^2 \sigma_Z^2(x)}{E[\pi_x^2 \sigma_Z^2(x)]}$$

- ▶ The weighting thus depends on the square of the local (to  $x$ ) complier share and instrument variance

## Abadie's (2003) $\kappa$

- ▶ For covariates (but  $D, Z$  binary) a more elegant approach
- ▶ Idea is to run regressions only on the compliers
- ▶ Compliers aren't directly observable, but they can be weighted
- ▶ Abadie showed that for any function  $G = g(Y, X, D)$

$$\mathbb{E}[G|T = c] = \frac{1}{\mathbb{P}[T = c]} \mathbb{E}[\kappa G], \kappa = 1 - \frac{D(1 - Z)}{\mathbb{P}[Z = 0|X]} - \frac{Z(1 - D)}{\mathbb{P}[Z = 1|X]}$$

### Intuition

- ▶ Complier = 1 – Always Taker – Never Taker
- ▶ On average,  $\kappa$  only applies positive weights to compliers:

$$\mathbb{E}[\kappa|T = t, X, D, Y] = \mathbb{1}[t = c]$$

- ▶ So on average,  $\kappa G$  is only positive for compliers

## IV with Covariates

- Abadie (2003) showed that

$$E[\kappa_0 g(Y, X)] = E[g(Y^0, X) | D^1 > D^0] \Pr(D^1 > D^0)$$

$$E[\kappa_1 g(Y, X)] = E[g(Y^1, X) | D^1 > D^0] \Pr(D^1 > D^0)$$

$$E[\kappa g(Y, D, X)] = E[g(Y, D, X) | D^1 > D^0] \Pr(D^1 > D^0)$$

where:

$$\kappa_0 = (1 - D) \frac{(1 - Z) - \Pr(Z = 0 | X)}{\Pr(Z = 0 | X) \Pr(Z = 1 | X)}$$

$$\kappa_1 = D \frac{Z - \Pr(Z = 1 | X)}{\Pr(Z = 0 | X) \Pr(Z = 1 | X)}$$

$$\begin{aligned} \kappa &= \kappa_0 \Pr(Z = 0 | X) + \kappa_1 \Pr(Z = 1 | X) \\ &= 1 - \frac{D(1 - Z)}{\Pr(Z = 0 | X)} - \frac{(1 - D)Z}{\Pr(Z = 1 | X)} \end{aligned}$$



# Using Abadie's (2003) $\kappa$

## Linear/nonlinear regression

- ▶ For example, take  $g(Y, X, D) = (Y - \alpha D - X' \beta)^2$  then:

$$\min_{\alpha, \beta} \mathbb{E}[(Y - \alpha D - X' \beta)^2 | T = c] = \min_{\alpha, \beta} \mathbb{E}[\kappa(Y - \alpha D - X' \beta)^2]$$

- ▶ Estimate  $\alpha, \beta$  by solving a sample analog of the second problem
- ▶ This is just a weighted regression, with estimated weights ( $\kappa$ )
- ▶ Result is general enough to use for many other estimators
- ▶ Specify  $X$  however you like - still picks out the compliers

# Using Abadie's (2003) $\kappa$

## Estimating $\kappa$

- ▶ To implement the result one must estimate  $\kappa$ , hence  $\mathbb{P}[Z = 1|X]$
- ▶ If  $\mathbb{P}[Z = 1|X]$  is linear, the  $\kappa$ -weighted regression equals TSLS
- ▶ Of course,  $Z$  is binary, so  $\mathbb{P}[Z = 1|X]$  typically won't be exactly linear
- ▶ Logit/probit often close to linear, so in practice may be close

# Empirical Example: Angrist and Evans (1998, “AE”)

## Motivation

- ▶ Relationship between fertility decisions and female labor supply?
- ▶ Strong negative correlation, but these are joint choices
- ▶ Leads to many possible endogeneity stories, here's just one:  
High earning women have fewer children due to higher opp. cost

# Empirical Example: Angrist and Evans (1998, “AE”)

## Empirical strategy

- ▶  $Y$  is a labor market outcome for the woman (or her husband)
- ▶ Restrict the sample to only women (or couples) with 2 or more children
- ▶  $D$  is an indicator for having more than 2 children (vs. exactly 2)
- ▶  $Z = 1$  if first two children had the same sex
  - Based on the idea that there is preference to have a mix of boys and girls
- ▶ Also consider  $Z = 1$  if the second birth was a twin
  - Twins are primarily for comparison - used before this paper

# Assumptions in AE

## Exogeneity

- ▶ Requires the assumption that sex at birth is randomly assigned
- ▶ Authors conduct balance tests to support this (next slide)
- ▶ The twins instrument is less compelling
- ▶ First, well-known that older women have twins more (see next slide)
  - More subtly, it impacts both the number and spacing of children

## Monotonicity

- ▶ Monotonicity restricts preference heterogeneity in unattractive ways
  - Some families may want two boys or girls (then stop)
- ▶ No discussion of this in the paper - unfortunately common practice
- ▶ Twins is effectively a one-sided non-compliance instrument
  - Twins compliers are the untreated since no twins never-takers

# Evidence in Support of Exogeneity

TABLE 4—DIFFERENCES IN MEANS FOR DEMOGRAPHIC VARIABLES  
BY SAME SEX AND TWINS-2

Variable	Difference in means (standard error)		
	By Same sex		By Twins-2
	1980 PUMS	1990 PUMS	1980 PUMS
<i>Age</i>	−0.0147 (0.0112)	0.0174 (0.0112)	0.2505 (0.0607)
<i>Age at first birth</i>	0.0162 (0.0094)	−0.0074 (0.0114)	0.2233 (0.0510)
<i>Black</i>	0.0003 (0.0010)	0.0021 (0.0011)	0.0300 (0.0056)
<i>White</i>	0.0003 (0.0012)	−0.0006 (0.0013)	−0.0210 (0.0066)
<i>Other race</i>	−0.0006 (0.0005)	−0.0014 (0.0009)	−0.0090 (0.0041)
<i>Hispanic</i>	−0.0014 (0.0009)	−0.0007 (0.0010)	−0.0069 (0.0047)
<i>Years of education</i>	−0.0028 (0.0076)	0.0100 (0.0074)	0.0940 (0.0415)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

- ▶ Same sex is uncorrelated with a variety of observed confounders
- ▶ Twins is well-known to be correlated with age (so, education) and race

