Gov 2002: 4. Observational Studies and Confounding

Matthew Blackwell

September 10, 2015

Where are we? Where are we going?

- Last two weeks: randomized experiments.
- From here on: observational studies.
  - What are they?
  - How do they admit the possiblity of confounding?
  - How can we adjust for confounding?

1/ Observational studies

Experiment review

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- A randomized experiment is an experiment with the following properties:

1. Positivity: assignment is probabilistic: $0 < p_i < 1$
   - No deterministic assignment.
2. Unconfoundedness: $\mathbb{P}[D_i = 1|\mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$
   - Treatment assignment does not depend on any potential outcomes.
   - Sometimes written as $D_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$

Observational studies

- Many different sets of identification assumptions that we'll cover.
- To start, focus on studies that are similar to experiments, just without a known and controlled treatment assignment.
  - No guarantee that the treatment and control groups are comparable.

1. Positivity: assignment is probabilistic:
   $0 < \mathbb{P}[D_i = 1|\mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)] < 1$
2. No unmeasured confounding:
   $\mathbb{P}[D_i = 1|\mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1|\mathbf{X}]$
   - For some observed $\mathbf{X}$
   - Also called: unconfoundedness, ignorability, selection on observables, no omitted variables, exogenous, conditional exchangeable, etc.

Designing observational studies

- Rubin (2008) argues that we should still "design" our observational studies:
  - Pick the ideal experiment to this observational study.
  - Hide the outcome data.
  - Try to estimate the randomization procedure.
  - Analyze this as an experiment with this estimated procedure.
- Tries to minimize "snooping" by picking the best modeling strategy before seeing the outcome.

Discrete covariates

- Suppose that we knew that D_i was unconfounded within levels of a binary X_i.
- Then we could always estimate the causal effect using iterated expectations as in a stratified randomized experiment:

$$\begin{aligned}
&\mathbb{E}_X\Big\{\mathbb{E}[Y_i|D_i=1,X_i]-\mathbb{E}[Y_i|D_i=0,X_i]\Big\} \\
&=\underbrace{\Big(\mathbb{E}[Y_i|D_i=1,X_i=1]-\mathbb{E}[Y_i|D_i=0,X_i=1]\Big)}_{\text{diff-in-means for } X_i=1}\underbrace{\mathbb{P}[X_i=1]}_{\text{share of } X_i=1} \\
&\quad+\underbrace{\Big(\mathbb{E}[Y_i|D_i=1,X_i=0]-\mathbb{E}[Y_i|D_i=0,X_i=0]\Big)}_{\text{diff-in-means for } X_i=0}\underbrace{\mathbb{P}[X_i=0]}_{\text{share of } X_i=0}
\end{aligned}$$

- Never used our knowledge of the randomization for this quantity.

Continuous covariates

- So, great, we can stratify. Why not do this all the time?
- What if X_i = income for unit i?
  - Each unit has its own value of X_i: \$54,134, \$123,043, \$23,842.
  - If $X_i = 54134$ is unique, will only observe 1 of these:

$$\mathbb{E}[Y_i|D_i = 1, X_i = 54134] - \mathbb{E}[Y_i|D_i = 0, X_i = 54134]$$

  - $\rightsquigarrow$ cannot stratify to each unique value of X_i:
- Practically, this is massively important: almost always have data with unique values.

Going to a superpopulation

- From here on out, we'll focus less on the finite population model.
  - Harder with (functionally) continuous covariates.
- Assume that each unit i is drawn from an infinite superpopulation,
  - implies that $(Y_i(0), Y_i(1), D_i, X_i)$ are a draw from their population joint distribution.
- Potential outcomes are now typical random variables.
  - $\mu_c(x) = \mathbb{E}[Y_i(0)|X_i = x]$ and $\mu_t(x) = \mathbb{E}[Y_i(1)|X_i = x]$
  - $\sigma_c^2(x) = \mathbb{V}[Y_i(0)|X_i = x]$ and $\sigma_t^2(x) = \mathbb{V}[Y_i(1)|X_i = x]$
  - $\tau = \mathbb{E}[\mu_t(x) - \mu_c(x)|X_i = x]$

Assumptions in the superpopulation

- With an infinite superpopulation, worry less about conditioning on the entire sample.
  - Units are now independent due to random sampling from an infinite population.
- No unmeasured confoudning implies that:

$$\mathbb{P}(D_i = 1|Y_i(0), Y_i(1), X_i) = \mathbb{P}(D_i = 1|X_i)$$

- Or, written using conditional independence:

$$D_i \perp\!\!\!\perp \Big(Y_i(0), Y_i(1)\Big)|X_i$$

- Positivity can be written: $0 < \mathbb{P}(D_i = 1|X_i = x) < 1$ for all x in the support of X_i.

