HANDOUT 12 – INSTRUMENTAL VARIABLES

AGENDA

- Do political protests matter? Evidence from the Tea Party Movement
- Instrumental variables (IV): Introduction
- Instrumental Variables in three steps
  - First stage regression
  - Reduced form regression
  - Second stage regression
- Other IV Examples
- Takeaways
- Vocabulary

BIBLIOGRAPHY FOR TODAY´S CLASS

- Madestam et al. (2013). Do political protests matter? Evidence from the Tea Party Movement. (*) (Focus on the abstract and introduction).
- Angrist and Pischke (2015), 3 (**)
- Stock and Watson (2007), 12.1, 12.3 (**)

DO POLITICAL PROTESTS MATTER? EVIDENCE FROM THE TEA PARTY MOVEMENT

Do Political Protests Matter? Evidence from the Tea Party Movement*

Andreas Madestam, Daniel Shoag, Stan Veuger, David Yanagizawa-Drott

The Quarterly Journal of Economics, Volume 128, Issue 4, 1 November 2013, Pages 1633–1685,
https://doi.org/10.1093/qje/qjt021
Published: 30 September 2013

Cite Permissions Share ▾

Abstract

Can protests cause political change, or are they merely symptoms of underlying shifts in policy preferences? We address this question by studying the Tea Party movement in the United States, which rose to prominence through coordinated rallies across the country on Tax Day, April 15, 2009. We exploit variation in rainfall on the day of these rallies as an exogenous source of variation in attendance. We show that good weather at this initial, coordinating event had significant consequences for the subsequent local strength of the movement, increased public support for Tea Party positions, and led to more Republican votes in the 2010 midterm elections. Policy making was also affected, as incumbents responded to large protests in their district by voting more conservatively in Congress. Our estimates suggest significant multiplier effects: an additional protester increased the number of Republican votes by a factor well above 1. Together our results show that protests can build political movements that ultimately affect policy making and that they do so by influencing political views rather than solely through the revelation of existing political preferences. *JEL* Code: D72.

Question: Can political protest affect elections and public policy?

Explored in a paper by former API-202 instructor Dan Shoag and former HKS professor David Yanagizawa-Drott.

On April 15 2009, Tea Party movement held nation-wide protests

- Rallies were held in 581 counties, with an average size of 166 protesters (standard deviation of 1,004)
- Studied if Tea Party protests impacted votes in the 2010 midterm elections and policy-making in the U.S. Congress

Specifically, authors asked whether size of protests affected how many votes Republican Party candidates got in the elections

The key regression of interest was:

$$rep\_votes = \beta_0 + \beta_1 protesters + u$$

Each observation is a U.S. county

- rep_votes is the number of Republican U.S. House votes in the country in 2010 (in '000s).
- $protesters$ is the number of Tea Party protesters (in '000s) in the county on the rally day.
-

Table 1.	Rep. Votes, 2010 '000
Number of Protesters, '000	27.5 (4.4)
Constant	10.2 (1.2)
Observations	2962

Robust standard errors in parentheses.

