

POLICY EVALUATION

1 TABLE OF CONTENTS

2	<i>Fundamentals</i>	2
2.1	Set Ups	2
2.2	Bias and Validity – “The Experimental Ideal” [See MHE Chp. 2 and bits of 3]	2
2.3	Treatment [see PDF in Readings]	4
2.4	CS: Krueger (1999): Project STAR Class Size Experiment [see MHE p29]	4
2.5	CS: Black et al. (2003) [see MHE p62]	5
3	<i>Instrumental Variables</i>	7
3.1	Basics	7
3.2	CS: Angrist and Krueger (1991)	8
3.3	IVs gone wrong	10
3.4	CS: Angrist (1990)	11
4	<i>Difference-In-Difference</i>	12
4.1	Set Up	12
4.2	CS: Card and Krueger (1994)	14
5	<i>Fixed Effects</i>	15
5.1	Set Up	15
5.2	Measurement Error and FE	15
5.3	CS: Ashenfelter & Krueger (1994)	16
6	<i>Regression Discontinuity</i>	17
6.1	Set Up	17
6.2	CS: Ludwig & Miller (2007): Head Start	18
6.3	CS: Ludwig & Miller (2007): Head Start	19

2 FUNDAMENTALS

Two things distinguish the discipline of Econometrics from our older sister field of Statistics

- One is a lack of shyness about causality
 - Statistician Paul Holland (1986) cautions that there can be “no causation without manipulation”
 - Less thoughtful observers fall back on the truism that “correlation is not causality”
- The second thing that distinguishes us from most statisticians - and indeed most other social scientists - is an arsenal of statistical tools

We will look at some of these:

- RCT (randomly allocates agents into treatment/lack-of-treatment)
- Basic Regressions (removes selection on observables)
- IV (removes selection on unobservables)
- Fixed Effects (removes constant unobservables)
- Regression Discontinuity (focuses on marginal people)

2.1 SET UPS

Regression

- Basic Regression: $Y_i = \alpha + \beta D_i + \gamma X_i + \epsilon_i$
 - D_i is treatment effect
 - β is the average distance between treatment and control
 - X_i are controls that remove selection on observables
 - Assume linearity so X_i does not affect β

Randomised Control Trials

- Specification: $Y_i = \alpha + \beta D_i + \epsilon_i$
 - Directly remove all selection issues. But internal vs. external validity
- Advantages:
 - Controls for all confounding factors as removes observable *and* unobservable selection bias
- Disadvantages:
 - Expensive as need to generate dataset through experiment and hard to implement (STAR cost \$12m and had high attrition in kindergardener cohort)
 - Moral concerns is refusing some people useful treatment
 - Inherently smaller sample and ‘unusual’ volunteers so limited external validity
 - Subject to Hawthorne and John Henry effects from participants

2.2 BIAS AND VALIDITY – “THE EXPERIMENTAL IDEAL” [SEE MHE CHP. 2 AND BITS OF 3]

2.2.1 SELECTION BIAS

We want to identify the ‘true’ causal effect of a treatment, on a given individual i , not its correlation/association

- True effect: $E[\Delta_i | D_i = 1] = E[Y_{1i} - Y_{0i} | D_i = 1] = E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1]$
- Ideally we would apply treatment and lack-of-treatment to same person/sample and observe both outcomes simultaneously. But only one at a time is possible.
 - In above only $E[Y_{1i} | D_i = 1]$ whilst $E[Y_{0i} | D_i = 1]$ is a counterfactual
 - Individual i is either $D_i = 1$ or $D_i = 0$. Cannot be both

- Instead we must compare different people who are respectively treated or not (i and k) to get the “observed effect”. However, this may be unjustified due to confounding variables [that causes people to self-select into groups]
 - Observed effect: $(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]) + (E[Y_{0i}|D_i = 1] - E[Y_{0k}|D_k = 0])$
 - First term is the true average treatment effect
 - Second term: selection bias (group differences in absence of treatment)
- The issue is that there may be selection on both observed and unobserved variables. This can be interpreted in two ways
 - Above only valid if $E[Y_{0k}|D_k = 0] = E[Y_{0i}|D_i = 1]$
 - Missing Data: do not observe counterfactual $E[Y_{0i}|D_i = 1]$ and no data to replace it
 - Selection Bias: people who are treated are different from those who are not
- We now interpret this intuition using a linear regression:
 - $Y = \text{Expected outcome in absence of treatment} + \text{treatment} + \text{error}$
 - $Y_i = E(Y_{0i}) + [Y_{1i} - Y_{0i}] + [Y_{0i} - E(Y_{0i})] = \alpha + \beta D_i + \epsilon_i$
 - Assumes treatment effect is constant and homogenous ($\Delta_j = Y_{1j} - Y_{0j} = \beta$)
 - ϵ_i is random part of Y_{0i}
 - Note the following:
 - $E(Y_i|D_i = 1) = \alpha + \beta + E(\epsilon_i|D_i = 1)$ and $E(Y_i|D_i = 0) = \alpha + E(\epsilon_i|D_i = 0)$
 - Thus $E(Y_i|D_i = 1) = \beta + E(\epsilon_i|D_i = 1) - E(\epsilon_i|D_i = 0)$
 - i.e. treatment + selection
- For a strategy to be valid we thus have the **Conditional Independence Assumption**
 - States that $E[Y_{0i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 0, X_i] = 0$
 - That is no selection/omitted-variable bias once we control for X_i
 - And even if estimate is already unbiased this helps lower standard errors
- If this does not hold we get **Omitted Variable Bias** (i.e. failing to control for X_i)
 - Suppose ‘true’ regression is $Y_i = \alpha + \Delta D_i + \beta X_i + \eta_i$
 - We run $Y_i = \alpha + \Delta D_i + \epsilon_i$ where $\epsilon_i = \beta X_i + \eta_i$
 - OVB thus $\hat{\Delta} = \frac{\text{Cov}(Y_i, D_i)}{\text{Var}(D_i)} = \Delta + \beta \frac{\text{Cov}(X_i, D_i)}{\text{Var}(D_i)} + \frac{\text{Cov}(\eta_i, D_i)}{\text{Var}(D_i)}$
 - Even in best case where $\text{Cov}(\eta_i, D_i) = 0$ so $\hat{\Delta} = \Delta + \beta \frac{\text{Cov}(X_i, D_i)}{\text{Var}(D_i)}$
 - Magnitude depends on β [i.e. $\text{Cov}(X_i, Y_i)$] and $\text{Cov}(X_i, D_i)$
 - Direction depends on potential sign of bias, which requires intuition
- This has been used to explain conflicting evidence for government subsidized job-training programs:
 - Non-experimental comparisons of participants and non-participants often show that after training, the trainees earn less than plausible comparison groups (Ashenfelter, 1978; Ashenfelter and Card, 1985; Lalonde 1995)
 - RCTs of training programs generate show positive effects (Lalonde, 1986; Orr et al., 1996). Makes sense that those who naturally seek help need it more to begin with

