

CAUSAL INFERENCE: INSTRUMENTAL VARIABLE ANALYSIS

Hye Young You (New York University)

RESEARCH

- There are many questions for which we do not have answers
- Performing research is a formal way to ask questions and search for answers

RESEARCH

- After we have asked a question, and (hopefully) found a convincing answer, we want to tell others what we have found
- Life is not easy. To convince others that the answer we have found is convincing, we can't just tell them our results; we need to describe the process that leads us to those results, so that they can judge whether or not we've come to a valid conclusion

OUR FOCUS

- Learn how to conduct empirical research
- How to do empirical research with observational data
- The questions that motivate most studies in the health, social and behavioral sciences are not associational but causal in nature

RANDOMIZED EXPERIMENT AND CAUSAL INFERENCE

- The gold standard of scientific research
 1. Randomly sample units from population
 2. Randomly assign treatment and control to the units
 3. Estimate the average treatment effect

SEARCHING FOR NATURAL EXPERIMENTS

- Randomized trials are expensive and have long duration
- We hope to find natural or quasi-experiments that mimic a randomized trial by changing the variable of interest while other factors are kept balanced
- The questions we are interested in may not even be feasible for randomized-trials (age or gender)
- How can we establish causal relationship with observational data?

CAUSAL INFERENCE WITH OBSERVATIONAL DATA

- Search for Identification: Identification refers generally to sufficiency for drawing a conclusion given the type of data that are available.
- Manski (1995, 4) : "Studies of identification seek to characterize the conclusions that could be drawn if one could...obtain an unlimited number of observations. Identification problems cannot be solved by gathering more of the same kind of data."
- Informally, a parameter is "identified" if, as the sample goes to infinity, the data come to require that the parameter equals one value.

CAUSAL IDENTIFICATION

- Causal identification refers to sufficiency for drawing a conclusion about a causal effect given the type of data at hand.

- Angrist and Krueger (1999):

The combination of a clearly labeled source of identifying variation in a causal variable and the use of a particular econometric technique to exploit this information is what we call an identification strategy.

MOST POPULAR ECONOMETRIC TECHNIQUES

1. Panel Data Analysis
2. Instrumental Variable Analysis
3. Regression Discontinuity Design

Instrumental Variable Analysis

WHY INSTRUMENTAL VARIABLE (IV)?

- Correlation can sometimes provide pretty good evidence of causal relationship
- Often, the regression we've got is not the regression we want
- IV methods solve the problem of missing or unknown control variables, much as a randomized trial obviates extensive controls in a regression
- The method of IV is a signature technique in the econometric toolkit

WHY INSTRUMENTAL VARIABLE (IV)?

- IV methods are typically used to address the following problems encountered in OLS regression:
 1. Omitted variable bias
 2. Measurement error
 3. Simultaneity or reverse causality

THE IDEA

- Example: earnings and college

$$\text{Earnings}_i = \beta_0 + \beta_1 * \text{Education}_i + \mu_i$$

- We suspect that μ_i could contain a number of omitted variables related to college attendance: motivation, parents' income, education, etc
- What if college were randomly assigned?
- Then $\text{Cov}(\text{Education}, \mu_i) = 0$

THE IDEA

- What if, instead, we randomly assign a subsidy ($z_i = 1$) to help pay for college?
- Some people, who would not have gone to college without the subsidy, might be induced to go to college
- Other people would have gone to college with or without the subsidy
- Still others would not go to college in either case

THE IDEA

- Suppose we estimate:

$$\text{Earnings}_i = \gamma_0 + \gamma_1 * \text{Subsidy}_i + \varepsilon_i$$

And find $\hat{\gamma}_1 > 0$

- What does this tell us?

THE IDEA

- It seems like people who get the subsidy eventually earn more
- Most likely explanation?

Subsidy \Rightarrow more likely to go to college

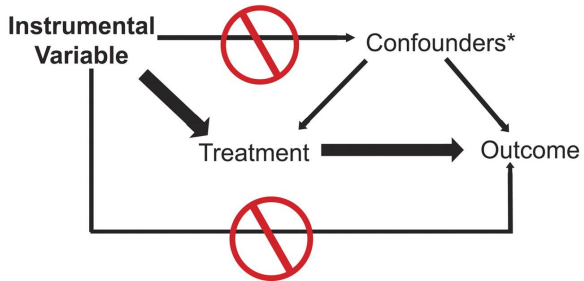
College \Rightarrow higher earnings

- This is the essence of IV

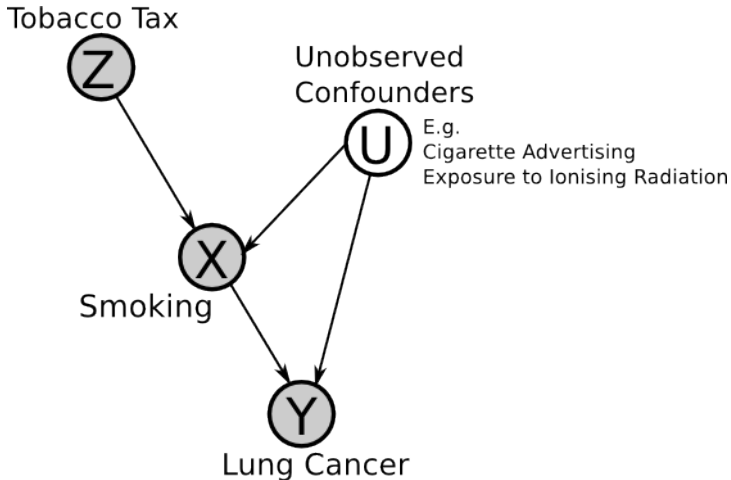
THE IDEA

- The key is that the subsidy causes changes in schooling that are exogenous
- In general, we are looking for some variable z_i which is
 1. Unrelated to omitted variables u_i (exogenous)
 2. Affects the regressor variables X_i (relevant)
- We call z_i an “instrumental variable”
- Using z_i , you create an exogenous variation in X_i and then exploit this exogenous variation to study the causal impact of X_i on your outcome variable

THE IDEA



THE IDEA



THE IDEA

- Intuitively, IV solves the OVB by using only part of the variability in the main interest variable - a part that is uncorrelated with the omitted variables - to estimate the relationship between X and Y
- IVs are sometimes derived from “natural experiment” - to exploit situations where the forces of nature or government policy have conspired to produce an environment somewhat akin to a randomized experiment

TWO DIFFERENT IVs

- First with constant effects
- Second with heterogeneous causal effects

Constant-Effects Models

IV GOES LONG

- Suppose the causal link between schooling and wages can be written $f_i(s) = \alpha + \rho s_i + \eta_i$
- Imagine a vector of control variables, A_i , called "ability"; write

$$\eta_i = A_i' \gamma + v_i$$

, where A_i are potentially observable control variables and γ are coefficients from the population OLS regression, in which case $\text{Cor}(A_i, v_i) = 0$ by definition

- For now, the variables A_i are assumed to be the only reason why η_i and s_i are correlated, so that $E[s_i v_i] = 0$

IV GOES LONG

- Given these assumptions, if we observe A_i , we can use OLS to estimate the slopes in the long regression,

$$Y_i = \alpha + \rho s_i + A_i' \gamma + v_i \quad (1)$$

which yields a consistent estimate of ρ , the causal effect of going from $S_i = s$ to $s + 1$.

- But what if we do not observe A_i ? Then we cannot use such a control strategy to estimate ρ .

