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

Andrew H. McCallum, Ph.D.

March 31, 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

End of MHE Ch. 3

- Regressions and causality
- When does a regression approximate an experiment?
- Casual effects and OVB
- Including too much

Start MHE Ch. 4

- IV regressions
- How do they get us to causality

Casual vs. causal

- *Casual* regressions happen for many reasons: exploratory or descriptive analysis, just having fun, no long-term commitment ...
- *Causal* regressions are more serious and enduring, describe counterfactual states of the world, useful for policy analysis
- Americans mortgage homes to send a child to elite private colleges. Does private pay? Denote private attendance by C_i. The causal relationship between private college attendance and earnings is

$$Y_i = \begin{cases} Y_{1i} & if\ C_i = 1 \\ Y_{0i} & if\ C_i = 0 \end{cases}$$

- $Y_{1i} - Y_{0i}$ is an individual causal effect. Alas, we only get to see one of Y_{1i} or Y_{0i}. The observed outcome, Y_i , is

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})\, C_i \tag{1}$$

  We hope to measure average $Y_{1i} - Y_{0i}$ for some group, say those who went private: $E\left[Y_{1i} - Y_{0i} \mid C_i = 1\right]$, i.e., TOT (treatment effect on those treated)

Casual vs. causal (cont.)

- Comparisons of those who did and didn't go private are biased:

$$\underbrace{E[Y_i \mid C_i = 1] - E[Y_i \mid C_i = 0]}_{\text{Observed difference in earnings}} = \underbrace{E[Y_{1i} - Y_{0i} \mid C_i = 1]}_{\text{TOT}}$$

$$+ \underbrace{E[Y_{0i} \mid C_i = 1] - E[Y_{0i} \mid C_i = 0]}_{\text{selection bias}}$$

- It seems likely that those who go to private college would have earned more anyway. The naive comparison, $E[Y_i \mid C_i = 1] - E[Y_i \mid C_i = 0]$, exaggerates the benefits of private college attendance
    - *Selection bias = OVB in a causal model*
- The *conditional independence assumption* (CIA) asserts that conditional on observed X_i, selection bias disappears:

$$\{Y_{0i}, Y_{1i}\} \coprod C_i \mid X_i$$

- Given the CIA, conditional-on-X_i comparisons are causal:

$$E[Y_i \mid X_i, C_i = 1] - E[Y_i \mid X_i, C_i = 0] = E[Y_{1i} - Y_{0i} \mid X_i]$$

Using the CIA

- The CIA means that C_i is "as good as randomly assigned," conditional on X_i
- A secondary implication: Given the CIA, the conditional on X_i causal effect of private college attendance on private graduates equals the average private effect at X_i:

$$E\left[Y_{1i} - Y_{0i} \mid X_i, C_i = 1\right] = E\left[Y_{1i} - Y_{0i} \mid X_i\right]$$

- This is important ... but less important than the elimination of selection bias
- Note also that the marginal average private college effect can be obtained by averaging over X_i:

$$\begin{aligned} & E\left\{E\left[Y_i \mid X_i, C_i = 1\right] - E\left[Y_i \mid X_i, C_i = 0\right]\right\} \\ = \quad & E\left\{E\left[Y_{1i} - Y_{0i} \mid X_i\right]\right\} \\ = \quad & E\left[Y_{1i} - Y_{0i}\right] \end{aligned}$$

- This suggests we compare people with the same X's ... like matching

Regression and the CIA

- The regression machine turns the CIA into causal effects
- Constant causal effects allow us to focus on selection issues (MHE 3.3 relaxes this). Suppose

$$\begin{aligned} Y_{0i} &= \quad \alpha + \eta_i \\ Y_{1i} &= \quad Y_{0i} + \rho \end{aligned} \tag{2}$$

- Using (1) and (2), we have

$$Y_i = \alpha + \rho C_i + \eta_i \tag{3}$$

- Equation (3) *looks* like a bivariate regression model, except that (2) associates the coefficients in (3) with a causal relationship
- This is not a regression (or should not be!), because C_i can be correlated with potential outcomes, in this case, the residual, η_i

Regression and the CIA (cont.)

