Georgetown University
Spring 2016
Advanced Applied Econometrics
(PPOL 754-20)

Andrew H. McCallum, Ph.D.

April 7, 2016

Disclaimer

Any opinions and conclusions expressed during this course are solely the responsibility of the author and do not necessarily represent the views of the Board of Governors or any other person associated with the Federal Reserve System.

Outline

MHE Ch. 4

- IV regressions
- How do they get us to causality?
- The Wald estimator (its so simple we can do it by hand)
- Using more than Wald
- IV examples
- Heterogeneous causal effects

The Wald estimator

Simplest IV with one dummy instrument, one endogenous regressor

- How were Vietnam-era vets earnings affected by their service?
- Let $D_i = 1$ indicate veterans. A causal constant-effects model is:

$$Y_i = \alpha + \rho D_i + \eta_i, \tag{1}$$

where η_i and D_i may be correlated. Because Z_i is a dummy,

$$\frac{\text{Cov}\left(Y_i, Z_i\right)}{V\left(Z_i\right)} = E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right],$$

with an analogous formula for $\frac{\text{Cov}(D_i, Z_i)}{V(Z_i)}$. It follows that,

$$\rho = \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(D_i, Z_i\right)} = \frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]} \tag{2}$$

A direct route uses (1) and $E\left[\eta_i \mid Z_i\right] = 0$ to get

$$E\left[Y_i \mid Z_i\right] = \alpha + \rho E\left[D_i \mid Z_i\right]$$

Then solving this for ρ produces (2)

Earnings Consequences of Vietnam-Era Military Service (Angrist, 1990)

- Key variables
  Z_i = randomly assigned draft-eligibility in the 1970-72 draft lotteries
  D_i = a dummy indicating Vietnam-era veterans
- The causal effect of Vietnam-era military service is the difference in average earnings by draft-eligibility status (RF) divided by the difference in the probability of service (first stage)

$$\begin{aligned} \frac{\operatorname{Cov}(D_i, Z_i)}{V(Z_i)} &= E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] \\ &= P[D_i = 1 \mid Z_i = 1] - P[D_i = 1 \mid Z_i = 0] \end{aligned}$$

- See RF, first stage, and IV in **Angrist (1990)**, Figures 1-2 and MHE Table 4.1.3 (based on Angrist 1990,Table 3)
- Draft lottery updates Angrist, Chen, and Song (2011)

Angrist (1990)

TABLE 3—WALD ESTIMATES

Cohort	Year	Draft-Eligibility Effects in Current $: FICA Earnings (1)	Adjusted FICA Earnings (2)	Total W-2 Earnings (3)	$\hat{p}^e - \hat{p}^n$ (4)	Service Effect in 1978 $ (5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8
		(281.9)	(330.1)	(376.0)		(1,959.3)

Notes: Standard errors in parentheses.
Columns (1) and (3) are taken from Table 1.
Column (2) reports draft-eligibility treatment effects on earnings adjusted for censoring

Multiple groups and 2SLS

- There's more to the draft lottery than draft-eligibility: Angrist and Chen (2008), Figure 1
- Let R_i denote draft lottery numbers. The draft-eligibility Wald estimator uses $1\left[R_i < 195\right]$ as an instrument in a just-identified setup The first stage that uses everything we know can be written:

$$E\left[Y_i \mid R_i\right] = \alpha + \rho P\left[D_i = 1 \mid R_i\right],$$

since $P\left[D_i = 1 \mid R_i\right] = E\left[D_i \mid R_i\right]$. Suppose $R_i \in j = 1, \ldots, J$. We can estimate ρ using J grouped obs by fitting

$$\overline{y}_j = \alpha + \rho \hat{p}_j + \overline{\eta}_j$$

- Efficient GLS for grouped data in a constant-effects linear model is weighted least squares, in this case weighted by the variance of $\overline{\eta}_j$ (Prais and Aitchison, 1954). If η_j has variance σ_η^2, the grouped variance is $\frac{\sigma_\eta^2}{n_j}$, where n_j is the group size

