Lab 6: Instrumental Variables

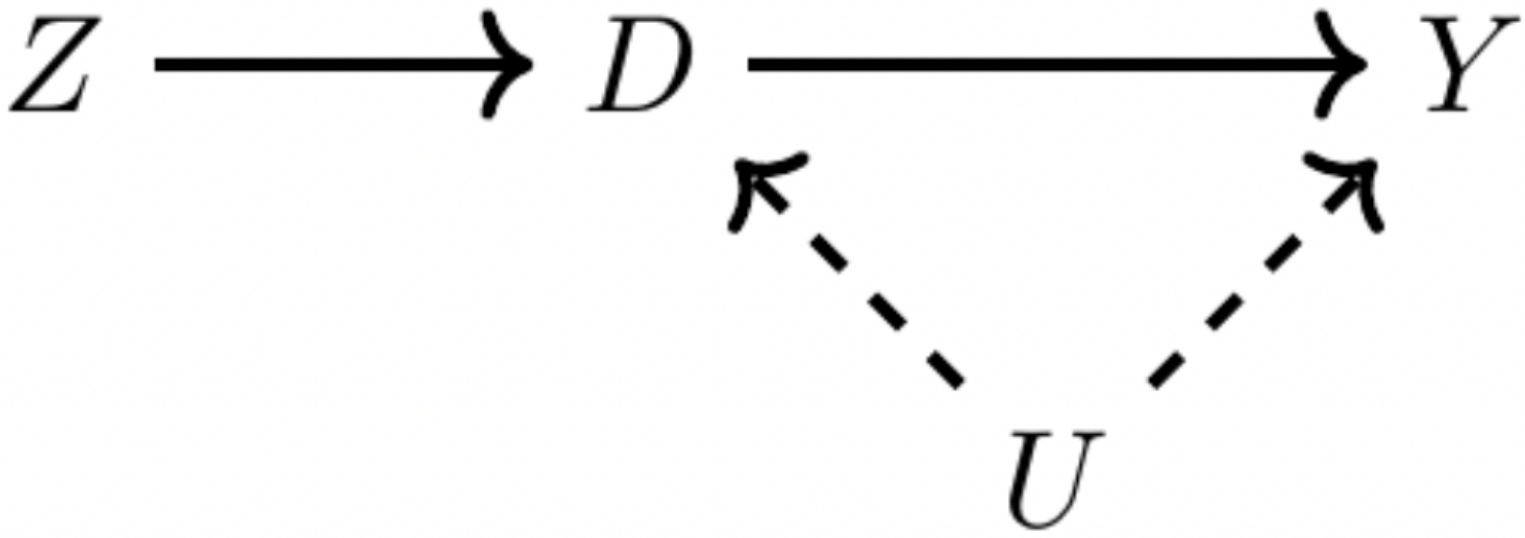
Dias Akhmetbekov

2024-03-08

Plan

- IV basics
- Characterizing Compliers
- Weak Instruments
- Practical Recommmendations

DAG



IV in practice: peasant unrest and representation

American Political Science Review (2018) 112, 1, 125–147
doi:10.1017/S0003055417000454

Collective Action and Representation in Autocracies: Evidence from Russia's Great Reforms

PAUL CASTAÑEDA DOWER *University of Wisconsin–Madison*
EVGENY FINKEL *George Washington University*
SCOTT GEHLBACH *University of Wisconsin–Madison*
STEVEN NAFZIGER *Williams College*

We explore the relationship between capacity for collective action and representation in autocracies with data from Imperial Russia. Our primary empirical exercise relates peasant representation in new institutions of local self-government to the frequency of peasant unrest in the decade prior to reform. To correct for measurement error in the unrest data and other sources of endogeneity, we exploit idiosyncratic variation in two determinants of peasant unrest: the historical incidence of serfdom and religious polarization. We find that peasants were granted less representation in districts with more frequent unrest in preceding years—a relationship consistent with the Acemoglu-Robinson model of political transitions and inconsistent with numerous other theories of institutional change. At the same time, we observe patterns of redistribution in subsequent years that are inconsistent with the commitment mechanism central to the Acemoglu-Robinson model. Building on these results, we discuss possible directions for future theoretical work.

