Econometrics II

Fabian Waldinger (LMU Munich)

Topics Covered in Lecture

1. Recap from last lecture: IV
2. IV with heterogeneous treatment effects

Short Description of Angrist (1990) Veteran Draft Lottery Example

- In the following we will refer to an example from Angrist's paper on the effects of military service on earnings
- Angrist (1990) uses the Vietnam draft lottery as in IV for military service
- In the 1960s and early 1970s, young American men were drafted for military service to serve in Vietnam
- Concerns about the fairness of the conscription policy lead to the introduction of a draft lottery in 1970
- From 1970 to 1972 random sequence numbers were assigned to each birth date in cohorts of 19-year-olds
- Men with lottery numbers below a cutoff were drafted while men with numbers above the cutoff could not be drafted
- The draft did not perfectly determinate military service:
    - Many draft-eligible men were exempted for health and other reasons
    - Exempted men volunteered for service

Summary of Findings on Vietnam Draft Lottery

1. First stage results:
   Having a low lottery number (being eligible for the draft) increases veteran status by about 16 percentage points (the mean of veteran status is about 27 percent)
2. Second stage results:
   Serving in the army lowers earnings by between $2,050 and $2,741 per year.

IV with Heterogeneous Treatment Effects

- Up to this point we only considered models where the causal effect was the same for all individuals (homogenous treatment effects):

$$Y_{1i} - Y_{0i} = \kappa \text{ for all } i$$

- We now try to understand what IV estimates if treatment effects are heterogeneous

IV with Heterogeneous 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} = is treatment status when $Z_i = 1$
  - D_{0i} = is treatment status when $Z_i = 0$
- Observed treatment status is therefore:

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \gamma_0 + \gamma_{1i}Z_i + \mu_i$$

  - $\gamma_0 = E[D_{0i}]$
  - $\gamma_{1i} = (D_{1i} - D_{0i})$ is the heterogeneous causal effect of the IV on D_i
  - The average causal effect of Z_i on D_i is $E[\gamma_{1i}]$

Key Assumptions in the Heterogeneous Effects Framework

1. **Independence assumption:**

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

$$Y_i(D_{1i},1), Y_i(D_{0i},0), D_{1i}, D_{0i} \perp Z_i$$

- The independence assumption is sufficient for a causal interpretation of the reduced form: $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$

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

- Independence also means that the first stage captures the causal effect of Z_i on D_i : $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$

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

Key Assumptions in the Heterogeneous Effects Framework

2. **Exclusion restriction:**
   - $Y_i(D, Z)$ is a function of D only. Or formally:

$$Y_i(D, 0) = Y_i(D, 1) \text{ for } D = 0, 1$$

- In the Vietnam draft lottery example: 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 reasons:
  1. through its effect on military service
  2. through its effect on educational attainment
- The fact that the lottery number is randomly assigned (and therefore satisfies the independence assumption) does not ensure that the exclusion restriction is satisfied

Key Assumptions in the Heterogeneous Effects Framework

- Using the exclusion restriction we can define potential outcomes indexed solely against treatment status:

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

- In terms of potential outcomes we can write:

$$Y_i = Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i$$
$$= Y_{0i} + [Y_{1i} - Y_{0i}]D_i$$

- Random coefficients notation for this is:

$$Y_i = \beta_1 + \kappa_i D_i$$

with $\beta_1 = E[Y_{0i}]$ and $\kappa_i = Y_{1i} - Y_{0i}$

3. **First Stage:**
   As in the constant coefficients model we also need that the instrument has to have a significant effect on treatment:
   $$E[D_{1i} - D_{0i}] \neq 0$$

Key Assumptions in the Heterogeneous Effects Framework

4. **Monotonicity:**
   Either $\gamma_{1i} \geq 0$ for all i or $\gamma_{1i} \leq 0$ for all i
   - While the instrument may have no effect on some people, all those who are affected are affected in the same way
   - In the draft lottery example: draft eligibility may have had no effect on the probability of military service. But there should also be no one who was kept out of the military by being draft eligible. $\rightarrow$ this is likely satisfied
   - In the quarter of birth example for schooling the assumption may not be satisfied (see 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, $Y_{1i} - Y_{0i}$

