Matching and Synthesis

EDLD 650: Week 9

David D. Liebowitz

Agenda

1. Roadmap and Goals (9:00–9:10)

2. Umansky & Dumont and DARE #4 (9:10–10:20)

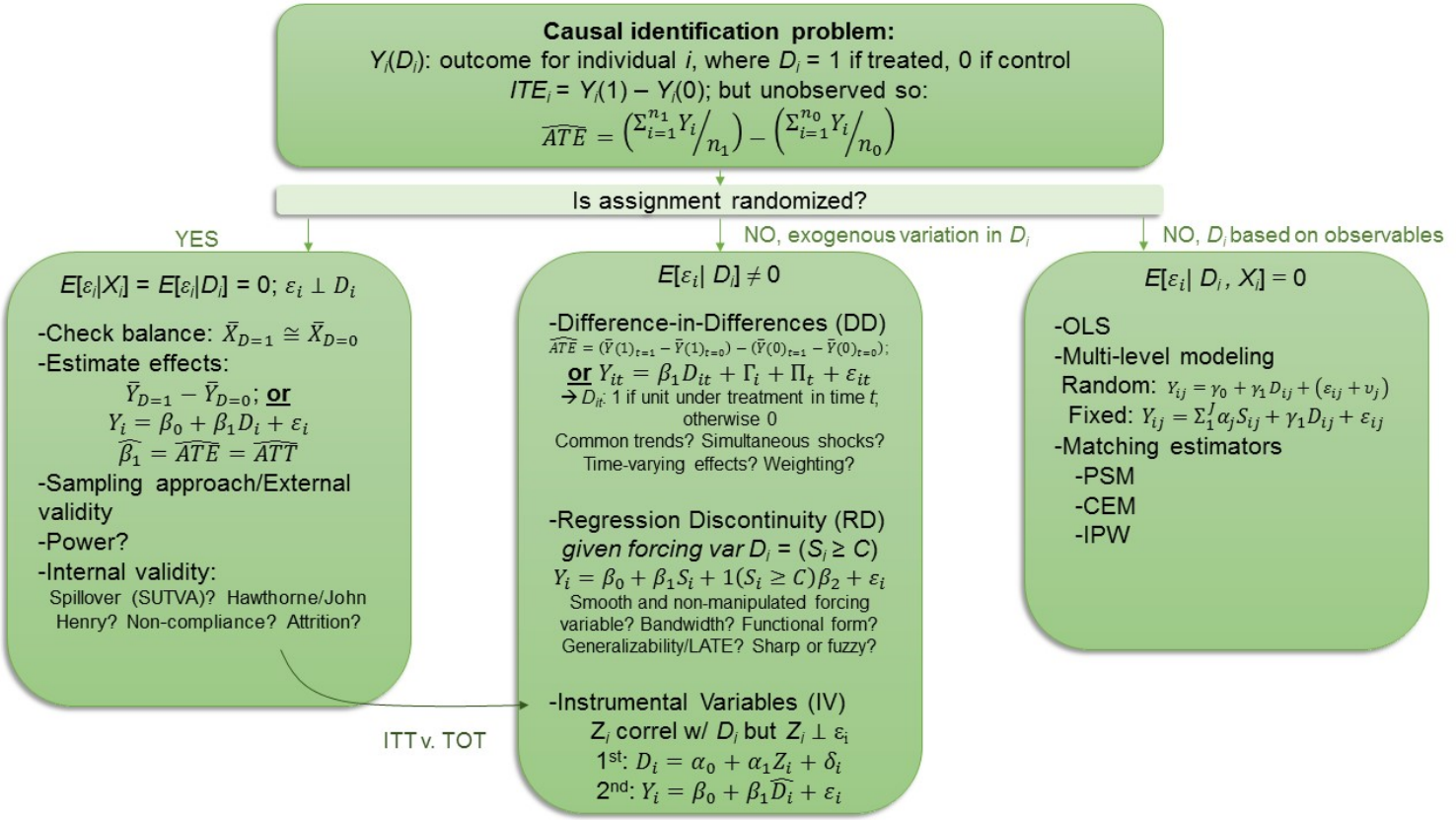
3. Break (10:20–10:30)