# Wald Estimates

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

Variable	Mean difference by Same sex	1980 PUMS		1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:		Mean difference by Same sex	Wald estimate using as covariate:		Wald estimate using as covariate:		
		More than 2 children	Number of children		More than 2 children	Number of children	Mean difference by Twins-2	More than 2 children	Number of children
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	-0.0080 (0.0016)	-0.133 (0.026)	-0.104 (0.021)	-0.0053 (0.0015)	-0.084 (0.024)	-0.063 (0.018)	-0.0459 (0.0086)	-0.076 (0.014)	-0.057 (0.011)
Weeks worked	-0.3826 (0.0709)	-6.38 (1.17)	-5.00 (0.92)	-0.3233 (0.0743)	-5.15 (1.17)	-3.87 (0.88)	-1.982 (0.386)	-3.28 (0.63)	-2.45 (0.47)
Hours/week	-0.3110 (0.0602)	-5.18 (1.00)	-4.07 (0.78)	-0.2363 (0.0620)	-3.76 (0.98)	-2.83 (0.73)	-1.979 (0.327)	-3.28 (0.54)	-2.44 (0.40)
Labor income	-132.5 (34.4)	-2208.8 (569.2)	-1732.4 (446.3)	-119.4 (42.4)	-1901.4 (670.3)	-1428.0 (502.6)	-570.8 (186.9)	-946.4 (308.6)	-705.2 (229.8)
ln(Family income)	-0.0018 (0.0041)	-0.029 (0.068)	-0.023 (0.054)	-0.0085 (0.0047)	-0.136 (0.074)	-0.102 (0.056)	-0.0341 (0.0223)	-0.057 (0.037)	-0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

- First stage (denominator of Wald) for two measures of fertility

# Wald Estimates

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

Variable	1980 PUMS			1990 PUMS			1980 PUMS		
	Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Twins-2	Wald estimate using as covariate:	
		More than 2 children	Number of children		More than 2 children	Number of children		More than 2 children	Number of children
<i>More than 2 children</i>	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
<i>Number of children</i>	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
<i>Worked for pay</i>	-0.0080 (0.0016)	-0.133 (0.026)	-0.104 (0.021)	-0.0053 (0.0015)	-0.084 (0.024)	-0.063 (0.018)	-0.0459 (0.0086)	-0.076 (0.014)	-0.057 (0.011)
<i>Weeks worked</i>	-0.3826 (0.0709)	-6.38 (1.17)	-5.00 (0.92)	-0.3233 (0.0743)	-5.15 (1.17)	-3.87 (0.88)	-1.982 (0.386)	-3.28 (0.63)	-2.45 (0.47)
<i>Hours/week</i>	-0.3110 (0.0602)	-5.18 (1.00)	-4.07 (0.78)	-0.2363 (0.0620)	-3.76 (0.98)	-2.83 (0.73)	-1.979 (0.327)	-3.28 (0.54)	-2.44 (0.40)
<i>Labor income</i>	-132.5 (34.4)	-2208.8 (569.2)	-1732.4 (446.3)	-119.4 (42.4)	-1901.4 (670.3)	-1428.0 (502.6)	-570.8 (186.9)	-946.4 (308.6)	-705.2 (229.8)
<i>ln(Family income)</i>	-0.0018 (0.0041)	-0.029 (0.068)	-0.023 (0.054)	-0.0085 (0.0047)	-0.136 (0.074)	-0.102 (0.056)	-0.0341 (0.0223)	-0.057 (0.037)	-0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

- First stage (numerator of Wald) for several labor market outcomes



# Wald Estimates

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

Variable	1980 PUMS			1990 PUMS			1980 PUMS		
	Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Same sex	Wald estimate using as covariate:		Mean difference by Twins-2	Wald estimate using as covariate:	
		More than 2 children	Number of children		More than 2 children	Number of children		More than 2 children	Number of children
<i>More than 2 children</i>	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
<i>Number of children</i>	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
<i>Worked for pay</i>	-0.0080 (0.0016)	-0.133 (0.026)	-0.104 (0.021)	-0.0053 (0.0015)	-0.084 (0.024)	-0.063 (0.018)	-0.0459 (0.0086)	-0.076 (0.014)	-0.057 (0.011)
<i>Weeks worked</i>	-0.3826 (0.0709)	-6.38 (1.17)	-5.00 (0.92)	-0.3233 (0.0743)	-5.15 (1.17)	-3.87 (0.88)	-1.982 (0.386)	-3.28 (0.63)	-2.45 (0.47)
<i>Hours/week</i>	-0.3110 (0.0602)	-5.18 (1.00)	-4.07 (0.78)	-0.2363 (0.0620)	-3.76 (0.98)	-2.83 (0.73)	-1.979 (0.327)	-3.28 (0.54)	-2.44 (0.40)
<i>Labor income</i>	-132.5 (34.4)	-2208.8 (569.2)	-1732.4 (446.3)	-119.4 (42.4)	-1901.4 (670.3)	-1428.0 (502.6)	-570.8 (186.9)	-946.4 (308.6)	-705.2 (229.8)
<i>ln(Family income)</i>	-0.0018 (0.0041)	-0.029 (0.068)	-0.023 (0.054)	-0.0085 (0.0047)	-0.136 (0.074)	-0.102 (0.056)	-0.0341 (0.0223)	-0.057 (0.037)	-0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

- IV (Wald) estimator, e.g.  $-0.133 \approx -0.008/0.060$  - these are LATEs

# Two Stage Least Squares Estimates

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for <i>More than 2 children</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>
Dependent variable:									
<i>Worked for pay</i>	-0.176 (0.002)	-0.120 (0.025)	-0.113 (0.025) [0.013]	-0.167 (0.002)	-0.120 (0.028)	-0.113 (0.028) [0.013]	-0.008 (0.001)	0.004 (0.009)	0.001 (0.008) [0.013]
<i>Weeks worked</i>	-8.97 (0.07)	-5.66 (1.11)	-5.37 (1.10) [0.017]	-8.05 (0.09)	-5.40 (1.20)	-5.16 (1.20) [0.071]	-0.82 (0.04)	0.59 (0.60)	0.45 (0.59) [0.030]
<i>Hours/week</i>	-6.66 (0.06)	-4.59 (0.95)	-4.37 (0.94) [0.030]	-6.02 (0.08)	-4.83 (1.02)	-4.61 (1.01) [0.049]	0.25 (0.05)	0.56 (0.70)	0.50 (0.69) [0.71]
<i>Labor income</i>	-3768.2 (35.4)	-1960.5 (541.5)	-1870.4 (538.5) [0.126]	-3165.7 (42.0)	-1344.8 (569.2)	-1321.2 (565.9) [0.703]	-1505.5 (103.5)	-1248.1 (1397.8)	-1382.3 (1388.9) (0.549)
<i>ln(Family income)</i>	-0.126 (0.004)	-0.038 (0.064)	-0.045 (0.064) [0.319]	-0.132 (0.004)	-0.051 (0.056)	-0.053 (0.056) [0.743]	—	—	—
<i>ln(Non-wife income)</i>	—	—	—	-0.053 (0.005)	0.023 (0.066)	0.016 (0.066) [0.297]	—	—	—

Notes: The table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text. Other covariates in the models are Age, Age at first birth, plus indicators for Boy 1st, Boy 2nd, Black, Hispanic, and Other race. The variable Boy 2nd is excluded from equation (6). The *p*-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

- OLS is quite different from IV - consistent with endogeneity (selection)