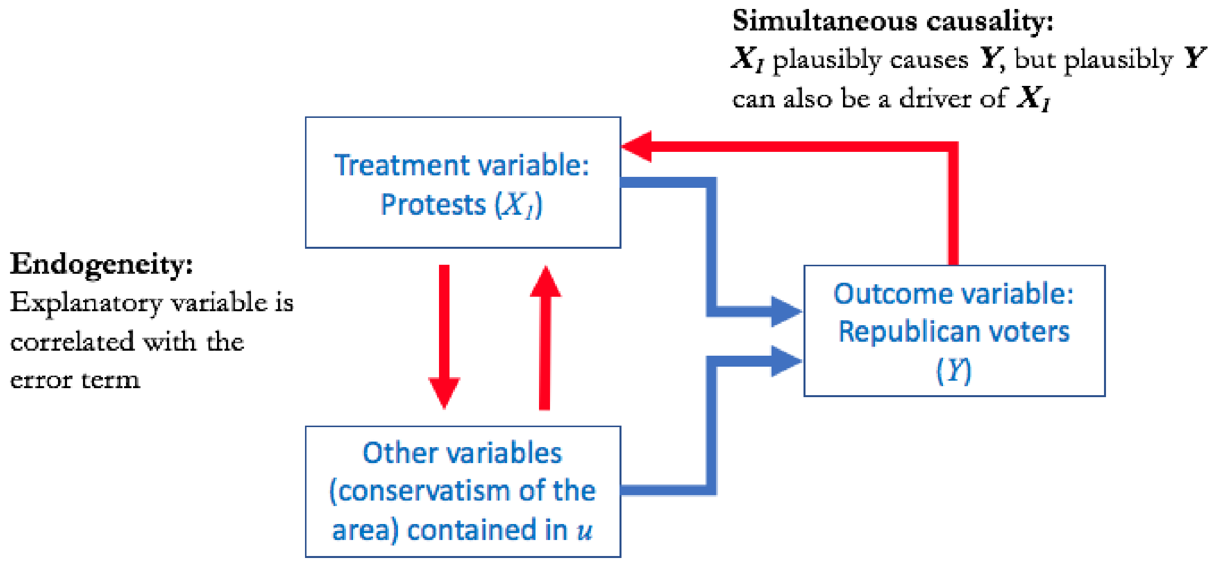
Interpret the $protesters$ coefficient:

Is the OLS estimate likely to be internally valid?

In what direction are the OLS estimates likely biased?

INSTRUMENTAL VARIABLES (IV) INTRODUCTION:

WHAT IS THE PROBLEM WE ARE TRYING TO SOLVE?



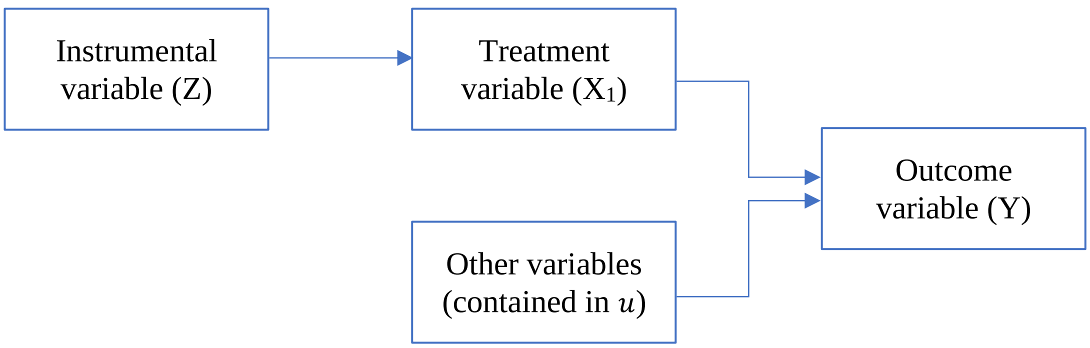
Instrumental variables (IV) estimation uses quasi-random variation in treatment assignment to mimic an experiment.

Ideally, we would like to randomly assign Tea Party protests to certain places with more intensity than others, and measure the difference in outcomes (Republican share of vote in those places, midterms of 2010):

- What if we find a natural way to randomly assign massive protests to some places, and smaller protests to other places?

It's a **quasi-experimental** technique (or a "natural experiment")

The challenge lies in finding an instrumental variable (Z) that is correlated with X_1 (protesters) but uncorrelated with u (everything else potentially causing OVB).