- The CIA applied to our constant-effects setup implies:

$$E\left[\eta \mid C_i, X_i\right] = E\left[\eta_i \mid X_i\right]$$

- Suppose also that

$$E\left[\eta_i \mid X_i\right] = X_i'\gamma$$

so that

$$E\left[Y_i \mid X_i, C_i\right] = \alpha + \rho C_i + E\left[\eta_i \mid X\right] = \alpha + \rho C_i + X_i'\gamma$$

- Mean-independence implies orthogonality, so

$$Y_i = \alpha + \rho C_i + X_i'\gamma + v_i$$

has error

$$\nu_i \equiv \eta_i - X_i'\gamma = \eta_i - E\left[\eta_i \mid C_i, X_i\right]$$

uncorrelated with regressors, C_i and X_i. The same ρ appears in the regression and causal models!

- Modified **Dale and Krueger (2002):** private proving ground

Dale and Krueger (2002)

ESTIMATING THE PAYOFF TO ATTENDING A MORE SELECTIVE COLLEGE: AN APPLICATION OF SELECTION ON OBSERVABLES AND UNOBSERVABLES*

STACY BERG DALE AND ALAN B. KRUEGER

Estimates of the effect of college selectivity on earnings may be biased because elite colleges admit students, in part, based on characteristics that are related to future earnings. We matched students who applied to, and were accepted by, similar colleges to try to eliminate this bias. Using the College and Beyond data set and National Longitudinal Survey of the High School Class of 1972, we find that students who attended more selective colleges earned about the same as students of seemingly comparable ability who attended less selective schools. Children from low-income families, however, earned more if they attended selective colleges.

A burgeoning literature has addressed the question, "Does the 'quality' of the college that students attend influence their subsequent earnings?"[1] Obtaining accurate estimates of the payoff to attending a highly selective undergraduate institution is of obvious importance to the parents of prospective students who foot the tuition bills, and to the students themselves. In addition, because college selectivity is typically measured by the average characteristics (e.g., average SAT score) of classmates, the literature is closely connected to theoretical and empirical studies of

Omitted Variables Bias

- The omitted variables bias (OVB) formula describes the relationship between regression estimates in models with different controls
- Go long: wages on schooling, S_i , controlling for ability (A_i)

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

- Ability is hard to measure. What if we leave it out? The result is

$$\frac{\mathrm{Cov}\left(Y_i, S_i\right)}{V\left(S_i\right)} = \rho + \gamma'\delta_{As},$$

  where δ_{As} is the vector of coeffcients from regressions of the elements of A_i on S_i ...
  - *Short equals long plus the effect of omitted times the regression of omitted on included*
- Short equals long when omitted and included are uncorrelated
- **Table 3.2.1** illustrates OVB (some controls are bad; the formula works for good and bad alike)

Example from the homework and book

TABLE 3.2.1
Estimates of the returns to education for men in the NLSY

	(1)	(2)	(3)	(4)	(5)
Controls:	None	Age Dummies	Col. (2) and Additional Controls*	Col. (3) and AFQT Score	Col. (4), with Occupation Dummies
	.132	.131	.114	.087	.066
	(.007)	(.007)	(.007)	(.009)	(.010)

Notes: Data are from the National Longitudinal Survey of Youth (1979 cohort, 2002 survey). The table reports the coefficient on years of schooling in a regression of log wages on years of schooling and the indicated controls. Standard errors are shown in parentheses. The sample is restricted to men and weighted by NLSY sampling weights. The sample size is 2,434.

*Additional controls are mother's and father's years of schooling, and dummy variables for race and census region.

Reverse Causality ("Endogeneity")

- Here we think that x_i may affect Y_i, but Y_i may also affect x_i
- Example: x_i is divorce, Y_i is income
- We are interested in causal model:

$$Y_i = \beta_0 + \beta x_i + \varepsilon_i$$

- But there is also a causal relationship in other direction:

$$x_i = \alpha_0 + \alpha Y_i + u_i$$

- 2 equations, 2 variables so you can solve for x_i and Y_i as functions of coefficients and errors
- You could then plug this into coefficient formula $\hat{\beta} = \mathrm{Cov}\left(Y_i, x_i\right) / V\left(x_i\right)$
- This will not be β
- In hospital example, being sick causes you to go to hospital...so you could never learn anything by regressing health outcomes on hospitalization regardless of the quality of your controls.