2.2.2 ASIDE: BAD CONTROLS [SEE MHE 3.2.3]

OVB is typically interpreted as highlighting danger from “lack of controls”. But can also have bad controls if these are themselves outcome variables in the notional experiment

- E.g. effect of education on outcome controlling for white collar job

Likewise proxy-controls are also bad that might partially control for omitted factors, but are themselves affected by the variable of interest

- E.g. effect of education on outcome using a post-school measure of intelligence as a proxy for IQ

2.2.3 VALIDITY

Internal validity

- Ability to get an unbiased estimate of the effect of interest

External validity

- Extrapolation/Generalisation: impact of treatment on the group studied not informative about other designs
- Randomisation/Bias: Individuals willing to participate in trial are inherently not representative
- General Equilibrium effects: As experiment rolls out may get new effects (e.g. demand for teachers itself impacts teacher quality)

2.3 TREATMENT [SEE [PDF](#) IN READINGS]

- Suppose have assigned control and treatment group. Some in treatment group may comply, other may not. There are thus two estimates to consider:
 - Z_i is assigned treatment (random) and D_i is compliance (not necessarily random)
- Intention-To-Treat (ITT): Out of people assigned control, how did treatment change outcome?
 - $\beta_{ITT} = E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$
 - This does not suffer from selection bias per se, since I am measuring the effect of “assignment”, which is random (and people cannot change)
- Treatment-On-Treated (ToT): Out of people assigned control *and* who complied, how did treatment change outcome?
 - $\beta_{ToT} = \frac{\tilde{\beta}_{ITT}}{\Pr(D_i=1|Z_i=1)} = \frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{E(D_i|Z_i=1) - E(D_i|Z_i=0)}$
 - This can suffer from selection bias, since people who do not comply may be due to a omitted variable
 - If compliance is random, then above formula is an IV!
 - Relevance: assignment and attendance are correlated if some compliers exist
 - Exclusion: Not correlated to error term as Z_i is randomly allocated
- Both estimates might of interest. Typically ITT is important for policy (since cannot force citizens to comply to programmes) and ToT for research (only care about compliance)
- Note also, that “true” treatment effect β_{True} may be heterogenous (see LATE and ATE)

2.4 CS: KRUEGER (1999): PROJECT STAR CLASS SIZE EXPERIMENT [SEE MHE P29]

See also Angrist & Lavvy on same question in RD-Fuzzy

Demonstrates advantages (and disadvantages) of RCT in education context

Context

Question: What is the effect of reducing class sizes on student achievement?

- Has implications for improving the “education production function” as class size is one of the most expensive inputs and thus want to know if it is justified

Previous literature estimated $Achievement_i = \beta_0 + \beta_1 Small_i + \beta_3 X_i + \epsilon_i$ and found conflicting results due to bias:

- Allocation to classes is not random as may have negative (place struggling students into small classes) or positive (place best students) selection bias
 - Boozer and Rouse (2001): Schools have compensatory allocation of resources where more is allocated to those who need it

- Cross-state results (positive) differ from within state results (insignificant)

Methodology

Tennessee STAR class size experiment saw 11,600 kindergarteners from 80 schools placed “randomly” into small (~15) or regular classes (22-25) through third grade

- Unusually ambitious study, costing \$12m
- Plausibly random as Krueger finds no differences between treatment and control in key variables (e.g. free school lunch), but do not have pre-treatment scores ☹
- Randomization occurred within schools not across to account for “group effects”
- Teachers also randomly assigned

Run regression $Achievement_{ics} = \beta_0 + \beta_1 Small_{cs} + \beta_2 \frac{REG}{Acs} + \beta_3 X_{ics} + \alpha_s + \epsilon_{ics}$

- Treatment effects are β_1 (small class) and β_2 (regular class with an aide)
- Control for student characteristics X (should not matter in RCT anyway)
- Control for school fixed effects as randomization occurs within school
- Control for correlation among residuals in class (e.g. unobserved teacher quality)

Results

Find that $\beta_1 = 5.37$ (i.e. >5% and significant) and $\beta_2 = 0.31$ (i.e. 0.3% and insignificant)

Criticisms

Limited internal validity due to non-compliance. Only half of students present in kindergarten present for full K-3 and class size was not stable (small 11-20; large 15-30)

- Does $\beta_{TOT} = \beta_{True}$? After first period, compliance is plausibly not random
 - Pupils who repeated or skipped a grade left the experiment (thus sample left improves $\beta_{TOT} > \beta_{True}$)
 - Rich more likely to enrol kids into private schools if assigned to regular (thus sample left worsens $\beta_{TOT} < \beta_{True}$)
- Limited external validity due to general equilibrium effects as per Jaspen & Rivkin (2009):
 - California wide small-class size reforms created demand for 25,000 new teachers
 - Hence (1) influx of lower-quality teachers and (2) reassortment of experienced teachers to newly created positions in affluent communities
 - Benefit of class reduction by ten pupils can be offset by a teacher with one (as opposed to two) years of experience

Another RCT example is The Education Maintenance Allowance, which pays British high school students in certain areas to attend school, is one such policy experiment (Dearden, et al, 2004).

2.5 CS: BLACK ET AL. (2003) [SEE MHE P62]

Demonstrates clever use of natural experiment (for CIA to hold) and matching

Context

Question: Does job search requirement help long-term unemployed get back to work?

- Long-term unemployment is costly for government and individuals so this may be a good solution for society (mention SCBA?)
- Title of paper is: Is the Threat of Reemployment Services More Effective Than the Services Themselves? As also look at when improvement happen

For literature review see Fredriksson and Holmlund (2006) Section 5.2

Methodology

WPRS introduced in 1993 and forced claimants with long UI/benefits spells to receive employment and training services

- WPRS assigned profiling scores 1-20 based on probability of benefit exhaustion (using data on past-employment/characteristics)

- WPRS wants to only give UI if people undertake re-employment services.
- But states/regions have limited capacity and allocate according to highest profiling scores (rest get UI regardless)
- Within the marginal profiling score, requirement to undertake re-employment services is thus randomly assigned. Hence natural experiment!
 - E.g. have 10 slots left to give out but 20 people with “marginal” profile score

Variation comes from 256 Profiling Tie Groups (PTG): group with a certain profiling score (difference priority, ranging 6-19) in a certain office/region (different capacity)

- In each PTG some have to participate in reemployment services (treatment)
- And some get it regardless (control)

Run regression $y_{ij} = \mu_j + \beta T_{ij} + v_{ij}$ where y_{ij} is outcome of interest, μ_j is fixed characteristics of i 's PTG j , and T_{ij} the treatment

Results

Reduced mean weeks of UI benefit receipt by 2.2 weeks, reduced amount claimed by \$143, and increases subsequent earning by \$1,000

- Treatment group has higher earnings in first two quarters after filing UI claims but none after. Suggests primarily benefit from earlier return to work

Criticisms

- But... measure of earning is not ideal as no earning records (1) outside Kentucky, especially relevant for commuting workers, and (2) informal jobs. Thus understate earnings
- Alternatively does not consider general equilibrium effects: increase in participation in formal sector reduces participation in informal sector. Thus overstate earnings