Intention to Treat

- Sometimes people refer to the "reduced form" as the "intention to treat"
- Think of an experiment, where treated group is offered a treatment
- Some of them accept the treatment some don't
- Can use instruments to model the acceptance of treatment as long as they are not related to the outcome

Some Examples of IV

- Same sex instrument for women's labor supply (Angrist and Evans)
- Vietnam draft lottery (Angrist)
  - Instrument for military service, used to estimate its impact on earnings
- Quarter of birth (Angrist Krueger)
  - Instrument for completion of high school used to school, estimate impact on earnings
- Settler mortality (Acemoglu Johnson Robinson) (Acemoglu, Johnson,
  - Instrument for quality of institutions, used to estimate impact on growth

Our Constant-Effects Benchmark

- The traditional IV framework is the linear, constant-effects world discussed last week with Bernoulli treatment, D_i, we have

$$\begin{aligned} Y_{0i} &= \alpha + \eta_i \\ Y_{1i} - Y_{0i} &= \rho \\ Y_i &= Y_{0i} + D_i (Y_{1i} - Y_{0i}) = \alpha + \rho D_i + \eta_i \end{aligned}$$

- η_j is not a regresson error (Y_{0i} is not independent of D_i), so OLS fails to capture causal effects
- Using an instrument, Z_i, that's independent of Y_{0i} but correlated with D_i, we have

$$\rho = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)}$$

- Constant FX focuses our attention on omitted variables bias, abstracting from more subtle concerns
- Time now to allow for the fact that $Y_{1i} - Y_{0i}$ need not be (in some cases, cannot be) the same for everyone

Sometimes You Get What You Need

- In a heterogeneous world, we distinguish between *internal validity and external validity*
- A good instrument — by definition — captures an internally valid causal effect. This is the causal effect on the group subject to (quasi-) experimental manipulation
- External validity is the predictive value of internally valid causal estimates in contexts beyond those generating the estimates
- Examples
  - Draft-lottery estimates of the effects of Vietnam-era military service
  - Quarter-of-birth estimates of the effects of schooling on earnings
  - Regression-discontinuity estimates of the effects of class size
- In a heterogeneous world:
  - Quasi-experimental designs capture causal effects for a well-defined subpopulation, usually a proper subset of the treated
  - In models with variable treatment intensity, we typically get effects over a limited but knowable range

Roadmap

1. An example: the effect of childbearing on labor supply
   - Two good instruments, two good answers
2. The theory of instrumental variables with heterogeneous potential outcomes
   - Notation and framework
   - The LATE Theorem
3. Implications for the design and analysis of field trials
   - The Bloom Result
   - Illustration: JTPA and MDVE
4. All about compliers: Kappa and QTE
5. Average causal response in models with variable treatment intensity
   - The ACR theorem and weighting function
   - A world of *continuous* activity
6. External Validity (first pass)

Children and Their Parents Labor Supply

- A causal model for the impact of a third child on mothers with at least two:

$$Y_i = Y_{0i} + D_i \left(Y_{1i} - Y_{0i}\right) = \alpha + \rho D_i + \eta_i$$

  Constant FX? Here, ρ is *the thing that must be named*
- Dependent variables = employment, hours worked, weeks worked, earnings
  - $D_i = 1\left[\text{kids} > 2\right]$ for samples of mothers mothers with at least two children
  - Z_i indicates twins or same-sex sibships at second birth
- With a single Bernoulli instrument and no covariates, the IV estimand is the Wald formula

$$\rho = \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(D_i, Z_i\right)} = \frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]}$$

- Results from your homework

The LATE Framework

- Let $Y_i(d, z)$ denote the potential outcome of individual i were this person to have treatment status $D_i = d$ and instrument value $Z_i = z$
- Note the double-indexing: candidate instruments *might* have a direct effect on outcomes
- We assume, however, that IV initiates a causal chain: the instrument, Z_i, affects D_i, which in turn affects Y_i
- To flesh this out, we first define *potential treatment status*, indexed against Z_i:
  - D_{1i} is i's treatment status when $Z_i = 1$
  - D_{0i} is i's treatment status when $Z_i = 0$
- Observed treatment status is therefore

$$D_i = D_{0i} + (D_{1i} - D_{0i}) Z_i$$

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

LATE Assumptions (Independence and First Stage)