4. Presenting (10:30–10:45)

5. Review and synthesis (10:45–11:40)

5. Wrap-up (11:40–11:50)

Roadmap


Causal identification problem:
$Y_i(D_i)$: outcome for individual i, where D_i = 1 if treated, 0 if control
$ITE_i = Y_i(1) - Y_i(0)$; but unobserved so:
$\widehat{ATE} = \left(\Sigma_{i=1}^{n_1} Y_i / n_1\right) - \left(\Sigma_{i=1}^{n_0} Y_i / n_0\right)$
Is assignment randomized?
YES
NO, exogenous variation in D_i
NO, D_i based on observables
$E[\varepsilon_i|X_i] = E[\varepsilon_i|D_i] = 0$; $\varepsilon_i \perp D_i$
-Check balance: $\bar{X}_{D=1} \cong \bar{X}_{D=0}$
-Estimate effects:
$\bar{Y}_{D=1} - \bar{Y}_{D=0}$; or
$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$
$\widehat{\beta_1} = \widehat{ATE} = \widehat{ATT}$
-Sampling approach/External validity
-Power?
-Internal validity:
Spillover (SUTVA)? Hawthorne/John Henry? Non-compliance? Attrition?
ITT v. TOT
$E[\varepsilon_i| D_i] \neq 0$
-Difference-in-Differences (DD)
$\widehat{ATE} = (\bar{Y}(1)_{t=1} - \bar{Y}(1)_{t=0}) - (\bar{Y}(0)_{t=1} - \bar{Y}(0)_{t=0})$;
or $Y_{it} = \beta_1 D_{it} + \Gamma_i + \Pi_t + \varepsilon_{it}$
→ D_{it}: 1 if unit under treatment in time t, otherwise 0
Common trends? Simultaneous shocks? Time-varying effects? Weighting?
-Regression Discontinuity (RD)
given forcing var D_i = ($S_i \geq C$)
$Y_i = \beta_0 + \beta_1 S_i + 1(S_i \geq C)\beta_2 + \varepsilon_i$
Smooth and non-manipulated forcing variable? Bandwidth? Functional form? Generalizability/LATE? Sharp or fuzzy?
-Instrumental Variables (IV)
Z_i correl w/ D_i but $Z_i \perp \varepsilon_i$
1st: $D_i = \alpha_0 + \alpha_1 Z_i + \delta_i$
2nd: $Y_i = \beta_0 + \beta_1 \widehat{D_i} + \varepsilon_i$
$E[\varepsilon_i| D_i, X_i] = 0$
-OLS
-Multi-level modeling
Random: $Y_{ij} = \gamma_0 + \gamma_1 D_{ij} + (\varepsilon_{ij} + \upsilon_j)$
Fixed: $Y_{ij} = \Sigma_1^J \alpha_j S_{ij} + \gamma_1 D_{ij} + \varepsilon_{ij}$
-Matching estimators
-PSM
-CEM
-IPW

Goals

1. Describe conceptual approach to matching analysis
2. Assess validity of matching approach and what selection on observable assumptions implies
3. Conduct matching analysis in simplified data using both coarsened-exact matching (CEM) and propensity-score matching (PSM)
4. Synthesize strategies for causal inference and articulate value of each strategy given a particular data-generating process

So random...

DARE-d to do it!

Student examples in class...

Break

Presenting

Causal inference presentations

With 10-15 minutes you have time for:

1. What is the focus and why is this important? (1-2 slides)
2. What are the research questions? (1 slide)
3. What are the key features of the sample and data (2 slides)
4. What is the methodology and research design? (2-3 slides)
5. What is the result? (3-4 slides)
6. What does this mean? (1 slide)

If you are only presenting on a proposal, cut results, but preserve/extend data and methodology (rather than motivation)

Are these any different than standard academic presentations? Yes and No

Is this particular structure just about disciplinary norms? Yes and No

Key insight: *For presentations to lay audience:* they trust you know what you are talking about, you need to convince them it is important. *For presentations to researchers:* they already believe your topic is important, you need to convince them you are right.

