Fiscal Policy and Inequality

20. Instrumental Variables

Elliott Ash

Monday, 26 November 2018

Schooling Example

- Our aim is still to estimate the causal effect of schooling on wages
- Lets assume we do not observe everything that affects both selection into schooling and earnings (ability)
- The relationship between earnings and schooling and earnings

$$Y_i = \alpha + \rho S_i + \eta_i$$

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

- The variables A_i are assumed the only reason why η_i and S_i are correlated, i.e.

$$\mathbb{E}[S_i \nu_i] = 0$$

Schooling Example

- If we could observe the variables A_i we could simply include them to the regressions and estimate

$$Y_i = \alpha + \rho S_i + A_i'\gamma + \nu_i$$

- How to estimate ρ without observing A_i?
  - **Instrumental Variable (IV)**: a variable Z_i, that is correlated with S_i, but not correlated with anything else affecting Y_i.

Instrumental variables: Main Intuition

- Regression based on observables:
  - The consistency of the estimate relies on the "hope" that any unobserved factor that might affect the outcome variable is balanced across the treatment and the control group.
  - Therefore, any difference in outcomes between the control and the treatment group can be attributed to the treatment.
- Instrumental variables:
  - We identify some source of variation in the assignment to the treatment which, for some reason, we know that it is orthogonal to any relevant unobserved variable which might be affecting the outcome variables.
  - We compare group of individuals that, due to the instrument, are assigned to the control and the treatment group. Any difference in outcomes between these two groups is attributed to the treatment.

- With a valid instrumental variable we can consistently estimate ρ in
$$Y_i = \alpha + \rho S_i + \eta_i$$
- We can write ρ in terms of the population moments
$$\text{Cov}[Z_i, Y_i] = \rho\text{Cov}[Z_i, S_i] + \text{Cov}[Z_i, \eta_i]$$
- The **exclusion restriction** is $\text{Cov}[Z_i, \eta_i] = 0$.
- Thus:
$$\rho = \frac{\text{Cov}[Z_i, Y_i]}{\text{Cov}[Z_i, S_i]}$$

What is a valid instrumental variable?

- Instrumental variable (IV) is a variable that:

1. Is correlated with causal variable of interest, S_i:

$$\text{Cov}[Z_i, S_i] \neq 0$$

2. Is uncorrelated with any other determinants of Y_i:

$$\text{Cov}[Z_i, \eta_i] = 0$$

- The second requirement can be decomposed in two:
  - 2.1: Exogeneity: None of the unobserved factors affects the instrument:

$$\eta_i \nrightarrow Z_i$$

  - 2.2 Exclusion restriction: Instrument only affects outcome through treatment variable:

$$Z_i \nrightarrow \eta_i$$

How does IV work?

- Intuitive idea behind IV is as follows:
  1. You found a variable (the instrument) that affects who is assigned to the treatment
  2. This variable is unrelated to other factors that affect the outcome
  3. And you know that your instrument has no direct impact on the outcome, it can only affect the outcome through its impact on the treatment.
- In sum, an IV strategy is equivalent to an RCT without full compliance

Can we test validity of IV?

- Is Z_i correlated with causal variable of interest, S_i?
  - YES: check for significance of first stage (first-stage F-statistic)
- Is Z_i uncorrelated with any other determinants of Y_i?
  - NO!
  - The validity of the instrument relies on theory!

Good instruments for schooling?

- Last digit of social security number?
- IQ?
- Month of birth
  - Angrist and Krueger (QJE 1991)
- Family background?
- Geographical proximity?
  - Altonji, Elder and Taber (JHR 2005)
- Working status?

Good instruments are hard to find

- Good instruments come from a combination of three ingredients:
  - Good institutional knowledge
  - Economic theory
  - Last but not least: Originality
- Some usual sources of instruments:
  - Nature (e.g. genes, weather)
  - Assignment rules
  - ‘Natural’ experiments (e.g. the quarter of birth, Vietnam lottery, electoral timing...)