IDENTIFICATION WITH AN INSTRUMENT

- IV allows us to estimate the long-regression coefficient ρ , when A_i is unobserved
- The instrument z_i is assumed to be: (1) correlated with the causal variable of interests, s_i ; and (2) uncorrelated with any other determinant of the outcome
- This is like saying $Cov(\eta_i, z_i) = 0$ or equivalently, z_i is uncorrelated with both A_i and v_i
- This is called an exclusion restriction since z_i can be said to be excluded from the causal model of interest

IDENTIFICATION WITH AN INSTRUMENT

- Suppose we have z_i such that $\text{Cov}(s_i, z_i) \neq 0$, but $\text{Cov}(\eta_i, z_i) = 0 \Rightarrow \text{Cov}(A_i, z_i) = \text{Cov}(v_i, z_i) = 0$
- Then z_i is an instrument for s_i and,

$$\text{Cov}(Y_i, z_i) = \text{Cov}(\alpha + \rho s_i + \eta_i, z_i) = \rho \text{Cov}(s_i, z_i)$$

IDENTIFICATION WITH AN INSTRUMENT

$$\rho = \frac{Cov(Y_i, z_i)}{Cov(s_i, z_i)} = \frac{\frac{Cov(Y_i, z_i)}{Var(z_i)}}{\frac{Cov(s_i, z_i)}{Var(z_i)}} = \frac{\text{Reduced Form}}{\text{First Stage}} \quad (2)$$

- The coefficient of interest, ρ , is the ratio of the population regression of Y_i on z_i (reduced form) to the population regression of s_i on z_i (first stage)
- The IV estimate is the sample analog of (2)

IDENTIFICATION WITH AN INSTRUMENT

$$\rho = \frac{Cov(Y_i, z_i)}{Cov(s_i, z_i)} = \frac{\frac{Cov(Y_i, z_i)}{Var(z_i)}}{\frac{Cov(s_i, z_i)}{Var(z_i)}} = \frac{\text{Reduced Form}}{\text{First Stage}} \quad (3)$$

- Note how $Cov(s_i, z_i) \approx 0$ would imply volatility ("weak instrument" or "weak first stage")
- The IV estimator is predicted on the notion that the first stage is not zero
- This is something you can check in the data
- "Strong first stage" is important - first stage F -stat is larger than 10 on (excluded) instruments (Stock & Watson, 2007)

IDENTIFICATION WITH AN INSTRUMENT

- If the instrument is “imperfect” $\text{Cov}(\eta_i, z_i) \neq 0$
- Then $\text{Cov}(Y_i, z_i) = \rho \text{Cov}(s_i, z_i) + \text{Cov}(\eta_i, z_i)$
- If we construct $\frac{\text{Cov}(Y_i, z_i)}{\text{Cov}(s_i, z_i)}$, we won't recover ρ but rather

$$\frac{\text{Cov}(Y_i, z_i)}{\text{Cov}(s_i, z_i)} = \rho + \frac{\text{Cov}(\eta_i, z_i)}{\text{Cov}(s_i, z_i)} = \rho + \frac{\text{Cor}(\eta_i, z_i)\sigma_\eta}{\text{Cor}(s_i, z_i)\sigma_s} \quad (4)$$

which is biased for ρ , where bias is large if $\text{Cor}(s_i, z_i)$ is small.

TWO-STATE LEAST SQUARES (2SLS)

- In practice, we do IV by doing 2SLS. This allows us to add exogenous covariates and to use multiple instruments. Returning to the schooling example, a causal model with covariates is

$$Y_i = \alpha X_i' + \rho s_i + \eta_i \quad (5)$$

where η_i is the compound error term, $A_i' \gamma + v_i$. The first stage and reduced form are

$$\text{FS:} \quad s_i = X_i' \pi_{10} + \pi_{11} z_i + \zeta_{1i} \quad (6)$$

$$\text{RF:} \quad Y_i = X_i' \pi_{20} + \pi_{21} z_i + \zeta_{2i} \quad (7)$$

- Reduced form is obtained by substituting (6) into (5):

$$\begin{aligned} Y_i &= \alpha X_i' + \rho [X_i' \pi_{10} + \pi_{11} z_i] + \rho \zeta_{1i} + \eta_i \\ &= X_i' [\alpha + \rho \pi_{10}] + \rho \pi_{11} z_i + [\rho \zeta_{1i} + \eta_i] \\ &= X_i' \pi_{20} + \pi_{21} z_i + \zeta_{2i} \end{aligned} \quad (8)$$

2SLS NOTES

- It's still all about the ratio of RF to FS:

$$\frac{\pi_{21}}{\pi_{11}} = \rho$$

- Where does two-stage least squares come from? Write the first stage as the sum of fitted values plus first-stage residuals:

$$s_i = X_i' \pi_{10} + \pi_{11} z_i + \zeta_{1i} = \hat{s}_i + \zeta_{1i}$$

2SLS estimates of (5) can be constructed by substituting first-stage fitted values for s_i in (5):

$$Y_i = \alpha X_i + \rho \hat{s}_i + [\eta_i + \rho \zeta_{1i}] \quad (9)$$

and using OLS to estimate the “second stage” (9)

- In practice, let Stata (`ivregress 2sls`) or R (`ivreg` in the AER package) do it

RECAP: 2SLS LINGO

- *endogenous variables* are the dependent variables and the independent variable(s) to be instrumented
- To treat an independent variable as endogenous is to instrument it, i.e., to replace with fitted values in the 2SLS second stage
- *exogenous variables* include *covariates* (by definition these are not instrumented) and the instruments themselves
- In any IV study, variables are either: dependent or endogenous variables, instruments, or covariates. No one is both.

WHERE IS A GOOD IV?!

- Good instruments come from a combination of institutional knowledge and ideas about the process determining the variable of interests
- For example, the economic model of education suggests that schooling decisions are based on the costs and benefits of alternative choices. Thus, one possible source of instruments for schooling is differences in costs due to loan policies or other subsidies that vary independently of ability or earnings potential

WHERE IS A GOOD IV?!

- A second source of variation in schooling is institutional constraints. A set of institutional constraints relevant for schooling is compulsory schooling laws
- Angrist and Krueger (1991, "Does Compulsory Schooling Attendance Affect Schooling and Earnings?") exploit the variation induced by compulsory schooling to examine the effect of schooling on incomes
- quarter-of-birth strategy

ANGRIST AND KRUEGER (1991)

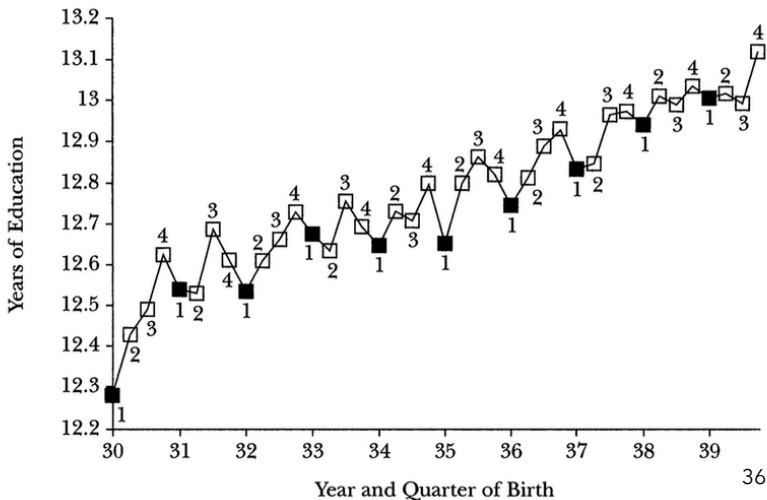
- Most states require students to enter school in the calendar year in which they turn 6. School start age is a function of date of birth
- Those born late in the year are young for their grade. In states with a December 31 birthday cutoff, children born in the fourth quarter enter school shortly before they turn 6 ($5\frac{3}{4}$), while those born in the first quarter enter school at around age $6\frac{3}{4}$

ANGRIST AND KRUEGER (1991)

- Because compulsory schooling laws typically require students to remain school only until their 16th birthday, these groups of students will be in different grades (or different degree) when they reach the legal dropout age
- The combination of school start-age policies and compulsory schooling laws create a natural experiment in which children are compelled to attend school for different lengths of time, depending on their birthdays
- Sample: men born 1930-1959

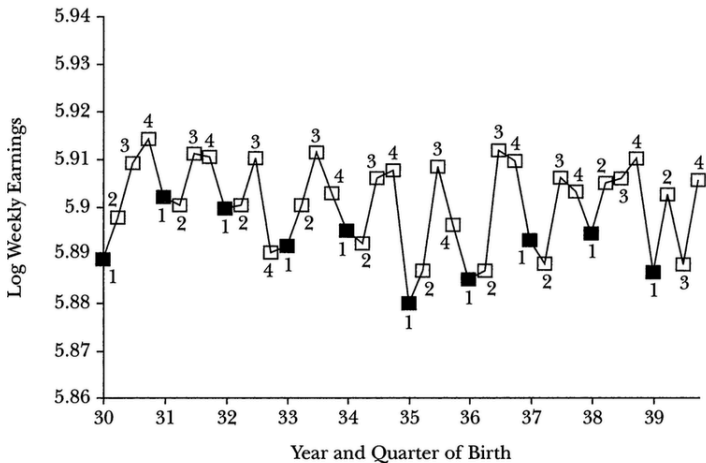
ANGRIST AND KRUEGER (1991)