Selection

- One explanation is sample selection: can you spot the precise problem here?
  - Only have earnings data on women who work
  - Women with small children who work tend to have high earnings (e.g. they can pay for childcare)
  - Employment rates of mothers with babies is 28%, of those with 5-year olds it is 50%

Bad control: an example

- The bad control problem is a version of selection bias
- Example:
  - Suppose we are interested in the effects of a college degree on earnings
  - People can work in one of two occupations: white collar and blue collar
  - College degree opens the door to higher-paying white collar jobs
  - Should occupation be seen as an omitted variable in a regression of wages on schooling?

Bad control example: formally

- Let's consider the following formal model using the potential outcomes framework:

$$
\begin{aligned}
Y_i &= C_i Y_{1i} + (1 - C_i)\, Y_{0i} \\
W_i &= C_i W_{1i} + (1 - C_i)\, W_{0i}
\end{aligned}
$$

- C_i is a indicator for college graduates
- $\{Y_{1i}, Y_{0i}\}$ denote earnings with/out college
- $\{W_{1i}, W_{0i}\}$ denote white collar status with/out college

Bad control example (cont'd)

- If we assume that C_i is (unconditionally) randomly assigned, we can get its causal effect of college on earning and occupation

$$
\begin{aligned}
E[Y_i \mid C_i = 1] - E[Y_i \mid C_i = 0] &= E[Y_{1i} - Y_{0i}] \\
E[W_i \mid C_i = 1] - E[W_i \mid C_i = 0] &= E[W_{1i} - W_{0i}]
\end{aligned}
$$

- How do we implement this?
- Now suppose we regress Y_i on C_i within the sample where $W_i = 1$ ...

Bad control example (cont'd)

- If we regress Y_i on C_i holding $W_i = 1$, the estimated coefficient on C_i will be:

$$E[Y_i \mid W_i = 1, C_i = 1] - E[Y_i \mid W_i = 1, C_i = 0] =$$
$$= E[Y_{1i} \mid W_{1i} = 1, C_i = 1] - E[Y_{0i} \mid W_{0i} = 1, C_i = 0]$$

- This comes from where?
  - Definitions of potential outcomes
- Since $\{Y_{1i}, W_{1i}, Y_{0i}, W_{0i}\} \coprod C_i$, the last expression is $= E[Y_{1i} \mid W_{1i} = 1] - E[Y_{0i} \mid W_{0i} = 1]$

Bad control example (cont'd)

- Finally, we add and subtract a term to get:

$$E[Y_{1i} \mid W_{1i}=1] - E[Y_{0i} \mid W_{0i}=1]$$
$$= \underbrace{E[Y_{1i} - Y_{0i} \mid W_{1i}=1]}_{\text{effect}} + \underbrace{\{E[Y_{0i} \mid W_{1i}=1] - E[Y_{0i} \mid W_{0i}=1]\}}_{\text{selection term}}$$

- So the comparison is the sum of: effect of college on earnings + selection term
- Selection term = diff in earnings without college for: (those who become white collars with college) minus (those who become white collars without college)

How To Surmount the Problems?

- More sophisticated econometric methods
  - E.g. later in the course we will talk about IV
- Better data — Griliches:
  - “since it is the ‘badness’ of the data that provides us with our living, perhaps it is not at all surprising that we have shown littl interest in improving it”

What Next?

- We're not always content to run regressions, of course, though this is usually where we start
  - Regression is our first line of attack on the identification problem; it's all about *control*
- If the regression you've got is not the one you want, that's because the underlying *relationship* is unsatisfactory
- Whats to be done with an unsatisfactory relationship?
  - Move on . . . to IV!

Organizing IV

I tell the IV story in two iterations, first with constant effects, then in a framework with heterogeneous potential outcomes.

- The constant effects framework focuses attention on the IV solution for selection bias and on essential IV mechanics
- But first: Why do IV?
  - I can't say "because the regressors are correlated with the errors."
  - As we've seen, regressors are uncorrelated with errors by definition
- The (short) regression of schooling on wages produces residuals uncorrelated with schooling (that's how they are defined)
- The problem, therefore, must be that the regression you've got is not the regression you want (and that's your fault!)

