Causal inference week 4: Instrumental variables

LATE, ITT, 2SLS

Andy Eggers

Oxford DPIR

HT 2018

Overview

Previous two weeks were about "selection-on-observables": how to estimate treatment effects by controlling for all relevant covariates.

Key assumption: **conditional independence assumption**, i.e. D_i as-if random conditional on covariates.

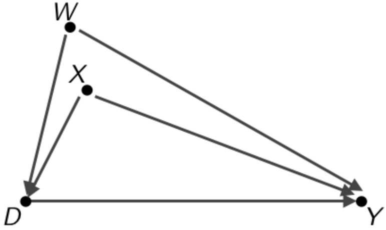
This week we consider situations where:

- Treatment depends on unobservables, i.e. CIA does not hold
- **But** treatment also depends on an as-if random variable Z_i that only affects the outcome through treatment (at least conditional on covariates).

This special variable Z_i is an **instrument**: it wiggles D_i, and we can use this wiggling to measure the effect of D_i.

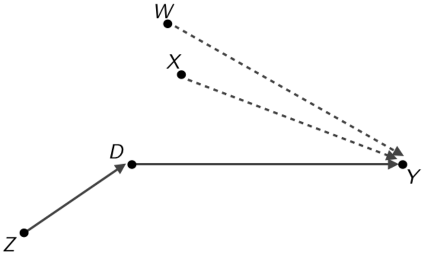
Graphical overview: selection on observables

To estimate the effect of D on Y, we must observe and control for X and W.



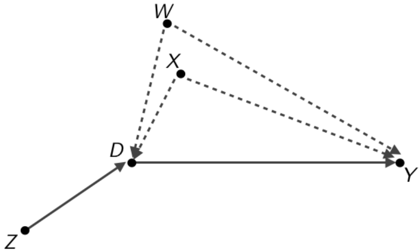
Graphical overview: randomized experiment

If *D* is completely determined by random *Z*, we can measure the effect of *D* on *Y* even if *X* and *W* are not observed (e.g. through DIGM).



Graphical overview: instrumental variables

If D is partly determined by random Z, and Z does not affect Y some other way, we can measure the effect of D on Y even if X and W are not observed (through IV techniques).



Treatment assignment vs. treatment received

In an experiment, we can distinguish between treatment assigned Z_i and treatment received D_i.

We previously (implicitly) assumed $D_i = Z_i$. But in practice there may be **non-compliance**:

- GOTV canvassing experiment in which some people don't answer the door **(one-sided non-compliance)**
- lottery for school places in which some lottery winners do not attend **(one-sided non-compliance)**
- draft lottery for military in which some are drafted but do not serve, some not drafted but serve **(two-sided non-compliance)**

Intention-to-treat (ITT)

Denote by $Y_{i,Z1}$ and $Y_{i,Z0}$ i's potential outcomes if *assigned* to treatment vs. control.

Intention-to-treat (ITT) effect defined as ITT $\equiv E\Big[Y_{i,Z1} - Y_{i,Z0}\Big]$.

If Z_i is randomized, $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$ is an unbiased estimator of the ITT.

If Z_i is randomized but there is non-compliance, $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$ (DIGM) will generally not be an unbiased estimator of the ATE. Why? Selection into treatment received $\Longrightarrow$ selection bias.

- GOTV canvassing experiment
- lottery for school places
- draft lottery for military

The ITT may be interesting on its own. But instrumental variables methods (IV) let us use ITT (effect of treatment assignment) to estimate an ATE (effect of treatment).

Compliance types

		Assigned to control ($Z_i = 0$)	
		Not treated ($D_i = 0$)	Treated ($D_i = 1$)
Assigned to treatment ($Z_i = 1$)	Not treated ($D_i = 0$)	Never taker (N)	Defier (D)
	Treated ($D_i = 1$)	Complier (C)	Always taker (A)

- Can we identify the compliance type of an individual?
- Can we identify the proportion of each compliance type (π_A, π_C, π_D, π_N)?

Estimating compliance frequencies

Who gets treated when $Z_i = 0$? *Always-takers* and *defiers*.
Who gets treated when $Z_i = 1$? *Always-takers* and *compliers*.

Assumption: treatment assignment Z_i is random.

This means the proportion of each compliance type is the same whether $Z_i = 0$ or $Z_i = 1$.

$$E\Big[D_i = 1|Z_i = 0\Big] = \pi_A + \pi_D$$
$$E\Big[D_i = 1|Z_i = 1\Big] = \pi_A + \pi_C$$

Can't estimate π_A, π_D, π_C, or π_N.