IV Estimates LATE

- If all 4 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
- 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. (This excludes volunteers and men who were exempted from military service for medical reasons for example)
- We have reviewed the properties of IV with heterogeneous treatment effects using a very simple dummy endogenous variable, dummy IV, and no additional controls example. The intuition of LATE generalizes to most cases where we have continuous endogenous variables and instruments, and additional control variables

Some LATE Framework Jargon The

- 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 treatment status is affected by the instrument in the "wrong" direction
- These terms come from an analogy to the medical literature where the treatment is taking a pill for example

Monotonicity Ensures That There Are No Deers Monotonicity ensures

- Monotonicity ensures that there are no defiers
- Why is it important to not have defiers?
  - If there were defiers, effects on compliers could be (partly) cancelled out by opposite effects on defiers
  - One could then observe a reduced form which is close to 0 even though treatment effects are positive for everyone (but the compliers are pushed in one direction by the instrument and the defiers in the other direction)

What Does IV Estimate and What Not?

- As outlined above, with all 4 assumptions satisfied, IV estimates the average treatment effect for compliers
- Without further assumptions (e.g. constant causal effects) LATE is not informative about effects on never-takers or always-takers because the instrument does not affect their treatment status
- In most applications we would be mostly interested in estimating the average treatment effect on the whole population (ATE):

$$E[Y_{1i} - Y_{0i}]$$

- This is usually not possible with IV

Other Potentially Interesting Treatment Effects

- Another effect which we may potentially be interested in estimating is the *average treatment effect on the treated* (ATT)
- Treatment status is: $D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$
- By monotonicity we cannot have $D_{0i} = 1$ and $(D_{1i} - D_{0i}) = 1$ since $D_{0i} = 1$ implies $D_{1i} = 1$
- The treated therefore either have $D_{0i} = 1$ (always-takers) or $(D_{1i} - D_{0i}) = 1$ and $Z_i = 1$ (compliers)
- It follows that LATE is not the same as ATT:

$$\underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{\text{Effect on treated}} = \underbrace{E[Y_{1i} - Y_{0i}|D_{0i} = 1]}_{\text{Effect on always takers}} P[D_{0i} = 1|D_i = 1]$$

$$+ \underbrace{E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]}_{\text{Effect on compliers}} P[D_{1i} > D_{0i}, Z_i = 1|D_i = 1]$$

- If there are no always takers we can, however, estimate ATT which is equal to LATE in that case

IV in Randomized Trials

- The use of IV methods can also be useful when evaluating a randomized trial
- In many randomized trials, participation is voluntary among those randomly assigned to treatment
- On the other hand people in the control group usually do not have access to treatment
  $\rightarrow$ only those who are particularly likely to benefit from treatment will actually take up treatment (leads almost always to positive selection bias)
  $\rightarrow$ if you just compare means between treated and untreated individuals (using OLS) you will obtain biased treatment effects
- Solution: Instrument for treatment with whether you were offered treatment (You estimate a LATE)

IV in Randomized Trials - Example: Training Programme

- David Autor calculates the different effects for the JTPA training programme
- The programme provides job training to people facing barriers for employment (e.g. dislocated workers, disadvantaged young adults)
- The programme was randomly offered. Only 60 percent of those who were offered the training actually received it
- 2 percent of people in the control group also received training
- Autor evaluates differences in earnings in the 30 month period after random assignment

Regression Results Training Programme

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	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

- Columns (1) and (2) show OLS estimates
- Columns (3) and (4) show ITT (reduced form) estimates
- Columns (5) and (6) show IV estimates
- Here we actually estimate something close to the ATT (not only LATE) because there are almost no always-takers

Weak Instruments

- IV is consistent but not unbiased
- For a long time researchers estimating IV models never cared much about the small sample bias
- In the early 1990s a number of papers, however, highlighted that IV can be severely biased in particular if instruments are weak (i.e. the first stage relationship is weak) and if you use many instruments to instrument for one endogenous variable (i.e. there are many overidentifying restrictions)
- In the worst case, if the instruments are so weak that there is no first stage, the 2SLS sampling distribution is centered on the probability limit of OLS