The effect of unrest on representation

```r
library(haven); library(AER); library(stargazer)
data <- read_dta("DFGN_cleaned.dta")

## OLS
olsfit <- lm(peasantrepresentation_1864 ~ afreq + distance_moscow +
               goodsoil + lnurban + lnpopn +
               province_capital, data)

## IV (1): serfdom
ivfit1 <- ivreg(peasantrepresentation_1864 ~ afreq + distance_moscow +
                  goodsoil + lnurban + lnpopn + province_capital | serfperc1 +
                  distance_moscow + goodsoil + lnurban + lnpopn +
                  province_capital, data=data)

## IV (2): religious polarization
ivfit2 <- ivreg(peasantrepresentation_1864 ~ afreq + distance_moscow +
                  goodsoil + lnurban + lnpopn + province_capital | religpolarf4_1870 +
                  distance_moscow + goodsoil + lnurban + lnpopn +
                  province_capital, data=data)

mod <- list(olsfit, ivfit1, ivfit2)
ses <- lapply(mod, function(x) coeftest(x, vcov = vcovHC(x, type = "HC1"))[,"Std. Error"])
labs <- c("", "Z: % serfs", "Z: religious pol.")
```

The effect of peasant unrest on representation

```
stargazer(mod, se = ses, column.labels = labs, omit.stat = c("f", "ser"), type = "text",omit=c("Constant")
```

```
##
## ===================================================
##                          Dependent variable:
##                    -------------------------------
##                          peasantrepresentation_1864
##                       OLS            instrumental
##                                        variable
##                                 Z: % serfs Z: religious pol.
##                       (1)          (2)            (3)
## ---------------------------------------------------
## afreq               -4.249**   -41.999***      -32.770*
##                      (1.830)    (8.509)        (17.352)
##
## distance_moscow      0.379     -7.222***        -5.401
##                      (1.288)    (2.203)         (3.733)
##
## goodsoil             1.127      3.860***        3.101*
##                      (0.811)    (1.317)         (1.801)
##
## lnurban             -2.605***  -1.901***       -2.086***
##                      (0.439)    (0.584)         (0.555)
##
## lnpopn              5.224***    8.291***       7.597***
##                      (1.092)    (1.243)         (1.777)
##
## province_capital    -3.345***  -5.177***       -4.689***
##                      (1.281)    (1.679)         (1.642)
##
## ---------------------------------------------------
## Observations          365        362             361
## R2                   0.396      -0.236          0.042
## Adjusted R2          0.386      -0.257          0.026
## ===================================================
```

Applications Discussion

- Randomized experiments with noncompliance
- Papers using a lagged version of the treatment as an instrument
- Miguel, Satyanath and Sergenti (2004), "Economic Shocks and Civil Conflict."
- Aronow, Carnegie and Marinov, "The Effects of Foreign Aid on Rights and Governance."
- Angrist (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery."
- Gerber, Green, Shachar (2003), "Voting May Be Habit-forming."
- Acemoglu et al. (2011), "The Consequences of Radical Reform: The French Revolution."

IV with Heterogeneous Treatment Effects

- Binary instrument $Z_i \in \{0, 1\}$
- Binary treatment $D_z \in \{0, 1\}$ is potential treatment status given $Z = z$
- Potential outcomes:
  $Y_i(D, Z) = \{Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0)\}$
- Heterogeneous treatment effects $\beta_i = Y_i(1) - Y_i(0)$
- Note that with constant treatment effect it is enough to use structural equations.

Compliance Types

- Four compliance types (or principal strata) in this setting:
  - Complier $D_i(1) = 1$ and $D_i(0) = 0$
  - Always-taker $D_i(1) = D_i(0) = 1$
  - Never-taker $D_i(1) = D_i(0) = 0$
  - Defier $D_i(1) = 0$ and $D_i(1) = 1$
- Connections between observed data and compliance types:

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Never-taker or Complier	Never-taker or Defier
$D_i = 1$	Always-taker or Defier	Always-taker or Complier

- Let π_{co}, π_{at}, π_{nt}, and π_{de} be the proportions of each type.