Robustness Check: Matching

THEORY [MHE 3.3 is on this but seems off syllabus]

- An alternative empirical strategy is matching, whereby $\hat{Y}_{0i} = \sum_{j \in C(i)} w_{ij} Y_{0j}^{obs}$
 - i 's outcome (who is treated) is compared j 's outcome (who is not treated), with the difference being an estimate of the treatment effect
 - This estimate is weighted w_{ij} based on how similar i and j are based on a set of observable characteristics X_i
 - This is done for all people in the control group
- To do this we need two assumptions:
 - Conditional Independence Assumption: Treatment is purely random conditioning on X_i i.e. $E[Y_{0i}|T_i = 1, X_i] = E[Y_{0i}|X_i]$
 - Intuitively, i 's counterfactual outcome $E[Y_{0i}|T_i = 1, X_i]$ is replaced with j 's real outcome $E[\hat{Y}_{0i}|X_i]$. This may not be valid, same as OLS
 - Common Support Condition: $0 < p_i = E[D_i = 1|X_i] < 1$
 - Intuitively, the assignment of treatment based on X_i must not be certain. Else for i (treated) there is no valid j (control) to compare to
 - These are called propensity scores by (Rosenbaum & Rubin, 1983)

APPLYING TO PAPER

- OLS estimator is inconsistent if impact of treatment varies by PTG (i.e. does not allow for arbitrary heterogeneity in causal effects $E[Y_{1i} - Y_{0i}|X_i]$)
- OLS is also a parametric technique that requires specification (e.g. functional form so x contributes linearly to y)
 - “In a world where the impacts do not vary among PTGs, the matching estimator remains consistent, but is inefficient relative to estimating equation by OLS.”

- By contrast matching treats each PTG as a separate experiment. These can then be aggregated to get an “average” treatment
 - In matching: $\beta = \sum_j \frac{p_j N_j}{\sum_k^{286} p_k N_k} \hat{\Delta}_j$
 - Where $\hat{\Delta}_j$ is the matching estimate for each PTG j
 - p_j is the probability of member of j receiving treatment and N_j the number of members in j . Thus each $\hat{\Delta}_j$ is weighted by its share of the treatment sample
 - Can be shown that for OLS $\beta = \sum_j \frac{(1-p_j)p_j N_j}{\sum_k^{286} (1-p_k)p_k N_k} \hat{\Delta}_j$ so more weight the closer to random experiment (large sample and random $p_j = 0.5$ assignment)
- Matching estimates “tell the same substantive story” (so results are robust) but matching reveals there is much heterogeneity based on profiling score, so good that authors check
 - E.g. annual earnings effect for score 16 is \$4,176 but 14-15 is -\$1,257 (see last slide of 14 for full table) [what matching does is only compare people 16 to other similar]
 - But... both OLS and matching still assume CIA

3 INSTRUMENTAL VARIABLES

First laid out by Wright (1928) and coined by Reiersol (1941)

- Good instruments come from institutional knowledge and your ideas about the processes determining the variable of interest

3.1 BASICS

Consider OVB situation: $y_i = \alpha_0 + \Delta D_i + \beta A_i + \eta_i$ where $\epsilon_i = \beta A_i + \eta_i$

Instrument Z_i is valid if

- Relevance Condition: it is a variable that predicts D_i so $cov(D_i, Z_i) \neq 0$
- Exclusion Restriction: it is uncorrelated with error ϵ_i so $cov(Z_i, \epsilon_i) = 0$

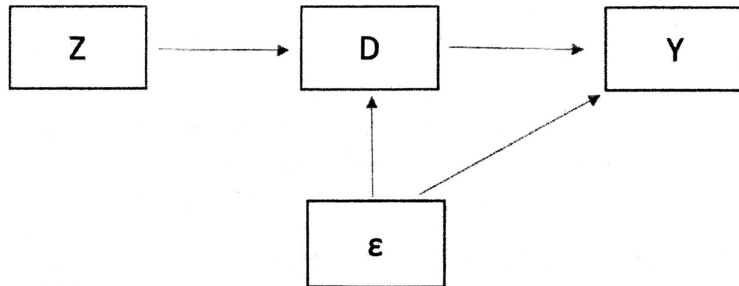
Thus note the following derivation

- $cov(y_i, Z_i) = cov(\alpha_0 + \Delta D_i + \epsilon_i, Z_i) = cov(\alpha_0, Z_i) + cov(\Delta D_i, Z_i) + cov(\epsilon_i, Z_i) = \Delta cov(D_i, Z_i) + cov(\epsilon_i, Z_i) = \Delta cov(D_i, Z_i)$ if valid
- $\hat{\Delta}_{IV} = \frac{cov(y_i, Z_i)}{cov(D_i, Z_i)} = \frac{cov(Y_i, Z_i)/var(Z_i)}{cov(D_i, Z_i)/var(Z_i)} = \frac{\text{regress } Y \text{ on } Z}{\text{regress } D \text{ on } Z}$

This is consistent: $plim \hat{\Delta}_{IV} = \beta + \frac{cov(Z_i, u)}{cov(D_i, Z_i)} = \beta$

The Intuition of IV Estimator

- A variable Z that is associated with D but not ε
- Z and y is correlated, but the only source of such correlation is the indirect path of z being correlated with D
- Therefore, An instrument Z that has the property that changes in Z are associated with changes in D but do not led to change in y (aside from the indirect route via D)



- When IV is Binary get Wald Estimator: $\hat{\Delta}_{IV} = \frac{cov(Y_i, Z_i)}{cov(D_i, Z_i)} = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)}$ [same as ToT!]
- Also eliminates measurement error: If x is noisy its effect on y goes to zero but IV cuts through this always as instrument correlates with true x not noise

2SLS

Can interpret as estimating using two stage least squares

- $D_i = \gamma + \beta Z_i + v_i$ so first stage $\hat{D}_i = \hat{\gamma} + \hat{\beta} Z_i$
 - $\hat{\beta} = \frac{cov(Z_i, D_i)}{var(Z_i)}$
- $Y_i = \alpha + \Delta D_i + \epsilon_i$ so second stage $Y_i = \alpha + \Delta \hat{D}_i + \epsilon_i$
 - $\hat{\Delta}_{2SLS} = \frac{cov(Y_i, \hat{D}_i)}{var(\hat{D}_i)} = \frac{cov(Y_i, \hat{\gamma} + \hat{\beta} Z_i)}{var(\hat{\gamma} + \hat{\beta} Z_i)} = \frac{\hat{\beta} cov(Y_i, Z_i)}{\hat{\beta}^2 var(Z_i)} = \frac{cov(Y_i, Z_i)}{var(D_i, Z_i)}$

3.2 CS: ANGRIST AND KRUEGER (1991)

Context

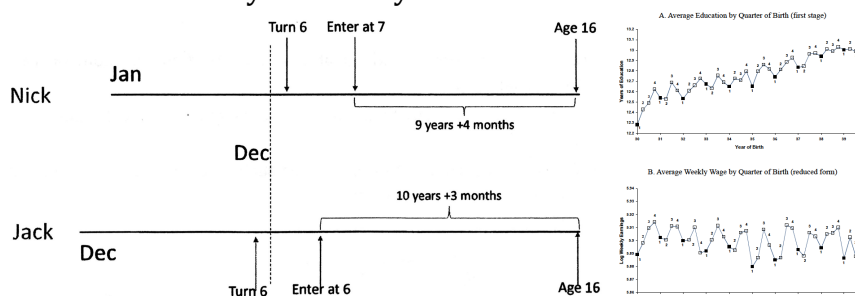
Question: Do years of education matter or is it only signalling? [RELEVANT FOR EDU!!]

