Causal Inference from Observational Data

1. Natural Experiments
   1. Instrumental Variables regression involves two equations. This is also two-stage least squares.
      1. First stage is the ITTD in the case of noncompliance. Equivalent to regressing the treatment dosage D on the instrument Z because it tells us how much higher the take-up rate in treatment than in control. This is the regression of .
         1. The first-stage regression of tells us that when , the average/“fitted” value of D is 0.787 and when , the fitted value of D is 0.046.
         2. Include covariates because they can help with statistical power in the estimated causal effect in the second stage. Can help us reduce omitted-variable bias in the first stage.
            1. Covariates you want to include in the final regression must also be included in the first stage.
      2. Second stage is regressing the test-score outcome on the fitted values of D from the first-stage regression. This is .
         1. Divide ITT/ITTD is the equivalent of running the second-stage regression ().
         2. The fitted values of D () retain only the “clean,” randomized variation in D, but get rid of all the dirty variation in D.
         3. Can put covariate to help eliminate selection bias resulting from the imperfect instrument.
      3. The reduced-form equation reduces these two equations down to a single equation: the regression of , estimates the intent-to-treat.

|  |  |
| --- | --- |
| **Advantages** | **Disadvantages** |
| Save the time of running an experiment | Administration of randomizing can be difficult (people select, randomization fails, etc.) |
| Estimation is straightforward. IV/2SLS is very similar to OLS. | No control over the design, data collection, execution, block, or attrition |
| Estimation packages like “ivreg” in R give correct standard errors. | Might not find a natural experiment on the topic you care about (good natural experiments can be as rare as unicorns) |

* 1. External and internal validity
     1. **External validity/generalizable**: whether the results of a study apply to a distinct subject pool
     2. **Internal validity/valid/unbiased**: whether the study produces an unbiased estimate of the desired causal effect.

1. **Regression Discontinuity**: regression slope changes dramatically at a certain day. Can use a quadratic regression or allow the regression to change slope right at the discontinuity. Care most about the discrete jump in Y at the age of discontinuity.
   1. Types of regression discontinuity
      1. **Parametric regression discontinuity**: we estimate regression lines or curves and have a shift due to the treatment dummy variable
      2. **Nonparametric regression discontinuity**: we avoid assuming some parametric form, in case that might introduce specification error. Instead, we look at the simple change in Y in a narrow window around the threshold
         1. **Bandwidth**: width of this window. Results are not very sensitive to the width of window. A narrow bandwidth gives a much cleaner regression discontinuity (we think all other things are equal on both sides of the threshold) but cuts down on the amount of data, which decreases precision.
   2. Strengths and Weaknesses of Regression Discontinuity Designs
      1. “Sorting” at cutpoint.
         1. In ante/before, people are aware of the threshold and can try to get above it. In post/after, there is manipulation which makes analysis difficult (e.g., say “just loss” instead of “just won”)
         2. Tests for sorting
            1. Covariate balance on either side of cut point.
            2. Smoothness (McCrary test)
            3. Check for proportion of observations in running variable
            4. Passage of regression discontinuity check doesn’t guarantee assumptions hold
2. Differences in Differences: technique used when a natural experiment takes place over time, in a before-after setting, and when we don’t have a randomized control group. We have a group that is similar to the treated group, but not guaranteed to be identical.
   1. DID Regressions: Subtract the before-after difference in the treatment group from the before-after difference in the pseudo-control group.
      1. Need to satisfy the “common trends” assumption.
   2. Strengths and Weaknesses of DID

|  |  |
| --- | --- |
| **Strengths** | **Weaknesses** |
| Simple tool: compute differences directly or do OLS regression with clustered standard errors | Causal inference hinges on assignment mechanism |
| If we have covariates that might predict differences in Y, we can include them to try to improve precision | Common/parallel trends assumption is a strong assumption |
| Causal effects are much stronger and more believable than with the simple before/after differences or simple differences across non-randomized comparison groups |  |

* 1. Synthetic Controls as Generalized Difference in Differences. Synthetic control attempts to use relatively simple ML framework to create untreated units that are similar to treated units. Weight other units so that it would look more similar to the treated unit.
     1. Learn a set of weights of untreated units that produce a close approximation to pre-treatment outcomes of the treated unit. Fix those weights, and then make predictions into the post-treatment time period.
     2. There are many weightings that can produce pre-treatment similarly. Use a reasoning system similar to randomization inference to choose the right weighting scheme. Compute many possible counterfactuals, and evaluate how many are more extreme than what is observed.
     3. Limitations of synthetic controls
        1. Do not solve the possibility of a lingering confounder
        2. Like credit default options
           1. If there is nothing systematically leading control units to be different than treatment, then we’re creating a composite counterfactual that manages confounding risk
           2. If there is something systematically leading controls to be different, then we’re in trouble.
        3. Limited to the matching algorithms capabilities.

|  |  |
| --- | --- |
| **Strengths** | **Weaknesses** |
| Get closer to parallel trends assumption in the pre-treatment time period | Method of finding weights overfits to sample data. This is controllable but must be consciously done |
| Follows a strategy that is familiar from machine learning based approaches | Possible confounding variables remain |
|  | No clear statement about who we are comparing to whom, and why |