IV Assumptions

- Canonical IV assumptions for Z_i to be a valid instrument:
  1. Randomization of Z_i
  2. Presence of some compliers $\pi_{co} \neq 0$ (first-stage)
  3. Exclusion restriction $Y_i(z, d) = Y_i(z', d)$
  4. Monotonicity: $D_i(1) \geq D_i(0)$ for all i (no defiers)
- Implies ITT effect on treatment equals proportion compliers: $ITT_D = \pi_{co}$
- Implies ITT for the outcome has the same interpretation:

$$ITT_Y = ITT_{Y,co}\pi_{co} + \underbrace{ITT_{Y,at}}_{=0 \text{ (ER)}}\pi_{at} + \underbrace{ITT_{Y,nt}}_{=0 \text{ (ER)}}\pi_{nt} + ITT_{Y,de}\underbrace{\pi_{de}}_{=0 \text{ (mono)}}$$

$$= ITT_{Y,co}\pi_{co}$$

- $\approx$ same identification result: $\tau_{LATE} = \frac{ITT_Y}{ITT_D}$

LATE Theorem

Theorem: Under assumptions 1 - 4:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

Proof.

Start with the first bit of the Wald estimator:

$$\begin{aligned} E[Y_i|Z_i = 1] &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_i|Z_i = 1] \\ &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}] \end{aligned}$$

LATE Theorem

Proof.

Similarly

$$E[Y_i|Z_i = 0] = E[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}]$$

So the numerator of the Wald estimator is

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})]$$

Monotonicity means $D_{1i} - D_{0i}$ equals one or zero, so

$$E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]P[D_{1i} > D_{0i}].$$

A similar argument shows

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = P[D_{1i} > D_{0i}].$$

Better LATE than never or vice versa?

- Angrist, Imbens, and others: The LATE cup is half full. We don't get the average effect on a stable population (like the average treatment effect on the treated) but we get something that still makes some sense (particularly for policy).
- Heckman, Deaton, and others: The LATE cup is half empty (or all empty!): We don't get what we want, the average treatment effect on the treated, so IV is somewhat useless, and we need to augment it with something else to produce economically meaningful parameters (e.g. more structural econometric models).

Getting under IV hood

It can often be useful to characterize the compliers of a given IV

- E.g., to hint at mechanisms, contextualize findings, or reconcile findings across different (quasi-)experiments

Of course we can't tell if $D_i(1) > D_i(0)$ for any given i. But it turns out we can still learn about compliers *on average*. We'll step through:

1. Average potential outcomes: $E[Y_i(d)|D_i(1) > D_i(0)]$ for $d \in \{0, 1\}$
2. Characteristics: $E[W_i|D_i(1) > D_i(0)]$ for baseline W_i
3. Abadie's Kappa

This is all easier than they might seem...

Separating Treated and Untreated Outcomes

Suppose we want to know $E[Y_i(1)|D_i(1) > D_i(0)]$ in the basic IA setup

Trick: Consider IV on the modified outcome $\tilde{Y}_i = Y_i D_i$ instead of Y_i

- Potential outcomes: $\tilde{Y}_i(1) = Y_i(1)$, and $\tilde{Y}_i(0) = 0$
- Hence "treatment effects": $\tilde{Y}_i(1) - \tilde{Y}_i(0) = Y_i(1) - 0 = Y_i(1)$
- Check: exclusion, independence, and monotonicity still hold with $\tilde{Y}_i$
- Hence IV identifies LATE: $E[\tilde{Y}_i(1)|D_i(1) > D_i(0)]$

Similar logic shows that IV of $\tilde{Y}_i = Y_i(1 - D_i)$ on $\tilde{D}_i = 1 - D_i$ identifies a LATE of
$E[\tilde{Y}_i(1) - \tilde{Y}_i(0)|D_i(1) > D_i(0)] = E[Y_i(0)|D_i(1) > D_i(0)]$

Such an easy bonus analysis to any *ivreg*!