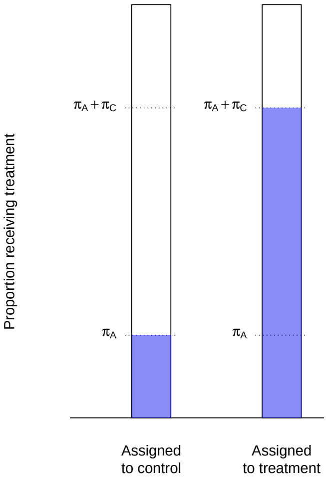
Estimating compliance frequencies (2)

If we assume $\pi_D = 0$ (**no defiers**), then

$$\pi_A = E\Big[D_i = 1 | Z_i = 0\Big]$$

and

$$\pi_C = E\Big[D_i = 1 | Z_i = 1\Big] - E\Big[D_i = 1 | Z_i = 0\Big]$$



ITT decomposition

We can decompose the ITT by **compliance type**.

Let π_G and ITT_G be proportion and ITT for compliance type $G \in \{C, A, N, D\}$.

ITT decomposition

We can decompose the ITT by **compliance type**.

Let π_G and ITT_G be proportion and ITT for compliance type $G \in \{C, A, N, D\}$.

Then by definition

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A \text{ITT}_A + \pi_N \text{ITT}_N + \pi_D \text{ITT}_D \quad (1)$$

ITT decomposition

We can decompose the ITT by **compliance type**.

Let π_G and ITT_G be proportion and ITT for compliance type $G \in \{C, A, N, D\}$.

Then by definition

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A \text{ITT}_A + \pi_N \text{ITT}_N + \pi_D \text{ITT}_D \tag{1}$$

Let's assume

- **No defiers** (monotonicity).
- **Exclusion restriction:** *Treatment assigned* only affects outcomes by affecting *treatment received*.

ITT decomposition

We can decompose the ITT by **compliance type**.

Let π_G and ITT_G be proportion and ITT for compliance type $G \in \{C, A, N, D\}$.

Then by definition

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A \text{ITT}_A + \pi_N \text{ITT}_N + \pi_D \text{ITT}_D \tag{1}$$

Let's assume

- **No defiers** (monotonicity).
- **Exclusion restriction:** *Treatment assigned* only affects outcomes by affecting *treatment received*.

No defiers tells us that $\pi_D = 0$. **Exclusion restriction** tells us that $\text{ITT}_A = \text{ITT}_N = 0$. So:

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A 0 + \pi_N 0 + 0 \text{ITT}_D. \tag{2}$$

ITT decomposition

We can decompose the ITT by **compliance type**.

Let π_G and ITT_G be proportion and ITT for compliance type $G \in \{C, A, N, D\}$.

Then by definition

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A \text{ITT}_A + \pi_N \text{ITT}_N + \pi_D \text{ITT}_D \tag{1}$$

Let's assume

- **No defiers** (monotonicity).
- **Exclusion restriction:** *Treatment assigned* only affects outcomes by affecting *treatment received*.

No defiers tells us that $\pi_D = 0$. **Exclusion restriction** tells us that $\text{ITT}_A = \text{ITT}_N = 0$. So:

$$\text{ITT} = \pi_C \text{ITT}_C + \pi_A 0 + \pi_N 0 + 0 \text{ITT}_D. \tag{2}$$

Exclusion also tells us that, for compliers, the effect of *treatment assignment* on outcomes is the same as the effect of *treatment* on outcomes:

$$\text{ITT} = \pi_C \text{CATE}_C, \tag{3}$$

where CATE_C is the conditional average treatment effect for compliers.

LATE and the Wald estimator

Using **no defiers** and **exclusion restriction**, we got

$$\text{ITT} = \pi_C \text{CATE}_C \tag{4}$$

Assuming $\pi_C > 0$ (**non-zero complier proportion**), the **conditional average treatment effect for compliers** or **local average treatment effect (LATE)** is

$$\text{LATE} = \text{CATE}_C = \frac{\text{ITT}}{\pi_C} \tag{5}$$

LATE and the Wald estimator

Using **no defiers** and **exclusion restriction**, we got

$$\text{ITT} = \pi_C \text{CATE}_C \tag{4}$$

Assuming $\pi_C > 0$ (**non-zero complier proportion**), the **conditional average treatment effect for compliers** or **local average treatment effect (LATE)** is

$$\text{LATE} = \text{CATE}_C = \frac{\text{ITT}}{\pi_C} \tag{5}$$

If in addition Z_i is randomly assigned, we have an unbiased estimator for the above - the **Wald estimator**:

$$\widehat{\text{CATE}}_C = \frac{E\left[Y_i|Z_i = 1\right] - E\left[Y_i|Z_i = 0\right]}{E\left[D_i|Z_i = 1\right] - E\left[D_i|Z_i = 0\right]} = \frac{\text{effect of } Z_i \text{ on } Y_i}{\text{effect of } Z_i \text{ on } D_i} = \frac{\text{ITT}_Y}{\text{ITT}_D} \tag{6}$$