Instrumental variables (**IV**) regression is aimed at breaking X_1 into two parts: one that might be correlated with u and another that is not

The part of X_1 that is uncorrelated with u gives an unbiased β_1 estimate

Authors use rainfall on the day of rally as instrumental variable

- The variable *rainy_rally* is a dummy variable equal to one if it rained in the county on the rally day (zero otherwise)

The authors claim that the instrument is valid because rainfall is a random event, as long as the probability of rain in the county is controlled for

What is it important to control for the probability of rain?

The instrument Z must must satisfy two conditions:

- **Instrument relevance**: corr $(Z, X_1) \neq 0$. The instrument is correlated with X_1, the endogenous (non-random) regressor
- **Instrument exogeneity (exclusion restriction)** corr $(Z, u) = 0$. If the instrument is correlated with Y, it is only through X_1 and not through any other channels.

Do you think this instrumental variable would be valid?

- Relevance:

- Exogeneity (Exclusion restriction)

INSTRUMENTAL VARIABLES IN THREE STEPS:

Step 1. First stage regression: test the instrument for **relevance**.

The endogenous regressor X (*protesters*) is regressed on instrument Z (*prob_rain* is the probability of rain):

$$protesters = \beta_0 + \beta_1\ rainy\_rally + \beta_2\ prob\_rain + \varepsilon$$

Table 2.	**First Stage**
Dependent Variables	Number of Protesters, '000
	(1)
Number of Protesters, '000	
Rainy Protest	-0.082
	(0.021)
Rain probability control	Yes
Observations	2962

Standard errors in parentheses.

Interpret the first stage coefficient on *rainy_rally*. Is the instrument **relevant**?

Step 2. Reduced form regression: the outcome Y (*rep_votes*) is regressed on instrument Z (*rainy_rally*):

$$rep\_votes = \beta_0 + \beta_1\, rainy\_rally + \beta_2\, prob\_rain + \varepsilon$$

Table 2.	**Reduced Form**
Dependent Variables	Rep. Votes, 2010 ‘000
	(2)
Number of Protesters, '000	
Rainy Protest	-1.145
	(0.503)
Rain probability control	Yes
Observations	2962

Standard errors in parentheses.

Does the instrument affect the outcome of interest? Does rain affect voting patterns?

Step 3. IV estimate or second stage regression: Run this regression (using Stata's *ivreg2 rep_votes (protesters=rainy rally)*):

$$rep\_votes = \beta_0 + \beta_1 \ \widehat{protesters} \ + \beta_2 \, prob\_rain + \varepsilon$$

Table 2.	**First Stage**	**Reduced Form**	**IV/2SLS**
Dependent Variables	Number of Protesters, '000	Rep. Votes, 2010 '000	Rep. Votes, 2010 '000
	(1)	(2)	(3)
Number of Protesters, '000			13.90
			(6.73)
Rainy Protest	-0.082	-1.145	
	(0.021)	(0.503)	
Rain probability control	Yes	Yes	Yes
Observations	2962	2962	2962

Standard errors in parentheses.

Interpret the IV (2SLS) coefficient on protesters:

Do you notice a relationship between the three coefficients?

Is that coefficient causal? Is the instrument exogenous? Is there another channel through which rain on April 15, 2009 could have affected election results in November 2010?

OTHER IV EXAMPLES

(always look for the instrumental variable Z, the "treatment" X, and the outcome Y)

- Settlers mortality rates in the colonies (Z) has been used as instrument for institutions (X), to study if institutions (X) have a positive impact on income per capita (Y) (Acemoglu-Robinson)
- Vietnam-era draft lottery numbers (Z) have been used as an instrument for military service (X) to ask whether military service (X) affects labor market earnings (Y) (Angrist)
- Cuban immigrants coming to Miami on Mariel lift-boat (Z) have been used as an instrument for the impact of immigration (X) on the equilibrium salaries of domestic workers (Y) (Borjas, Clemens, and Pritchett – not together!)
- Major discoveries of oil and gas (Z) have been used as an instrument for the share of natural resources in exports (X), in studying how higher natural resources impact the diversification of the non/resource export basket (Bahar, Santos)