Illustration: Angrist, Pathak, and Walters (2013)

American Economic Journal: Applied Economics 2013, 5(4): 1–27
http://dx.doi.org/10.1257/app.5.4.1

Explaining Charter School Effectiveness†

By JOSHUA D. ANGRIST, PARAG A. PATHAK, AND CHRISTOPHER R. WALTERS*

Lottery estimates suggest Massachusetts' urban charter schools boost achievement well beyond that of traditional urban public schools students, while nonurban charters reduce achievement from a higher baseline. The fact that urban charters are most effective for poor nonwhites and low-baseline achievers contributes to, but does not fully explain, these differences. We therefore link school-level charter impacts to school inputs and practices. The relative efficacy of urban lottery sample charters is accounted for by these schools' embrace of the No Excuses approach to urban education. In our Massachusetts sample, Non-No-Excuses urban charters are no more effective than nonurban charters. (*JEL* H75, I21, I28)

Illustration: Angrist, Pathak, and Walters (2013)

	Urban				Nonurban			
Subject	Treatment effect (1)	$E_u[Y_0 \mid D=0]$ (2)	λ_0^u (3)	λ_1^u (4)	Treatment effect (5)	$E_n[Y_0 \mid D=0]$ (6)	λ_0^n (7)	λ_1^n (8)
Panel A. Middle school								
Math	0.483***	−0.399***	0.077	0.560***	−0.177**	0.236***	0.010	−0.143***
	(0.074)	(0.011)	(0.049)	(0.054)	(0.074)	(0.007)	(0.061)	(0.042)
N	4,858				2,239			
ELA	0.188***	−0.422***	0.118**	0.306***	−0.148***	0.260***	0.102**	−0.086***
	(0.064)	(0.012)	(0.054)	(0.049)	(0.048)	(0.007)	(0.050)	(0.030)
N	4,551				2,323			

Decomposing

$$LATE = \underbrace{E[Y_i(1)|D_i(1) > D_i(0)] - E[Y_i(0)|D_i = 0]}_{\lambda_1}$$

$$- \underbrace{(E[Y_i(0)|D_i(1) > D_i(0)] - E[Y_i(0)|D_i = 0])}_{\lambda_0}$$

shows that charter compliers have typical counterfactual achievement

Complier Summary Statistics

We can use the same trick to estimate $E[W_i|D_i(1) > D_i(0)]$:

- IV of $W_i D_i$ on D_i, instrumenting with Z_i
- IV of $W_i(1 - D_i)$ on $(1 - D_i)$, instrumenting with Z_i
- Some weighted average of the two

Testing that these two approaches indeed estimate the same thing can be shown to be equivalent to a balance regression of W_i on Z_i

- "Stacking" the two specifications up, and estimating a single $E[W_i|D_i(1) > D_i(0)]$ with two IVs, automates the weights+overid. test

Fun to compare with $E[W_i|D_i(1) = D_i(0) = 1] = E[W_i|D_i = 1, Z_i = 0]$ and $E[W_i|D_i(1) = D_i(0) = 0] = E[W_i|D_i = 0, Z_i = 1]$

Illustration: Angrist, Hull, and Walters (2023)

CHAPTER

Methods for measuring school effectiveness

1

Joshua Angrist[a,b], **Peter Hull**[b,c], **and Christopher Walters**[b,d]

[a]*MIT, Cambridge, MA, United States*
[b]*NBER, Cambridge, MA, United States*
[c]*Brown University, Providence, RI, United States*
[d]*UC Berkeley, Berkeley, CA, United States*

1 Introduction

Many personal and policy decisions turn on perceptions of school quality. Families choose schools and neighborhoods by balancing perceived school effectiveness against other factors, like housing costs and commute times. Questions of school quality also drive high-stakes policy decisions related to school closures, restructuring, and expansion. In response to the demand for quality information, achievement-based measures of public school effectiveness have proliferated. Such measures include "school report cards" distributed by some states and districts, as well as school quality ratings published by private organizations like GreatSchools.org.

How should school quality be measured? This chapter reviews econometric strategies for estimating school effectiveness, defined as the causal effect of attending a particular school or set of similar schools (like charter schools) on student outcomes. Efforts to gauge school quality must confront the fundamental challenge of selection bias: school-to-school comparisons can reflect student ability and family background as much as or more than school effectiveness. Economists have devised an array of solutions to this problem; increasingly, these empirical strategies use elements of randomness in modern school assignment schemes to devise convincing natural experiments.