Mean Years of Completed Education, by Quarter of Birth



ANGRIST AND KRUEGER (1991)

Mean Log Weekly Earnings, by Quarter of Birth



ANGRIST AND KRUEGER (1991)

- First stage: men born earlier in the calendar year tended to have lower average schooling levels
- Reduced form: men born in early quarters almost always earned less, on average, than those born later in the year
- Importantly, this reduced-form relationship parallels the quarter-of-birth pattern in schooling, suggesting that two patterns are closely related
- An individual's date of birth is probably unrelated to his or her innate ability, motivation, or family connections, it seems credible to assert that the only reason for the up-and-down quarter of birth pattern in earnings is the up-and-down quarter of both pattern in schooling

ANGRIST AND KRUEGER (1991)

- They estimated the following TSLS model:

$$(1) \quad E_i = X_i\pi + \sum_c Y_{ic}\delta_c + \sum_c \sum_j Y_{ic}Q_{ij}\theta_{jc} + \varepsilon_i$$

$$(2) \quad (\ln)W_i = X_i\beta + \sum_c Y_{ic}\xi_c + \rho E_i + \mu_i$$

, where E_i is the education of i th individual, X_i is a vector of covariates, Q_{ij} is a dummy variable indicating whether the individual was born in quarter j ($j = 1, 2, 3$), and Y_{ic} is a dummy variable indicating whether the individual was born in year c ($c = 1, \dots, 10$), and W_i is the weakly wage.

- The coefficient ρ is the return to education

ANGRIST AND KRUEGER (1991)

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS*

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)	0.0711 (0.0003)	0.0760 (0.0290)	0.0632 (0.0003)	0.0806 (0.0164)	0.0632 (0.0003)	0.0600 (0.0299)
Race (1 = black)	—	—	—	—	−0.2575 (0.0040)	−0.2302 (0.0261)	−0.2575 (0.0040)	−0.2626 (0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763 (0.0029)	0.1581 (0.0174)	0.1763 (0.0029)	0.1797 (0.0305)
Married (1 = married)	—	—	—	—	0.2479 (0.0032)	0.2440 (0.0049)	0.2479 (0.0032)	0.2486 (0.0073)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	−0.0772 (0.0621)	−0.0801 (0.0645)	—	—	−0.0760 (0.0604)	−0.0741 (0.0626)
Age-squared	—	—	0.0008 (0.0007)	0.0008 (0.0007)	—	—	0.0008 (0.0007)	0.0007 (0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

ANGRIST AND KRUEGER (1991)

- As it turns out, the TSLS estimate of the change in earning due to additional education differs little from a simple OLS regression
- This suggests that there is little bias from omitted ability variable in the OLS
- In other applications, the IV estimates and OLS estimates are quite different (e.g., Angrist and Lavy (1999) on the class size effect on student achievement)

VALIDITY OF IV

- The key assumptions are:
 1. Exclusion, which embeds both exogeneity and “no-alternative-pathway” assumptions, and
 2. Valid first stage
- Any IV application needs to make the case for both of these!
- First stage can be checked directly. Exclusion cannot

VALIDITY OF IV

- MHE (p. 131) provide some useful indirect tests, known as “placebo tests” for exclusion:
 1. Look for associations between z_i and things that predict Y_i but should not be sensitive to z_i (e.g., things that were determined prior to z_i)
 2. Find subpopulations for which there should be no relationship between z_i and s_i . If z_i predicts Y_i in this subsample, then you may have an exclusion violation

Causal Effects in a Heterogeneous World

CONSTANT-EFFECTS BENCHMARK

- The traditional IV framework is the linear, constant-effect world discussed in Part I
- The causal effect was the same for all individuals (homogeneous treatment effects): $Y_{1i} - Y_{0i} = \rho$ for all i
- Time now to allow for the fact that $Y_{1i} - Y_{0i}$ is not likely to be the same for everyone

IV WITH HETEROGENOUS TREATMENT EFFECTS

- Variables used in this setup: $Y_i(d, z)$ = potential outcome of individual i ; D_i = treatment dummy, Z_i = instrument dummy
- Causal chain is: $Z_i \rightarrow D_i \rightarrow Y_i$
- Notation for D_i :
 D_{1i} = i 's treatment status when $Z_i = 1$
 D_{0i} = i 's treatment status when $Z_i = 0$
- Observed treatment status is therefore

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_1 Z_i + \zeta_i$$

- The causal effect of Z_i on D_i is $D_{1i} - D_{0i}$

IV WITH HETEROGENOUS TREATMENT EFFECTS

- Gerber and Green (2000), "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment"
- Reports the results of a randomized field experiment involving approximately 30,000 registered voters in New Haven, CT
- Through a random assignment, the sample was divided into control and treatment groups
- The treatment group for personal canvassing contained 5,794 people, the control group 23,586
- Canvassers were able to contact only 1,615 (28%) of the 5,794 people in the personal canvassing treatment group

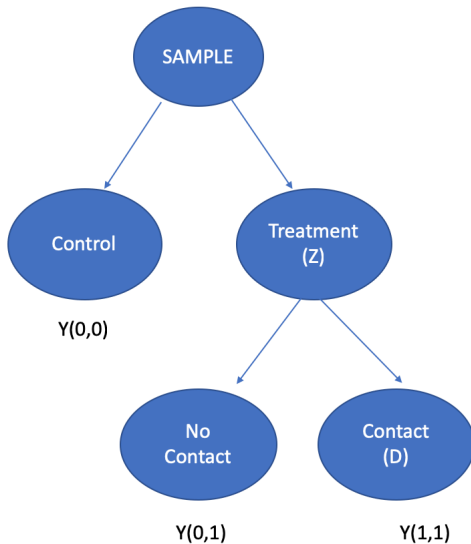
IV WITH HETEROGENOUS TREATMENT EFFECTS

- Do we have to worry that this study was a failure since the take-up among those assigned to treatment is only 28%?
- No. Most of the randomized experiments or natural experiments involve this kind of situation
- IV solves this problem!

IV WITH HETEROGENOUS TREATMENT EFFECTS

- Gerber and Green (2000):
- $Y_i(d, z) = \text{voting}$, $D_i = \text{contact}$, $Z_i = \text{treatment}$
- $D_{1i} = 0 \text{ or } 1$ given $Z_i = 1$, $D_{0i} = 0$ given $Z_i = 0$
- Potential Outcomes: $Y(0, 0)$, $Y(1, 1)$, $Y(0, 1)$
- Treatment (D_i) based comparison: $Y(1, 1) - \text{mean}[Y(0, 1), Y(0, 0)]$
- Intent to treated (Z_i) based comparison: $\text{mean}[Y(1, 1), Y(0, 1)] - Y(0, 0)$

GERBER AND GREEN (2000)



KEY ASSUMPTION 1: INDEPENDENCE

- The IV is independent of the vector of potential outcomes and potential treatment assignment (i.e. as good as randomly assigned):

$$[Y_i(d, z), D_{1i}, D_{0i}] \perp\!\!\!\perp Z_i$$

- Independence means assignment into treatment and control group is independent of turn out (potential outcomes) and probability of being contacted (potential treatment) in Gerber and Green (2000)
- Independence means the first-stage is the average causal effect of Z_i on D_i :

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

KEY ASSUMPTION 1: INDEPENDENCE

- Independence is sufficient for a causal interpretation of the reduced form. Specifically,

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

- This is sometimes called the “intention to treatment effect” (ITT)
- This is the causal effect of the instrument on the dependent variable, but we have yet to link this to treatment

KEY ASSUMPTION 2: EXCLUSION

- Exclusion: The instrument affects Y_i only through D_i :

$$Y_i(1, 1) = Y_i(1, 0) = Y_{1i}$$

$$Y_i(0, 1) = Y_i(0, 0) = Y_{0i}$$

- The exclusion restriction means Y_i can be written solely by treatment status:

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

for Y_{1i} and Y_{0i} that satisfy the independence assumption

KEY ASSUMPTION 2: EXCLUSION

- In the Vietnam draft lottery example (Angrist, 1998): an individual's earnings potential as a veteran or non-veteran are assumed to be unchanged by draft eligibility status
- The exclusion restriction would be violated if low lottery numbers may have affected schooling (e.g. to avoid the draft). If this was the case the lottery number would be correlated with earnings for at least two cases:
 1. through its effect on military service
 2. through its effect on educational attainment
- The fact that lottery number is randomly assigned (and therefore satisfies the independence assumption) does not ensure that the exclusion restriction is satisfied

KEY ASSUMPTION 2: EXCLUSION

- Exclusion means draft lottery numbers affect earnings only via veteran status; quarter of birth affects earnings only through schooling; sex composition of children affects labor supply only via family size
- This assumption is not testable. Requires faith

KEY ASSUMPTION 3: MONOTONICITY

- A necessary technical assumption:
- Monotonicity. $D_{1i} \geq D_{0i}$ for everyone (or vice versa)
- Remember, $D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \zeta_i$
- Monotonicity means either $\pi_{1i} \geq 0$ for all i or $\pi_{1i} \leq 0$ for all i

KEY ASSUMPTION 3: MONOTONICITY

- While the instrument may have no effect on some people, all those who are affected are affected in the same way
- In the quarter of birth example for schooling the assumption may not be satisfied (Barua and Lang, 2009): Being born in the 4th quarter (which typically increases schooling) may have reduced schooling for some because their school enrollment was held back by their parents
- Without monotonicity, IV estimators are not guaranteed to estimate a weighted average of the underlying causal effects of the affected group, $Y_{1i} - Y_{0i}$

HETEROGENOUS TREATMENT EFFECTS ASSUMPTION RECAP

1. The independence assumption is sufficient for identification of the causal effects of the instrument
2. The exclusion restriction means that the causal effect of the instrument on the dependent variable is due solely to the effect of the instrument on D_i (treatment). Exclusion is more controversial than independence
3. We also assume there is a first-stage; by virtue of monotonicity, this is the proportion of the population for which D_i is changed by Z_i

THE LATE THEOREM

- If all assumptions are satisfied, IV estimates LATE (Local Average Treatment Effect)
- LATE is the average effect of X on Y for those whose treatment status has been changed by the instrument Z:

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

Proof: See MHE 4.4.1

- In the draft lottery example: IV estimates the average effect of military service on earnings for the subpopulation who enrolled in military service because of the draft but would not have served otherwise

THE LATE THEOREM

- The LATE framework partitions any population with an instrument into potentially 4 groups:
 1. Compliers: the subpopulation with $D_{1i} = 1$ and $D_{0i} = 0$. Their treatment status is affected by the instrument in the right direction
 2. Always-takers: the subpopulation with $D_{1i} = D_{0i} = 1$. They always take the treatment independently of Z
 3. Never-takers: the subpopulation with $D_{1i} = D_{0i} = 0$. They never take the treatment independently of Z
 4. Defiers: The subpopulation with $D_{1i} = 0$ and $D_{0i} = 1$. Their treatments status is affected by the instrument in the "wrong" direction

THE LATE THEOREM

- Monotonicity ensures that there are no defiers
- Why is it important to not have defiers?
 1. If there were defiers, effects on compliers could be (partly) cancelled out by opposite effects on defiers
 2. One could then observe a reduced form which is close to 0 even though treatment effects are positive

THE LATE THEOREM

- IV is uninformative for always-takers and never-takers because treatment status for these types is unchanged by the instrument
- With all assumptions satisfied IV estimates the average treatment effect for compliers
- Without further assumptions (e.g., constant causal effects) IV does not give us the average treatment effect on the whole population (ATE)
- Of course, we can assume effects are the same for all three groups (a version of the constant-effect-model)

THE LATE THEOREM

- Another effect which we may potentially be interested in estimating is the average treatment effect on the treated (TOT)
- Treatment status is: $D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$
- We see that

$$\underbrace{\{D_i = 1\}}_{\text{treated}} = \underbrace{\{D_{0i} = D_{1i} = 1\}}_{\text{always-takers}} \cup \underbrace{\{\{D_{1i} - D_{0i} = 1\} \cap \{Z_i = 1\}\}}_{\text{compliers assigned } Z_i = 1}$$

- TOT is therefore a weighted average of effects on always-takers and compliers

IV IN RANDOMIZED TRIALS

- The compliance problem in RCTs: Some randomly assigned to the treatment group are untreated
- When compliance is voluntary, an *as-treated* analysis is contaminated by selection bias
- Intention-to-treat (ITT) analyses preserve independence but are diluted by non-compliance
- Treatment-compliers are systematically different from treatment-non compliers (Gerber and Green, 2000)
- If you just compare means between treated and untreated individuals (using OLS) you will obtain biased treatment effect, even you do a randomized experiment!

IV IN RANDOMIZED TRIALS

- IV using randomly assigned treatment intended as an instrumental variable for treatment received solves this sort of compliance problem
- There are no always-takers (no controls treated), so $TOT = LATE$

THE BLOOM RESULT

- Z_i is a dummy variable indicating random assignment to the treatment group; D_i is a dummy indicating whether treatment received or taken
- In practice, because of noncompliance, D_i is not equal to Z_i

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i]} = \frac{\text{ITT (RF)}}{\text{compliance rate}} = \underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{TOT}$$

Direct proof (Bomm, 1984; MHE 4.4.3)

IV IN RANDOMIZED TRIALS - EXAMPLE 1

- Chetty, Hendren, and Katz (2016), “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment”
- The Moving to Opportunity (MTO) experiment offered randomly selected families housing vouchers to move from high-poverty housing projects to lower-poverty neighborhoods
- Estimate the MTO-experiment’s long-term impacts on children who were young when their families moved to better neighborhoods

IV IN RANDOMIZED TRIALS - EXAMPLE 1

- 48 percent of those in the experimental voucher group took up the voucher to move to a low-poverty area
- Estimating “intent-to-treat” (ITT) effects of the MTO treatments, which are essentially differences between treatment and control group means:

$$y_i = \alpha + \beta^{ITT} Exp_i + \gamma X_i + \delta s_i + \varepsilon_i$$

- The estimate of β^{ITT} identifies the causal impact of being *offered* a voucher to move through MTO
- Since not all the families offered vouchers actually took them up, these ITT estimates understate the causal effect of actually moving to a different neighborhood

IV IN RANDOMIZED TRIALS - EXAMPLE 1

- Estimating the impacts of moving through MTO - the impact of “treatment on the treated” (TOT) - by instrumenting for MTO voucher take-up with the treatment assignment indicators:

$$y_i = \alpha_T + \beta^{TOT} TakeExp_i + \gamma_T X_i + \delta_T s_i + \varepsilon_i^T$$

Since *TakeExp* is an endogenous variable, we instrument for it using the randomly-assigned MTO treatment group indicator (*Exp*)

- Under the assumption that MTO voucher offers only affect outcomes through the actual use of the voucher to lease a new residence, β_E^T can be interpreted as the causal effect of taking up the experimental and moving to a lower-poverty neighborhood (LATE)

MOVING TO OPPORTUNITY

TABLE 3—IMPACTS OF MTO ON CHILDREN'S INCOME IN ADULTHOOD

	W-2 earn- ings (\$) 2008–2012	Individual earnings 2008–2012 (\$)			Individual earnings (\$)		Employed (%) 2008– 2012	Hhold. inc. (\$) 2008–2012	Inc. growth (\$) 2008–2012
	ITT (1)	ITT (2)	ITT w/ controls (3)	TOT (4)	Age 26 ITT (5)	2012 ITT (6)	2008– 2012 ITT (7)	ITT (8)	ITT (9)
<i>Panel A. Children < age 13 at random assignment</i>									
Exp. versus control	1,339.8** (671.3)	1,624.0** (662.4)	1,298.9** (636.9)	3,476.8** (1,418.2)	1,751.4* (917.4)	1,443.8** (665.8)	1.824 (2.083)	2,231.1*** (771.3)	1,309.4** (518.5)
Sec. 8 versus control	687.4 (698.7)	1,109.3 (676.1)	908.6 (655.8)	1,723.2 (1051.5)	551.5 (888.1)	1,157.7* (690.1)	1.352 (2.294)	1,452.4** (735.5)	800.2 (517.0)
Observations	8,420	8,420	8,420	8,420	1,625	2,922	8,420	8,420	8,420
Control group mean	9,548.6	11,270.3	11,270.3	11,270.3	11,398.3	11,302.9	61.8	12,702.4	4,002.2

CONSTANT V.S. HETEROGENOUS EFFECT

Compared to last class (constant effect model), what's different and what's the same?