- Independence: The instrument is as good as randomly assigned

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

- Independence means that draft lottery numbers are independent of potential outcomes and potential treatments
- Independence implies that the **first-stage** is the average causal effect of Z_i on D_i

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

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

$$E\left[Y_i \mid Z_i=1\right]-E\left[Y_i \mid Z_i=0\right]=E\left[Y_i\left(D_{1i},1\right)-Y_i\left(D_{0i},0\right)\right]$$

  RF is the causal effect of the *instrument* on the dependent variable, but we have yet to link this to treatment

LATE Assumptions (Exclusion)

- Our journey from RF to treatment effects starts here. The exclusion restriction means the instrument affects Y_i only through D_i

$$\begin{aligned} Y_i(1,1) &= Y_i(1,0) \equiv Y_{1i} \\ Y_i(0,1) &= Y_i(0,0) \equiv Y_{0i} \end{aligned}$$

- The exclusion restriction means Y_i can be written:

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

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

- Exclusion means draft lottery numbers affect earnings only via veteran status; quarter of birth affects earnings only through schooling; sex comp affects labor supply only by changing family size

LATE assumptions (Monotonicity)

A necessary technical assumption:
Monotonicity $D_{1i} \geq D_{0i}$ for everyone (or vice versa).

- By virtue of monotonicity, $E\left[D_{1i} - D_{0i}\right] = P\left[D_{1i} > D_{0i}\right]$
- Interpreting monotonicity in latent-index models:

$$D_i = \begin{cases} 1 & \text{if} \quad \gamma_0 + \gamma_1 Z_i > v_i \\ 0 & \text{otherwise} \end{cases}$$

  where v_i is a random factor.
- This model characterizes potential treatment assignments as:

$$\begin{aligned} D_{0i} &= 1\left[\gamma_0 > v_i\right] \\ D_{1i} &= 1\left[\gamma_0 + \gamma_1 > v_i\right], \end{aligned}$$

  which clearly satisfy monotonicity

The LATE Theorem

Assumption recap

- The independence assumption is sufficient for identification of the causal effects of the *instrument*
- 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
  - Exclusion is (or should be) more controversial than independence
- 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
- Given these assumptions, we have:
  THE LATE THEOREM

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

Proof — See MHE 4.4.1

The Compliant Subpopulation

LATE compliers are subjects with $D_{1i} > D_{0i}$

- This language comes from randomized trials where Z_i is treatment assigned and D_i is treatment received (more on this soon)
- LATE assumptions partition the world
  - Compliers $D_{1i} > D_{0i}$
  - Always-takers $D_{1i} = D_{0i} = 1$
  - Never-takers $D_{1i} = D_{0i} = 0$
- IV is uninformative for always-takers and never-takers because treatment status for these types is unchanged by the instrument
  - An analogy: panel models with fixed effects identify treatment effects only for "changers"
- Of course, we can assume effects are the same for all three groups (a version of the constant-effects model)

The Compliant Subpopulation (cont.)

- From the fact that

$$D_i = D_{0i} + (D_{1i} - D_{0i})\, Z_i,$$

we see that:

$$\{D_i = 1\} = \{D_{0i} = D_{1i} = 1\} \cup \{\{D_{1i} - D_{0i} = 1\} \cap \{Z_i = 1\}\}$$

- In other words . . .

$$\{\text{treated}\} = \{\text{always-takers}\} + \{\text{compliers assigned } Z_i = 1\}$$

- TOT is therefore a weighted average of effects on always-takers and compliers (compliers rolling $Z_i = 1$ are representative of all 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* analyses preserve independence but are diluted by non-compliance
- IV solves this problem: Z_i, is a dummy variable indicating random assignment to the treatment group; D_i is a dummy indicating whether treatment received or taken
- No always-takers! (no controls are treated), so LATE = TOT
  THE BLOOM RESULT

$$\frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right]} = \frac{\text{ITT effect}}{\text{compliance rate}} = E\left[Y_{1i} - Y_{0i} \mid D_i = 1\right]$$