For literature review see basic Zheng (2010) and advanced Card (1999) section 4 [but old]

Methodology

Examine 300,000+ men born 1930-39. Use two features of US schooling and hence quarter of birth matters

- Compulsory schooling laws is in terms of age, not years completed
- School entry is once a year and cut-offs are based on birthdays



Age is important to control for as those born earlier in the year may have higher earnings because of the upward sloping portion of the age-earnings profile (or retirement)

- Thus focus on 40-49 year-old men, where wages are hardly related to age
- Set up Wald Estimator: $\Delta = \frac{cov(\log Y, Q_1)}{cov(D, Q_1)} = \frac{E(\log Y | Q_1=1) - E(\log Y | Q_1=0)}{E(D | Q_1=1) - E(D | Q_1=0)}$

Also set up 2SLS:

- First stage: regress $E_i = \gamma_t + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \epsilon_i$. Find
 - E_i is education years of individual i
 - γ_t is year fixed effects (i.e. control trend). This is important to control for because of increasing trend of education years
 - Q is birth quarter. Find this does impact (see relevance condition)
- Second stage: regress $Y_i = \alpha_t + \Delta_i \hat{E}_i + \delta X_i + \epsilon_i$
 - X_i is vector of controls e.g. race, centre city, married etc.
- Differs slightly from the Wald estimates: Include covariates and identified by variation in education across each quarter in each year (not just 1st and rest)
 - Using all four quarters as instruments means overidentification so can perform Hausman to test exogeneity

Results

- The estimated monetary return to an additional year of schooling for those who are compelled to attend by law is 7.5% (by OLS and 10.2% for Wald)
- Wald is hardly any different to OLS and 2SLS is 30% larger!!
 - “Economists have devoted a great deal of attention to correcting for bias in the return to education due to omitted ability [...]. This type of a bias would occur, for example, in Spence's [1973] signalling model [...] In contrast to this prediction, estimates based on season of birth indicate that, if anything, conventional OLS estimates are biased slightly downward.”

VALIDITY & CRITICISMS

- If the fraction of students who want to drop out prior to the legal dropout age is independent of season of birth, then the observed seasonal pattern in education is consistent with the view that compulsory schooling constrains some students born later in the year to stay in school longer
- **Author justification for Exclusion:**
 - Relationship between quarter of birth and educational attainment is weaker for more recent cohorts that are less likely to have been constrained by the law
 - Relationship between quarter of birth and education is weaker for better- educated individuals (none for high-school, barely any for college)
 - Test same for subsample of college graduates (who are not constrained by compulsory schooling and thus natural control) and find effect disappears
 - Relationship between quarter of birth and educational attainment varies across states, depending on when each state requires children to start school
 - Test with Hausman (see above). But, this only shows combinations of instruments give same estimate, not that it is necessarily correct...
- **Criticisms of Exclusion:** But is it really the only one?
 - Age at school entry can affect academic performance (more mature so perform better)
 - Bedard and Dhuey (2006) find youngest score 4-12 percentiles lower than oldest across OECD. “Summer babies” penalty in British context
 - Also Carroll (1992) and Mortimore et al. (1988)
 - Still... Elder and Lubotsky (2008) argue gap is only in first months of kindergarten and sharply declines. Thus reflects prior skill accumulation

- Age at school entry can be produce of parents socioeconomic status
 - AR cite Lam and Miron (1987) who showed season of birth is unrelated to the socioeconomic status but...
 - Bound, Jaeger, and Baker (1995) criticise A&R by showing difference in father earnings by quarters can alone explain 1/3 of association between quarter of birth and educational attainment
 - Hungerman and Buckles (2013): winter births more common by unmarried women with no high school degree as others plan away
- Could also be associated with mental/physical health due to shocks (e.g. prenatal exposure to 1957 A2 influenza epidemic and schizophrenia, O'Callaghan, 1991)
- **Author Justification for Relevance:**
 - Angrist and Krueger [1990] formally model link between age at school entry and compulsory schooling. Agents want to drop out but are stopped by law
 - Direct: Average number of completed years one tenth of a year lower for men born in first quarter vs last
 - Indirect: In 1960s states where 16yo could drop out, enrolment is 4.5%-points lower boundary versus 0.6%-points in other states. Total dropout rate is 12% so implies 1/3 of potential dropouts are stopped by compulsory laws (lower for 1980s as total rate falls to 5%)
- **Criticisms of Relevance:** Bound, Jaeger, and Baker (1995)
 - BJB: Note correlation is very week (R^2 .0001-.0002) which authors say doesn't matter because sample is large (300,000).
 - But if exogeneity fails IV-equation shows lead to large inconsistencies (even if violation is small)
 - "Finite-sample biases" in IV are same direction as OLS and magnitude approaches OLS as R^2 falls. This explains small difference between the two estimates
 - OLS is unbiased and consistent, 2SLS only consistent so "finite-sample bias"
 - But defended by... Cruz and Moreira (2005): Apply the conditional method of Moreira (2003) to AK specification with 180-instrument (quarters# states) and controls for age[-squared]
 - Returns-to-schooling 8-25% (4-14%) in 1970 (1980). Less accurate than AK initially suggest (6-9.9% for 1930-39 cohort) but similar story
- **Criticisms: External Validity**
 - Does this estimate of education RoR apply to all years?
 - Card and Krueger (1992) find little evidence of nonlinearity ☺
 - But others disagree
 - Rosenbaum (2005): Schooling years affected by compulsory years is a marginal disadvantage group. For general population date of birth should not matter
- Other points:
 - Are all dropouts counted? Can dropout before legal age if have mental/physical health issues, live far from a school, disruptive to other students, already got degree, attend private/home school
- Other Education papers:
 - Welch (1973) find return to education are lower for black men than for white men
 - Return to education surveys by Griliches [1977] and Willis [1986]

3.3 IVs GONE WRONG

3.3.1 BAD INSTRUMENTS

Exclusion Restriction is Violated

When exclusion restriction is violated $cov(Z_i, \epsilon_i) \neq 0$ IV estimator becomes inconsistent

- When Z_i and ϵ_i are strongly correlated (i.e. $|cov(Z_i, \epsilon_i)|$ is large), the bias of IV estimator can be much greater than OLS (i.e. $E[|\hat{\Delta}_{IV} - \Delta|] \gg E[|\hat{\Delta}_{OLS} - \Delta|]$)

Weak Instruments

Problems occur when $cov(D_i, Z_i)$ is small (i.e. $cov(D_i, Z_i) \sim 0$)