IV Goes Long

- Suppose the causal link between schooling and wages can be written

$$f_i(s) = \alpha + \rho s + \eta_i$$

- Imagine a vector of control variables, A_i , called "ability"; write

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

  where γ is a vector of pop. reg. coefficients, so v_i and A_i are uncorrelated by *construction*

- We'd happily include ability in the regression of wages on schooling, producing this long regression:

$$Y_i = \alpha + \rho s_i + A_i'\gamma + v_i$$

  The error term here is the random part of potential outcomes, v_i, left over after controlling for A_i

- If $E[S_i v_i] = 0$, a version of the CIA, the population regression of Y_i on S_i and A_i identifies ρ. That's like saying: "A_i is the only reason schooling is correlated with potential outcomes."

IV and OVB

- *IV allows us to estimate the long-regression coefficient, ρ, when A_i is unobserved.*
  The instrument, Z_i, is assumed to be (1) correlated with the causal variable of interest, S_i, and (2) uncorrelated with potential outcomes
- Here, "uncorrelated with potentials" means $\text{Cov}\left(\eta_i, Z_i\right) = 0$, or, equivalently, Z_i is uncorrelated with both A_i and v_i
  - The is a version of the *exclusion restriction:* Z_i can be said to be excluded from the causal model of interest

  Given the exclusion restriction, it follows from equation (1) that

$$\rho = \frac{\text{Cov}\left(Y_i, Z_i\right)}{\text{Cov}\left(S_i, Z_i\right)} = \frac{\text{Cov}\left(Y_i, Z_i\right)/V\left(Z_i\right)}{\text{Cov}\left(S_i, Z_i\right)/V\left(Z_i\right)}$$

$$= \frac{\text{"RF"}}{\text{"1st"}}$$

- The IV *estimator* is the sample analog of (2)

Two-stage least squares (2SLS)

- In practice, we do IV by doing 2SLS. This allows us to add covariates (controls) and combine multiple instruments. Returning to the schooling example, a causal model with covariates is

$$Y_i = \alpha' X_i + \rho S_i + \eta_i,$$

where η_i is the compound error term, $A_i'\gamma + v_i$. The first stage and reduced form are

$$\begin{aligned} S_i &= X_i'\pi_{10} + \pi_{11} Z_i + \xi_{1i} \\ Y_i &= X_i'\pi_{20} + \pi_{21} Z_i + \xi_{2i} \end{aligned}$$

- The reduced form is obtained by substituting (4) into (3):

$$\begin{aligned} Y_i &= \alpha' X_i + \rho \left[X_i'\pi_{10} + \pi_{11} Z_i \right] + \rho \xi_{1i} + \eta_i \\ &= X_i' \left[\alpha + \rho \pi_{10} \right] + \rho \pi_{11} Z_i + \left[\rho \xi_{1i} + \eta_i \right] \\ &= X_i'\pi_{20} + \pi_{21} Z_i + \xi_{2i} \end{aligned}$$

2SLS Notes

- Again, it's all about the ratio of RF to 1st:

$$\frac{\pi_{21}}{\pi_{11}} = \rho$$

  In simultaneous squations models, the sample analog of this ratio is called an *Indirect Least Squares* (ILS) estimator of ρ

- Where does *two-stage least squares* come from? Write the first stage as the sum of fitted values plus first-stage residuals:

$$S_i = X_i'\pi_{10} + \pi_{11}Z_i + \xi_{1i} = \widehat{s}_i + \xi_{1j}$$

  2SLS estimates of (3) can be constructed by substituting first-stage fitted values for S_i in (3)

$$Y_i = \alpha' X_i + \rho\widehat{s}_i + [\eta_i + \rho\xi_{1j}]$$

  and using OLS to estimate this "second stage" (a version of eq. 6)

- In practice, we let Stata do it "manual 2SLS" doesn't get the standard errors right

2SLS example: Angrist and Krueger (1991)

- AK-91 argue that because children born in late-quarters start school younger, they are kept in school longer by birthday-based compulsory schooling laws
- There's a powerful first stage supporting this: Schooling tends to be higher for late-quarter births; this is driven by high school and not college, consistent with the CSL story
- The QOB first stage and reduced form are plotted in Figure 4.1.1
- The corresponding 2SLS estimates appear in Table 4.1.1
  - 2SLS matches the QOB pattern earnings (the RF) to the QOB pattern in schooling (the first stage).
  - The exogenous covariates include year-of-birth and state-of-birth dummies, as well as linear and quadratic functions of age in quarters
- QOB Questioned: Bound, Jaeger, and Baker (1995) and Buckles and Hungerman (2008) argued QOB is correlated with maternal characteristics. Allowing for this fails to overturn AK conclusions