Illustration: Angrist, Hull, and Walters (2023)

	Compliers			Always-takers	Never-takers
	Untreated	Treated	Pooled		
	(1)	(2)	(3)	(4)	(5)
Female	0.506	0.510	0.508	0.539	0.463
	(0.023)	(0.021)	(0.016)	(0.024)	(0.017)
Black	0.401	0.380	0.390	0.623	0.490
	(0.022)	(0.021)	(0.016)	(0.023)	(0.017)
Hispanic	0.250	0.300	0.275	0.183	0.228
	(0.02)	(0.018)	(0.013)	(0.019)	(0.014)
Asian	0.022	0.024	0.023	0.004	0.024
	(0.007)	(0.005)	(0.004)	(0.003)	(0.005)
White	0.229	0.216	0.223	0.154	0.215
	(0.018)	(0.016)	(0.012)	(0.016)	(0.014)
Special education	0.190	0.181	0.186	0.158	0.177
	(0.018)	(0.016)	(0.012)	(0.018)	(0.013)
English language learner	0.143	0.148	0.145	0.054	0.088
	(0.015)	(0.013)	(0.010)	(0.011)	(0.010)
Subsidized lunch	0.689	0.705	0.697	0.698	0.666
	(0.021)	(0.019)	(0.014)	(0.022)	(0.016)
Baseline math score	−0.274	−0.312	−0.293	−0.394	−0.301
	(0.047)	(0.041)	(0.032)	(0.045)	(0.036)
Baseline English score	−0.352	−0.349	−0.350	−0.362	−0.299
	(0.050)	(0.043)	(0.033)	(0.046)	(0.038)
Share of sample			0.546	0.197	0.257

Abadie's Kappa (2003)

Suppose assumptions of LATE theorem hold conditional on covariates X. Let $g(\cdot)$ be any measurable real function of Y, D, X with finite expectation. We can show that the expectation of g is a weighted sum of the expectation in the three groups

$$E[g|X] = \underbrace{E[g|X, D_1 > D_0]Pr(D_1 > D_0|X)}_{\text{Compliers}} +$$

$$+ \underbrace{E[g|X, D_1 = D_0 = 1]Pr(D_1 = D_0 = 1|X)}_{\text{Always Takers}}$$

$$+ \underbrace{E[g|X, D_1 = D_0 = 0]Pr(D_1 = D_0 = 0|X)}_{\text{Never Takers}}$$

Abadie's Kappa (2003)

Rearranging terms gives us then,

$$E[g(Y,D,X)|D_1 > D_0] = \frac{E[k \cdot g(Y,D,X)]}{Pr(D_1 > D_0)} = \frac{E[k \cdot g(Y,D,X)]}{E[k]}$$

where

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

- This result can be applied to *any characteristic or outcome and get its mean for compliers by removing the means for never and always takers*.
- Standard example: average covariate value among compliers: $E[X|D_1 > D_0] = \frac{E[kX]}{E[k]}$

Illustration: Mullainathan, Washington, and Azari (2009)

- Do debates truly provide citizens with information that influences their opinions or choices?
- During the 2005 election season, in the days leading up to the final debate between mayoral incumbent Republican Michael Bloomberg and Democratic challenger Fernando Ferrer, the authors interviewed a random sample of 1,000 New York City voters
- They randomly assigned these 1,000 individuals to one of two groups: they asked the treatment group to watch the November 1 debate, and they asked the control group to watch a "placebo" program, PBS's The NewsHour with Jim Lehrer, which aired opposite WNBC's debate broadcast.
- They find that those in the watch group were 6 percentage points more likely to report that their opinions of one or both candidates had changed from the first to the second interview.

Illustration: Mullainathan, Washington, and Azari (2009)

```
# watch is Z, actwatch is D, age is X
# using a standard logit/probit for Pr(Z=1|X).
mod <- glm(watch ~ age, family = binomial(link = "logit"),
           data = debate, na.action = na.exclude)
debate$watch_hat <- predict(mod, type = "response")
# generate kappa
debate <-debate %>%mutate(
k=1-((actwatch*(1-watch))/(1-watch_hat))-((1-actwatch)*watch/watch_hat))
# compute avg age of complier
avg_complier_age = weighted.mean(debate$age, debate$k,na.rm = T)
avg_complier_age
```

```
## [1] 66.1528
```

Profiling Compliers and Noncompliers (Marbach & Hangartner 2020)

- Subjects assigned to the control group who take the treatment are "observable" always-takers, and subjects assigned to the treatment group who do not take the treatment are "observable" never-takers;
- Observable and nonobservable always-takers and nevertakers, respectively, have the same covariate distribution (the instrument is independently assigned) $\rightarrow$ can directly estimate the covariate means for these two subpopulations;
- By subtracting the weighted covariate mean of observable always-takers and never-takers from the covariate mean of the entire sample, we can back out the covariate mean for compliers.