- We dropped the constant effects assumption, which is usually unrealistic
- We added a monotonicity assumption
- We clarify that effects are identified only for a particular subpopulation - the "complier" subpopulation
- With effect heterogeneity, different instruments do not identify the same causal effect

SOME IV "GREATEST HITS"

- Draft lottery numbers → military service → income (Angrist 1990)
- Election year → number of police → crime (Levitt 1997)
- Sibling sex composition → number of children → labor supply (Angrist & Evans 1998)
- Settler mortality → investment in institutions → average income (Acemoglu et al. 2001)
- Rainfall → economic growth → civil war (Miguel et al. 2004)
- Density of railroads → segregation → inequality (Ananat 2011)

IV EXAMPLE 1



Cities as Lobbyists

Rebecca Goldstein Harvard University
Hye Young You Vanderbilt University

Abstract: *Individual cities are active interest groups in lobbying the federal government, and yet the dynamics of this intergovernmental lobbying are poorly understood. We argue that preference incongruence between a city and its parent state government leads to underprovision of public goods, and cities need to appeal to the federal government for additional resources. We provide evidence for this theory using a data set of over 13,800 lobbying disclosures filed by cities with populations over 25,000 between 1999 and 2012. Income inequality and ethnic fragmentation are also highly related to federal lobbying activities. Using an instrumental variables analysis of earmark and Recovery Act grant data, we show that each dollar a city spends on lobbying generates substantial returns.*

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <http://dx.doi.org/10.7910/DVN/RSD5BV>.

CITIES AS LOBBYISTS

- State and local governments are active interest groups that lobby the Congress and executive branches in Washington D.C.
- Study the effect of municipal government lobbying on earmarks and grants from the federal government
- The difficulties in identifying the causal effect of lobbying on outcomes?

DATA: CITY LOBBYING ACTIVITIES

- Sample: 1,262 U.S. cities with population greater than 25,000
- Period: 1999 to 2012 (106th through 112th Congresses)
- Cities submitted 13,858 lobbying reports and spent over 367 million dollars
- Collect data on earmarks (2008, 2009) and grants from the Recovery Act (2009-2010) given to cities
- Challenge: Lobbying decision (spending) is not random

DIRECT FLIGHT AS AN INSTRUMENT

- Using a direct flight from a city i to DC as an instrumental variable of city lobbying
- City lobbying is usually done when city officials travel to Washington, DC, to meet their lobbyists, and lobbyists arrange meetings for city officials with House representatives or senators from the state where the city is located
- The existence of a direct flight captures the convenience of travel to Washington, DC, from each city

DIRECT FLIGHT AS AN INSTRUMENT

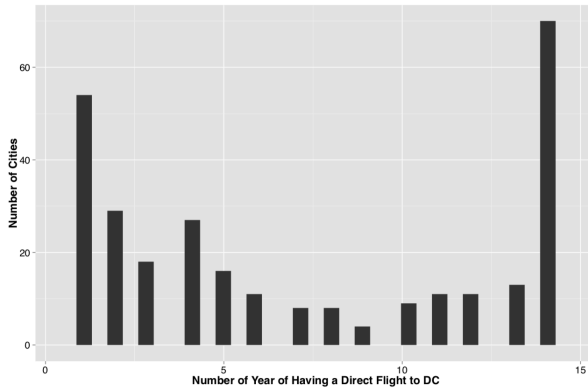
- Airline companies choose new destinations mainly based on market dynamics: long- term growth, market competition, and profitability
- The Airline Deregulation Act of 1978 has allowed freedom for air carriers to set their own fares and routes
- After conditioning on all factors (e.g., existence of an airport, population, and income level) in the first-stage regression, we argue that a direct flight to Washington, DC, is a valid instrument for lobbying decisions and lobbying spending

DATA ON DIRECT FLIGHT

- Collect data on a direct flight from city i to Washington, DC (all three airports in the area: Baltimore (BWI), Dulles (IAD), and Reagan (DCA)) for each year for all cities between 1999 to 2012
- Cities in Virginia and Maryland are tricky cases in terms of binary indicators of a direct flight, so we run regressions both including them and excluding them, and the results are similar

DISTRIBUTION OF YEARS HAVING A DIRECT FLIGHT TO WASHINGTON, DC

Figure 1: Distribution of Number of A Direct Flight to Washington DC, 1999-2012



FIRST-STAGE

Variable	(ln) Lobbying Spending _t	(ln) Lobbying Spending _t
A Direct Flight _{t-1}	1.59*** (4.35)	0.67*** (3.62)
Controls	Y	N
City FE	N	Y
Observations	3,827	3,827
adj. R^2	0.197	0.692

Variable	Lobbying Participation	(ln) Lobbying Spending
Average Flight Fare (\$)	-0.01*** (-3.07)	-0.02*** (-2.68)
Controls	Y	Y
Observations	1,802	1,802

ROBUSTNESS OF INSTRUMENT

- We also present evidence that the decision to have direct flight at time t is mainly an independent decision by airline companies and the decision is not associated with city i 's previous years' lobbying spending
- The Department of Transportation provides the Essential Air Services (EAS) program that guarantees that small communities are served by air carriers since the Airline Deregulation Act (ADA) passed in 1978
- Among those cities, 41 cities are included in our data and 21 cities among them had a direct flight to Washington, D.C. for at least one year during the period of study. We ran the same analysis excluding those 21 cities and the result is consistent with the previous analyses

DOES LOBBYING WORK?

Variable	(ln) Earmark		(ln) Recovery Grant	
	(1) Tobit	(2) IV	(3) OLS	(4) IV
<i>Panel A.</i>				
(ln) Lobbying Spending	0.51*** (8.45)	1.02*** (3.98)	0.06*** (2.94)	0.47*** (2.68)
<i>Panel B. First-stage</i>				
<i>DV=(ln) Lobbying Spending</i>				
Direct Flight to D.C.		3.13*** (4.85)		3.06*** (4.74)
F-statistic		23.62		22.51
Controls:	Y	Y	Y	Y
State Fixed Effect	Y	Y	Y	Y
Observations	1,262	1,262	1,262	1,262

REPLICATION FILES

are available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/RSD5BV>

US Food Aid and Civil Conflict[†]

By NATHAN NUNN AND NANCY QIAN*

We study the effect of US food aid on conflict in recipient countries. Our analysis exploits time variation in food aid shipments due to changes in US wheat production and cross-sectional variation in a country's tendency to receive any US food aid. According to our estimates, an increase in US food aid increases the incidence and duration of civil conflicts, but has no robust effect on interstate conflicts or the onset of civil conflicts. We also provide suggestive evidence that the effects are most pronounced in countries with a recent history of civil conflict. (JEL D74, F35, O17, O19, Q11, Q18)

US FOOD AID AND CIVIL CONFLICT

- Study the effect of US food aid on conflict in recipient countries
- Aid workers, human rights observers, and journalists have accused humanitarian aid of being not only ineffective, but of actually promoting conflict
- The difficulties in identifying the causal effect of food aid on conflict?