2SLS is a many-splendored thing

- 2SLS is the same as IV where the instrument is $\hat{s}_i^*$, the residual from a regression of $\hat{s}_i$ on X_i
- One-instrument 2SLS equals IV, where the instrument is $\tilde{z}_i$, the residual from a regression of Z_i on the covs, X_i
- One-instrument 2SLS equals indirect least squares (ILS), that is, the ratio of reduced form to first stage coefficients on the instrument. In other words,

$$\begin{aligned} \frac{\mathrm{Cov}\left(Y_i \hat{s}_i^*\right)}{V\left(\hat{s}_i^*\right)} &= \frac{\mathrm{Cov}\left(Y_i, \hat{s}_i^*\right)}{\mathrm{Cov}\left(S_i, \hat{s}_i^*\right)} \\ &= \frac{\mathrm{Cov}\left(Y_i, \tilde{z}_i\right)}{\mathrm{Cov}\left(S_i, \tilde{z}_i\right)} = \frac{\pi_{21}}{\pi_{11}} \end{aligned}$$

- With more than one instrument, 2SLS is a weighted average of the one-at-time (just-identified) estimates (In a linear homoskedastic constant-effects model, this is efficient)

Multi-Instrument 2SLS (details; mistakes)

- Let

$$\rho_j = \frac{\text{Cov}\left(Y_i, Z_{ji}\right)}{\text{Cov}\left(D_i, Z_{ji}\right)};\ j = 1, 2$$

denote two IV estimands using Z_{1i} and Z_{2i} to instrument D_i.

- The 2SLS estimand is

$$\rho_{2SLS} = \psi\rho_1 + (1-\psi)\,\rho_2,$$

where ψ is a number between zero and one that depends on the relative strength of the instruments in the first stage.

- Angrist and Evans (1998) use twins and sex-mix instruments
  - Using a twins-2 instrument alone, the IV estimate of the effect of a third child on female labor force participation is $-.084$ (s.e. $= .017$). The corresponding samesex estimate is $-.138$ (s.e. $= .029$).
  - Using both instruments produces a 2SLS estimate of $-.098$ $(.015)$.
  - The 2SLS weight in this case is $.74$ for twins, $.26$ for samesex, due to the stronger twins first stage.

2SLS Mistakes

2SLS . . . so simple a fool can do it . . .

and many do!

What can go wrong?

- As explained in MHE 4.6.1, three mistakes have yet to be relegated to the dustbin of IV history:
  - Manual 2SLS
  - Covariate ambivalence
  - Forbidden regressions (from the left and the right)
- These can be interpreted as the result of failed attempts to get round hard-wired 2SLS protocols
- Avoid temptation: let Stata do it!

2SLS Lingo

- These terms come to us from simultaneous equations modeling, the intellectual birthplace of IV:
  - *Endogenous variables* are the dependent variable and the independent variable(s) to be instrumented; in a simultaneous equations model, endogenous variables are determined by solving the system
  - *To treat an independent variable as endogenous* is to instrument it, i.e., to replace it with fitted values in the 2SLS second stage
  - Exogenous variables include covariates (not instrumented) and the excluded instruments themselves. In a simultaneous equations model, exogenous variables are determined outside the system
- In any IV study, variables are either: dependent or (other) endogenous variables, instruments, or covariates
- If you're unsure what's what, or find yourself asking variables to play more than one role . . . seek counseling

The Wald estimator

- How were Vietnam-era vets affected by their service?
- Let D_1 indicate veterans. A causal constant-effects model is:

$$Y_i = \alpha + \rho D_i + \eta_i,$$

where η_i and D_i may be correlated. B/c 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,

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

A direct route uses (8) and $E\left[\eta_i \mid Z_i\right] = 0$:

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

Solving this for ρ produces (9)

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)