Illustration: Gerber, Karlan, and Bergan (2009)

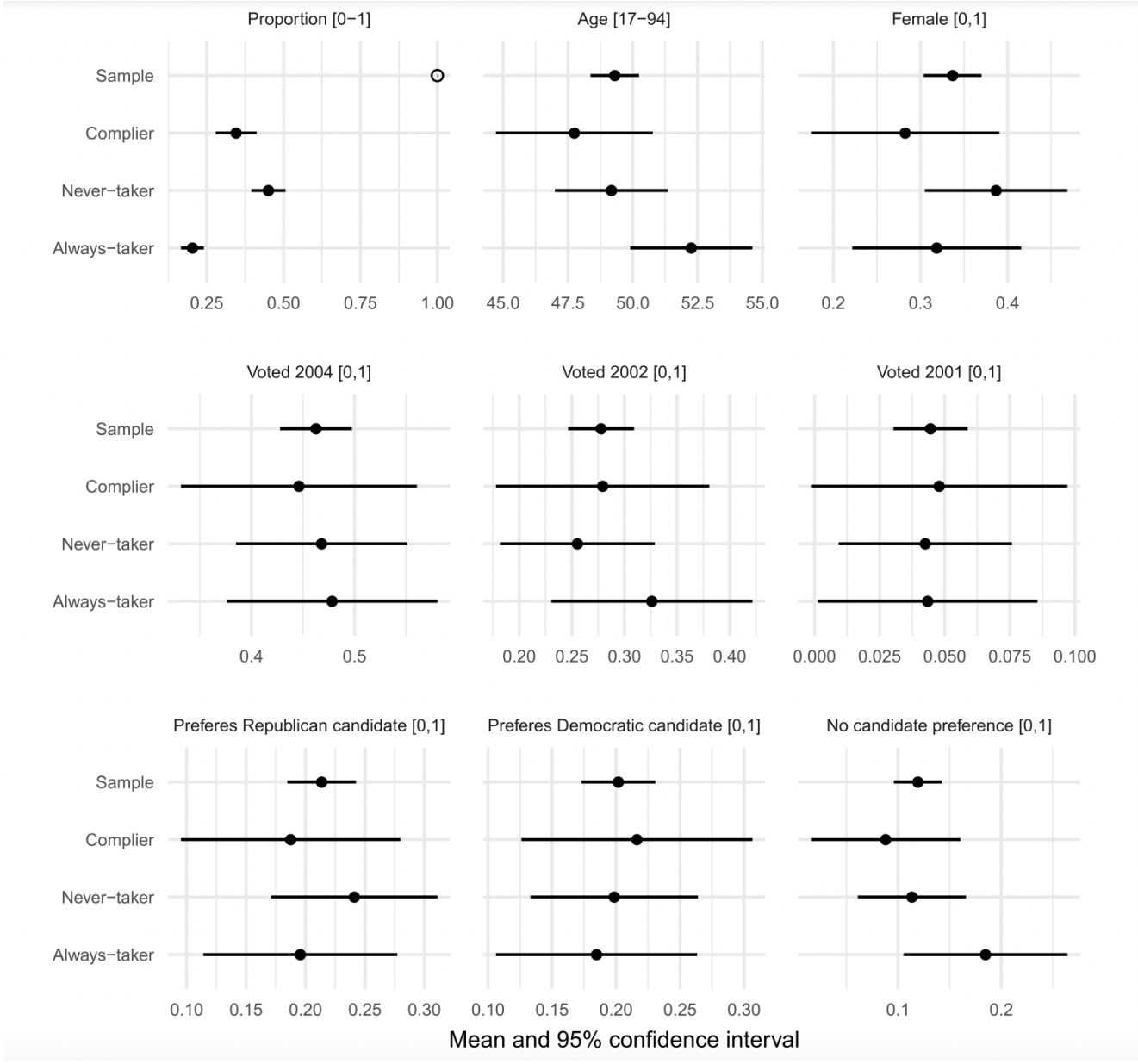
American Economic Journal: Applied Economics 2009, 1:2, 35–52
http://www.aeaweb.org/articles.php?doi=10.1257/app.1.2.35

Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions†

By Alan S. Gerber, Dean Karlan, and Daniel Bergan*

We conducted a field experiment to measure the effect of exposure to newspapers on political behavior and opinion. Before the 2005 Virginia gubernatorial election, we randomly assigned individuals to a Washington Post *free subscription treatment, a* Washington Times *free subscription treatment, or a control treatment. We find no effect of either paper on political knowledge, stated opinions, or turnout in post-election survey and voter data. However, receiving either paper led to more support for the Democratic candidate, suggesting that media slant mattered less in this case than media exposure. Some evidence from voting records also suggests that receiving either paper led to increased 2006 voter turnout.* (*JEL* D72, L82)

Illustration: Gerber, Karlan, and Bergan (2009)



Weak Instruments

The probability limit of the IV estimator is given by:

$$\text{plim } \hat{\alpha}_{IV} = \frac{\text{Cov}[Y, Z]}{\text{Cov}[Z, D]} + \frac{\text{Cov}[Z, u_2]}{\text{Cov}[Z, D]} = \alpha_D + \frac{\text{Cov}[Z, u_2]}{\text{Cov}[Z, D]}$$

The second term is non-zero if the instrument is not exogenous. Let $\sigma^2_{u_1}$ be the variance of the first stage error and F be the F statistic of the first-stage. Then, the bias in IV is

$$\text{E}[\hat{\alpha}_{IV} - \alpha] = \frac{\sigma_{u_1 u_2}}{\sigma^2_{u_2}} \left(\frac{1}{F + 1} \right)$$

If the first stage is weak, the bias approaches $\frac{\sigma_{u_1 u_2}}{\sigma^2_{u_2}}$. As F approaches infinity, B_{IV} approaches zero.

Weak Instrument: Simulation

```r
set.seed(123) # For reproducibility
n <- 1000 # Number of observations
beta_true <- 2 # True effect of the treatment on the outcome
simulate_2SLS <- function(n, beta_true, corr_strength) {
  Z <- rnorm(n)
  X <- rnorm(n, mean = corr_strength * Z)
  epsilon <- rnorm(n) #
  Y <- X * beta_true + epsilon
  first_stage <- lm(X ~ Z)
  second_stage <- ivreg(Y ~ X | Z)
  return(summary(second_stage)$coef[2, 1])
}
beta_hat_strong <- simulate_2SLS(n, beta_true, corr_strength = 0.9)
beta_hat_weak <- simulate_2SLS(n, beta_true, corr_strength = 0.1)
```

```
## Estimated treatment effect with strong instrument: 1.9807
```

```
## Estimated treatment effect with weak instrument: 2.439578
```

Test for weak instruments

- Use the F-statistic of the first-stage regression
- Stock & Yogo (2005): provide critical values for rejecting the null of weak instrument
- Results based on assumption of homoskedastic errors, hardly satisfied (Andrews, Stock & Sun 2019)
- Solutions: Robust F statistic. Equivalent to Kleibergen-Paap statistic for the case of a single regressor (Andrews, Stock & Sun 2019)
- Reported automatically by `ivreg2` in Stata
- Solutions: Effective first-stage F-statistic (Montiel Olea & Pflueger 2013)
- Equivalent to robust F in the just-identified case
- With one instrument, Effective F can be compared to Stock & Yogo critical values
- With multiple instruments, use the critical values in Montiel Olea & Pflueger

Recommendations (Lal et al. 2023): Design

- Prior to using an IV strategy, consider how selection bias may be affecting treatment effect estimates obtained through OLS. If the main concern is underestimating an already statistically significant treatment effect, an IV strategy may not be necessary.
- During the research design phase, consider whether the chosen instrument can realistically create random or quasi-random variations in treatment assignment while remaining excluded from the outcome equation.

Recommendations (Lal et al. 2023): Characterizing the first-stage

- Calculate and report F_{Eff} for the first stage, taking into account heteroscedasticity and clustering structure as needed. However, do not discard a design simply because $F_{Eff} < 0$.
- If d and z are continuous, plot d against its predicted values $\hat{d}$ (with covariates and fixed effects already partialled out from both) and visually verify whether their relationship aligns with theoretical expectations.