IV EXAMPLE 2

Empirical strategy:

- Exploit plausibly exogenous time variation in US wheat production, which is primarily driven by changes in US weather condition; US wheat production is positively correlated with US food aid shipments in the following year

IV EXAMPLE 2

$$C_{irt} = \beta F_{irt} + X_{irt}\Gamma + \delta_r Y_t + \psi_{ir} + \mu_{irt} : \text{Second Stage}$$

$$F_{irt} = \alpha P_{t-1} + X_{irt}\Gamma + \delta_r Y_t + \psi_{ir} + \varepsilon_{irt} : \text{First Stage}$$

, where C = conflict, F = quantity of wheat aid shipped from the US, P = the amount of US wheat production in the previous year, serves as an instrument; i denotes countries, r denotes six geographical regions, and t denotes years; sample = 125 non-OECD countries between 1971 - 2006

IV EXAMPLE 2

- Conceptually, the identification strategy compares conflict in developing countries in years after US wheat production is high to the years after it is low
- Changes in us production have larger effects on the aid received by regular air recipients, we can strengthen the fit of the first stage by allowing for this form of heterogeneity ("creating more variation")
- Interaction between lagged US wheat production with the propensity to receive food from US solves the problem

IV EXAMPLE 2

$$F_{irt} = \alpha(P_{t-1} \times \bar{D}_{ir}) + X_{irt}\Gamma + \psi_{rt} + \varphi_{ir} + \varepsilon_{irt} : \text{First Stage}$$

$$C_{irt} = \beta F_{irt} + X_{irt}\Gamma + \psi_{rt} + \varphi_{ir} + \mu_{irt} : \text{Second Stage}$$

, where D = how many years a country i received UF food aid between 1971 and 2006 (propensity to receive US food aid)

COMPLIERS

- Compliers are observations that receive more US food aid following increases in US wheat production
- Instrumental variable estimates are not driven by the effect of US food aid for the countries whose food aid receipts are unaffected by changes in US wheat production over time

THREATS TO VALIDITY: EXCLUSION RESTRICTIONS

- The underlying driver of the variation in US wheat production, US weather conditions, may be correlated with weather conditions in aid-recipient countries → control for weather conditions in recipient countries
- US wheat production may affect foreign conflict through its influence on the world price of wheat → US price stabilization policy, region-year fixed effect to control for all region-specific changes over time, controlling a country's average per capita net imports of cereals and the average per capita production of cereals over the sample period
- Regular recipients of US food aid could differ from irregular recipients → controlling for US military aid and other economic aids

CORRELATION BETWEEN WHEAT PRODUCTION AND FOOD AID

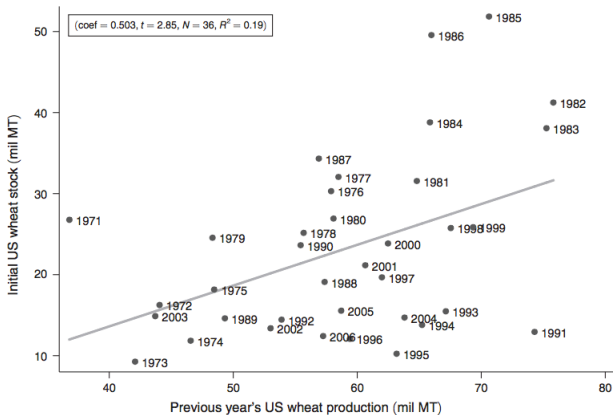
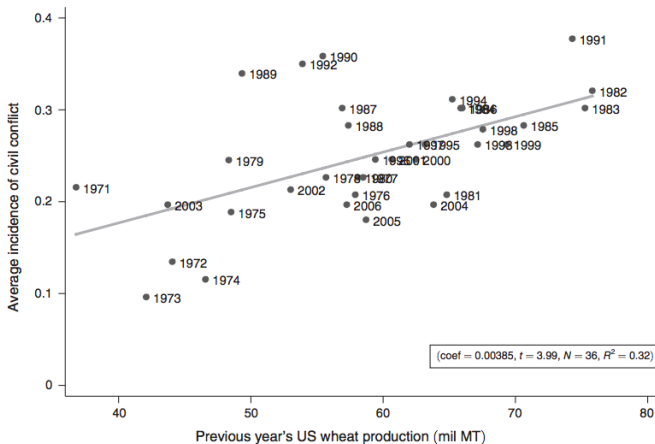


FIGURE 1. US WHEAT RESERVES AND LAGGED US WHEAT PRODUCTION

CORRELATION BETWEEN WHEAT PRODUCTION AND CONFLICT



FIRST STAGE RESULT

Dependent variable (panel D):	US wheat aid (1,000 MT)						
<i>Panel D. First-stage estimates</i>							
Lag US wheat production (1,000 MT) × avg. prob. of any US food aid	0.00227 (0.00094)	0.00343 (0.00126)	0.00343 (0.00120)	0.00330 (0.00092)	0.00358 (0.00103)	0.00358 (0.00103)	0.00358 (0.00103)
Kleibergen-Paap <i>F</i> -statistic	5.84	7.37	8.24	12.76	12.10	12.10	12.10
Controls (for all panels):							
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
US real per capita GDP × avg. prob. of any US food aid	No	Yes	Yes	Yes	Yes	Yes	Yes
US democratic president × avg. prob. of any US food aid	No	Yes	Yes	Yes	Yes	Yes	Yes
Oil price × avg. prob. of any US food aid	No	Yes	Yes	Yes	Yes	Yes	Yes
Monthly recipient temperature and precipitation	No	No	Yes	Yes	Yes	Yes	Yes
Monthly weather × avg. prob. of any US food aid	No	No	Yes	Yes	Yes	Yes	Yes
Avg. US military aid × year FE	No	No	No	Yes	Yes	Yes	Yes
Avg. US economic aid (net of food aid) × year FE	No	No	No	Yes	Yes	Yes	Yes
Avg. recipient cereal imports × year FE	No	No	No	No	Yes	Yes	Yes
Avg. recipient cereal production × year FE	No	No	No	No	Yes	Yes	Yes
Observations (for all panels)	4,089	4,089	4,089	4,089	4,089	4,089	4,089

SECOND STAGE RESULT

Dependent variable (panels A, B, and C):	Parsimonious specifications				Baseline specification		
	Any conflict (1)	Any conflict (2)	Any conflict (3)	Any conflict (4)	Any conflict (5)	Intrastate (6)	Interstate (7)
<i>Panel A. OLS estimates</i>							
US wheat aid (1000 MT)	-0.00000 (0.00019)	0.00000 (0.00019)	0.00000 (0.00019)	0.00000 (0.00019)	-0.00000 (0.00020)	0.00006 (0.00019)	-0.00004 (0.00003)
R^2	0.477	0.477	0.481	0.483	0.485	0.460	0.245
<i>Panel B. Reduced form estimates ($\times 1,000$)**</i>							
Lag US wheat production (1,000 MT) \times avg. prob. of any US food aid	0.00224 (0.00078)	0.00254 (0.00087)	0.00254 (0.00086)	0.00251 (0.00086)	0.00255 (0.00086)	0.00183 (0.00081)	0.00087 (0.00042)
R^2	0.479	0.480	0.483	0.485	0.488	0.461	0.246
<i>Panel C. 2SLS estimates</i>							
US wheat aid (1,000 MT)	0.00507 (0.00386)	0.00405 (0.00227)	0.00366 (0.00205)	0.00354 (0.00200)	0.00366 (0.00209)	0.00263 (0.00160)	0.00124 (0.00093)
Anderson-Rubin confidence interval	[0.00186, 0.01778]	[0.00156, 0.01151]	[0.00155, 0.01037]	[0.00148, 0.01013]	[0.00165, 0.01053]	[0.00088, 0.00788]	[0.00288, 0.00429]

REPLICATION FILES

are available at `https:`

`//www.aeaweb.org/articles?id=10.1257/aer.104.6.1630`

SUMMARY

- The IV paradigm provides a powerful and flexible framework for causal inference:
 1. An alternative to random assignment with a strong claim on internal validity
 2. A solution to the compliance problem in randomized trials
 3. A flexible strategy for the analysis of observational designs; Fuzzy RD is IV (a point to which we'll return)

PRACTICAL TIPS FOR IV PAPERS

1. Report the first stage
 - Does it make sense?
 - Do the coefficients have the right magnitude and sign?
2. If you have many IVs pick your best instrument and report the just identified model (weak instrument problem is much less problematic)
3. Look at the Reduced Form
 - The reduced form is estimated with OLS and is therefore unbiased
 - If you can't see the causal relationship of interest in the reduced form, it is probably not there

Thank You!