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

Visual IV, Grouping, and GLS

- Equation (12) in action: Angrist (1990). Figure 3 This illustrates visual instrumental variables (VIV)
- GLS (weighted least squares) applied to equation (12) *is* 2SLS
  - The instruments in this case are dummies for each lottery-number cell. Define $Z_i \equiv \{r_{ji} = 1\left[R_i = j\right]; j = 1, \ldots, J-1\}$. The first stage for D_i on Z_i plus a constant is saturated so fitted values are cond means $\hat{p}_i$ repeated n_j times for each j. The second stage slope estimate is therefore weighted least squares on the grouped equation (12) weighted by the cell size, n_j
  - Because GLS is efficient, 2SLS is also the efficient linear combination of the underlying just-identified IV (Wald) estimates (earlier, we saw that 2SLS is a weighted average of just-identified estimates in a two-instrument example)
- That's why we call Figure 3 "VIV"

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
- For example, you will discuss a negative income tax example in class
- I this case, we normalize difference in outcome $(E[y \mid \text{treatment}] - E[y \mid \text{control}])$ by the diff in takeup $(E[x \mid \text{treatment}] - E[x \mid \text{control}])$ where $E[x \mid \text{control}] = 0$

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 (see exercise)

Our Constant-Effects Benchmark

- The traditional IV framework is the linear, constant-effects world discussed in Part 1 With Bernoulli treatment, D_i, we have

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

- 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

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

- 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

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

- Results

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 Exclusion 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, $E[D_{1i} - D_{0i}] = P[D_{1i} > D_{0i}]$
- Interpreting monotonicity in latent-index models:

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

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

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

  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

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

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

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

  ($.62$ 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 (coddled-delivered)		0.178 (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, $g\left(Y_{ji}\right)$,

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

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

$$= \quad E\left[g\left(Y_{0i}\right) \mid D_{1i} > D_{0i}\right]$$

- 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**

All About Kompliers

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:

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

  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, $h[\alpha D_i + X_i'\beta]$
- Suppose, for example,

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

  This gives "best ProbIt 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!

The Bias of 2SLS

- Cross-section OLS estimates are typically unbiased for the pop BLP, as well as consistent, but this might not be the regression you want
- 2SLS estimates are consistent for causal effects but biased towards OLS estimates
- Endogenous var is vector x, dep var is vector y, no covs:

  $$y = \beta x + \eta$$

  The $N \times Q$ matrix of instruments is Z , with first-stage

  $$x = Z\pi + \xi$$

  Outcome error η_i is correlated with ξ_i. Instruments are uncorrelated with ξ_i by construction and with η_i by assumption
- The 2SLS estimator is

  $$\widehat{\beta}_{2SLS} = \left(x' P_Z x\right)^{-1} x' P_Z y = \beta + \left(x' P_Z x\right)^{-1} x' P_Z \eta$$

  where $P_Z = Z\left(Z'Z\right)^{-1} Z'$ is the projection matrix that produces fitted values

The Bias of 2SLS (cont.)

- Substituting for x in $x'P_Z\eta$, we get

$$\begin{aligned}\widehat{\beta}_{2SLS}-\beta &= \left(x'P_Zx\right)^{-1}\left(\pi'Z'+\xi'\right)P_Z\eta \\ &= \left(x'P_Zx\right)^{-1}\pi'Z'\eta+\left(x'P_Zx\right)^{-1}\xi'P_Z\eta\end{aligned}$$

- Expectation of the ratios on the right hand side of (16) are closely approximated by the ratio of expectations

$$E\left[\widehat{\beta}_{2SLS}-\beta\right]\approx\left(E\left[x'P_Zx\right]\right)^{-1}E\left[\pi'Z'\eta\right]+\left(E\left[x'P_Zx\right]\right)^{-1}E\left[\xi'P_Z\eta\right]$$

  This Bekker (1994) approximation ("group asymptotics" in AK-95) gives a good account of finite-sample behavior

- Using the fact that $E\left[\pi'Z'\xi\right]=0$ and $E\left[\pi'Z'\eta\right]=0$, we have

$$E\left[\widehat{\beta}_{2SLS}-\beta\right]\approx\left[E\ \left(\pi'Z'Z\pi\right)+E\left(\xi'P_Z\xi\right)\right]^{-1}E\left(\xi'P_Z\eta\right)$$