- Relevance condition is problematic or IV does not induce much variance

Creates bias either way but still consistent in any finite sample

One solution is to use multiple instruments (where rule of thumb $F\text{-Stat} > 10$)

3.3.2 HETEROGENOUS EFFECTS

- In an IV framework, driving causal inference is instrument, Z_i , but the variable of interest is D_i . Are these equivalent? No! Not all Z_i may comply...
 - IV does not capture effect on always/never takers (except if special) (see p134, 142)
- What we get is the Local Average Treatment Effect (LATE): Average treatment effect on those whose status has changed because of the instrument i.e. compliers
 - That is $LATE = Y_{D_i, Z=1} - Y_{D_i, Z=0}$ and equivalent to a binary IV!
 - OLS captures average treatment effect on treated (ATE). IV captures local average treatment effect where local is those who “comply” with the IV (LATE)
 - i.e. IV measures effect only for those people’s who’s behaviour changed by nature of the instrument
- Is small compliant subpopulation worrying? Depends on context...
 - For many policy interventions we care about a marginal group specifically
 - E.g. Raising compulsory attendance to 18 is designed to help dropouts
 - Veteran status: non-draft IV needed for effect on volunteers. (Angrist, 1998)
 - Acemoglu and Angrist (2000) argue that quarter-of-birth instruments and state compulsory attendance laws essentially affect same people. Thus two IVs give similar estimates

3.4 CS: ANGRIST (1990)

Context

- Question: Does veteran status hurt or benefit civilian earnings?
- Existing literature is conflicted:
 - Social security admin records indicate that long after Vietnam war (1980s) white veterans earned 15% less. This is concerning
 - Berger and Hirsh (1983) found no effect on 1977 weekly earnings
 - Crane and Wise (1987) found 11% reduction in 1979 weekly earnings
- Want to allow for some men benefits of military service whilst it hurt others. Use fact that some volunteered and others didn’t...

Methodology

- During Vietnam war there was national draft lottery system with changing ceilings. This was completely random and thus provides natural experiment (relevance condition!)
- Set up 2SLS
 - First Stage: Effect of draft eligibility on veteran status in 1950-52 was 0.1-0.16
 - RELEVANCE: Veteran status was not completely determined by draft, but those men eligible were $\sim 16\%$ -points more likely to serve
 - Second Stage: Since treatment is a dummy use Wald Estimator

Result

- White veterans suffer annual earnings loss of ~\$3,500 or 15% then average salary
 - Military experience is only a partial substitute for civilian labour market experience lost while in armed forces

Validity: Exclusion?

- Draft is clearly random, so Z_i is. But is D_i ?
- Rely that the only reason for draft-eligibility affects on earnings is veteran status:
 - Treatment effect on 1969 earnings is zero (predates 1970 draft lottery)
 - Treatment effect for 1953 cohort is zero (never called to serve)

Validity: LATE? Angrist, Imbens and Rubin (1996)

- Can define male population into four groups. Population whose treatment[veteran] status can be manipulated by IV[draft] can be seen as the set of compliers
 - Compliers: $D_{i,z=1} = 1$ and $D_{i,z=0} = 0$ (i.e. causal effect)
 - Always Takers: $D_{i,z=1} = D_{i,z=0} = 1$
 - Never Takers: $D_{i,z=1} = D_{i,z=0} = 0$
 - Defiers: $D_{i,z=1} = 0$ and $D_{i,z=0} = 1$ (i.e. opposite treatment effect by instrument)
- Recall $LATE = Y_{D_{i,z=1}} - Y_{D_{i,z=0}}$
 - [Compliers + always takers + never takers + defiers] who are drafted – [Compliers + always takers + never takers + defiers] who are never drafted
- Assume that (1) Instrument is randomized and (2) Instrument no effect on outcome once fix value of treatment. Thus never-takers and always-takers cancel out
 - Formally, (2) is (i.e. $E[Y_{1i,z}] = E[Y_{1i}], E[Y_{0i,z}] = E[Y_{0i}]$)
 - Thus [Compliers who are drafted] – [Compliers who are not drafted]. LATE!
- Assume that (3) no defiers – Unlikely that people join army because not drafted
 - Formally, the monotonicity assumption (Imbens and Angrist, 1994): while IV may have no effect on some, all of those who are affected are affected in the same way
- This simplifies down to $LATE = [\text{Compliers who are drafted}] - [\text{Compliers who are not}]$

Validity: Causal Interpretation?

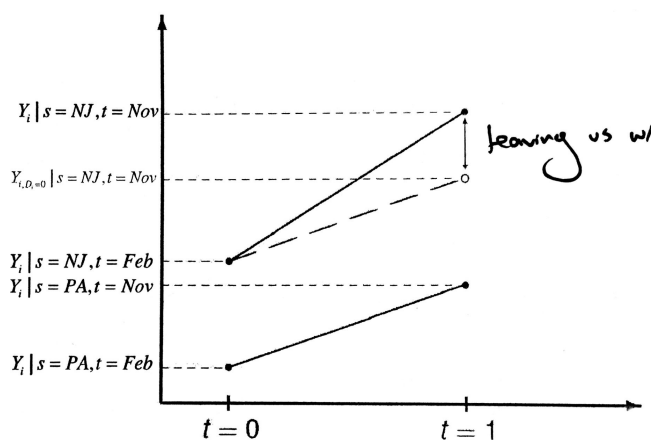
- Estimate finds that there is a earnings penalty for people being “forced” to be Vietnam soldiers. But why is this? IV cannot explain this, need a theory
 - Angrist (1990) interprets draft lottery estimates as penalty for lost labour market experience (so should have external validity, e.g. WWII draftees AK, 1994)
 - Angrist and Krueger (1992) looked for an association between draft lottery numbers and schooling (as stay in college longer to avoid service) but find increase in college attendance is small

4 DIFFERENCE-IN-DIFFERENCE

4.1 SET UP

- First Difference takes out state fixed effect to “identify” time trend
 - $E(Y_{ist}|s = PA, t = Nov) - E(Y_{ist}|s = PA, t = Feb) = \mu_{Nov} - \mu_{Feb}$ (1)
 - $E(Y_{ist}|s = NJ, t = Nov) - E(Y_{ist}|s = NJ, t = Feb) = \mu_{Nov} - \mu_{Feb} + \Delta$ (2)
- Second Difference takes away time fixed effect
 - $(2) - (1) = \Delta$

	Feb (before)	Nov (after)	Diff. (Nov-Feb)
Control (PA)	α_0	$\alpha_0 + \alpha_2$	α_2
Treatment (NJ)	$\alpha_0 + \alpha_1$	$\alpha_0 + \alpha_1 + \alpha_2 + \Delta$	$\alpha_2 + \Delta$
Diff (NJ-PA)	α_1	$\alpha_1 + \Delta$	Δ



- Requires Common Trend Assumption: Satisfied if unobserved confounding is time-invariant and additive (i.e. can compare the outcomes before the treatment happens)
 - $E(Y_{i,D_i=0} | s = NJ, t = Nov) - E(Y_i | s = NJ, t = Feb) = E(Y_i | s = PA, t = Nov) - E(Y_i | s = PA, t = Feb)$