Direct proof (Bloom 1984; See MHE 4.4.3)

Bloom Example 1: Training

The Job Training Partnership 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
- IV setup
  - D_i indicates those who received JTPA services
  - Z_i indicates the random offer of treatment
  - Y_i is earnings in the 30 months since random assignment
- The first-stage here is approximately the compliance rate

$$E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right] \cong P\left[D_i = 1 \mid Z_i = 1\right] = .60$$

  ($.62$ of $Z_i = 1$ group trained; $.02$ of $Z_i = 0$ group also trained)
- Table 4.4.1 Selection bias in OLS (as delivered), ITT (as assigned) is diluted, IV (TOT) is . . . just right!

Bloom Example 2: Battered Wives

What's the best police response to domestic violence? The Minneapolis Domestic Violence Experiment (MDVE; Sherman and Berk, 1984) boldly tried to find out ...

- Police were randomly assigned to advise, separate, or arrest
- Substantial compliance problems as officers made their own judgements in the field

Table: Assigned and Delivered Treatments in Spousal Assault Cases

Assigned Treatment	Delivered Treatment			
		Coddled		
	Arrest	Advise	Separate	Total
Arrest	98.9 (91)	0.0 (0)	1.1 (1)	29.3 (92)
Advise	17.6 (19)	77.8 (84)	4.6 (5)	34.4 (108)
Separate	22.8 (26)	4.4 (5)	72.8 (83)	36.3 (114)
Total	43.4 (136)	28.3 (89)	28.3 (89)	100.0(314)

Notes: The table shows statistics from Sherman and Berk (1984), Table 1.

MDVE First-Stage and Reduced Forms

- Analysis in Angrist (2006)

Table: First Stage and Reduced Forms for Model 1

Endogenous Variable is Coddled				
	First-Stage		Reduced Form (ITT)	
	(1)	(2)*	(3)	(4)*
Coddled-assigned	0.786	0.773	0.114	0.108
	(0.043)	(0.043)	(0.047)	(0.041)
Weapon		-0.064		-0.004
		(0.045)		(0.042)
Chem. Influence		-0.088		0.052
		(0.040)		(0.038)
Dep. Var. mean	0.567		0.178	
	(coddled-delivered)		(failed)	

Notes: The table reports OLS estimates of the first-stage and reduced form for Model 1 in the text. *Other covariates include year and quarter dummies, and or non-white and mixed.

MDVE OLS and 2SLS

Table: First Stage and Reduced Forms for Model 1

Endogenous Variable is Coddled				
	OLS		IV/2SLS	
	(1)	(2)*	(3)	(4)*
Coddled-assigned	0.087	0.070	0.145	0.140
	(0.044)	(0.038)	(0.060)	(0.053)
Weapon		0.010		0.005
		(0.043)		(0.043)
Chem. Influence		0.057		0.064
		(0.039)		(0.039)

Notes: The Table reports OLS and 2SLS estimates of the structural equation in Model 1. *Other covariates include year and quarter dummies, and dummies for non-white and mixed race.

How Many Compliers You Got?

- Given monotonicity, we have

$$\begin{aligned} P\left[D_{1i} > D_{0i}\right] &= E\left[D_{1i} - D_{0i}\right] = E\left[D_{1i}\right] - E\left[D_{0i}\right] \\ &= E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right] \end{aligned}$$

  The first stage tells us how many!
- And among the treated?
  - Start with the definition of conditional probability:

$$\begin{aligned} P\left[D_{1i}\right. &> \left. D_{0i} \mid D_i = 1\right] \\ &= \frac{P\left[D_i = 1 \mid D_{1i} > D_{0i}\right] P\left[D_{1i} > D_{0i}\right]}{P\left[D_i = 1\right]} \\ &= \frac{P\left[Z_i = 1\right]\left(E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]\right)}{P\left[D_i = 1\right]} \end{aligned}$$

    An easy calculation, proportional to the first stage
- **Sample complier counts**

Characterizing Compliers

- Are sex-comp compliers more or less likely to be college graduates (indicated by $x_{1i} = 1$) than other women?