Examples 1

- Immigration
  - Networks of immigrants (Card 1991)
- Does police decrease crime?
  - Electoral cycles (Levitt 1997)
- The impact of violent movies on crime
  - Blockbuster movies (Dahl and DellaVigna 2009)

Examples 2

- The effect of preschool television exposure on standardized test scores during adolescence:
  - Gentzkow and Shapiro 2008
- The Potato's Contribution to Population and Urbanization:
  - Nunn and Nancy Qian 2011
- Influence of mass media on U.S. government response to natural disasters
  - Eisensee and Strömberg 2007

Examples of Bad Instruments

- Parental socioeconomic characteristics as an instrument for children education
- 'South of Italy' as an instrument for CEO's gender

First Stage and Reduced Form

- Recall the formula for IV estimate:

$$\rho = \frac{\text{Cov}[Z_i, Y_i]}{\text{Cov}[Z_i, S_i]}$$

- Note that this is equal to

$$\rho = \frac{\frac{\text{Cov}[Z_i, Y_i]}{\text{Var}[Z_i]}}{\frac{\text{Cov}[Z_i, S_i]}{\text{Var}[Z_i]}} = \frac{\phi}{\gamma}$$

  where ϕ is estimated from the reduced form:

$$Y_i = \alpha + \phi Z_i + \epsilon_i$$

  and γ is estimated from the first stage

$$S_i = \alpha + \gamma Z_i + \nu_i$$

2SLS (1)

- In a model with a single endogenous variable and a single instrument, IV estimates are equivalent to a two stage procedure.
- First stage
$$S_i = \gamma Z_i + \nu_i$$
- Second stage
$$Y_i = \rho S_i + \eta_i$$

2SLS (2)

- Regress and predict first stage:

$$\hat{S}_i = \hat{\gamma} Z_i$$

- Plug into second stage and regress:

$$Y_i = \rho(\hat{\gamma} Z_i) + \eta_i$$

In Python

- In Python, a nice IV package is linearmodels.IV2SLS.
  - not as fast or fully features as ivreghdfe in Stata, but gets the job done.
- See accompanying Jupyter Notebook.

Validity of the instrument

1. Power of the instrument?
2. Exogeneity?
3. Exclusion restriction?

Exogeneity vs. Exclusion Restriction

- The requirement than $E[Z_i'\eta_i] = 0$ can be decomposed in two:
  - 2.1: Exogeneity: None of the unobserved factors affects the instrument:
$$\eta_i \nrightarrow Z_i$$
  - 2.2 Exclusion restriction: Instrument only affects outcome through treatment variable:
$$Z_i \nrightarrow \eta_i$$
- Exogeneity is sufficient for a causal interpretation of the reduced form.
- The exclusion restriction is distinct from the claim that the instrument is (as good as) randomly assigned. Rather, it is a claim about a unique channel for causal effects of the instrument.

Weak Instruments

- The bias of 2SLS can be written as:

$$\text{plim}\hat{\rho} = \rho + \frac{\text{Corr}[Z, \eta]}{\text{Cov}[S, Z]} \cdot \frac{\sigma_\eta}{\sigma_S}$$

  - When the instrument is weakly correlated with the endogenous regressor, the bias increases.
- Can check for a weak instrument with first-stage F-statistic: it should be higher than 10.

Matrix Notation, and Comparison to OLS

- With model $Y = X'\beta + U$ and instrument Z, we have

$$\beta_{OLS} = (X'X)^{-1}(X'Y)$$
$$\beta_{IV} = (Z'X)^{-1}(Z'Y)$$

$$\begin{aligned}\mathbb{E}[\beta_{OLS}] &= \mathbb{E}[(X'X)^{-1}(X'Y)] = \mathbb{E}[(X'X)^{-1}(X'(X'\beta + U))] \\ &= \beta + \mathbb{E}[(X'X)^{-1}(X'U)]\end{aligned}$$

$$\begin{aligned}\mathbb{E}[IV] &= \mathbb{E}[(Z'X)^{-1}(Z'Y)] = \mathbb{E}[(Z'X)^{-1}(Z'(X'\beta + U))] \\ &= \beta + \mathbb{E}[(Z'X)^{-1}(Z'U)]\end{aligned}$$

$$\mathbb{E}[(X'X)^{-1}(X'U)] \gtrless \mathbb{E}[(Z'X)^{-1}(Z'U)]?$$

Local average treatment effects

- Do instruments affect all individuals equally?
  - usually not.
  - If not, then the effect is driven by individuals who respond to the instrument ("compliers").
  - therefore, we can say that IV estimates a "local average treatment effect" rather than "average treatment effect".