Recommendations (Lal et al. 2023): Hypothesis testing and inference

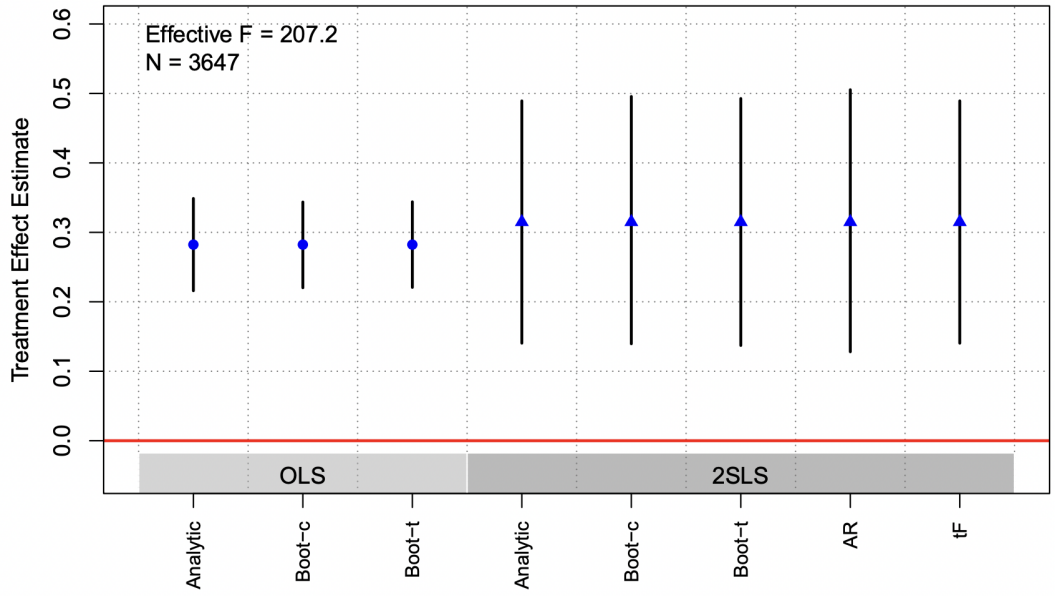
- **Option 1**. *t-test with F_{Eff} pretesting*. If $F_{Eff} < 10$, choose Options 2 or 3. Utilize conservative methods like *bootstrap-t* and *bootstrap-c* if outliers or group structures are present.
- **Option 2**. *tF procedure*. For single treatment and instrument cases, adjust t-test critical values based on FEff.
- **Option 3**. *Direct testing*. Apply weak-instrument-robust procedures, such as the AR test.

Recommendations (Lal et al. 2023): Communicating your findings

- Present OLS and IV estimates alongside CIs from various inferential methods in a graphical format, like in figure below These CIs may not concur on statistical significance, but they collectively convey the findings' robustness to different inferential approaches.
- Remember to report first-stage and reduced-form estimation results, including 95% CIs for coefficients, as they offer insight into both instrument strength and statistical power.

Recommendations (Lal et al. 2023): Communicating your findings (cont.)

Example of McClendon (2014):



Note: The treatment is reading an email with a promise of social esteem. The instrument is being encouraged to take the treatment. The outcome is attending LGBTQ events. The AR test does not rely on the first-stage F. Similar figures for each of the 70 IV designs are shown in the SM. This plot is made by ivDiag.

Recommendations (Lal et al. 2023): Additional diagnostics

- If you expect the OLS results to be upward biased, be concerned if the 2SLS estimator yields much larger estimates.
- If there is good reason to believe that treatment effects on compliers are significantly larger in magnitude than those on non-compliers, explain this through profiling of these principal strata (Abadie, 2003; Marbach and Hangartner, 2020).
- If it is possible to identify an observational analogue of "never takers" or a subset of them, conduct a placebo test by estimating the effect of the instrument on the outcome of interest in this ZFS (Zero-first-stage) sample. Using results from the ZFS test, obtain local-to-zero IV estimates and CIs and compare them to the original estimates and CIs.