$$\begin{aligned} & \frac{P\left[x_{1i} = 1 \mid D_{1i} > D_{0i}\right]}{P\left[x_{1i} = 1\right]} \\ = & \frac{P\left[D_{1i} > D_{0i} \mid x_{1i} = 1\right]}{P\left[D_{1i} > D_{0i}\right]} \\ = & \frac{E\left[D_i \mid Z_i = 1, x_{1i} = 1\right] - E\left[D_i \mid Z_i = 0, x_{1i} = 1\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]} \end{aligned}$$

- The relative likelihood a complier is a college grad Is given by the ratio of the first stage for college grads to the overall first stage
- **Sample complier characteristics**

Distribution Treatment Effects

- LATE, $E\left[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}\right]$, is an average causal effect. We turn now to the *distribution* of potential outcomes for compliers.
- Abadle (2002) showed that, for any measurable function, $g\left(Y_{ji}\right)$,

$$\begin{aligned} & \frac{E\left[D_i g\left(Y_i\right) \mid Z_i = 1\right] - E\left[D_i g\left(Y_i\right) \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]} = E\left[g\left(Y_{1i}\right) \mid D_{1i} > D_{0i}\right] \\ & \frac{E\left[\left(1 - D_i\right) g\left(Y_i\right) \mid Z_i = 1\right] - E\left[\left(1 - D_i\right) g\left(Y_i\right) \mid Z_i = 0\right]}{E\left[1 - D_i \mid Z_i = 1\right] - E\left[1 - D_i \mid Z_i = 0\right]} \\ = \quad & E\left[g\left(Y_{0i}\right) \mid D_{1i} > D_{0i}\right] \end{aligned}$$

- Set $g\left(Y_{ji}\right) = Y_{ji}$ to capture marginal mean outcomes; set $g\left(Y_{ji}\right) = 1\left[Y_{ji} < c\right]$ to capture distributions:

$$E\left\{1\left[Y_{ji} < c\right] \mid D_{1i} > D_{0i}\right\} = P\ \left[Y_{ji} < c\right] \mid D_{1i} > D_{0i}$$

- **Charter school IV and the distribution of test scores**

Theorem

ABADIE KAPPA. Suppose the assumptions of the LATE theorem hold conditional on covariates, X_i. Let $g(Y_i, D_i, X_i)$ be any measurable function of (Y_i, D_i, X_i) with finite expectation. Define

$$\kappa_i = 1 - \frac{D_i(1 - Z_i)}{1 - P(Z_i = 1 \mid X_i)} - \frac{(1 - D_i) Z_i}{P(Z_i = 1 \mid X_i)}.$$

Then

$$E[g(Y_i, D_i, X_i) \mid D_{1i} > D_{0i}] = \frac{E[\kappa_i g(Y_i, D_i, X_i)]}{E[\kappa_i]}$$

Proof.

By monotonicity, those with $D_i(1 - Z_i) = 1$ are always-takers because they have $D_{0i} = 1$, while those with $(1 - D_i) Z_i = 1$ are never-takers because they have $D_{1i} = 0$. Kappa removes means for always-takers and never-takers from marginal means, leaving the average for compliers. □

Using Kappa

- Sketch of proof: Kappa uses this relation, true by monotonicity:

$$E\left[Y \mid c\right] = \frac{1}{P\left(c\right)}\left\{E\left[Y\right] - E\left[Y \mid AT\right]P\left(AT\right) - E\left[Y \mid NT\right]P\left(NT\right)\right\}$$

  Who cares? *Conditional on compliance, treatment is ignorable*:

$$\{Y_{1i}, Y_{0i}\} \perp D_i \mid D_{1i} > D_{0i},$$

  so we can use κ to approximate a causal CEF, by solving:

$$(\alpha, \beta) = \arg\min_{a,b} E\left\{\kappa_i \left(Y_i - h\left[aD_i + X_i'b\right]\right)^2\right\}$$

  for any linear *or nonlinear* approx function, $h\left[\alpha D_i + X_i'\beta\right]$

- Suppose, for example,

$$h\left[\alpha D_i + X_i'\beta\right] = \Phi\left[\alpha D_i + X_i'\beta\right]$$