Recap

- IV estimates are a powerful tool to identify causal links. But IV power relies on the quality of the instruments. Always discuss instrument plausibility.
- Three dimensions:
  1. Power
     - Always report the first stage (F-test above 10)
     - Weak instruments have very unpleasant consequences
  2. Exogeneity
     - Does it make sense to believe that the instrument is randomly assigned?
     - To be sure: check if the instrument is correlated with predetermined variables
  3. Exclusion restriction
     - Cannot be tested, but discuss the possible links between Z and η
     - Specify the group which is affected by the instrument (local average treatment effect)

Elasticity of Taxable Income (ETI): Estimation

- Recall Gruber and Saez's (2002) choice function for taxable income:

$$y_{it} = y_{it}^0 \cdot (1 - \tau_{it})^\varepsilon$$

- Taking logs:

$$\log\, y_{it} = \log\left(y_{it}^0\right) + \varepsilon \cdot \log\left(1 - \tau_{it}\right)$$

  and want to estimate ϵ.

- OLS estimation is biased because corr $\left(\tau_{it}, y_{it}^0\right) > 0$ due to progressivity of tax system
  - People with positive income shock $\left(y_{it}^0 \uparrow\right)$ face higher tax rate $(\tau_{it} \uparrow)$
- Need to find instrumental variable (IV) for τ_{it}:
  1. Correlated with τ_{it}
  2. Uncorrelated with potential income y_{it}^0

$\Rightarrow$ Hard to find good IVs

Methodology

- Generalization of Feldstein
  - Use instrumental variables (IV) regression in a diff-in-diff setting
  - Panel dataset from 1979-1990, larger sample ($N \approx 60,000$)
  - Multiple reforms, some tax increases ($\tau \uparrow$) and some tax cuts ($\tau \downarrow$)

Assumptions

Underlying assumptions:

1. ETI (ε) is constant over time and identical across people
2. Short-term and long-term responses are identical
3. Individuals have perfect knowledge of tax system
4. Income effects are allowed (in some specifications)

Empirical Strategy 1

- **Idea:** relate *several* pairs of years t and $t+k$:
  - % change in income = $\log y_{it+k} - \log y_{it}$
  - % change in mgl. tax rate = $\log(1-\tau_{it+k}) - \log(1-\tau_{it})$
    - where $k = 1, 2, 3...$ years
- Main regression equation in differences:

$$\underbrace{\log\left(\frac{y_{it+k}}{y_{it}}\right)}_{\text{\% change in income}} = \alpha + \varepsilon \cdot \underbrace{\log\left(\frac{1-\tau_{it+k}}{1-\tau_{it}}\right)}_{\text{\% change in tax}} + \underbrace{u_{it}}_{\text{error}} \quad (1)$$

Empirical Strategy 2

- Endogeneity problem in (1):
  - If positive shock to income ($u_{it} > 0$), the marginal tax rate increases mechanically due to progressivity $\left(\log\left(\frac{1-\tau_{it+k}}{1-\tau_{it}}\right)\downarrow\right)$
  - Hence, $\text{corr}\left(\log\left(\frac{1-\tau_{it+k}}{1-\tau_{it}}\right), u_{it}\right) < 0$
  - By construction, $\text{corr}\left(\log\left(\frac{y_{it+k}}{y_{it}}\right), u_{it}\right) > 0$
  - OLS estimate of ε is biased downwards.
- Solution: find instrumental variable (IV) for τ_{it}

Instrumental Variables (IV)

- Let τ^p be the marginal tax rate that an individual would face in year $t+k$ if her real income was the same as in year t
  - Example:

$$\tau_{it} = \begin{cases} 0.2, & y_{it} \in [0,\infty] \end{cases} \text{ and } \tau_{it+k} = \begin{cases} 0.2, & y_{it+k} \in [0,100] \\ 0.4, & y_{it+k} \in (100,\infty) \end{cases}$$

Let $y_{it} = 110$, such that $\tau_{it} = 0.2$. *If the individual does not change her labor supply behavior*, then $\tau_i^p = 0.4$, but in reality she might change her income to $y_{it+k} = 99$, and then $\tau_{it+k} = 0.2$.