Issues and how to resolve these

- Self-selection and targeting
 - Self-selection: participants in worker training programs experience a decrease in earnings before they enter the program
 - Targeting: policies may be targeted at units that are currently performing the worst
- Compositional differences across time
 - May confound any DD estimate since “effect” may be attributable to change in population between periods (e.g. workers can commute)
- Long-term effects versus reliability
 - Parallel trends assumption most likely to hold over a short time period
 - In long run many things confound treatment effect
- Functional form dependence
 - Magnitude/sign of DD effect may be sensitive to functional form when average outcomes for controls and treated are different at baseline
 - Employment young increases from 20 to 30%
 - Employment old increases from 5 to 10%
 - DD is positive: $(30 - 20) - (10 - 5) = 5 > 0$
 - Log changes is negative: $[\log(30) - \log(20)] - [\log(1.5) - \log(2)] < 0$
 - CTA can be applied to transformed data: but... if there is a common trend in logs, there will not be one in levels and vice versa (e.g. Athey and Imbens, 2006)

Triple Differences

- “Placebo DD” makes parallel trends more plausible (e.g. min wage affects only low-wage)
 - Eliminates time-varying confounding that are common for same-state workers but different for different-state workers
- Formally, $[E(Y_i | NJ, Nov, low) - E(Y_i | NJ, Feb, low)] - [E(Y_i | PA, Nov, low) - E(Y_i | PA, Feb, low)] - [E(Y_i | NJ, Nov, high) - E(Y_i | NJ, Feb, high)]$

- Note three terms not four! – go over this!
- Or $Y_{istG} = \alpha_0 + \alpha_1 NJ_s + \alpha_2 Nov_t + \alpha_3 G_i + \alpha_4 NJ_s * G_i + \alpha_5 Nov_t * G_i + \alpha_6 Nov_t * NJ_s + \Delta NJ_s * Nov_t * G_i + \epsilon_{istG}$

4.2 CS: CARD AND KRUEGER (1994)

Context

- Question: What is the effect of a minimum wage rise on employment?
- For literature review see Neumark et al (2014) intro
 - Minimum wage reduces employment if competitive industry (as moves above market clearing rate) but... if monopsonist firm it can be beneficial

Methodology

- In 1992 New Jersey raised it from \$4.25 to \$5.05 but neighbouring Pennsylvania did not. Thus looking at fast-food restaurants (i.e. ones affected) creates natural experiment
 - Use observation in state NJ before policy change as control for unobservable state characteristics; use other state as control for unobservable time effects
- Regress $Y_{ist} = \gamma_s + \mu_t + \Delta D_{st} + \epsilon_{ist}$ and observe $E[Y_{ist}] = \gamma_s + \mu_t + \Delta D_{st}$
 - Y_{ist} : Employment at (fast-food) restaurant i in state s at time t
 - γ_s : State fixed effects
 - μ_t : Time fixed effects
 - Δ : Policy effect of minimum wage
- See equations and graph in Set Up above

Result

- The relative gain of NJ relative to Pennsylvania (i.e. "difference in differences" of the changes in employment) is 2.76 FTE employees or 13%
- Within NJ employment expanded at low-wage stores and contracted at high wage stores
- After tax meal prices rose 3.2% faster in NJ than Pennsylvania

Criticisms

- Requires two assumptions: effect of time is the same across different states; difference between states does not change over time
- Machine replacement?

Criticisms: Common Trends Assumption

- Card and Krueger (2000) obtained administrative payroll data for restaurants in New Jersey and Pennsylvania for a number of years:
 - Fairly substantial year-to-year employment variation in other periods ☹
 - Similar at the end of 1991 but for next 3y P fell more than NJ (especially in the 14-county group), critically before 1996 change in Federal minimum
 - P thus not good measure of counterfactual employment rates in New Jersey in the absence of a policy change, and vice versa.

Criticisms: Implicit treatment and control groups changes as a result of treatment.

- Moffitt (1992): poor people who would in any case have weak labour force attachment might move to states with more generous welfare benefits. [Likewise min. wage]
- Migration problems can usually be fixed if we know where an individual starts out (i.e. use birthplace/previous-residence as IV for current location)

Alternative: regression-DD

- (1) Makes it easy to add additional control states and pre-periods and (2) facilitates work with regressions rather than on/off-dummy (min wage variable with differing treatment intensity)
- Also, when sample has many years can test Granger causality : conditional on state and year effects, past Dst predicts Yist while future Dst does not.
- Card (1992) exploits regional variation in the impact of the federal minimum wage before and after 1989 and 1992 across all 51 states
 - Wages increased more in states where min. wage increase has more 'bite'. Verifies notion that fraction affected variable is valid to test federal minimum effect.
 - Addition of an adult employment control (although have reason to believe this is a bad control) has little effect on Card's estimates

5 FIXED EFFECTS

5.1 SET UP

- Essentially extend difference-in-difference using repeated cross-sectional data
- $Y_{it} = \alpha + \beta X_{it} + \Delta D_{it} + \gamma A_i + \epsilon_{it}$
 - γA_i is time-invariant unobservable
 - Assume A_i has linear (additive) functional form and is time-invariant
- Can take difference over time: $Y_{it} - Y_{it-1} = \beta(X_{it} - X_{it-1}) + \Delta(D_{it} - D_{it-1}) + (\epsilon_{it} - \epsilon_{it-1})$
- Can take difference over sample: $Y_{it} - Y_{jt} = \beta(X_{it} - X_{jt}) + \Delta(D_{it} - D_{jt}) + (\epsilon_{it} - \epsilon_{jt})$

5.2 MEASUREMENT ERROR AND FE

- Classical measurement error causes attenuation bias in OLS regression but even worse under FE/DiD
 - E.g. In Freeman (1985) union status may be misreported/coded for few workers but can imply observed year-to-year changes in union status are mostly noise
- A variant interpretation: FE typically remove both good and bad variation. This means whatever errors are left no matter more ("throw the baby out with the bathwater")

5.2.1 SET UP

OLS

- Let true $Y_i = \beta X_i^* + \epsilon$ but observe $X_i = X_i^* - \zeta_i$ thus regress $Y_i = \beta X_i^* + \epsilon - \beta \zeta_i$
- Then estimator $\hat{\beta}_{OLS} = \frac{cov(Y_i, X_i)}{var(X_i)} = \frac{cov(\beta X_i^* + \epsilon, X_i^* - \zeta_i)}{var(X_i^* - \zeta_i)} = \beta \frac{var(X_i^*)}{var(X_i^*) + var(\zeta_i)}$
 - Magnitude of attenuation bias decreases as $var(\zeta_i)$ increases
 - Issue that X_i^* and thus $var(X_i^*)$ is unknown

Difference-In-Difference

- Let true: $\Delta Y_i = \beta \Delta Z_i^* + \Delta \epsilon$ but observe $Z_{ti} = Z_{ti}^* - \zeta_{ti}$ $\Delta Y_i = \beta \Delta Z_i^* + \Delta \epsilon + \beta \Delta \zeta_i$
- Then estimator $\hat{\beta}_{FD} = \beta \frac{var(\Delta Z_i^*)}{var(\Delta Z_i^*) + var(\Delta \zeta_i)}$