LATE and the Wald estimator

Using **no defiers** and **exclusion restriction**, we got

$$\text{ITT} = \pi_C \text{CATE}_C \tag{4}$$

Assuming $\pi_C > 0$ (**non-zero complier proportion**), the **conditional average treatment effect for compliers** or **local average treatment effect (LATE)** is

$$\text{LATE} = \text{CATE}_C = \frac{\text{ITT}}{\pi_C} \tag{5}$$

If in addition Z_i is randomly assigned, we have an unbiased estimator for the above - the **Wald estimator**:

$$\hat{\text{CATE}}_C = \frac{E\left[Y_i|Z_i = 1\right] - E\left[Y_i|Z_i = 0\right]}{E\left[D_i|Z_i = 1\right] - E\left[D_i|Z_i = 0\right]} = \frac{\text{effect of } Z_i \text{ on } Y_i}{\text{effect of } Z_i \text{ on } D_i} = \frac{\text{ITT}_Y}{\text{ITT}_D} \tag{6}$$

Four assumptions used (not including SUTVA):

- No defiers (monotonicity)
- Exclusion restriction (Z_i affects Y_i only through D_i)
- Non-zero complier proportion
- Random assignment of Z_i

Additional terminology

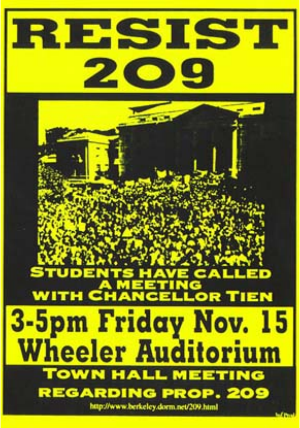
IV methods can be seen as a remedy for a **broken experiment**, i.e. failure to obtain 100% compliance.

More positively, IV methods part of the design of **downstream experiments** or **encouragement designs** in which researcher randomly varies Z_i to create some variation in D_i and then (given exclusion restriction) measures effect of D_i on some outcome Y_i.

Encouragement design example

Proposition 209: 1996 ballot proposition to end race-based preferences (affirmative action) in California government policies

Research question (Albertson and Lawrence 2009): Could watching a TV program affect citizens' attitudes toward Prop. 209?



Albertson and Lawrence 2009: Design

- Representative sample of households in Orange County, CA, interviewed by phone in October 1996
- All respondents told there will be a follow-up interview after the election
- Random subset of respondents told to watch upcoming TV debate on Prop. 209
- In follow-up, asked if they watched the debate; supported Prop. 209; felt knowledgeable about Prop. 209

In this design:

- What are Z_i, D_i, Y_i?
- What does intention-to-treat (ITT) effect mean?
- Is the no-defier assumption reasonable?
- What is the exclusion restriction?
- What does the LATE ($CATE_C$) measure?

Albertson and Lawrence 2009: Data

	$Z_i = 0$	$Z_i = 1$	Difference
Watched TV program	0.052	0.48	0.428
Know about Prop. 209	3.251	3.293	0.041
Support Prop. 209	0.654	0.651	-0.003

- What is π_C?
- What is the ITT?
- What is LATE i.e. CATE_C?

Taking stock

We assumed **binary treatment assignment** and **binary treatment**.

Given random treatment assignment, we can

- estimate the **intention-to-treat** effect (ITT) by comparing average Y_i among units assigned to treatment and units assigned to control
- estimate the **proportion of compliers** (π_C) by comparing average D_i among units assigned to treatment and units assigned to control
- estimate the **LATE** (CATE_C) by dividing the ITT by the proportion of compliers

Taking stock

We assumed **binary treatment assignment** and **binary treatment**.

Given random treatment assignment, we can

- estimate the **intention-to-treat** effect (ITT) by comparing average Y_i among units assigned to treatment and units assigned to control
- estimate the **proportion of compliers** (π_C) by comparing average D_i among units assigned to treatment and units assigned to control
- estimate the **LATE** (CATE_C) by dividing the ITT by the proportion of compliers

Can we generalize this somehow?