2/ Confounding

What is confounding?

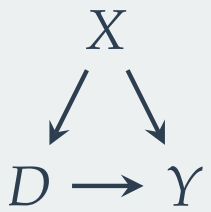
- Confounding is the bias caused by common causes of the treatment and outcome.
  - Leads to "spurious correlation."
- In observational studies, the goal is to avoid confounding inherent in the data.
- Pervasive in the social sciences:
  - effect of income on voting (confounding: age)
  - effect of job training program on employment (confounding: motivation)
  - effect of political institutions on economic development (confounding: previous economic development)
- No unmeasured confounding assumes that we've measured all sources of confounding.

Big problem

- How can we determine if no unmeasured confounding holds if we didn't assign the treatment?
- Put differently:
  - What covariates do we need to condition on?
  - What covariates do we need to match on?
  - What covaraites do we need to include in our regressions?
- One way, from the assumption itself:
  - $\mathbb{P}[D_i = 1|\mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1|\mathbf{X}]$
  - Include covariates such that, conditional on them, the treatment assignment does not depend on the potential outcomes.
- Another way: use DAGs and look at back-door paths.

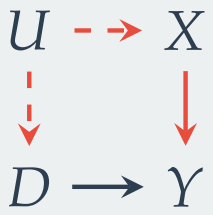
Backdoor paths and blocking paths

- Backdoor path: is a non-causal path from D to Y.
  - Would remain if we removed any arrows pointing out of D.
- Backdoor paths between D and $Y \rightsquigarrow$ common causes of D and Y:



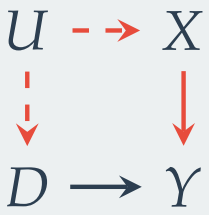
- Here there is a backdoor path $D \leftarrow X \rightarrow Y$, where X is a common cause for the treatment and the outcome.

Other types of confounding



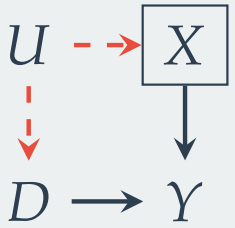
- D is enrolling in a job training program.
- Y is getting a job.
- U is being motivated
- X is number of job applications sent out.
- Big assumption here: no arrow from U to Y.

Other types of confounding



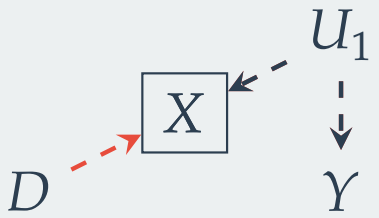
- D is exercise.
- Y is having a disease.
- U is lifestyle.
- X is smoking
- Big assumption here: no arrow from U to Y.

What's the problem with backdoor paths?



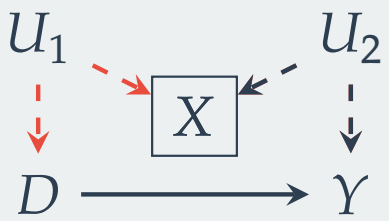
- A path is blocked if:
  1. we control for or stratify a non-collider on that path OR
  2. we do not control for a collider.
- Unblocked backdoor paths $\rightsquigarrow$ confounding.
- In the DAG here, if we condition on X, then the backdoor path is blocked.

Not all backdoor paths



- Conditioning on the posttreatment covariates opens the non-causal path.
  - $\rightsquigarrow$ selection bias.

M-bias

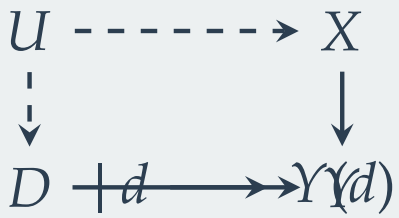


- Not all backdoor paths induce confounding.
- This backdoor path is blocked by the collider X_i that we don't control for.
- If we control for $X_i \rightsquigarrow$ opens the path and induces confounding.
  - Sometimes called M-bias.
- Controversial because of differing views on what to control for:
  - Rubin thinks that M-bias is a "mathematical curiosity" and we should control for all pretreatment variables
  - Pearl and others think M-bias is a real threat.