# Two Stage Least Squares Estimates

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for <i>More than 2 children</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>
Dependent variable:									
<i>Worked for pay</i>	−0.176 (0.002)	−0.120 (0.025)	−0.113 (0.025) [0.013]	−0.167 (0.002)	−0.120 (0.028)	−0.113 (0.028) [0.013]	−0.008 (0.001)	0.004 (0.009)	0.001 (0.008) [0.013]
<i>Weeks worked</i>	−8.97 (0.07)	−5.66 (1.11)	−5.37 (1.10) [0.017]	−8.05 (0.09)	−5.40 (1.20)	−5.16 (1.20) [0.071]	−0.82 (0.04)	0.59 (0.60)	0.45 (0.59) [0.030]
<i>Hours/week</i>	−6.66 (0.06)	−4.59 (0.95)	−4.37 (0.94) [0.030]	−6.02 (0.08)	−4.83 (1.02)	−4.61 (1.01) [0.049]	0.25 (0.05)	0.56 (0.70)	0.50 (0.69) [0.71]
<i>Labor income</i>	−3768.2 (35.4)	−1960.5 (541.5)	−1870.4 (538.5) [0.126]	−3165.7 (42.0)	−1344.8 (569.2)	−1321.2 (565.9) [0.703]	−1505.5 (103.5)	−1248.1 (1397.8)	−1382.3 (1388.9) (0.549)
<i>ln(Family income)</i>	−0.126 (0.004)	−0.038 (0.064)	−0.045 (0.064) [0.319]	−0.132 (0.004)	−0.051 (0.056)	−0.053 (0.056) [0.743]	—	—	—
<i>ln(Non-wife income)</i>	—	—	—	−0.053 (0.005)	0.023 (0.066)	0.016 (0.066) [0.297]	—	—	—

Notes: The table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text. Other covariates in the models are *Age*, *Age at first birth*, plus indicators for *Boy 1st*, *Boy 2nd*, *Black*, *Hispanic*, and *Other race*. The variable *Boy 2nd* is excluded from equation (6). The *p*-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

- Break same-sex into two instruments - two boys vs two girls

# Two Stage Least Squares Estimates

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for <i>More than 2 children</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>	—	<i>Same sex</i>	<i>Two boys, Two girls</i>
Dependent variable:									
<i>Worked for pay</i>	-0.176 (0.002)	-0.120 (0.025)	-0.113 (0.025) [0.013]	-0.167 (0.002)	-0.120 (0.028)	-0.113 <del>(0.008)</del> [0.013]	-0.008 (0.001)	0.004 (0.009)	0.001 (0.008) [0.013]
<i>Weeks worked</i>	-8.97 (0.07)	-5.66 (1.11)	-5.37 (1.10) [0.017]	-8.05 (0.09)	-5.40 (1.20)	-5.16 (1.20) [0.071]	-0.82 (0.04)	0.59 (0.60)	0.45 (0.59) [0.030]
<i>Hours/week</i>	-6.66 (0.06)	-4.59 (0.95)	-4.37 (0.94) [0.030]	-6.02 (0.08)	-4.83 (1.02)	-4.61 (1.01) [0.049]	0.25 (0.05)	0.56 (0.70)	0.50 (0.69) [0.71]
<i>Labor income</i>	-3768.2 (35.4)	-1960.5 (541.5)	-1870.4 (538.5) [0.126]	-3165.7 (42.0)	-1344.8 (569.2)	-1321.2 (565.9) [0.703]	-1505.5 (103.5)	-1248.1 (1397.8)	-1382.3 (1388.9) (0.549)
<i>ln(Family income)</i>	-0.126 (0.004)	-0.038 (0.064)	-0.045 (0.064) [0.319]	-0.132 (0.004)	-0.051 (0.056)	-0.053 (0.056) [0.743]	—	—	—
<i>ln(Non-wife income)</i>	—	—	—	-0.053 (0.005)	0.023 (0.066)	0.016 (0.066) [0.297]	—	—	—

Notes: The table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) and (6) in the text. Other covariates in the models are *Age*, *Age at first birth*, plus indicators for *Boy 1st*, *Boy 2nd*, *Black*, *Hispanic*, and *Other race*. The variable *Boy 2nd* is excluded from equation (6). The *p*-value for the test of overidentifying restrictions associated with equation (6) is shown in brackets. Standard errors are reported in parentheses.

- Overid test p-values - many interpretations with heterogeneity

# Comparison to Abadie's $\kappa$ (Angrist 2001)

Dependent variable	Linear models			Nonlinear models			Structural models			
	2SLS (1)	2PM (2)	Causal-IV linear (3)	Mullahy (4)	Causal-IV probit (5)	Causal-IV expon. (6)	Bivar. probit (7)	Endog. tobit (8)	Mills ratio (9)	2SLS benchmark (10)
A. With covariates										
Employment	-.088 (.017)	—	-.089 (.017)	—	-.088 (.016)	—	-.124 (.016)	—	—	-.089 (.017)
Hours worked	-3.55 (.592)	-3.54 (.598)	-3.55 (.592)	-3.82 (.598)	—	-3.21 (.694)	—	-3.81 (.580)	-4.51 (.549)	-3.60 (.599)
B. No covariates										
Employment	-.084 (.017)	—	-.084 (.017)	—	-.084 (.017)	—	-.086 (.017)	—	—	-.084 (.018)
Hours worked	-3.47 (.617)	-3.37 (.614)	-3.47 (.617)	-3.10 (.561)	—	-3.12 (.616)	—	-3.35 (.642)	-3.48 (.641)	-3.52 (.624)

NOTE: Sample and covariates are the same as in Table 1. Results for nonlinear models are derivative-based approximations to effect on the treated. Causal-IV estimates are based on a procedure discussed by Abadie (2000b). Standard errors are shown in parentheses.

- Illustration of Abadie's  $\kappa$  (and other methods) using the AE data
- Results are almost identical to TSLS - uses this to promote TSLS
- Logic is strange - we know that in general this is not the case
- In fact, Abadie's (2003) paper has an application where it is not

## Multiple unordered treatments

## Estimating equation: Example with 3 field choice

- ▶ Individuals are often choosing between multiple unordered treatments:  
Education types, occupations, locations, etc.
- ▶ MHS is completely silent about multiple unordered treatment
- ▶ What does 2SLS identify in this case?
- ▶ Kirkeboen et al. (2016, QJE) discusses this in the context of educational choices
- ▶ See also Kline and Walters (2016), Heckman and Pinto (2019) and Mountjoy (2019).

## Estimating equation: Example with 3 field choice

- ▶ Students choose between three fields,  $D \in \{0, 1, 2\}$
- ▶ Our interest is centered on how to interpret IV (and OLS) estimates of

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \epsilon$$

- ▶  $Y$  is observed earnings
- ▶  $D_j \equiv 1(D = j)$  is an indicator variables that equals 1 if individual chooses field  $j$
- ▶  $\epsilon$  is the residual which is potentially correlated with  $D_j$



## Potential earnings and field choices

- ▶ Individuals are assigned to one of three groups,  $Z \in \{0, 1, 2\}$
- ▶ Linking observed and potential earnings and field choices

$$\begin{aligned}Y &= Y^0 + (Y^1 - Y^0)D_1 + (Y^2 - Y^0)D_2 \\D_1 &= D_1^0 + (D_1^1 - D_1^0)Z_1 + (D_1^2 - D_1^0)Z_2 \\D_2 &= D_2^0 + (D_2^1 - D_2^0)Z_1 + (D_2^2 - D_2^0)Z_2\end{aligned}$$