- non-binary treatment (D_i)
- non-binary instrument (Z_i)
- covariates (e.g. because non-random Z_i)
- more than one instrument

Another way to get LATE

We estimated the LATE with

$$\frac{E\Big[Y_i|Z_i = 1\Big] - E\Big[Y_i|Z_i = 0\Big]}{E\Big[D_i|Z_i = 1\Big] - E\Big[D_i|Z_i = 0\Big]} = \frac{\text{ITT}_Y}{\text{ITT}_D}$$

Here is another way:

- Regress D_i on Z_i, get fitted values $\hat{D}_i$
- Regress Y_i on $\hat{D}_i$

This is called **two-stage least squares**.

Two-stage least squares: illustration

Estimating LATE via $\text{ITT}_Y/\text{ITT}_D$:

```
# define the sample -- no missing data on key vars
use = !is.na(d$infopro2) & !is.na(d$watchpro) & !is.na(d$conditn) & !is.na(d$support3)

itt.y.reg = lm(infopro2 ~ conditn, data = d[use, ]) # conditn is Z, infopro2 is Y
itt.d.reg = lm(watchpro ~ conditn, data = d[use, ]) # watchpro is D
coef(itt.y.reg)["conditn"]/coef(itt.d.reg)["conditn"]  # IV estimate by hand
```

```
   conditn
0.09666038
```

Same thing via two-stage least squares:

```
d$watchpro.fit = NA  # store fitted treatment status here
d$watchpro.fit[use] = predict(lm(watchpro ~ conditn, data = d[use,], na.action = na.exclude))

summary(lm(infopro2 ~ watchpro.fit, data = d[use,]))  # 2sls estimate
```

```
Call:
lm(formula = infopro2 ~ watchpro.fit, data = d[use, ])

Residuals:
    Min      1Q  Median      3Q     Max
-2.2926 -0.2926 -0.2512  0.7074  0.7488

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.24615    0.06242   52.01   <2e-16 ***
watchpro.fit  0.09666    0.17915    0.54     0.59
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8039 on 438 degrees of freedom
Multiple R-squared:  0.0006642,	Adjusted R-squared:  -0.001617
F-statistic: 0.2911 on 1 and 438 DF,  p-value: 0.5898
```

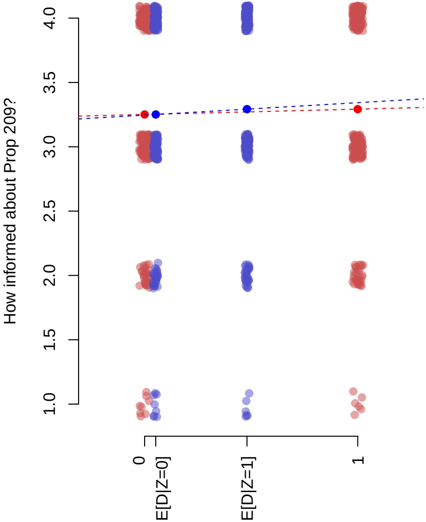
Why does 2SLS work? Intuition

Wald estimator:

- Regressing Y on Z gives you ITT_Y.
- Dividing ITT_Y by $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$ inflates to give you CATE_C.

2SLS:

- Regressing D on Z gives you fitted values $E[D_i|Z_i = 1]$ and $E[D_i|Z_i = 0]$
- Regressing Y on $\hat{D}$ inflates to give you CATE_C.



Why does 2SLS work? Math

Regression fact: The slope coefficient from the regression of Y on X is

$$\frac{\text{Cov}(Y,X)}{\text{Var}(X)}$$

Let

- $\rho = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$ (ITT_Y)
- $\phi = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$ (ITT_D)

Can get ρ from regression of Y_i on Z_i.
Can get ϕ from regression of D_i on Z_i.
Wald estimator is ρ/ϕ.

Regressing D_i on Z_i, we get

$$\hat{D}_i = \alpha + \phi Z_i.$$

Regressing Y_i on $\hat{D}_i$, slope coefficient is

$$\begin{aligned} &= \frac{\text{Cov}(Y_i, \alpha + \phi Z_i)}{\text{Var}(\alpha + \phi Z_i)} \\ &= \frac{\phi\text{Cov}(Y_i, Z_i)}{\phi^2\text{Var}(Z_i)} \\ &= \frac{\rho}{\phi} \end{aligned}$$

Now we can generalize

Wald estimator is limited to binary D_i and Z_i:

$$\lambda = \frac{E\left[Y_i|Z_i = 1\right] - E\left[Y_i|Z_i = 0\right]}{E\left[D_i|Z_i = 1\right] - E\left[D_i|Z_i = 0\right]} = \frac{\text{ITT}_Y}{\text{ITT}_D}$$

Two-stage least squares is a much more general procedure:

$$\begin{aligned} \text{First stage:} \quad D_i &= \alpha_1 + \phi Z_i + \beta_1 X_{1i} + \gamma_1 X_{2i} + e_{1i} \\ \text{Second stage:} \quad Y_i &= \alpha_2 + \lambda \hat{D}_i + \beta_2 X_{1i} + \gamma_2 X_{2i} + e_{2i} \end{aligned}$$

where Z_i and D_i might not be binary and you can include covariates e.g. X_{1i}, X_{2i}.