MULTI-INSTRUMENT 2SLS

- With more than one instrument, 2SLS is a weighted average of the one-at-time (just-identified) estimates
- Let

$$\rho_j = \frac{Cov(Y_i, z_{ji})}{Cov(D_i, z_{ji})}; j = 1, 2$$

denote two IV estimands using z_{1i} and z_{2i} to instrument D_i

- The 2SLS estimand is:

$$\rho_{2SLS} = \psi \rho_1 + (1 - \psi) \rho_2$$

where ψ is a number between zero and one that depends on the relative strength of the instruments in the first stage

MULTI-INSTRUMENT 2SLS

- Angrist and Evans (1998) study the effect of the third child on women's labor force participation
- Use twins and sex-mix instruments; Using a twins instrument alone, the IV estimate of the effect of a third child on female labor force participation is -0.084 ($se = 0.017$). The corresponding same sex estimate is -0.138 ($se = 0.029$)
- Using both instruments produces a 2SLS estimate of -0.098 ($se = 0.015$)
- The 2SLS weight in this case is 0.74 for twins, 0.26 for same sex, due to the stronger twins first stage

BIAS AND CONSISTENCY

- Unbiasedness means the estimator has a sampling distribution centered on the parameter of interest in a sample of any size
- Consistency means that the estimator converges to the population parameter as the sample size grows
- Since IV estimates are consistent, but not unbiased, researchers using IV should aspire to work with large samples

THE BIAS OF 2SLS

- Cross-section OLS estimates are typically unbiased for the population BLP, as well as consistent, though BLP may not be the regression you want
- 2SLS estimates are consistent for causal effects but biased towards OLS estimates
- Endogenous variable is vector x ; dependent variable is vector y ; assume no covariates:

$$y = \beta x + \eta \quad (10)$$

The $N \times Q$ matrix of instruments is Z , with first-stage:

$$x = Z\pi + \zeta \quad (11)$$

Outcome error η_i is correlated with ζ_i . Instruments are uncorrelated with ζ_i by construction and with η_i by assumption

THE BIAS OF 2SLS

- The 2SLS estimator is

$$\hat{\beta}_{2SLS} = (x'P_zx)^{-1}x'P_zy = \beta + (x'P_zx)^{-1}x'P_z\eta$$

where $P_z = Z(Z'Z)^{-1}Z'$ is the projection matrix that produces fitted values, \hat{x}

- Substituting x in $x'P_z\eta$, we get:

$$\hat{\beta}_{2SLS} - \beta = (x'P_zx)^{-1}(\pi'Z' + \zeta')P_z\eta \quad (12)$$

$$= (x'P_zx)^{-1}\pi'Z'P_z\eta + (x'P_zx)^{-1}\zeta'P_z\eta \quad (13)$$

THE BIAS OF 2SLS

- Expectation of the ratios on the right hand side of (11) are closely approximated by the ratio of expectations:

$$E[\hat{\beta}_{2SLS} - \beta] = (E[(x'P_zx)])^{-1}E[\pi'Z'P_z\eta] + (E[(x'P_zx)])^{-1}E[\zeta'P_z\eta]$$

This Bekker (1994) approximation gives a good account of finite-sample behavior

- Using the fact that $E[\pi'Z'\zeta] = 0$, $E[\pi'Z'\eta] = 0$, and substituting $x = Z\pi + \delta$, we have

$$E[\hat{\beta}_{2SLS} - \beta] \approx [E(\pi'Z'Z\pi) + E(\eta'P_z\zeta)]^{-1}E(\zeta'P_z\eta) \quad (14)$$

- 2SLS is biased because $E(\zeta'P_z\eta) \neq 0$ unless η_i and ζ_i are uncorrelated

THE BIAS OF 2SLS: FIRST-STAGE F

- Manipulating of (12) generates:

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\zeta}}{\sigma_{\zeta}^2} \left[\frac{E(\pi' Z' Z \pi)/Q}{\sigma_{\zeta}^2} + 1 \right]^{-1}$$

$\frac{E(\pi' Z' Z \pi)/Q}{\sigma_{\zeta}^2}$ is the “population F” for joint significance of instruments in first-stage, so we can write

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\zeta}}{\sigma_{\zeta}^2} \frac{1}{F + 1} \quad (15)$$

THE BIAS OF 2SLS: FIRST-STAGE F

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\zeta}}{\sigma_{\zeta}^2} \frac{1}{F + 1}$$

- As F gets small, the bias of 2SLS approaches $\frac{\sigma_{\eta\zeta}}{\sigma_{\zeta}^2}$. The bias of the OLS estimator is $\frac{\sigma_{\eta\zeta}}{\sigma_x^2}$, which also equals $\frac{\sigma_{\eta\zeta}}{\sigma_{\zeta}^2}$ if $\pi = 0$ (remember $x = \pi z + \zeta$)
- 2SLS estimates are therefore said to be “biased toward OLS estimates” when there isn’t much of a first stage
- On the other hand, the bias of 2SLS vanishes when F gets large, as it should happen in large sample when $\pi \neq 0$

THE BIAS OF 2SLS: FIRST-STAGE F

- When the instruments are weak, the F-statistics varies inversely with the number of instruments. To see why, consider adding useless instruments to your 2SLS model, that is, instruments with no effect on the first-stage R-squared. The model sum of squares, $E(\pi'Z'Z\pi)$, and the residual variance, σ_ζ^2 will both stay the same while Q (= number of IV) goes up. The F-statistic becomes smaller as a result. From this we learn that the addition of weak instruments increases bias
- Holding the first-stage sum of squares fixed, bias is least in the just-identified case when the number of instruments is as low as it can get

WEAK INSTRUMENT PROBLEM

- Strong first stage is important for unbiased result
- Stock, Wright, and Yogo (2002), Stock & Watson (2007)
- “F-stat larger than 10” as a criteria

WHAT CAN YOU DO IF YOU HAVE WEAK INSTRUMENTS?

- Use a just identified model with your strongest IV. If the instrument is very weak, however, your standard errors will probably be very large
- Use a limited information maximum likelihood estimator (LIML). It provides the same asymptotic distribution as 2SLS (under constant effects) but provides a finite-sample bias reduction (`ivregress liml`)
- Find stronger instruments

TESTING OVERIDENTIFYING RESTRICTIONS

- IV should be correlated with the endogenous regressors, and must also be uncorrelated with the structural error term
- If we have more than one instrument, we can effectively test whether some of them are uncorrelated with the structural error
- Consider the example:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Z_1 and Z_2 are instruments for X_i
- We want $Cov(\varepsilon, Z_1) = 0, Cov(\varepsilon, Z_2) = 0$

TESTING OVERIDENTIFYING RESTRICTIONS

- Sargan-Hansen Test or Sargan's J-test
- (1) Run and get $Y_i = \hat{\alpha} + \hat{\beta}X_i + \hat{\varepsilon}_i$
- (2) Estimate $\hat{\varepsilon}_i = \delta_0 + \delta_1Z_1 + \delta_2Z_2$; get R^2
- (3) $NR^2 \sim \tilde{\chi}_{m-k}^2$, where m = number of IV, k = number of endogenous variable
- Null-hypothesis ($\delta_1 = \delta_2 = 0$), the value of NR^2 will be small and so test will not reject the null-hypothesis

TESTING OVERIDENTIFYING RESTRICTIONS

- That indicates that your IVs jointly matter. But it does not tell whether individually they matter
- `estat overid` in Stata

TESTING OVERIDENTIFYING RESTRICTIONS: AN EXAMPLE

```
. ivregress 2sls rent pcturban (hsngval = faminc i.region), vce(robust)
(output omitted)

. estat overid

Test of overidentifying restrictions:

Score chi2(3)          =    6.8364   (p = 0.0773)
```

IV EXAMPLE 3

- Thomas Eisensee and David Stromberg, "News Droughts, News Floods, and U.S. Disaster Relief," *Quarterly Journal of Economics* 2007
- Study the influence of mass media on U.S. government response to 5,000 natural disasters occurring between 1968 and 2002
- It is challenging to to show that news coverage influences government policy. Why?