Backdoor criterion

- Can we use a DAG to evaluate no unmeasured confounders?
- Pearl answered yes, with the backdoor criterion, which states that the effect of D on Y is identified if:
  1. No backdoor paths from D to Y OR
  2. Measured covariates are sufficient to block all backdoor paths from D to Y.
- First is really only valid for randomized experiments.
- The backdoor criterion is fairly powerful. Tells us:
  - if there confounding given this DAG,
  - if it is possible to removing the confounding, and
  - what variables to condition on to eliminate the confounding.

SWIGs



- It's a little hard to see how the backdoor criterion implies no unmeasured confounders.
  - No potential outcomes on this graph!
- Richardson and Robins: Single World Intervention Graphs
  - Split D node into natural value (D) and intervention value d.
  - Let all effects of D take their potential value under intervention $Y(d)$.
- Now can see: are D and $Y(d)$ related?
  - $D \leftarrow U \rightarrow X \rightarrow Y(d)$ implies not independent
  - Conditioning on X blocks that backdoor path $\rightsquigarrow D \perp\!\!\!\perp Y(d)|X$

No unmeasured confounders is not testable

- No unmeasured confounding places no restrictions on the observed data.

$$\underbrace{\left(Y_i(0)\middle|D_i = 1, X_i\right)}_{\text{unobserved}} \overset{d}{=} \underbrace{\left(Y_i(0)\middle|D_i = 0, X_i\right)}_{\text{observed}}$$

- Here, $\overset{d}{=}$ means equal in distribution.
- No way to directly test this assumption without the counterfactual data, which is missing by definition!
- With backdoor criterion, you must have the correct DAG.

Assessing no unmeasured confounders

TABLE VI
THE FOX NEWS EFFECT: INTERACTIONS AND PLACEBO SPECIFICATIONS

	Interactions		Placebo specifications		
	Presid. Rep. vote share 2000–1996		Presidential Republican vote share		
			2000–1996	1996–1992	1992–1988
Dep. var.	(1)	(2)	(3)	(4)	(5)
Availability of Fox News via cable in 2000	0.0109 (0.0042)***	0.0105 (0.0039)***	0.0036 (0.0016)**	−0.0024 (0.0031)	0.0026 (0.0026)
Availability of Fox News via cable in 2003			−0.0001 (0.0012)		

- Can do "placebo" tests, where D_i cannot have an effect (lagged outcomes, etc)
- Della Vigna and Kaplan (2007, QJE): effect of Fox News availability on Republican vote share
  - Availability in 2000/2003 can't affect past vote shares.
- Unconfoundedness could still be violated even if you pass this test!

Alternatives to no unmeasured confounding

- Without explicit randomization, we need some way of identifying causal effects.
- No unmeasured confounders $\approx$ randomized experiment.
  - Indentification results very similar to experiments.
- With unmeasured confounding are we doomed? Maybe not!
- Other approaches rely on finding plausibly exogenous variation in assignment of D_i:
  - Instrumental variables (randomization + exclusion restriction)
  - Over-time variation (diff-in-diff, fixed effects)
  - Arbitrary thresholds for treatment assignment (RDD)

3/ No unmeasured confounders and OLS

Justifying regression

- We know how randomized experiments imply that differences-in-means identify the ATE.
- In the next few weeks, we'll work through how no unmeasured confounding justifies a number of estimation strategies.
- Today, it's useful to walk through what no unmeasured confounding can buy us in a familiar setting: OLS.
    - We'll cover regression more formally later.

Constant effects set up

- Assume a constant effects setup:

$$Y_i(0) = \alpha + X_i'\beta + u_i$$
$$Y_i(1) = \alpha + \tau + X_i'\beta + u_i$$

- Constant effects because $Y_i(1) - Y_i(0) = \tau$ for all units.
- Use consistency to get the usual regression formula:

$$\begin{aligned} Y_i &= Y_i(1)D_i + Y_i(0)(1 - D_i) \\ &= Y_i(0) + \big(Y_i(1) - Y_i(0)\big) \cdot D_i \\ &= \alpha + \tau \cdot D_i + X_i'\beta + u_i \end{aligned}$$

- Does no unmeasured confounding help us identify the causal parameter τ?