Two-stage least squares: terminology and assumptions

Terminology:

$$\begin{aligned}
\text{Reduced form:} \quad Y_i &= \alpha_0 + \rho Z_i + \beta_0 X_{1i} + \gamma_0 X_{2i} + e_{0i} \\
\text{First stage:} \quad D_i &= \alpha_1 + \phi Z_i + \beta_1 X_{1i} + \gamma_1 X_{2i} + e_{1i} \\
\text{Second stage:} \quad Y_i &= \alpha_2 + \lambda \hat{D}_i + \beta_2 X_{1i} + \gamma_2 X_{2i} + e_{2i}
\end{aligned}$$

Key assumptions (Wald assumptions with covariates and without “complier” terminology):

- **Non-zero first-stage**: instrument affects treatment, conditional on covariates ($\phi \neq 0$ in first stage)
- **Independence** (exogeneity, ignorability): instrument unrelated to potential outcomes, conditional on covariates (no OVB on ρ in reduced form or ϕ in first stage)
- **Exclusion restriction**: instrument only affects outcome through treatment, conditional on covariates
- **Monotonicity**: instrument’s effect on treatment is weakly positive or weakly negative for all units

NB: Must use same covariates in first stage and second stage.

Kern and Hainmueller (2009) on impact of West German TV in East Germany

- **Question**: How did watching West German TV affect East Germans' political attitudes?
- **Treatment**: Watching West German TV (as measured in surveys conducted by the Zentralinstitut für Jugendforschung in late 1988-early 1989)
- **Outcome**: Regime support (measured in same surveys)

To consider:

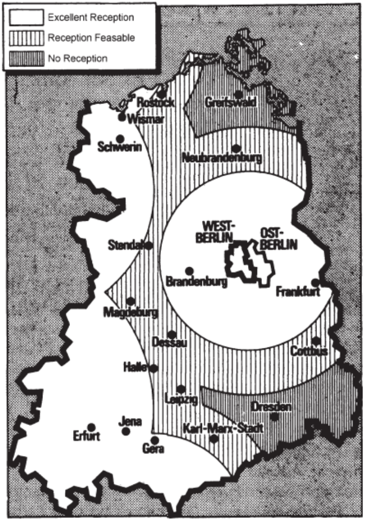
- What about just regressing outcome on treatment?
- What covariates might remove the bias in that regression?
- How might an IV approach help?

Kern & Hainmueller (2)

Instrument: living in a place where West German TV signals could reach

Evaluate:

- **Independence** (exogeneity, ignorability): instrument unrelated to potential outcomes, conditional on covariates
- **Exclusion restriction**: instrument only affects outcome through treatment, conditional on covariates
- **Monotonicity**: instrument's effect on treatment is weakly positive or weakly negative for all units



Kern & Hainmueller: results

Table 3 Effect of West German television exposure on regime support

Estimator	*Diff.*	*LATE*	*2SLS*	*LARF*	*2SLS*	*LARF*
Covariate set	—	—	*Limited*	*Limited*	*Extensive*	*Extensive*
Convinced of Leninist/Marxist worldview						
West German TV	−0.079	0.147	0.205	0.204	0.198	0.204
	(0.053)	(0.083)	(0.084)	(0.084)	(0.087)	(0.108)
Feel closely attached to East Germany						
West German TV	−0.013	0.217	0.258	0.255	0.256	0.251
	(0.044)	(0.067)	(0.072)	(0.075)	(0.073)	(0.090)
Political power exercised in ways consistent with my views						
West German TV	−0.014	0.158	0.193	0.191	0.186	0.185
	(0.047)	(0.078)	(0.082)	(0.083)	(0.081)	(0.106)

Note. $N = 3441$. The table shows treatment effect estimates for each specification with cluster-adjusted standard errors in parentheses. Diff. denotes the difference in means between exposed and unexposed respondents. LATE is the unconditional average treatment effect for compliers. 2SLS is the two-stage least squares estimator. LARF is the local average response function estimator. The limited covariate set includes age, gender, and father's and mother's occupational classification. The extensive covariate set adds marriage status, living situation, number of children, highest educational attainment, occupational classification, net monthly income, and employment status to the limited set. Response categories for the outcome variables are coded as fully disagree = 1, largely disagree = 2, largely agree = 3, and fully agree = 4.