Weak Instruments - Bias Towards OLS

- Lets consider a model with a single endogenous regressor and a simple constant treatment effect
- The causal model of interest is:

$$Y = \beta_2 X_2 + \varepsilon \quad (1)$$

- The N x Q matrix of instrumental variables is Z with the first stage equation:

$$X_2 = \boldsymbol{Z}\gamma + \xi \quad (2)$$

- If ε and ξ are correlated, estimating (1) by OLS would lead to biased results
- The OLS bias is:

$$E[\beta_2^{OLS} - \beta_2] = \frac{Cov[\varepsilon, X_2]}{Var[X_2]}$$

- If ε and ξ are correlated the OLS bias is therefore: $\frac{\sigma_{\varepsilon\xi}}{\sigma^2_{X_2}}$

Weak Instruments - Bias Towards OLS

- It can be shown that the bias of 2SLS is approximately:

$$E[\beta_2^{2SLS} - \beta_2] \approx \frac{\sigma_{\varepsilon\xi}}{\sigma_\xi^2} \frac{1}{F+1}$$

- F is the population analogue of the F-statistic for the joint significance of the instruments in the first stage regression. See MHE pp. 206-208 for a derivation (only for interested students)
- If the first-stage is weak (i.e. $F \to 0$) the bias of 2SLS approaches $\frac{\sigma_{\varepsilon\xi}}{\sigma_\xi^2}$
- This is the same as the OLS bias as for $\gamma = 0$ in equation (2) (i.e. there is no first stage relationship) $\sigma_{X_2}^2 = \sigma_\xi^2$ and therefore the OLS bias becomes $\frac{\sigma_{\varepsilon\xi}}{\sigma_\xi^2}$
- If the first stage is very strong (i.e. $F \to \infty$) the IV bias goes to 0

Weak Instruments - Adding More Instruments

- Adding more weak instruments will increase the bias of 2SLS
- By adding further instruments without predictive power the first stage F-statistic goes towards 0 and the bias increases
- If the model is just identified, weak instrument bias is less of a problem: (in MHE p. 209 they write it is approximately unbiased - this is only true if the first stage is not 0; see http://econ.lse.ac.uk/staff/spischke/mhe/josh/solon_justid_April14.pdf)
- Bound, Jaeger, and Baker (1995) highlighted this problem for the Angrist & Krueger study. A&K present findings from using different sets of instruments:
    1. quarter of birth dummies $\rightarrow$ 3 instruments
    2. quarter of birth + (quarter of birth) x (year of birth) dummies $\rightarrow$ 30 instruments
    3. quarter of birth + (quarter of birth) x (year of birth) + (quarter of birth) x (state of birth) $\rightarrow$ 180 instruments.

Adding Instruments in Angrist & Krueger

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063	.142	.063	.081	.063	.060
	(.000)	(.033)	(.000)	(.016)	(.000)	(.029)
F (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, ×100)		.012		.043		.014
F (overidentification)		.932		.775		.725
	Age Control Variables					
Age, Age^2	x	x			x	x
9 Year of birth dummies			x	x	x	x
	Excluded Instruments					
Quarter of birth		x		x		x
Quarter of birth × year of birth				x		x
Number of excluded instruments		3		30		28

- Adding more weak instruments reduced the first stage F-statistic and moves the coefficient towards the OLS coefficient

Adding Instruments in Angrist & Krueger

	(1) OLS	(2) IV
Coefficient	.063	.083
	(.000)	(.009)
F (excluded instruments)		2.428
Partial R^2 (excluded instruments, ×100)		.133
F (overidentification)		.919
Age Control Variables		
Age, Age^2		
9 Year of birth dummies	x	x
Excluded Instruments		
Quarter of birth		x
Quarter of birth × year of birth		x
Quarter of birth × state of birth		x
Number of excluded instruments		180