- ▶  $Y^j$  is potential earnings if individual chooses field  $j$
- ▶  $Z_k = 1(Z = k)$  is an indicator variable that equals 1 if  $Z$  is equal to  $k$
- ▶  $D_j^Z \equiv 1(D^Z = j)$  is indicator variables that equals 1 if individual chooses field  $j$  for a given value of  $Z$

## Standard IV assumptions

- ▶ ASSUMPTION 1: (EXCLUSION):  $Y^{d,z} = Y^d$  for all  $d, z$
- ▶ ASSUMPTION 2: (INDEPENDENCE):  $Y^0, Y^1, Y^2, D^0, D^1, D^2 \perp Z$
- ▶ ASSUMPTION 3: (RANK):  $\text{Rank } E(Z'D) = 3$
- ▶ ASSUMPTION 4: (MONOTONICITY):  $D_1^1 \geq D_1^0$  and  $D_2^2 \geq D_2^0$

# Moment conditions

- ▶ IV uses the following moment conditions:

$$E[\epsilon Z_1] = E[\epsilon Z_2] = E[\epsilon] = 0$$

- ▶ Expressing these conditions in potential earnings and choices gives:

$$E[(\Delta^1 - \beta_1)(D_1^1 - D_1^0) + (\Delta^2 - \beta_2)(D_2^1 - D_2^0)] = 0 \quad (1)$$

$$E[(\Delta^1 - \beta_1)(D_1^2 - D_1^0) + (\Delta^2 - \beta_2)(D_2^2 - D_2^0)] = 0 \quad (2)$$

where

$$\Delta^j \equiv Y^j - Y^0$$

- ▶ To understand what IV can and cannot identify, we solve these equations for  $\beta_1$  and  $\beta_2$

# What IV cannot identify

## PROPOSITION 1

- ▶ Suppose Assumptions 1-4 hold
- ▶ Solving equations (1)-(2) for  $\beta_1$  and  $\beta_2$ , it follows that  $\beta_j$  for  $j = 1, 2$  is a linear combination of the following three payoffs:
  1.  $\Delta^1$ : Payoff of field 1 compared to 0
  2.  $\Delta^2$ : Payoff of field 2 compared to 0
  3.  $\Delta^2 - \Delta^1 \equiv Y^2 - Y^1$ : Payoff of field 2 compared to 1

# Constant effects

- ▶ Suppose Assumptions 1-4 hold.
- ▶ Solving equations (1)-(2) for  $\beta_1$  and  $\beta_2$ :
  - ▶ If  $\Delta^1$  and  $\Delta^2$  are common across all individuals (Constant effects):

$$\beta_1 = \Delta^1$$

$$\beta_2 = \Delta^2$$

- ▶ Alternatively, move to goal post to estimating effect of, say, field 1 versus next best (combination of 2 and 3)
- ▶ Back to binary treatment but hard to interpret and requires strong exogeneity assumption

# Data on Second Choices

- ▶ In certain circumstances, one might plausibly observe next best options
  - ▶ Kirkeboen et al (2016) show one can point identify

$$\beta_1 = E[\Delta^1 | D_1^1 - D_1^0 = 1, D_2^0 = 0]$$

$$\beta_2 = E[\Delta^2 | D_2^2 - D_2^0 = 1, D_1^0 = 0]$$

- ▶ Kirkeboen et al (2016) do this with Norwegian admissions data
- ▶ Students apply with a list of desired fields and universities
- ▶ Assigned based on preference and merit rankings

# Data on Second Choices

- ▶ Strategy proof mechanism, so stated preferences should be actual
  - ▶ Conditional exogeneity uses a local type of argument
  - ▶ Compare students with similar rankings and stated preferences  $j, k$
  - ▶ One is slightly above the cutoff, gets  $j$  - other slightly below gets  $k$
  - ▶ An example of a (fuzzy) RDD — we will discuss these more soon

Weak and many instruments



## Weak instruments

An instrumental variable is **weak** if its correlation with the included endogenous regressor is small.

- ▶ “small” depends on inference problem at hand, and on sample size

Why is weak instruments a problem?

- ▶ Weak instrument is a “divide by (almost) zero” problem (recall IV = reduced form/first stage)

For the usual asymptotic approximation to be “good”, we would like to effectively treat the denominator as a constant

- ▶ In other words, we would like the mean to be much larger than the standard deviation of the denominator
- ▶ Otherwise, the finite-sample distribution can be very different from the asymptotic one (even in relatively “large” samples)
- ▶ And remember that 2SLS’s justification is asymptotic!

For details, see Azeem’s lecture notes

# What (not) to do about weak instruments

Large literature on (how to detect) weak instruments

- ▶ Useful summary of theory and practice in Andrews et al. (2019); see also their NBER lecture slides

Standard practice is to report the usual F-stat for instruments, and proceed as usual if F exceeds 10 (or some other arbitrary number)

Increasingly people instead report the “Effective first-stage F statistic” of Montiel Olea and Ploger (2013)

- ▶ Robust to the worst type of heteroscedasticity, serial correlation, and clustering in the second stage

The idea behind this practice is to decide if instruments are strong (TSLS “works”) or weak (use weak-instrument robust methods)

- ▶ But screening on F-statistics induces size distortions

# What to do about weak instruments (con't)

To me, it makes more sense to

1. report and interpret reduced form
2. think hard about why your instrument could be weak  
(instruments comes from knowledge about treatment assignment)
3. (also) report weak instrument robust confidence sets

Weak instrument robust confidence sets:

- ▶ Ensure correct coverage regardless of instrument strength
- ▶ No need to screen on first stage
- ▶ Avoids pretesting bias
- ▶ Avoids throwing away applications with valid instruments just because weak
- ▶ Confidence sets can be informative even with weak instruments

## Many instruments and overfitting

At seminars (and in referee reports), people often talk about many instruments and weak instruments as if they are the same problem

Very confusing (at least to me)

Confusion may stem from Angrist and Krueger (1991)

- ▶ Looked at how years of schooling ( $S$ ) affects wages ( $Y$ ), and uses the instrument quarter of birth ( $Z$ )
- ▶ Problem: quarter of birth only produces very small variation in the years of schooling
- ▶ Thus people worry it is a weak instrument.

To overcome this issue, they interacted the instrument with many control variables (assumed to be exogenous)

They found that the estimate for the coefficient on years of schooling from the IV regression was very similar to that from the OLS

## Many instruments and overfitting (con't)

The re-analysis of Bound et al (1993) suggests the similarity was due to **overfitting**

They take the data that Angrist and Krueger (1991) used and added many randomly generated variables

- ▶ Find that running IV regression with these variables leads to a coefficient estimate that is similar to that using OLS
- ▶ Intuitively, the problem here is that when we have many instruments,  $S$  and  $\hat{S}$ , are essentially the “same”
- ▶ Since the true  $S$  is endogenous, this means that  $\hat{S}$  is also endogenous
  - ▶ results in IV having a bias towards the OLS

## Many instruments and overfitting (con't)

In response to the many instrument problem and overfitting, recent work on how to select the “optimal” instruments (e.g. using Lasso)

- ▶ Not clear what optimal means with heterogeneous effects
- ▶ Most settings, hard to find even one good instrument
- ▶ Thus, many instruments usually involves implicit exclusion restrictions (from interacting  $X$  and  $Z$  but not  $S$  and  $Z$ )
  - ▶ Effectively solving an estimation/ inference issue by violating exclusion restriction

Taking stock

# Summary

## IV

- ▶ The IV estimand in the binary  $D$ , binary  $Z$  case is the LATE
- ▶ Easy to interpret as the average effect for compliers
- ▶ Could be relevant for a policy intervention that affects compliers

## Extensions

- ▶ 2SLS used in general cases  $\rightarrow$  interpretation is complicated
- ▶ At best, a weighted average of several different (complier) groups
- ▶ When would these weights be useful to inform a counterfactual?