Review

Correlation and causation

- Causal, correlational and descriptive research are all important, but they are distinct and should be approached differently
- If you encounter a research study (or embark on your own research project) an important first consideration to ask yourself is:
  - Is this study attempting to an answer an explicitly or implicitly causal question? If so, what are its identifying assumptions?
- One framework for considering these identifying assumptions : **the potential outcomes framework**

Causal inference: Platonic ideal

Y_i^1 = potential value of outcome for i^{th} person, when treated $(D_i = 1)$

Y_i^0 = potential value of outcome for i^{th} person, when NOT treated $(D_i = 0)$

The Individual Treatment Effect (ITE) is the difference in potential outcome values between treatment and control conditions, for each individual:

$$ITE_i = Y_i^1 - Y_i^0$$

We never actually observe this!!!

The Average Treatment Effect (ATE) is the average of the individual treatment effects across all participants:

$$\hat{ATE}_i = \frac{1}{n}\sum_{i}^{n} ITE_i$$

If the ATE differed from zero, we could claim that the treatment *caused* the effect because there would be no other explanation for the differences detected between the treatment and control conditions!

Conditions of causal claims

1. Cause must precede effect in time
2. Systematic variation in levels of cause must result in corresponding variation in the effect
3. Must be able to discount all other plausible explanations

RCTs: Gold Standard

Randomly assign each participant to the Treatment (where we measure their value of Y_i^1) or Control (where we measure their value of Y_i^0) condition.

$$\hat{ATE}_i = \frac{1}{n_1}\sum_{i}^{n_1} ITE_i - \frac{1}{n_0}\sum_{i}^{n_0} ITE_i$$

- Treatment variation is exogenously and randomly assigned.
- Members of the treatment and control groups are then equivalent, on average, in the population ("equal in expectation") before the experiment begins, on every possible dimension, $\bar{\mathbf{X}}_{D=1} \approxeq \bar{\mathbf{X}}_{D=0}$
- The values of treatment variable, D, will also be completely uncorrelated with all characteristics of participants, observed and unobserved, in the population.

RCTs: Issues & assumptions

- **Randomization**: Was it successful Check balance at variable level and with omnibus F-test
- **Sample**: Representative? Sufficiently powered? for tests of heterogeneity? Pre-registered?
- **Threats**:
  - Spillover
  - Hawthorne/John Henry
  - Non-compliance
  - Attrition
  - SUTVA...the lurking monster

Questions?