  This gives "best Problt approximation" to a causal CEF with endogenous treatment

Quantile Treatment Effects

- QR models conditional distributions. Assume:

$$Q_\tau\left(Y_i \mid X_i\right) = \gamma_\tau' X_i$$

  Then γ_τ solves

$$\gamma_\tau = \arg\min_c E\left\{\rho_\tau\left(Y_i - X_i' c\right)\right\}$$

  where $\rho_\tau(u) = (\tau - 1(u \leq 0))\, u$ is the "check function."
  - If the CQF is nonlinear, QR provides a regression-like minimum weighted MSE approx to it; see Angrist, Chernozhukov and Fernandez-Val, 2006)
- Kappa captures a causal *quantile treatment effec*, α_τ, in

$$Q_\tau\left(Y_i \mid X_i, D_i, D_{1i} > D_{0i}\right) = Q_\tau\left(Y_{Di} \mid X_i, D_{1i} > D_{0i}\right) = \alpha_\tau D_i + X_i'\beta_\tau,$$

  by solving:

$$(\alpha_\tau, \beta_\tau) = \arg\min_{a,b} E\left\{\kappa_i \rho_\tau\left(Y_i - aD_i - X_i' b\right)\right\}$$

- **QR 'n QTE**

Average Causal Response

Variable intensity S_i takes on values in the set $\{0, 1, \ldots, \overline{s}\}$ There are $\overline{s}$ unit causal effects, $Y_{si} - Y_{s-1,i}$.

- A linear model assumes these are the same for all s and for all i, obviously unrealistic assumptions
- Fear not! 2SLS generates a weighted average of unit causal effects
  - Suppose a single binary instrument Z_i (say a dummy for late quarter births) is used to estimate the returns to schooling
  - Let S_{1i} denote the schooling i would get if $Z_i = 1$ let; S_{0i} denote the schooling i would get $Z_i = 0$
  - We observe $S_i = S_{0i}(1 - Z_i) + Z_i S_{1i}$
- Key assumptions
  - Independence and Exclusion. $\{Y_{0i}, Y_{1i}, \ldots, Y_{\overline{s}i}; S_{0i}, S_{1i}\} \perp Z_i$
  - First Stage. $E[S_{1i} - S_{0i}] \neq 0$
  - Monotonicity. $S_{1i} - S_{0i} \geq 0 \quad \forall\, i$ (or vice versa)

The ACR Theorem

- Angrist and Imbens (1995) show

$$\frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[S_i \mid Z_i = 1\right] - E\left[S_i \mid Z_i = 0\right]} = \sum_{s=1}^{\bar{s}} \omega_s E\left[Y_{si} - Y_{s-1,i} \mid S_{1i} \geq s > S_{0i}\right]$$

where

$$\omega_s = \frac{P\left[S_{1i} \geq s > S_{0i}\right]}{\sum_{j=1}^{\bar{s}} P\left[S_{1i} \geq j > S_{01}\right]}$$

The weights, ω_s, are non-negative and sum to 1.

- The Wald estimator is a weighted average of the *unit causal response* along the length of a potentially nonlinear causal relation
- $E\left[Y_{si} - Y_{s-1,i} \mid S_{1i} \geq s > S_{0i}\right]$, is the average difference in potential outcomes for *compliers at point* s
- Here, compliers are subjects the instrument moves from treatment Intensity less than s to at least s

The ACR Weighting Function

- The size of the group of compliers at point s is

$$\begin{aligned} P\left[S_{1i} \geq s > S_{0i}\right] &= P\left[S_{1i} \geq s\right] - P\left[S_{0i} \geq s\right] \\ &= P\left[S_{0i} < s\right] - P\left[S_{1i} < s\right] \end{aligned}$$

- By Independence, this is an observed CDF difference:

$$P\left[S_{0i} < s\right] - P\left[S_{1i} < s\right] = P\left[S_i < s \mid Z_i = 0\right] - P\left[S_i < s \mid Z_i = 1\right]$$