IV EXAMPLE 3

- Use the availability of other newsworthy material as an instrument for whether the disaster was in the news
- Natural disaster is less likely to receive relief because news about this disaster was crowded out by, for example, the Olympics, or O.J. Simpson trial
- Empirical specification:

$$\text{relif}_i = \alpha_1 * \text{news}_i + \gamma' \theta_i + \varepsilon_i : \text{RF}$$

$$\text{news}_i = \beta_1 * \text{news pressure} + \beta_2 * \text{Olympics} + \delta' \theta_i + \mu_i : \text{FS}$$

, where i is disaster

FIRST STAGE AND REDUCED FORM RESULTS

EFFECT OF THE PRESSURE FOR NEWS TIME ON DISASTER NEWS AND RELIEF

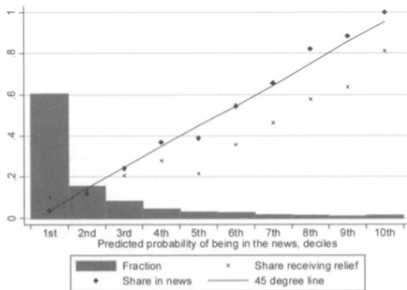
	Dependent variable: <i>News</i>				Dependent variable: <i>Relief</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>News Pressure</i>	-0.0162 (0.0041)***	-0.0163 (0.0041)***	-0.0177 (0.0057)***	-0.0142 (0.0037)***	-0.0117 (0.0045)***	-0.0119 (0.0045)***	-0.0094 (0.0058)	-0.0078 (0.0040)**
<i>Olympics</i>	-0.1078 (0.0470)**	-0.1079 (0.0470)**	-0.0871 (-0.0628)	-0.111 (0.0413)***	-0.1231 (0.0521)**	-0.1232 (0.0521)**	-0.1071 (0.0763)	-0.1098 (0.0479)**
<i>World Series</i>	-0.1133 (-0.1065)				-0.1324 (0.1031)			
<i>log Killed</i>			0.0605 (0.0040)***				0.0582 (0.0044)***	
<i>log Affected</i>			0.0123 (0.0024)***			0.0376		
<i>Imputed log Killed</i>				0.0491 (0.0034)***				0.0442 (0.0037)***
<i>Imputed log Affected</i>				0.0151 (0.0020)***				0.0394 (0.0020)***
Observations	5,212	5,212	2,926	5,212	5,212	5,212	2,926	5,212
R-squared	0.1799	0.1797	0.3624	0.2875	0.1991	0.1989	0.4115	0.3726

HETEROGENOUS EFFECT

- There are strong reasons to believe that the effect of news is heterogenous across disasters. For some disasters, news coverage has little effect (2004 Indian Ocean earthquake).
- The effect of news on relief is greater for disasters that are marginal in the news decision.
- The reason is that these disasters are also more likely to be marginal in the relief decision, in the sense of receiving relief if and only if they receive news coverage

HETEROGENOUS EFFECT

- Calculate the predicted probability of a disaster being in the news, based on the number of killed and affected, disaster type, year, month, and country. See the relationship between predicted pr. and actual media coverage (and disaster relief receiving)



HETEROGENOUS EFFECT

- Figure shows that news and relief are driven in a similar way by observables
- Thus, some disasters are highly likely both to be in the news and to receive relief
- Others are very unlikely both to be in the news and receive relief
- Those close to 50 percent in the news decision are also close to 50 percent in the relief decision

HETEROGENOUS EFFECT

- If there are heterogenous effects of news on relief, then consistent OLS measures the average effect of news on relief across all disasters, while 2SLS estimates the average effect in the subgroup of disasters that are marginally newsworthy in the sense of being in the news if and only if there is little other news around (LATE!)
- Since the effect of news on relief is presumably greater for disasters that are marginal in the news decision, the 2SLS estimates are likely to be larger than those of consistent OLS

HETEROGENOUS EFFECT

- If the effect of news on relief is greater for disasters that are marginal in the news decision, the correlation between news and relief would be higher for disasters with close to a 50 percent probability of being in the news
- Include the interaction between news and the absolute distance of the predicted probability of the disaster being in the news from 0.5 in the regression of relief on news
- See Column (4)
- IV estimates the effects in this subgroup (compliers)

EFFECTS OF NEWS ON RELIEF

	OLS					IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News	0.2886 (0.0200)***	0.158 (0.0232)***	0.1309 (0.0178)***	0.2323 (0.0328)***	0.2611 (0.0569)***	0.8237 (0.2528)***	0.6341 (0.3341)*	0.6769 (0.2554)***
News*abs(Pr(news)-0.5)				-0.4922 (0.1059)***	-0.302 (0.0840)***			
abs(Pr(news)-0.5)				0.5374 (0.0943)***	0.2959 (0.0831)***			
log Killed		0.0486 (0.0046)***					0.0198 (0.0208)	
log Affected		0.0358 (0.0024)***					0.0299 (0.0048)***	
Imputed log Killed			0.0378 (0.0038)***	0.0546 (0.0049)***	0.0307 (0.0046)***			0.0109 (0.0132)
Imputed log Affected			0.0375 (0.0020)***	0.0445 (0.0023)***	0.0345 (0.0026)***			0.0292 (0.0045)***
F-stat, instruments, 1st stage						11.0	6.1	11.1
Over-id restrictions, $\chi^2_{df}(p\text{-value})$						0.51 ₁ (0.47)		0.64 ₁ (0.42)
Observations	5,212	2,926	5,212	5,212	5,027	5,212	2,926	5,212
R-squared	0.2443	0.4225	0.3800	0.3860				

THREAT TO VALIDITY: EXCLUSION RESTRICTIONS

- Dates of the Olympic Games are exogenous with respect to disaster relief and news coverage, controlling for month and year effects
- News like O.J. Simpson scandal or 9/11 could distract policy makers from thinking about natural disasters; Some Olympics are political → Remove all observations when news pressure was in the highest one-third that year & Remove Olympics (1968, 1972, 1980, 1984) that created political issues and rerun the analysis

IDENTIFICATION

- In IV regression, whether the coefficients are identified depends on the relation between the number of instruments (m) and the number of endogenous regressors (k)
- Intuitively, if there are fewer instruments than endogenous regressors, we can't estimate β
- The coefficients $\beta_1 \dots \beta_k$ are said to be **Just (or Exactly) identified** if $m = k$; **Overidentified** if $m > k$

SPECIFICATION TESTS

- Testing for Endogeneity: Durbin-Wu-Hausman Test
- Since OLS is preferred to IV if we do not have an endogeneity problem, we'd like to be able to test for endogeneity
- If we do not have endogeneity, both OLS and IV are consistent, but IV is inefficient
- Idea of Hausman test is to see if the estimates from OLS and IV are different
- `estat endogenous` in Stata

SPECIFICATION TESTS: AN EXAMPLE

```
. use http://www.stata-press.com/data/r13/hsng
(1980 Census housing data)

. ivregress 2sls rent pcturban (hsngval = faminc i.region), vce(robust)
(output omitted)

. estat endogenous

Tests of endogeneity
Ho: variables are exogenous

Robust score chi2(1)          = 2.10428   (p = 0.1469)
Robust regression F(1,46)     = 4.31101   (p = 0.0435)
```


IV IN RANDOMIZED TRIALS - EXAMPLE 2

- The Job Training Program Act (JTPA) included a large randomized trial to evaluate the effect of training on earnings
- The JTPA offered treatment randomly; participation was voluntary
- Roughly 60 percent of those offered training received it
- 2 percent of people in the control group also received training
- David Autor (2003) evaluates differences in earnings in the 30 month period after random assignment

IV IN RANDOMIZED TRIALS - EXAMPLE 2

- IV set up:
 1. D_i indicates those who received JTPA services
 2. Z_i indicates the random offer of treatment
 3. Y_i is earnings in the 30 months since random assignment
- The first stage here is approximately the compliance rate:

$$\underbrace{E[D_i|Z_i = 1]}_{0.62} - \underbrace{E[D_i|Z_i = 0]}_{0.02} = P[D_i = 1|Z_i = 1] = 0.60$$

IV IN RANDOMIZED TRIALS - EXAMPLE 2

Table 4.4.1: Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970 (555)	3,754 (536)	1,117 (569)	970 (546)	1,825 (928)	1,593 (895)
B. Women	2,133 (345)	2,215 (334)	1,243 (359)	1,139 (341)	1,942 (560)	1,780 (532)

- Columns (1) and (2) show OLS estimates
- Columns (3) and (4) show ITT (reduced form) estimates
- Columns (5) and (6) show IV estimates

IV IN RANDOMIZED TRIALS - EXAMPLE 2

- Columns (1) and (2) show OLS estimates: These estimates are misleading because they compare individuals according to D_i , the actual treatment received. Since individuals assigned to the treatment group were free to decline (40% did so), this comparison throws away the random assignment unless the decision to comply was itself independent of potential outcomes
- Columns (3) and (4) show ITT (reduced form) estimates: Since Z_i was randomly assigned, ITT effects have a causal interpretation - they tell us the causal effect of the *offer of treatment*, but given that many of those offered have declined to participate, the ITT effect is too small relative to the average causal effect on those who were in fact treated
- Columns (5) and (6) show IV estimates: Put the pieces together and give us the effect of treatment on the treated