5.2.2 WORSE IN DiD

Assume observations are drawn from same distribution [$var(Z_{1i}^*) = var(Z_{2i}^*)$] and measurement errors are classical so uncorrelated [$\rho_\zeta = 0$]

Need $var(\Delta Z_i^*)$ and $var(\Delta \zeta_i^*)$ to know extent of attenuation bias. But can rewrite:

- $var(\Delta Z_i^*) = var(Z_{1i}^*) + var(Z_{2i}^*) - 2cov(Z_{1i}^*, Z_{2i}^*) = 2var(Z_{1i}^*) - 2\rho_Z var(Z_{1i}^*, Z_{2i}^*) = 2(1 - \rho_Z)var(Z_{1i}^*)$
 - Likewise derive $var(\Delta \zeta_i^*) = 2(1 - \rho_\zeta)var(\zeta_{1i})$
 - Measurement error is classical so $\rho_\zeta = 0$
- We can show that attenuation bias is more severe under first differences
 - $\hat{\beta}_{OLS} = \beta \frac{var(Z_{1i}^*)}{var(Z_{1i}^*) + var(\zeta_{1i})}$
 - $\hat{\beta}_{FD} = \beta \frac{var(\Delta Z_{1i}^*)}{var(\Delta Z_{1i}^*) + var(\Delta \zeta_{1i})} = \beta \frac{2var(Z_{1i}^*)}{var(Z_{1i}^*) + var(\zeta_{1i}) \frac{(1 - \rho_\zeta)}{(1 - \rho_Z)}}$
- Thus if $\rho_\epsilon < \rho_Z$ then $|\hat{\beta}_{FD}| < |\hat{\beta}_{OLS}|$. Indeed likely that correlation between twin education (high) is larger than correlation of measurement error (low)
- See 4.3 CS: ASHENFELTER & KRUEGER (1994) for IV solution

5.3 CS: ASHENFELTER & KRUEGER (1994)

Context

- What are the returns to education? See 2.2 Angrist and Krueger (1991)
- See also Ashenfelter & Rouse (1998) who use same methodology of samples of twins, controlling for family FE and find with-family estimates (9%) come out larger than OLS
 - Finding bit of heterogeneity where estimated returns appear to be slightly higher for less able individuals and thus cross-section is upwards biased

Methodology

- Data on twins to estimate returns on schooling
- General model $Y_i = \alpha X_i + \beta Z_i + A_i + \epsilon_i$
 - Where Y_i is wage; X_i set of observed individual fixed effects; Z_i years of schooling; A_i unobserved component that is time invariant

Endogeneity Approach

- Note many unobserved effects are same for twins e.g. genes, family support. Running naïve regression results in OVB because have to omit A_i
- Formally, see this as follows:
 - For twin 1: $Y_{1i} = \alpha X_i + \beta Z_1 + A_i + \epsilon_{1i}$
 - For twin 2: $Y_{2i} = \alpha X_i + \beta Z_1 + A_i + \epsilon_{2i}$
 - Thus $\Delta Y_i = \beta \Delta Z_i + \Delta \epsilon_i$ and we eliminate source of endogeneity A_i
- Key is that A_i is the same for both twins but... cannot eliminate all e.g. sibling rivalry

Measurement Error Approach

- Schooling is often self-reported and may hence cause measurement error
- Bound and Solon (1999) point out that there are small differences between twins, with first-borns typically having higher birth weight and higher IQ scores. While these are not large, neither is the difference in their schooling.
 - Small amount of unobserved ability differences among twins could be responsible for substantial bias in the resulting estimates.
- Thus ask twins to report on their siblings to construct instruments. Presumably less/no incentive to miss report and thus sensible: $Z_1^2 = Z_{1i}^* + \zeta_1^2$ instruments for $Z_1^1 = Z_{1i}^* + \zeta_1^1$
 - Alternative approach is to use external information on extent of ME to adjust naïve estimates (Card, 1996 in context of union wage effects)
- Again assume that measurement errors are classical and twins drawn from same distribution (see 4.2). Thus get following:
 - OLS: $cov(Z_1^1, Z_1^2) = cov(Z_1^* + \zeta_1^1, Z_1^* + \zeta_1^2) = var(Z_1^*) + cov(\zeta_1^1, \zeta_1^2) = var(Z_1^*)$

- FD: $cov(\Delta Z_1^1, \Delta Z_1^2) = cov(\Delta Z_1^* + \Delta \zeta_1^1, \Delta Z_1^* + \Delta \zeta_1^2) = var(\Delta Z_1^*)$
- Substitute this back into our coefficient terms:
 - $\hat{\beta}_{OLS} = \beta \frac{var(Z_{1i}^*)}{var(Z_{1i}^*) + var(\zeta_{1i})} = \beta \frac{var(Z_{1i}^*)}{var(Z_{1i})} = \beta \frac{cov(Z_{1i}^1, Z_{1i}^2)}{var(Z_{1i})}$
 - $\hat{\beta}_{FD} = \beta \frac{var(\Delta Z_{1i}^*)}{var(\Delta Z_{1i}^*) + var(\Delta \zeta_{1i})} = \beta \frac{var(\Delta Z_{1i}^*)}{var(\Delta Z_{1i})} = \beta \frac{covar(\Delta Z_{1i}, \Delta Z_{2i})}{var(\Delta Z_{1i})}$

Results

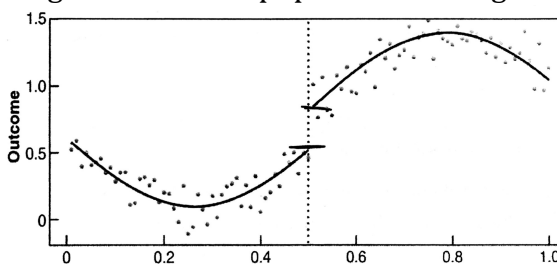
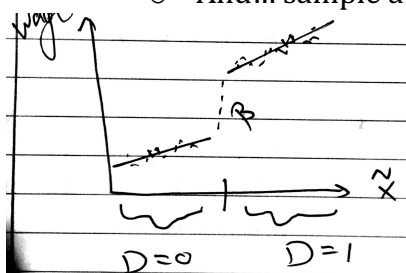
- Extra year of schooling increases wage by 12-16%. Almost double previous estimates
- $\hat{\beta}_{FD} = 9.2\%$ (bigger but not much) versus $\hat{\beta}_{FD,IV} = 16.7\%$ (much bigger and expected)
 - IV fixes endogeneity and measurement error. FD only endogeneity. Thus difference between these is indicative of which matters more
 - In literature IV results are typically higher than OLS. Puzzle because expect ability biases upwards. This shows measurement error is the culprit!
- Card (2001): 10% due to measurement error. Rest is due to LATE vs. ATE effect

6 REGRESSION DISCONTINUITY

6.1 SET UP

Simple [Linear and Non-Linear Sharp]