- Finally, because the mean of a non-negative random variable is one minus the CDF,

$$\begin{aligned} & E\ \left[S_i \mid Z_i = 1\right] - E\ \left[S_i \mid Z_i = 0\right] \\ = & \sum_{j=1}^{\bar{s}} \left(P\left[S_i < j \mid Z_i = 0\right] - P\left[S_i < j \mid Z_i = 1\right]\right) = \sum_{j=1}^{\bar{s}} P\left[S_{1i} \geq j > S_{0i}\right] \end{aligned}$$

- The ACR weighting function is given by the difference in the CDFs of treatment intensity with the instrument switched off and on, normalized by the first-stage

More Variable Treatment Intensities

- Returns to schooling again, identified using compulsory attendance and child labor laws (Acemoglu and Angrist, 2000)
- Class size (Angrist and Lavy, 1999; Krueger, 1999)
  - Y_i is test score; S_i is class size
  - Z_i is Maimonides Rule (regression-discontinuity) or random assignment
- GRE test preparation (Powers and Swinton, 1984)
  - Y_i is GRE analytical score; S_i is hours of study
  - Z_i is randomly assigned letter of encouragement
- Maternal smoking (Permutt and Hebel, 1989)
  - Y_i is birthweight; S_i is mother's pre-natal smoking
  - Z_i is randomly assigned offer of anti-smoking counseling
- Quantity-quality trade-offs (Angrist, Lavy, and Schlosser, 2010)
  - Y_i is schooling, earnings, etc.; S_i is sibship size
  - Z_i is derived from twins and sibling-sex composition

So Long and Thanks for All the Fish

- Let $q_i(p)$ denote quantity demanded in market i at hypothetical price p, a continuous function
- The slope of this demand curve is $q_i'(p)$, with quantity and price measured in logs, this is an elasticity
- The Wald estimator using a $stormy_i$ instrument is

$$\frac{E[q_i \mid stormy_i = 1] - E[q_i \mid stormy_i = 0]}{E[p_i \mid stormy_i = 1] - E[p_i \mid stormy_i = 0]} = \frac{\int E[q_i'(t) \mid p_{1i} \geq t > p_{0i}] P[p_{1i} \geq t > p_{0i}] dt}{\int P[p_{1i} \geq t > p_{0i}] dt},$$

  where p_{1i} and p_{0i} are potential prices prices by $stormy_i$
- This is a weighted average derivative with weighting function $P[p_{1i} \geq t > p_{0i}] = P[p_i \leq t \mid z_i = 0] - P[p_i \leq t \mid z_i = 1]$ at price t

Continuous Special Cases

1. *Linear:* $q_i(p) = \alpha_{0j} + \alpha_{1j} p$, for random coefficients, α_{0j} and α_{1j} Then, we have,

$$\frac{E[q_i \mid stormy_i = 1] - E[q_i \mid stormy_i = 0]}{E[p_i \mid stormy_i = 1] - E[p_i \mid stormy_i = 0]} = \frac{E[\alpha_{1j}(p_{1i} - p_{0i})]}{E[p_{1i} - p_{0i}]},$$

   a weighted average of α_{1j}, with weights proportional to the price change induced by the weather in market i

2. *Additive nonlinear*

$$q_i(p) = Q(p) + \eta_i$$

   By this we mean $q_i'(p) = Q'(p)$ every day or in every market ACR becomes,

$$\int Q'(t)\,\omega(t)\,dt, \quad \text{where} \quad \omega(t) = \frac{P[p_{1i} \geq t > p_{0i}]}{\int P[p_{1i} \geq r > p_{0i}]\,dr}$$

3. Case 1 emphasizes heterogeneity, Case 2 focuses on nonlinearity
4. Y'allah, let's fish!

Summary

- The IV paradigm provides a powerful and flexible framework for causal inference:
  - An alternative to random assignment with a strong claim on internal validity (when the instruments are good)
  - A solution to the compliance problem in randomized trials (the biomed RCT world has been slow to absorb this; e.g., AIDS vaccine trials)
  - A flexible strategy for the analysis of observational designs
- Kappa weighting extends the LATE framework to nonlinear and quantile models
- IV produces weighted averages of ordered and contnuous treatment effects; the weighting function describes the range contributing to a particular estimate
- Up next: DD and RD ... these too are often IV!