Regression on residuals

- First estimate the residuals of regression of the treatment and outcome on the covariates:

$$\tilde{Y}_i = Y_i - \mathbb{E}[Y_i|X_i]$$
$$\tilde{D}_i = D_i - \mathbb{E}[D_i|X_i]$$

- Running a regression of $\tilde{Y}_i$ on $\tilde{D}_i$ is equivalent to controlling for X_i:

$$Y_i = \alpha + \tau \cdot D_i + X_i'\beta + u_i$$
$$\tilde{Y}_i = \alpha + \tau \cdot \tilde{D}_i + \tilde{u}_i$$

- Here, $\tilde{u}_i = u_i - \mathbb{E}[u_i|X_i]$.

What does OLS estimate?

- Using the usual OLS theory, we can show that the probability limit of the OLS estimator of τ is:

$$\begin{aligned}
\text{plim } \hat{\tau}_{\text{OLS}} &= \frac{\text{Cov}(\tilde{D}_i, \tilde{Y}_i)}{\text{Var}(\tilde{D}_i)} \\
&= \frac{\text{Cov}(\tilde{D}_i, \alpha + \tau \cdot \tilde{D}_i + \tilde{u}_i)}{\text{Var}(\tilde{D}_i)} \\
&= \frac{\tau \cdot \text{Cov}(\tilde{D}_i, \tilde{D}_i) + \text{Cov}(\tilde{D}_i, \tilde{u}_i)}{\text{Var}(\tilde{D}_i)} \\
&= \tau + \frac{\text{Cov}(\tilde{D}_i, \tilde{u}_i)}{\text{Var}(\tilde{D}_i)}
\end{aligned}$$

Key OLS assumption

$$\text{plim}\ \hat{\tau}_{\text{OLS}} = \tau + \frac{\text{Cov}(\tilde{D}_i, \tilde{u}_i)}{\text{Var}(\tilde{D}_i)}$$

- Key identification comes from: $\text{Cov}(\tilde{D}_i, \tilde{u}_i) = 0$
  - Conditional on X_i, no relationship between D_i and u_i.
- Note: u_i is a function of X_i and $Y_i(d)$.
  - $u_i = Y_i(0) - \alpha - X_i'\beta$ when $D_i = 0$
  - $u_i = Y_i(1) - \alpha - \tau - X_i'\beta$ when $D_i = 1$
  - $\rightsquigarrow$ condition on X_i, only variation in u_i comes from $Y_i(d)$
- No unmeasured confounding implies this assumption:

$$D_i \perp\!\!\!\perp \big(Y_i(1), Y_i(0)\big) | X_i \implies D_i \perp\!\!\!\perp u_i | X_i \implies \text{Cov}(\tilde{D}_i, \tilde{u}_i) = 0$$

Omitted variable bias

- What happens when this is violated? Suppose that there is one omitted variable (residualized from X_i):

$$\tilde{u}_i = \lambda \tilde{L}_i + \omega_i$$

- We'll assume that if we could measure L_i, then no unmeasured confounding would hold.
- Leads to inconsistency in the OLS estimator:

$$\text{plim } \hat{\tau}_{\text{OLS}} = \tau + \lambda \frac{\text{Cov}(\tilde{D}_i, \tilde{L}_i)}{\text{Var}(\tilde{D}_i)}$$

- Bias here is terms multiplied together:
  1. coefficient on L_i, (λ)
  2. the coefficient of regression of D_i on L_i also controlling for X_i

4/ Estimating causal effects under no unmeasured confounders

Basic approach to estimation

- Remember the usual approach to estimating the ATE with covariates.
- Stratification:
  - Stratify the units by the covariates
  - Calculate CATE within these strata
- Standardization/direct adjustment:
  - Average the CATEs across the strata to get ATE
- How to create strata when X has continuous components?
  - If X is discrete with only a few levels, can use the exact values of X.
  - Otherwise, we may have to subclassify/coarsen the data.