## Reverse engineering

- ▶ These results are motivated by a backward thought process
- ▶ Start with a common estimator, then interpret the estimand
- ▶ Why not start with a parameter of interest  $\rightarrow$  create an estimator?
  - ▶ More on that later!



# Practical advice when doing IV

## 1. Motivate your instruments

- ▶ Motivate exclusion and independence
  - ▶ how is  $Z$  generated? What do I need to control for to make it as good as randomly assigned?
  - ▶ why is  $Z$  not in the outcome equation? what are the distinct channels through which  $Z$  can affect  $Y$ ?
- ▶ Specification: what control variables should be included?
  - ▶ conditional exclusion restrictions can be more credible
  - ▶ assess by regressing instrument on other pre-determined variables
- ▶ Interpretation: what is the complier group?
  - ▶ is the instrument policy relevant?

# Practical advice when doing IV

## 2. Check your instruments

- ▶ Always report the first stage and
  - ▶ discuss whether the magnitude and signs are as expected
  - ▶ report the (relevant) F-statistic on instruments
    - ▶ larger is better (rule-of-thumb:  $F > 10$ .... but who knows what's large enough)
    - ▶ consider also reporting weak instrument robust confidence intervals
- ▶ Inspect the reduced-form regression of dependent variables on instruments
  - ▶ both first stage and reduced form; sign, magnitude, etc.
  - ▶ remember that the reduced form is proportional to the causal effect of interest
  - ▶ the reduced-form is unbiased (and not only consistent) because this is OLS

# How do I find instruments?

- ▶ There is no "recipe" that guarantees success
- ▶ But often necessary ingredients: Detailed knowledge of
  1. the *economic mechanisms*, and
  2. *institutions* determining the endogenous regressor
  3. restrictions from *economic theory*
- ▶ Examples:
  1. Naturally occurring random events (like weather, twin birth, etc)
  2. Policy reforms (which conditional on something are as good as random)
  3. Random assignment to individuals deciding treatment (e.g. judges)
  4. Cutoff rules for admission to programs — more next week on using such discontinuities
- ▶ Randomized experiments with imperfect compliance
  - ▶ gives a LATE interpretation of RCT

## Application: Judge design

# Family welfare cultures: Opposing views

Two opposing views:

1. Welfare use reinforces itself through the family, because parents on welfare may
  - ▶ Provide information about program to their children
  - ▶ Reduce stigma of participation
  - ▶ Invest differentially in child development
2. The determinants of health and poverty are correlated across generations, so that
  - ▶ Child welfare dependency is associated with
    - but not caused by –
    - a parent's use of welfare

# What do we do?

1. We investigate existence and importance of family welfare cultures
  - ▶ In a setting with no correlated unobservables
2. We explore breadth and nature of welfare cultures
  - ▶ Spillover effects in other social networks
  - ▶ Explore channels of welfare culture
3. We illustrate the policy relevance of intergenerational spillovers
  - ▶ Use estimates to simulate direct and indirect effects of policy

# Empirical Challenges: Statistical Model

- ▶ Characterize child's latent demand/qualification ( $P_i^{c*}$ ) as a function of
  1. parent's actual participation ( $P_i^p$ )
  2. other observed traits ( $x_i^c$ )
  3. unobserved taste/health/etc. ( $\varepsilon_i^c$ )

$$P_i^{c*} = \alpha^c + \beta^c P_i^p + \delta^c x_i^c + \varepsilon_i^c \quad (3)$$

- ▶ Similar equation for parents and grandparents

$$P_i^{p*} = \alpha^p + \beta^p P_i^g + \delta^p x_i^p + \varepsilon_i^p \quad (4)$$

# Empirical Challenges: Sources to Bias

- ▶ Substitution of parent's choice yields

$$P^{c*} = \alpha^c + \beta^c I(\alpha^p + \beta^p P_i^g + \delta^p x_i^p + \varepsilon_i^p > 0) + \delta^c x_i^c + \varepsilon_i^c. \quad (5)$$

where child participates if  $P_i^{c*} > 0$

1. This equation illustrates that if unobservables are correlated across generations

$$\text{cov}(\varepsilon_i^p, \varepsilon_i^c | x_i^c, x_i^p) \neq 0$$

2. Similarly, unobservables common to grandparent and child:

$$\text{cov}(\varepsilon_i^g, \varepsilon_i^c | x_i^c, x_i^p, x_i^g) \neq 0$$

→ *Family welfare culture* parameter will be biased



# Empirical Challenges: Correlations and Bias

**Table:** OLS Estimates of Intergenerational Welfare Transmission

	Child DI use ( $P_i^c$ )		
	(1)	(2)	(3)
Parent DI use ( $P_i^p$ )	0.036*** (0.001)	0.035*** (0.001)	0.025*** (0.001)
Grandparent DI use ( $P_i^g$ )		0.005*** (0.000)	0.004*** (0.000)
Additional controls?	NO	NO	YES
Obs.	1,022,507	1,022,507	1,022,507
Dep. mean	0.03	0.03	0.03

*Notes: Data come from 2008 and are restricted to parents age 60 or below with children age 23 and above and a grandparent who is alive during the period 1967-2010. DI use in each generation defined to be equal to 1 if the individual is currently receiving DI benefits (except for grandparents, which is defined as having ever received DI benefits). Column (3) controls flexibly for child, parent and grandparent characteristics (age, gender, education, foreign born, marital status, earnings history, and region fixed effects). Standard errors clustered at the family level.*

# Research design and setting

## ► *Research design*

1. Exploit a policy which randomizes probability that parents receive welfare
2. Use a unique source of population panel data, linking welfare use of members in social networks

## ► *Setting:*

Disability insurance (DI) system in Norway

# Identification: Random assignment of judges

- ▶ Denied DI applicants may decide to appeal the decision:
  1. Cases are randomly assigned to judges
  2. Some appeal judges systematically more lenient

⇒ random variation in probability a parent receives DI

- ▶ Exploit this exogenous variation to examine intergenerational links
- ▶ Since variation driven by difficult-to-verify cases
  - ▶ Randomization picks out the more marginal applicants
- ▶ Policy relevant group
  1. Driving the recent rise in DI rolls
  2. Affected by policy proposals to tighten screening