Martin and Yurukoglu (2017) on impact of Fox News in USA

- **Question**: “how much does consuming slanted news, like the Fox News Channel, change individuals’ partisan voting preferences?”
- **Treatment**: Minutes spent watching Fox News Channel, based on surveys
- **Outcome**: Voting in presidential election, based on aggregate zip code-level results

To consider:

- What about just regressing outcome on treatment?
- What covariates might remove the bias in that regression?
- How might an IV approach help?

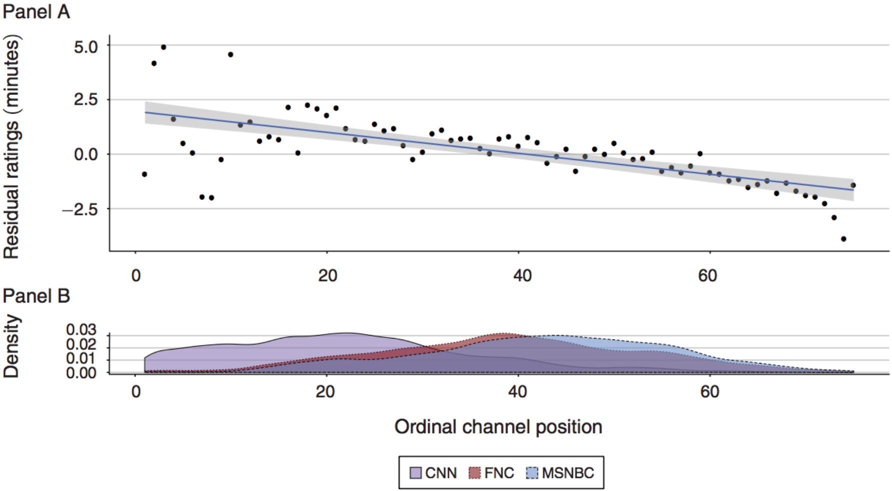
Martin & Yurukoglu (2)

Instrument: channel position of Fox News on the cable lineup

Evaluate:

- **Independence** (exogeneity, ignorability): instrument unrelated to potential outcomes, conditional on covariates
- **Exclusion restriction**: instrument only affects outcome through treatment, conditional on covariates
- **Monotonicity**: instrument's effect on treatment is weakly positive or weakly negative for all units

Martin & Yurukoglu: first-stage



Martin & Yurukoglu: first-stage (2)

TABLE 2—FIRST-STAGE REGRESSIONS: NIELSEN DATA

	FNC minutes per week					
	(1)	(2)	(3)	(4)	(5)	(6)
FNC position	−0.146	−0.075	−0.174	−0.167	−0.097	−0.111
	(0.043)	(0.039)	(0.028)	(0.025)	(0.033)	(0.030)
MSNBC position	0.078	0.073	0.064	0.070	0.019	0.020
	(0.036)	(0.032)	(0.025)	(0.022)	(0.034)	(0.035)
Has MSNBC only	1.904	1.137	−3.954	−2.804	−1.220	−1.562
	(3.697)	(3.713)	(4.255)	(3.416)	(6.180)	(5.397)
Has FNC only	31.423	26.526	23.460	22.011	15.141	15.069
	(2.677)	(2.546)	(2.278)	(1.864)	(2.697)	(2.314)
Has both	24.859	23.118	18.338	16.168	15.159	14.486
	(2.919)	(2.687)	(2.361)	(1.991)	(3.216)	(2.842)
Satellite FNC minutes				0.197		0.173
				(0.013)		(0.015)
Fixed effects	Year	State-year	State-year	State-year	County-year	County-year
Cable controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	None	None	Extended	Extended	Extended	Extended
Robust *F*-stat	11.39	3.72	39.02	44.7	8.86	13.43
Number of clusters	5,789	5,789	4,830	4,761	4,839	4,770
Observations	71,150	71,150	59,541	52,053	59,684	52,165
R^2	0.030	0.074	0.213	0.377	0.428	0.544

Notes: Cluster-robust standard errors in parentheses (clustered by cable system). Instrument is the ordinal position of FNC on the local system. The omitted category for the availability dummies is systems where neither FNC nor