- 2SLS is biased b/c $E\left(\xi'P_Z\eta\right)\neq 0$ unless η_i and ξ_i are uncorrelated

The Bias of 2SLS: First-stage F

- Manipulation of (18) generates:

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}} \left[\frac{E\left(\pi' Z' Z \pi\right)/Q}{\sigma_{\xi}^{2}} + 1\right]^{-1}$$

$\left(1/\sigma_{\xi}^{2}\right) E\left(\pi' Z' Z \pi\right)/Q$ is the "population F" for joint significance of instruments in first-stage, so we can write

$$E\left[\widehat{\beta}_{2SLS} - \beta\right] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}} \frac{1}{F+1}$$

- As F gets small, the bias of 2SLS approaches $\frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}}$. The bias of the OLS estimator is $\frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}}$, which also equals $\frac{\sigma_{\eta\xi}}{\sigma_{\xi}^{2}}$ if $\pi = 0$. estimates are therefore said to be "biased towards OLS estimates" when there isn't much of a first stage. On the other hand, the bias of 2SLS vanishes when F gets large, as it should happen in large samples when $\pi \neq 0$

The Bias of 2SLS: First-stage F (cont.)

- First-stage F varies inversely with the number of instruments if they're weak.
  - Adding instruments with no effect on the first-stage R-squared, the model sum of squares, $E\left(\pi' Z' Z\pi\right)$, and the residual variance, σ_{ξ}^{2}, are fixed while Q increases
  - The F-statistic shrinks as a result. From this we learn that the addition of weak instruments increases bias
- Holding the first-stage sum of squares fixed, bias is least in the just-ID case when the number of instruments is as low as it can get
- 2SLS bias is a consequence of first-stage estimation error. We'd like to use $\widehat{x}_{pop} = Z\pi$ as IVs since these fits are uncorrelated with the second stage error
  - In pracrice, we use $\widehat{x} = P_Z x = Z\pi + P_Z\xi$, which differs from $\widehat{x}_{pop}$ by the term $P_Z\xi$
  - 2SLS bias arises from the corr between $P_Z\xi$ and η

IV without bias or tears

- **Just-identified 2SLS** (say, the Wald estimator) is *approximately* unbiased (this isn's clear from the Bekker sequence). The just-ID sampling distribution has no moments, yet it's approximately centered where it should be unless the instruments are *really* weak
- The **Reduced Form** is unbiased: if you can't see the relationship you're after in the reduced form, it ain't there! In just-identified models, the p-value for the reduced-form effect of the instrument is approximately the p-value from the second stage. (Chernozhukov and Hansen, 2008, use this to do bias-free inference)
- **LIML** is approximately median-unbiased for over-identified constant-effects models, and therefore provides an attractive alternative to just-identifed estimation using one instrument at a time (see, e.g., Davidson and MacKinnon, 1993, and Mariano, 2001). (LIML=2SLS in just-identifed models)

What does this mean in practice? Besides retaining a vague sense of worry about your first stage, we recommend the following:

1. Report the first stage and think about whether it makes sense. Are the magnitude and sign as you would expect, or are the estimates too big or large but wrong-signed? If so, perhaps your hypothesized first-stage mechanism isn't really there, rather, you simply got lucky.
2. Report the F-statistic on the excluded instruments. The bigger this is, the better. Stock, Wright, and Yogo (2002) suggest that F-statistics above about 10 put you in the safe zone though obviously this cannot be a theorem.
3. Pick your best single instrument and report just-identified estimates using this one only. Just-identified IV is median-unbiased and therefore unlikely to be subject to a weak-instruments critique.
4. Check over-identified 2SLS estimates with LIML. LIML is less precise than 2SLS but also less biased. If the results come out similar, be happy. If not, worry, and try to find stronger instruments.
5. Look at the coefficients, t-statistics, and F-statistics for excluded instruments in the reduced-form regression of dependent variables on instruments. Remember that the reduced form is proportional to the causal effect of interest. Most importantly, the reduced-form estimates, since they are OLS, are unbiased. As Angrist and Krueger (2001) note, if you can't see the causal relation of interest in the reduced form, it's probably not there.

Bound, Jaeger, and Baker (1995) argued that bias is a major concern when using quarter birth even though N = 300k