# Research design: Baseline Regression Model

- ▶ First and second stage of IV model:

$$P_i^p = \alpha^p + \gamma^p Z_i^p + X_i \delta^p + \varepsilon_i^p \quad (6)$$

$$P_i^c = \alpha^c + \beta^c P_i^p + X_i \delta^c + \varepsilon_i^c \quad (7)$$

- ▶ Due to randomization,  $Z_i^p$  (judge leniency)  $\perp \varepsilon_i^c$  and  $\varepsilon_i^p$ 
  - ▶ Correlated unobservables do not bias the estimate
  - ▶  $X_i$  always includes year of appeal  $\times$  department fixed effects

# Research design: Baseline Regression Model

- First and second stage of IV model:

$$P_i^p = \alpha^p + \gamma^p Z_i^p + X_i \delta^p + \varepsilon_i^p \quad (6)$$

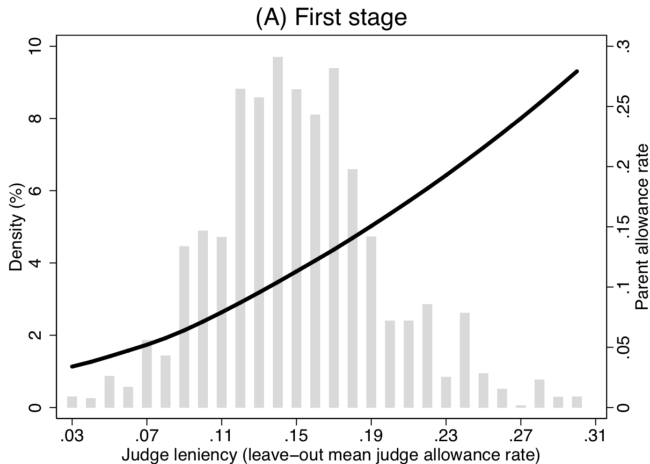
$$P_i^c = \alpha^c + \beta^c P_i^p + X_i \delta^c + \varepsilon_i^c \quad (7)$$

- Due to randomization,  $Z_i^p$  (judge leniency)  $\perp \varepsilon_i^c$  and  $\varepsilon_i^p$ 
  - Correlated unobservables do not bias the estimate
  - $X_i$  always includes year of appeal  $\times$  department fixed effects
- First stage:  $\gamma^p$  identified from a regression of  $P_i^p$  on  $Z_i^p$
- Reduced form: Regression of  $P_i^c$  on  $Z_i^p$
- Second stage: Intergenerational transmission coefficient  $\beta^c$  given by ratio of reduced form and first stage

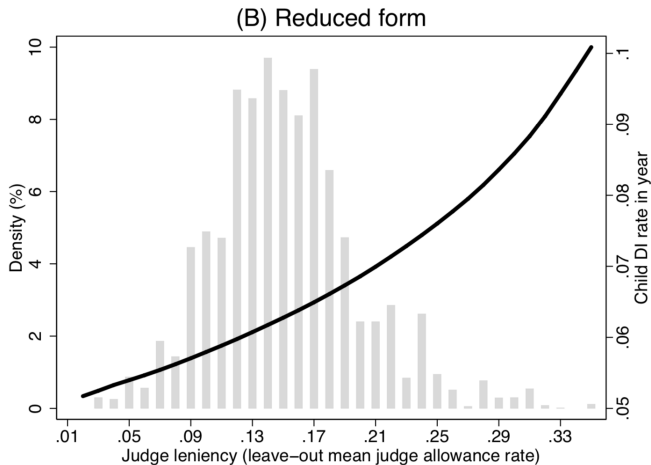
# Testing Random Assignment

	Case Allowed		Judge Leniency	
Age	0.0054***	(0.0009)	0.0003*	(0.0002)
Female	0.0109	(0.0096)	0.0002	(0.0019)
Married	0.0041	(0.0076)	0.0013	(0.0019)
Foreign born	-0.0271***	(0.0114)	0.0009	(0.0025)
High school degree	-0.01670***	(0.0070)	-0.0002	(0.0017)
Some college	0.01317*	(0.0070)	0.00041	(0.0014)
College graduate	0.02282	(0.0161)	-0.00073	(0.0033)
One child	-0.1033***	(0.0199)	0.00389	(0.0094)
Two children	-0.0052	(0.0087)	-0.00097	(0.0020)
Three or more children	-0.0159	(0.0132)	0.00103	(0.0016)
Previous earnings	-0.0355***	(0.0146)	0.00319	(0.0021)
Years of work	0.0000***	(0.0000)	0.0000	(0.0000)
Mental disorders	0.0357***	(0.0105)	0.00005	(0.0038)
Musculoskeletal disorders	0.0026	(0.0086)	0.0018	(0.00256)
Test for joint significance	F: 9.25	p-value: .001	F: .77	p-value: .723

# Graphical evidence: first stage

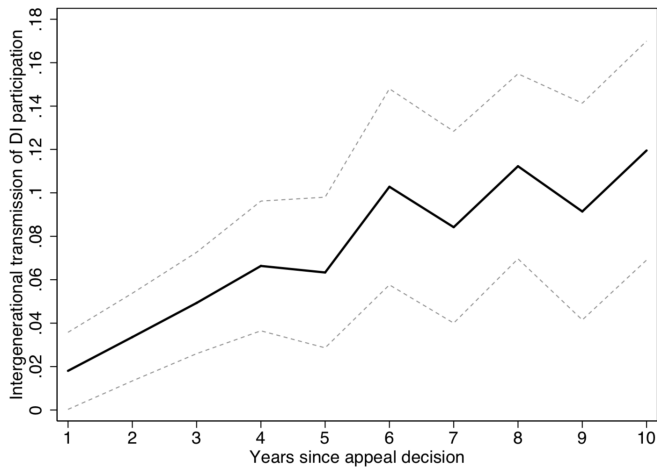


# Graphical evidence: reduced form





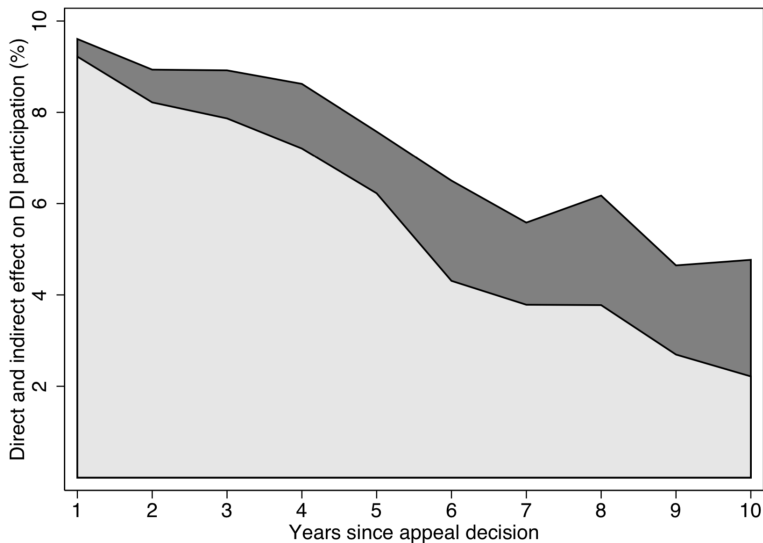
# Time profile in IV estimates



# Why welfare cultures matter for policy

- ▶ Intergenerational links could be important for policy design
- ▶ In particular, making the disability screening more stringent:
  1. Directly reduce DI participation among parents
  2. Further reduce DI participation in next generation
- ▶ Policy simulation
  1. Make judges 1/5 std dev stricter  
(10% less likely to grant an appeal on average)
  2. Combine with estimates of how parent's judge leniency affect parent and child participation over time