TAKEAWAYS

- Instrumental variables are useful when we do not have a randomized experiment, but nature or public policy has caused two similar groups of individuals to be treated differently
- Finding a good instrument is hard!
  - Many fail the exogeneity test
  - An instrumental variables regression can adjust for imperfect compliance to the treatment in a randomized experiment
- Thinking about external validity:
  - Estimates generated from instrumental variables are based on the individuals whose behavior is affected by the instrument
  - People refer to instrumental variable estimates as local average treatment effects (or LATE)
  - This is because the estimates are average effects for a subset (i.e. local part) of the population

VOCABULARY

- Instrumental variable: a variable that is correlated with an endogenous regressor (instrument relevance) and is uncorrelated with the regression error (instrument exogeneity).
- Quasi experiments or natural experiment: a circumstance in which randomness is introduced by variations in individual circumstances that make it appear as if the treatment is randomly assigned.
- Instrument relevance: first condition a valid instrumental variable must satisfy. If an instrument is relevant, then variations in the instrument must be correlated to variations in X_i.
- Instrument exogeneity: second condition a valid instrumental variable must satisfy. If the instrument is exogenous, then the variation of X_i captured by the instrumental variable is exogenous. This exogenous variation can in turn be used to estimate the population coefficient β_1.
- First stage regression: the regression of the X variable we are instrumenting for, and the instrument (Z). It can include other controls, and is used to test for relevance of the instrument.
- Reduced form regression: Regression between the outcome variable (Y) and all the available exogenous variables, including the instrumental variable (Z) and control variables.
- Second stage regression: Regress Y_i on the predicted values of the endogenous variables (using Z as an instrumental variable) and the included exogenous variables using OLS.

BONUS SLIDES – ESTIMATION WITH IV

We want to measure the effect of X_1 on Y controlling for X_2, using Z as an instrument for X_1, i.e., estimate β_1 from $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, using Z as an instrument for X_1.

Suppose for now that we have such a Z. How can we use Z to estimate β_1?

The standard econometric method for estimating the effect of X on Y using an Instrumental Variable (IV) approach is called "**Two-stage Least Squares (2SLS)**". It is called "Two-stage" because it involves two stages (actually two regressions).

- First Stage: isolate the part of X that is uncorrelated with u by running the regression: $X_1 = \delta_0 + \delta_1 Z + \delta_2 X_2 + \delta_3 X_3 + \eta$ and get predicted values for X_1. Call these predicted values $\hat{X}_1$.

  Note that if Z is uncorrelated with u, then $\hat{X}_1$ is uncorrelated with u.

- Second Stage: replace X_1 with $\hat{X}_1$ in the regression of interest by running the regression: $Y = \alpha_0 + \alpha_1 \hat{X}_1 + \alpha_2 X_2 + \alpha_3 X_3 + \nu$

  If Z is a valid instrument then $\hat{\alpha}_1$ is an unbiased estimator of the effect of X_1 on Y.

Can you test the validity of the instruments?

- **Relevance**: this assumption can be tested by examining the first stage regression. Stock and Watson suggest using a rule of thumb indicating that if the F test for the joint significance of the instruments is greater than 10, then the instrument is relevant. If not, the instrument is weak. Weak instruments can be problematic.

- **Exogeneity**: this assumption needs to be assessed conceptually. While there are statistical tests (called over-identifying tests) to test this assumption, they give us limited information that is applicable only when the number of instruments is greater than the number of endogenous variables (see Stock and Watson).

Note that you can have more than one instrument. This method works just as described above, but you need each instrument to comply with the two conditions, and in the first stage you would include all the instruments as explanatory variables.

Most statistical packages, including Stata, do both stages together, if you do them separately there is a problem with the standard errors.

IV estimates can be interpreted as the effect of X on Y for those whose treatment status was changed by the instrument. IV estimates measure ***marginal effects***.