Martin & Yurukoglu: reduced form

TABLE 3—REDUCED-FORM REGRESSIONS: ZIP CODE VOTING DATA

	2008 McCain vote percentage			
	(1)	(2)	(3)	(4)
FNC cable position	−0.011	0.004	−0.027	−0.015
	(0.023)	(0.020)	(0.008)	(0.008)
MSNBC cable position	0.054	0.041	0.008	0.003
	(0.019)	(0.016)	(0.005)	(0.006)
Has MSNBC only	−2.118	−0.465	0.749	1.374
	(1.585)	(1.306)	(1.002)	(1.219)
Has FNC only	7.557	5.500	2.262	1.061
	(1.175)	(0.975)	(0.547)	(0.504)
Has both	4.223	4.351	1.358	0.814
	(1.521)	(1.269)	(0.661)	(0.609)
Fixed effects	None	State	State	County
Cable system controls	Yes	Yes	Yes	Yes
Demographics	None	None	Extended	Extended
Number of clusters	6,035	6,035	4,814	4,814
Observations	22,584	22,584	17,400	17,400
R^2	0.148	0.294	0.833	0.907

Notes: Cluster-robust standard errors in parentheses (clustered by cable system). See first-stage tables for description of control variables.

Martin & Yurukoglu: second-stage

TABLE 4—SECOND STAGE REGRESSIONS: ZIP CODE VOTING DATA

	2008 McCain vote percentage			
	(1)	(2)	(3)	(4)
Predicted FNC minutes	0.152	0.120	0.157	0.098
	(0.056, 0.277)	(0.005, 0.248)	(−0.126, 0.938)	(−0.121, 0.429)
Satellite FNC minutes		−0.021		−0.015
		(−0.047, 0.001)		(−0.073, 0.022)
Fixed effects	State	State	County	County
Cable system controls	Yes	Yes	Yes	Yes
Demographics	Extended	Extended	Extended	Extended
Number of clusters	4,814	3,993	4,729	4,001
Observations	17,400	12,417	17,283	12,443
R^2	0.833	0.841	0.907	0.919

Notes: The first stage is estimated using viewership data for all Nielsen TV households. See first-stage tables for description of instruments and control variables. Observations in the first stage are weighted by the number of survey individuals in the zip code according to Nielsen. Confidence intervals are generated from 1,000 independent STID-block-bootstraps of the first and second stage datasets. Reported lower and upper bounds give the central 95 percent interval of the relevant bootstrapped statistic.

Dinas et al (2018) on political impact of refugees

- **Question**: Did the influx of refugees in Greece increase support for the right-wing Golden Dawn party in 2015?
- **Treatment**: Number of refugees arriving per capita in locality
- **Outcome**: Golden Dawn vote share in locality

To consider:

- What about just regressing outcome on treatment?
- What covariates might remove the bias in that regression?
- How might an IV approach help?

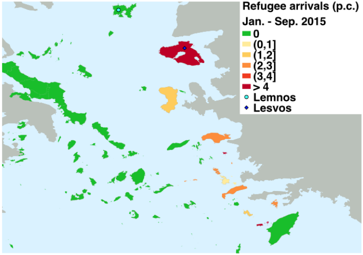
Dinas et al on the Golden Dawn (2)

They use two strategies: diff-in-diff (wait until week 6) and IV.

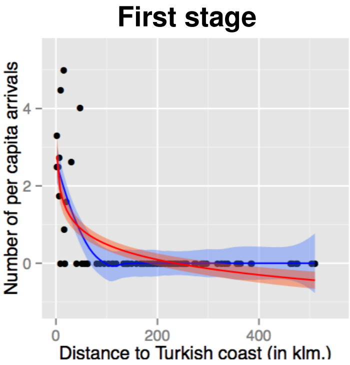
Instrument: distance to the Turkish coast.

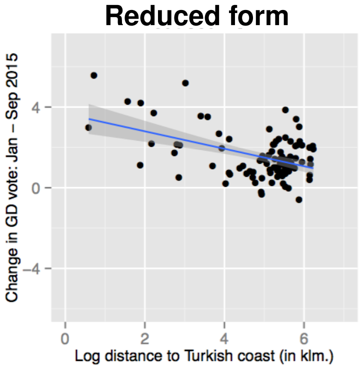
Evaluate:

- **Independence** (exogeneity, ignorability): instrument unrelated to potential outcomes, conditional on covariates
- **Exclusion restriction**: instrument only affects outcome through treatment, conditional on covariates
- **Monotonicity**: instrument's effect on treatment is weakly positive or weakly negative for all units



Dinas et al on the Golden Dawn (3)





Why I am skeptical of most IV designs

IV designs must convince me of two key untestable assumptions:

- Your instrument Z_i satisfies **independence**, i.e. the CIA is met: conditional on covariates, Z_i is as-if random.
- Your instrument Z_i satisfies **exclusion**, i.e. it only affects Y_i through some D_i.