- RD is used when treatment is deterministic and discontinuous function of x_i
 - Regression $Y_i = \alpha + \beta x_i + \Delta D_i + \epsilon_i$ where $D_i = 1(x_i > x_0)$
 - i.e. $D_i = 1$ if $x_i \geq x_0$ and $D_i = 0$ if $x_i < x_0$
 - Also locally approximate any smooth function $Y_i = f(x_i) + \Delta D_i + \eta_i$
- Imbens and Lemieux (2008): there is no value of x_i at which we get to observe both treatment and control observations. Instead extrapolate across covariate values, at least in a neighbourhood of the discontinuity
- Sharp RD is akin to RCT (i.e. non-parametric) where treatment is essentially “randomly” assigned. Requires good estimates of Y in small neighbourhoods to right and left of x_0
 - But... Working in a small neighbourhood of the cutoff means not much data
 - And... sample average is biased for population in neighbourhood of boundary



Complicated [General Model by Angrist and Pischke]

- Let $\tilde{x}_i = x_i - x_0$
- Allow for completely different trends...
 - $E[Y_{0i}] = f_0(\tilde{x}_i) = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 \dots \beta_{0n}\tilde{x}_i^n$
 - $E[Y_{1i}] = f_1(\tilde{x}_i) = \Delta + \alpha + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 \dots \beta_{1n}\tilde{x}_i^n$
- ... and combine two trends in one model:
 - $Y_i = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 \dots \beta_{0n}\tilde{x}_i^n + \Delta D_i + \beta_1^* D_i \tilde{x}_i + \beta_2^* D_i \tilde{x}_i^2 \dots \beta_n^* D_i \tilde{x}_i^n + \eta_i$
 - Where $\beta_j^* = \beta_{1j} - \beta_{0j}$
- Can use "discontinuity sample" as robustness check

- RD estimates get less precise as window gets smaller (less observations)
- Number of polynomial terms needed to model $f(x_i)$ should go down (less controls)
- Thus overall, estimated effect should remain stable as narrow in on x_0

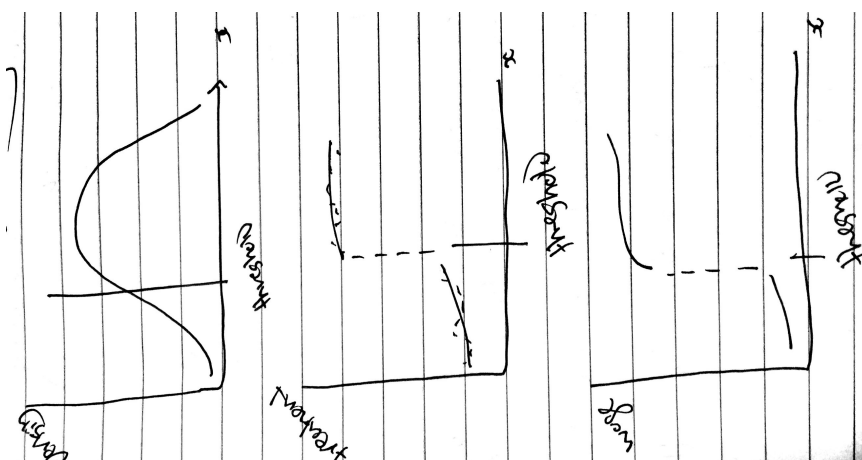
Fuzzy regression discontinuity

- Fuzzy RD exploits discontinuities in the probability or expected value of treatment conditional on a covariate (Discontinuity is IV for treatment status, not on/off switch)
- Used when treatment not clearly deterministic function of x_i but a probability function
- $\Pr(D_i = 1) = p(x_i)$ where $\lim_{x_i \rightarrow x_0} p(x_i) \neq \lim_{x_i \rightarrow x_0^+} p(x_i)$
- Special case of an IV. The IV is the cut-off dummy variable; the RD estimator is LATE
- RD has one cut-off: $\tilde{x}_i = x_i - c_i$
- Fuzzy RD has many: $\tilde{x} = x - c$

6.1.1 CONCERNS AND SOLUTIONS

Reduce Selection Bias

- Flexible enough to get functional form of $E[Y_i|x_0]$ right away from x_0
- Choose window wisely so RD provide very local information on treatment around x_0 (as small as possible)
- Enough number of observations (as large as possible)
- Another program is active with similar threshold.
 - Like RCT controlling for observables should not matter for point estimates
- Underlying outcome function is jumpy
 - Check values of other (predetermined) characteristics. Do they vary discontinuously around threshold?
- Individuals are able to manipulate X to push themselves over threshold
 - Check for bunching at the threshold



6.2 CS: LUDWIG & MILLER (2007): HEAD START

Context

- Question: Does Head Start (a two generation program targeting young children and parents) improve long-run health and schooling outcomes?
 - Introduced 1965. If earn less than 130% federal poverty line become eligible to receive preschool education, medical services, and parental involvement
- Reason to believe it is necessary and effective:
 - Tremblay et al. (2004): large disparities in cognitive and noncognitive skills along race and class lines observed well before children start school
 - Heckman and Krueger (2003): human capital interventions may be particularly promising for disadvantaged children during the early years of life
- Existing literature is positive but not robust
 - Currie & Thomas (1995): Use within-family comparisons of siblings who have and have not participated in the program. Positive effect on tests and immunization but (2000) fade out for black
 - Currie (2001): “The jury is still out on Head Start” as measurement error and spillover effects across siblings may bias their impact estimates toward zero

Methodology

- Make use of fact that 300 poorest countries got Head Start funding proposal assistance. Allows for use of fuzzy regression discontinuity
 - Counties with poverty rates just above cut-off have 50-100% higher funding and participation rates
- Reason to believe it is robust:
 - No concern over strategic behaviours as treatment eligibility is based on poverty rates of five years ago
 - No jumps in governmental spending or other federal funding programs (i.e. not confounding variables)
 - Significant results using several specification and proximity to threshold. Estimated differences persisted over 70s
 - No jumps in non-Head Start susceptible health outcomes or unrelated age groups

Results

- A 50–100% increase in Head Start funding reduces mortality rates from relevant causes by 33–50% of the control mean. Enough to drive back down to national average!
- But... weaker evidence of achievement effects

6.3 CS: LUDWIG & MILLER (2007): HEAD START

Context

- See STAR for motivation

Methodology

- Use natural experiment in Israel of “Maimonides Rule”, whereby class size is capped at 40. Thus cohort of 40 creates big class, but cohort of 41 small class (i.e. split 20 & 21)
- Not strict split as because some schools split grades at enrolments lower than 40, but works as fuzzy for multiple points (40, 80, 120, ...)

Results

- Find small class size helps, similar to STAR study: a 7-student reduction in class size (as in Tennessee STAR) raises Math scores by about 1.75 points, for an effect size of .18\sigma.
 - Naïve analysis (no RD or controls) suggests students from smaller classes do worse on standardized tests
 - Most of naïve bias vanishes when %-disadvantaged in school is included as a control and becomes insignificant when enrolment is added as an additional control