# Direct and indirect effects of stringent screening



Application combining theory and instrument

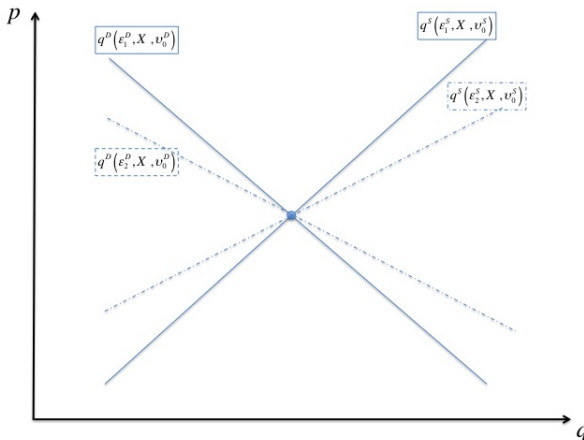
# The Model: Supply and Demand

- ▶ Quantity traded and price are equilibrium outcomes from a system of simultaneous equations:

$$\begin{aligned}q_i^S &= \epsilon^S p_i + \Gamma^S X_i + \nu_i^S \\q_i^D &= \epsilon^D p_i + \Gamma^D X_i + \nu_i^D\end{aligned}$$

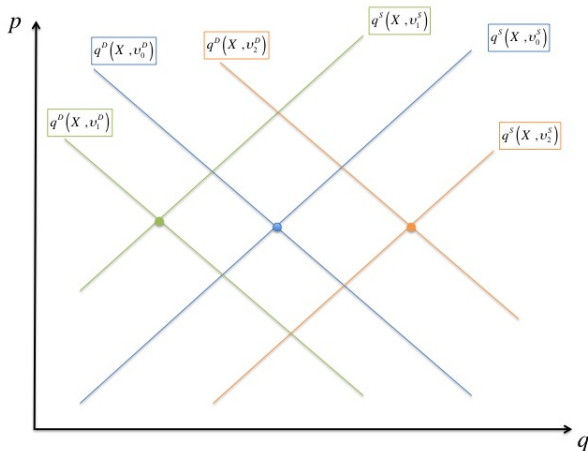
- ▶ Where:
  - ▶  $i$  indexes different markets,  $S$  indexes supply,  $D$  indexes demand
  - ▶  $q$  is log quantity,  $p$  is log price
  - ▶  $X$  is a vector of (pre-determined) observable determinants of demand and supply (including a constant term)
  - ▶  $\{\nu^S, \nu^D\}$  are unobservable determinants of supply and demand.
- ▶ Target parameters:  $\epsilon^S$  and  $-\epsilon^D$

# We only observe the equilibrium, not supply/demand



Solid and dashed lines represent two different supply/demand systems with different elasticities  $\epsilon_1^D \neq \epsilon_2^D$  and  $\epsilon_1^S \neq \epsilon_2^S$  yet observed equilibrium can be rationalized by both systems

# Endogeneity



Endogeneity - equilibria across multiple markets  $i \in \{1, 2, 3\}$  do not trace out either supply or demand

# Exclusion Restrictions - Supply shifter

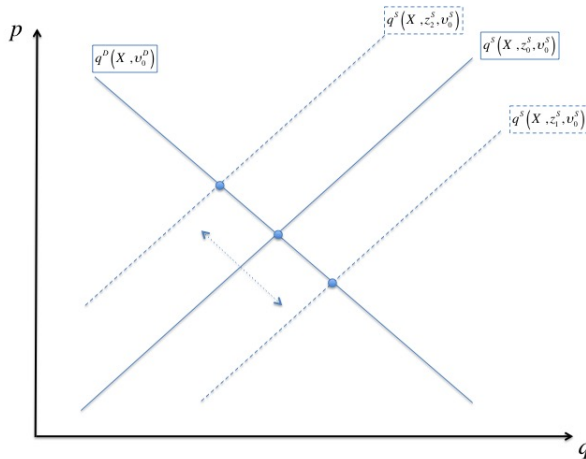
- ▶ Assume that we observe a variable ( $Z_i^S$ ) that enters the supply equation but is excluded from the demand equation:

$$\begin{aligned}q_i^S &= \epsilon^S p_i + \Gamma^S X_i + \theta^S Z_i^S + \nu_i^S \\q_i^D &= \epsilon^D p_i + \Gamma^D X_i + \nu_i^D\end{aligned}$$

- ▶ We further assume:
  - ▶  $\theta^S \neq 0$  so that quantity supplied is a nontrivial function of  $Z_i^S$
  - ▶  $Z_i^S \perp\!\!\!\perp \nu_i^S, \nu_i^D \mid X_i$



# Exclusion Restrictions - Supply shifter



Using variation in  $Z_i^S$  identifies the elasticity of demand by shifting supply along the demand curve.

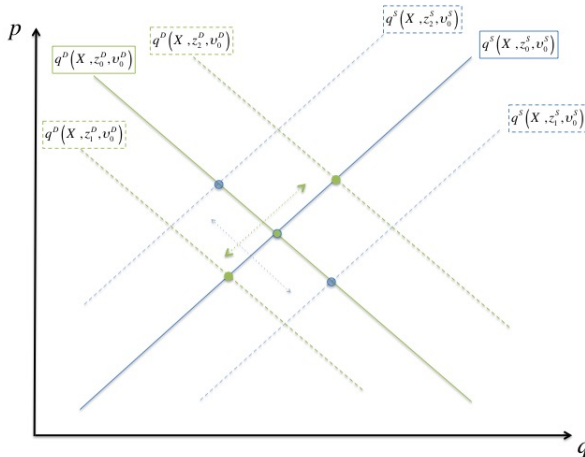
# Exclusion Restrictions - Supply and Demand shifters

- ▶ Assume that in addition to the supply shifter ( $Z_i^S$ ), we observe a variable ( $Z_i^D$ ) that enters the demand equation but is excluded from the supply equation:

$$\begin{aligned}q_i^S &= \epsilon^S p_i + \Gamma^S X_i + \theta^S Z_i^S + \nu_i^S \\q_i^D &= \epsilon^D p_i + \Gamma^D X_i + \theta^D Z_i^D + \nu_i^D\end{aligned}$$

- ▶ We further assume:
  - ▶  $\theta^D \neq 0$  so that quantity demanded is a nontrivial function of  $Z_i^D$
  - ▶  $Z_i^D \perp\!\!\!\perp \nu_i^S, \nu_i^D \mid X_i$

# Exclusion Restrictions - Supply and Demand shifters



Variation in  $Z_i^D$  (holding  $Z_i^S$  constant) identifies the elasticity of supply.  
 Variation in  $Z_i^S$  (holding  $Z_i^D$  constant) identifies the elasticity of demand.