When Z_i is randomly determined in an experiment, I'll accept **independence** and think hard about **exclusion**.

In an observational study, very hard to convince me of either one.

- Is the CIA really satisfied in the reduced form?
- Is D_i really the only channel through which Z_i affects Y_i?

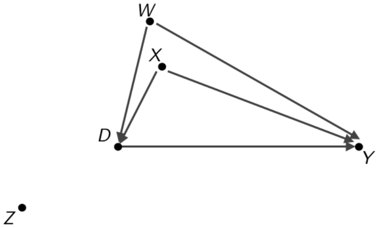
Instrument or covariate?

Consider a study of media effects, where D_i is exposure to some media and Y_i is voting.

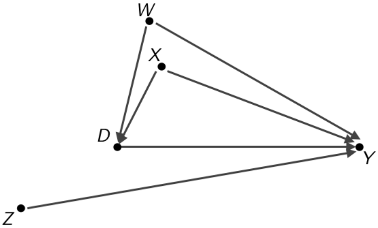
		Affects Y_i? No	Affects Y_i? Yes
Affects D_i?	No	**Noise** e.g. month of birth	**Irrelevant determinant** e.g. electoral campaign
	Yes	**Instrument** e.g. channel position	**Covariate** e.g. political orientation

Instrument or covariate? (2)

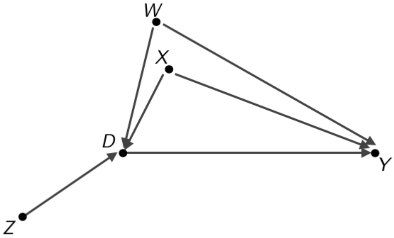
Noise



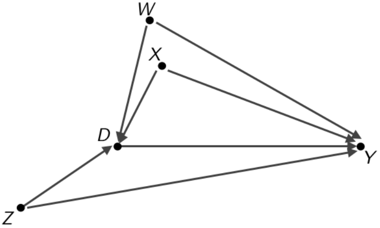
Irrelevant determinant



Instrument



Covariate



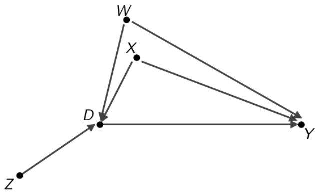
Can we test the exclusion restriction?

What about

- regress Y on D and Z
- conclude exclusion is valid if coefficient on Z is 0

Unfortunately this doesn't work, because D is also affected by X and W:

- if X and W are observed, you don't need IV to estimate effect of D on Y
- if they are *not* observed, by controlling for D you induce an association between Z and X and/or W, leading to **collider bias**



Collider bias: intuition and examples

Suppose Z_i and X_i are not correlated with each other, but both increase the probability of $D_i = 1$.

Conditional on D_i, Z_i and X_i may be negatively correlated:

- e.g. height (Z_i) and athletic ability (X_i) among basketball players who become pros ($D_i = 1$)
- e.g. low channel position of Fox (Z_i) and political conservatism (X_i) among people who watch Fox News ($D_i = 1$)

Therefore, regressing some Y on D_i and Z_i (but not X_i), coefficient on Z_i is contaminated by effect of omitted X_i on Y_i.

You could find an effect of Z_i even if the exclusion restriction actually holds (Gerber & Green pg. 199).

See also: Acharya, Blackwell, Sen 2016 for requirements for estimating controlled direct effect: exclusion restriction is that CDE $= 0$, but assumptions necessary to estimate CDE are equivalent to CIA for estimating effect of D_i on Y_i

Can we test the exclusion restriction? (2)

Angrist & Pischke recommend the following indirect test:

- Identify a subset of observations for which the first stage (or complier proportion) should be zero
- Show that the 2SLS (LATE) estimate for this subset is zero

How would this work in e.g. the Greek islands example?

Further thoughts on IV

- In observational studies, a variable that satisfies **independence** (CIA) is a rare and wonderful thing. Usually **exclusion** is doubtful, but you can measure its effect and speculate about channels.
- If exclusion is also plausible, you are truly blessed. But do not expect this.
- Now that you know about instrumental variables, you should not refer to an independent variable in a regression as an “IV”: say “treatment”, “control variable”, “covariate”, “regressor”, “RHS variable”