Classic example: cigars/pipes versus cigarettes

- $D_i = 1$ for pipe/cigar smokers, $D_i = 0$ for cigarette smokers.
- Y_i = death in the first year of follow-up
- Naive positive effect: cigar/pipe smokers more likely to die.
  - What's the confounder here? Age!
  - Pipe/cigar smokers much older than cigarette smokers.
- Cochran's approach: stratify based on coarsened age:
  - Divide age into k strata: $S_i \in s_1, s_2, \ldots, s_k$
  - s_1 might be 18-25, s_2 might be 26-35, and so on.
  - Calculate effect within strata and aggregate.
- Key assumption: no unmeasured confounders using stratified version of age.

$$D_i \perp\!\!\!\perp \Big(Y_i(0), Y_i(1)\Big) | S_i$$

Stratification on the propensity score

- What about when X has has many dimensions?
- Curse of dimensionality: there will be very few, if any, units in a given stratum of X_i.
- Stratify on a low-dimensional summary, the propensity score:

$$e(x) = \mathbb{P}[D_i = 1 | X_i = x]$$

  - PS $=$ unit's probability of being treated, conditional on X_i

- For a particular unit, this is $e(X_i) = \mathbb{P}[D_i = 1 | X_i]$
- Rosenbaum and Rubin (1983) showed that:

$$D_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big) \,|\, X_i \implies D_i \perp\!\!\!\perp \big(Y_i(0), Y_i(1)\big) \,|\, e(X_i)$$

  - $\rightsquigarrow$ stratifying on e_i is the same as stratifying on the full X_i.

Propensity score as balancing score

- The propensity score is actually a balancing score, which means that

$$D_i \perp\!\!\!\perp X_i \,|\, e(X_i)$$

- Conditional on the propensity score, treatment is independent of the covariates.
  - Treatment status is said to be balanced
  - $f(X_i|D_i = 1, e(X_i)) = f(X_i|D_i = 0, e(X_i))$
- Of course, we have to know the true PS to have all these results work!

Estimating the propensity score

- Of course, in observational studies, we don't know the propensity score.
- We would run a parametric model with parameters γ to estimate the propensity scores:

1. Estimate $\hat{\gamma}$
2. Create $\hat{e}_i = \Pr[D_i = 1 | X_i; \hat{\gamma}]$

- For instance, in R, we could easily calculate the propensity scores using the `glm` function:

```
pscores <- glm(treat ~ var1 + var2 + var3, data = mydata,
    family = binomial())$fitted.values
```

Propensity score specifics

- What variables do we include in the propensity score model?
  - Any set of variables that blocks all the backdoor paths from D_i to Y_i.
- Check balance within strata of $\hat{e}_i$. Covariates should be balanced:

$$f(X_i|D_i = 1, \hat{e}_i) = f(X_i|D_i = 0, \hat{e}_i)$$

- Can also use automated/nonparametric tools for estimating $\hat{e}_i$.
  - Covariate Balancing Propensity Scores (Imai and Ratkovic)

Stratifying by the propensity score

- How will we use the propensity score?
  - Matching (next week), Weighting (two weeks), Regression (three weeks)
- Today: coarsening the propensity score and stratifying
- Choose boundary points: $0 = b_0 < b_1 < \ldots < b_{K-1} < b_K = 1$.
- Create block indicators:

$$B_i(k) = \begin{cases} 1 & \text{if } b_{k-1} < \hat{e}(X_i) < b_k, \\ 0 & \text{otherwise} \end{cases}$$

- Calculate within-strata effect estimates:

$$\tau_k = \mathbb{E}[Y_i | D_i = 1, B_i(k) = 1] - \mathbb{E}[Y_i | D_i = 0, B_i(k) = 1]$$

Standardization/direct adjustment

- We calculated the CATEs for each strata of the PS, τ_k.
- We can use law of iterated expectations to back out the ATE.
- Take the average of the CATEs over the distribution of X:

$$\tau = \sum_{k=1}^{K} \tau_k \mathbb{P}[B_i(k) = 1]$$

- Note that $\mathbb{P}[B_i(k) = 1]$ is just the proportion of units in block k:

$$\mathbb{P}[B_i(k) = 1] = \frac{\sum_{i=1}^{N} B_i(k)}{N}$$

5/ Wrapping Up

Summary

- Defined observational studies
- Defined confounding and assessed when no unmeasured confounding holds
- Saw how no unmeasured confounding helps with OLS
- Saw how to estimate causal effects under no unmeasured confounding using the propensity score.

Next few weeks

- Learn how to estimate causal effects under no unmeasured confounders via:
  - Matching
  - Weighting
  - Regression
- Then we move onto situations where no unmeasured confounders is violated.