# Supply and Demand Shifters - Reduced Form

- Solving equations for the equilibrium quantity and price on each market  $i$ , we obtain:

$$q_i = \frac{\epsilon^S \Gamma^D - \epsilon^D \Gamma^S}{\epsilon^S - \epsilon^D} X_i + \frac{\epsilon^S \theta^D Z_i^D - \epsilon^D \theta^S Z_i^S}{\epsilon^S - \epsilon^D} + \frac{\epsilon^S \nu_i^D - \epsilon^D \nu_i^S}{\epsilon^S - \epsilon^D}$$
$$p_i = \frac{\Gamma^D - \Gamma^S}{\epsilon^S - \epsilon^D} X_i + \frac{\theta^D Z_i^D - \theta^S Z_i^S}{\epsilon^S - \epsilon^D} + \frac{\nu_i^D - \nu_i^S}{\epsilon^S - \epsilon^D}$$

- Denote by  $q^*$  and  $p^*$  the residual variation in  $q$  and  $p$  after partialling out variation in  $X_i$ .
- Note:  $q_i^* = \frac{\epsilon^S \theta^D Z_i^D - \epsilon^D \theta^S Z_i^S}{\epsilon^S - \epsilon^D} + \frac{\epsilon^S \nu_i^D - \epsilon^D \nu_i^S}{\epsilon^S - \epsilon^D}$  and  $p_i^* = \frac{\theta^D Z_i^D - \theta^S Z_i^S}{\epsilon^S - \epsilon^D} + \frac{\nu_i^D - \nu_i^S}{\epsilon^S - \epsilon^D}$

## IV estimates

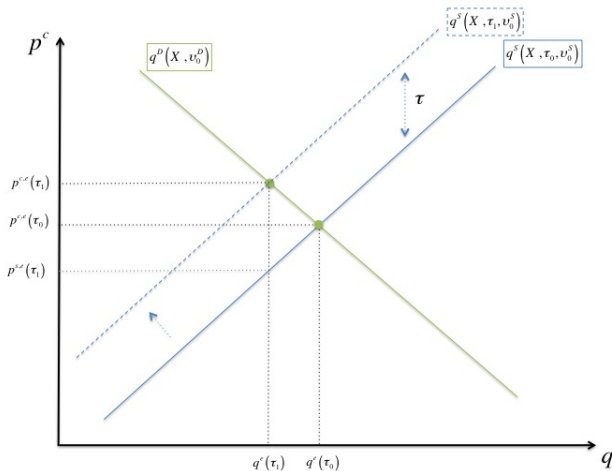
$$\beta^{IV,D} = \frac{\text{Cov}(q_i^*, Z_i^S)}{\text{Cov}(p_i^*, Z_i^S)} = \frac{-\epsilon^D \theta^S}{-\theta^S} = \epsilon^D$$
$$\beta^{IV,S} = \frac{\text{Cov}(q_i^*, Z_i^D)}{\text{Cov}(p_i^*, Z_i^D)} = \frac{\epsilon^S \theta^D}{\theta^D} = \epsilon^S$$

- ▶ IV recovers the elasticities. In general, we need one instrument for each elasticity.
- ▶ An interesting exception: when Tax rate is an instrument  $\Rightarrow$  a single instrument (tax rate) recovers both elasticities (Gavrilova, Zoutman and Hopland 2018)

# Using tax rates as an instrument

- ▶ Assume that there is an ad valorem tax rate  $t_i$  imposed on producers. We define  $\tau_i = \log(1 + t_i)$ .
- ▶ We also denote by  $p_i^C$  the price paid by consumers and by  $p_i^S = p_i^C - \tau_i$  the price received by suppliers.
- ▶ We assume  $\tau_i \perp\!\!\!\perp \nu_i^S, \nu_i^D \mid X_i$
- ▶ Because the tax is on producers, it does not enter the demand equation  $\Rightarrow \epsilon^D$  is identified via standard exclusion restriction
- ▶ Economic theory generates an additional exclusion restriction: Ramsey Exclusion Restriction (see GZH 2018)

# Identification of Demand



The tax is a “supply shifter” - it allows identification of  $\epsilon^D$

# Tax Rate as an Instrument

- ▶ The system of equations becomes:

$$\begin{aligned}q_i^D &= \epsilon^D p_i^c + \Gamma^D X_i + \nu_i^D \\q_i^S &= \underbrace{\epsilon^S p_i^c + \theta^S Z_i^S}_{= -\epsilon^S \tau_i} + \Gamma^S X_i + \nu_i^S \\&= \epsilon^S (p_i^c - \tau_i) + \Gamma^S X_i + \nu_i^S\end{aligned}$$

- ▶ Note: we impose an additional restriction- extremely common in public finance - that suppliers respond to the tax the same way they would respond to a cost shock ( $\theta^S = -\epsilon^S$ ). This directly follows from assumption of profit maximization.



# Tax Rate as an Instrument - Reduced Form

- ▶ Solving previous system of equations for the equilibrium quantity and price on each market  $i$ , we obtain:

$$q_i = \frac{\epsilon^S \Gamma^D - \epsilon^D \Gamma^S}{\epsilon^S - \epsilon^D} X_i + \frac{\epsilon^S \epsilon^D}{\epsilon^S - \epsilon^D} \tau_i + \frac{\epsilon^S \nu_i^D - \epsilon^D \nu_i^S}{\epsilon^S - \epsilon^D}$$
$$p_i^c = \frac{\Gamma^D - \Gamma^S}{\epsilon^S - \epsilon^D} X_i + \frac{\epsilon^S}{\epsilon^S - \epsilon^D} \tau_i + \frac{\nu_i^D - \nu_i^S}{\epsilon^S - \epsilon^D}$$

- ▶ Denote by  $q^*$  and  $p^{s*}$  the residual variation in  $q$  and  $p^c$  after partialling out variation in  $X_i$ .

## Tax Rate as an instrument - IV estimate

$$\beta_{\tau}^{IV,D} = \frac{Cov(q_i^*, \tau_i)}{Cov(p_i^{c*}, \tau_i)} = \epsilon^D$$

- ▶ This directly follows from slide 102 and fact that the tax is excluded from Demand equation (Standard Exclusion Restriction)
- ▶ Can we identify more than just  $\epsilon^D$ ?
- ▶ Yes, it is the role of the additional restriction that suppliers respond to the tax the same way they would respond to an increase in marginal cost ( $\theta^S = -\epsilon^S$ ).  $\Rightarrow$  Key implication is that the passthrough of the tax (to consumers) is  $\frac{dp^c}{d\tau} = \frac{\epsilon^S}{\epsilon^S - \epsilon^D}$

## Tax Rate as an instrument - Identifying $\epsilon^S$

- ▶ Because 1)  $\epsilon^D$  is identified and 2) we can estimate the passthrough  $\frac{dp^c}{d\tau}$  which is a function of the two elasticities, we can recover  $\epsilon^S$ .
- ▶ GZH 2018 recommend using the following IV estimator:

$$\beta_{\tau}^{IV,S} = \frac{Cov(q_i^*, \tau_i)}{Cov(p_i^{S*}, \tau_i)} = \epsilon^S$$