- **Intuition:** instrument captures only the mechanical variation in tax rates due to exogenous changes in the law, not to *endogenous* behavioral responses

IV Strategy

- First-stage regression:

$$\log\left(\frac{1-\tau_{t+k}}{1-\tau_t}\right) = \delta_0 + \delta_1 \log\left(\frac{1-\tau^p}{1-\tau_t}\right) + v_{it}$$

  Where τ^p is the tax rate that the individual would have faced in year $t+k$ if her real income had not changed

- Second-stage regression:

$$\log\left(\frac{y_{it+k}}{y_{it}}\right) = \alpha + \varepsilon \cdot \log\widehat{\left(\frac{1-\tau_{it+k}}{1-\tau_{it}}\right)} + u_{it}$$

- also, control for base-year income with polynomial $f(y_{it})$

Outcomes

Two income measures:

- ***Broad income***: includes all types of income (wages, interest, business, etc) without applying any adjustments from the tax code.
- ***Taxable income***: amount over which the tax rate is calculated. Result of applying all the deductions and exemptions from the tax law to *Broad income.*

Results

Table 4

Income measure:	Broad	Taxable	Broad	Taxable
	(3)	(4)	(5)	(6)
Elasticity	**0.170**	**0.611**	**0.120**	**0.400**
	(0.106)	(0.144)	(0.106)	(0.144)
Married	0.045	0.049	0.050	0.055
	(0.014)	(0.023)	(0.012)	(0.055)
Single	-0.034	-0.032	-0.036	-0.027
	(0.013)	(0.022)	(0.013)	(0.021)
log income	log income		log income splines	
Obs.	69,129	59,199	69,129	59,199

Adapted from Gruber and Saez (JPubEc, 2002)

Results

- Elasticity of taxable income (ETI) around 0.4-0.6
  - Significant, but far below Feldstein's estimates
  - In line with other studies post-Feldstein
- Elasticity of broad income is lower, 0.1-0.2, for two reasons:
  - Mechanical: the base is larger for broad income, so same change in $ is smaller in %
  - Behavioral: taxable income includes deductions, which are likely to respond to changes in τ
- Small income effects

Caveats

Some caveats:

- Results sensitive to exclusion of low incomes (minimum income thresold)
- Weber (2014) critique: instrument not exogenous in practice
  - Use further lags of income to obtain a consistent IV (unbiased in large samples)
- Later studies find smaller elasticities using data from other countries (eg, Denmark)
- Bundles together small tax changes and large tax changes
  - Chetty (2009) shows that if individuals respond only in short-medium term, the estimated elasticity is too low

Does biased media affect voting?
Yurukoglu and Martin (2017)

- First stage:

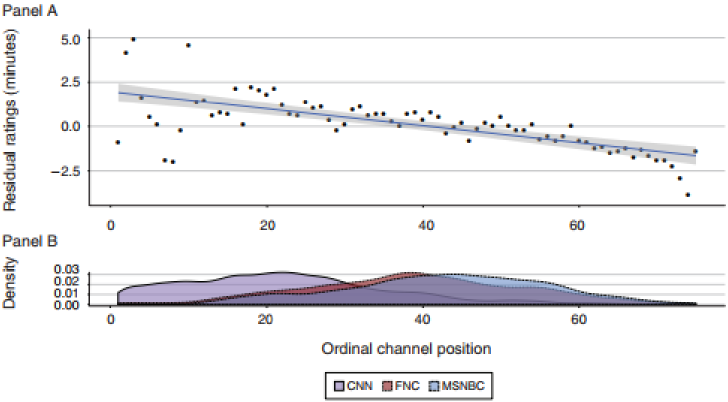
$$FoxNewsRatings_{it} = \alpha_{it} + \beta FoxChannelNumber_{it} + \eta_{it}$$

- Second stage:

$$RepubVote_{it} = \alpha_{it} + \beta FoxNewsRatings_{it} + \epsilon_{it}$$

First stage effect of channel numbers

Yurukoglu and Martin (2017)



First Stage Statistics

Yurukoglu and Martin (2017)

Table 2—First-Stage Regressions: Nielsen Data

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

Notes: Cluster-robust standard errors in parentheses (clustered by cable system). Instrument is the ordinal position of FNC on the local system. The omitted category for the availability dummies is systems where neither FNC nor MSNBC are available. In columns 4 and 6, the specification conditions on the average FNC ratings among satellite subscribers in the same zip code. Cable system controls include the total number of channels on the system and the number of broadcast channels on the system, as well as an indicator for Nielsen collection mode (diary versus set-top). Basic demographics include the racial, gender, age, income, educational, and urban/rural makeup of the zip code. Extended demographics include the Basic set plus information on the percentage of homeowners; median

Second Stage Results

Yurukoglu and Martin (2017)

Table 4—Second Stage Regressions: Zip Code Voting Data

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

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