DD: Classic two-period

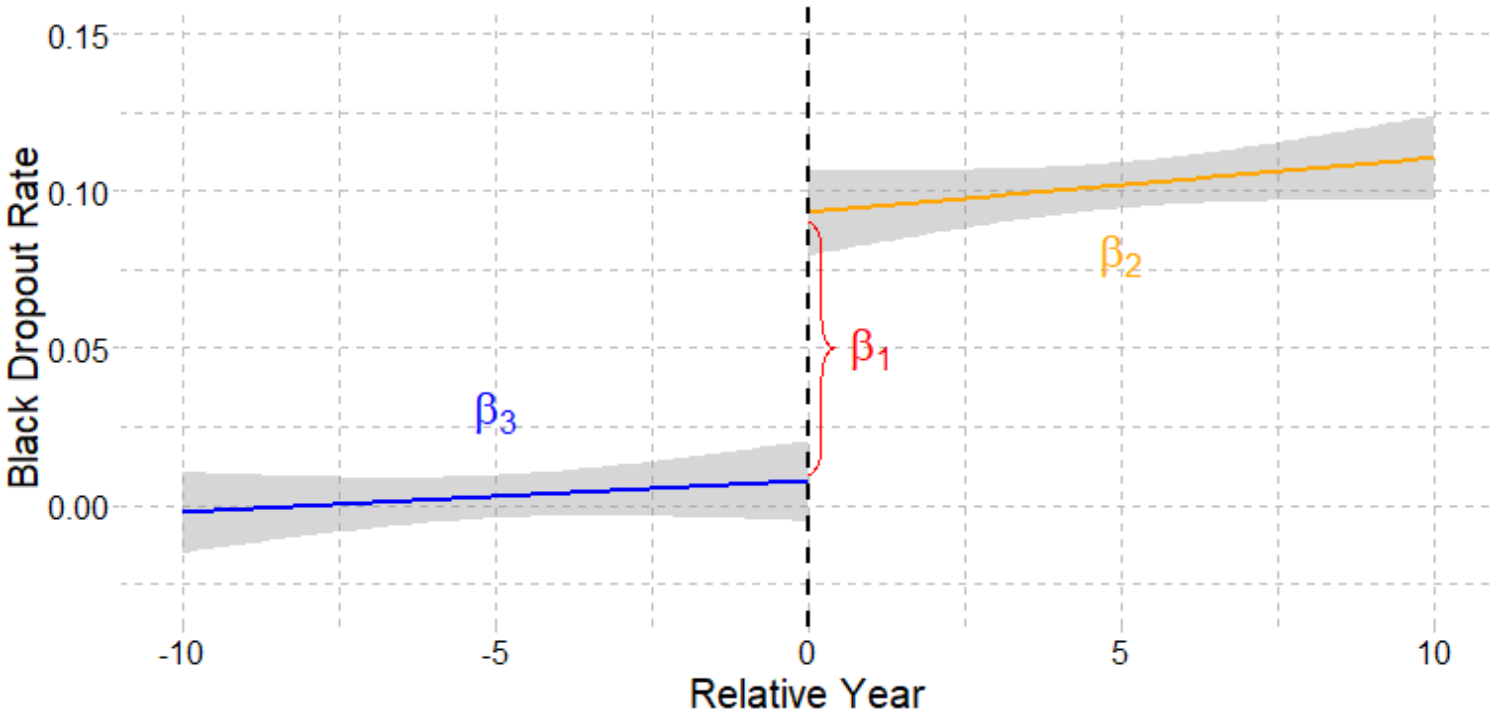
$$y_i = \alpha + \beta(\text{FATHERDEC}_i \times \text{OFFER}_i) + \delta\text{FATHERDEC}_i + \theta\text{OFFER}_i + \upsilon_i$$

DD: Two-way fixed effects

$$\text{DROPOUT\_BLACK}_{jt} = \beta_1 \text{UNITARY}_{jt} + \Gamma_j + \Pi_t + \epsilon_j$$

DD: Time-variant effects

$$\text{DROPOUT\_BLACK}_{jt} = \beta_1 \text{UNITARY}_{jt} + \beta_2 (\text{UNITARY} \times \text{YEAR\_CENT})_{jt} + \beta_3 \text{YEAR\_CENT}_{jt} + \Gamma_j + \Pi_t + \epsilon_j$$



DD: Event Study

$$\begin{aligned}\text{DROPOUT\_BLACK}_{jt} =& \beta_1 \text{pre}_{jt}^{-n} + \beta_2 \text{pre8} + \beta_3 \text{pre7}_{jt} + \dots \\ &+ \beta_m \text{post0}_{jt} + \dots + \beta_n \text{post}_{jt}^{n} + \Gamma_j + \Pi_t + \epsilon_j\end{aligned}$$

Could also write as:

$$\text{DROPOUT\_BLACK}_{jt} = \sum_{t=-10}^{n} 1(\text{t} = \text{t}_j^*)\beta_t + \Gamma_j + \Pi_t + \epsilon_j$$

the assumptions and design structure are the same across all these!

DD: Assumptions

1. Not-treated (or not-yet-treated) units are **valid counterfactuals**
    - Parallel trends?
    - Selection into treatment?
2. There are no **simultaneous shocks** or unobserved **secular trends**
    - Other observed and unobserved events or patterns?
3. Appropriate weighting
    - See "further reading" for latest

Questions?

Regression Discontinuity

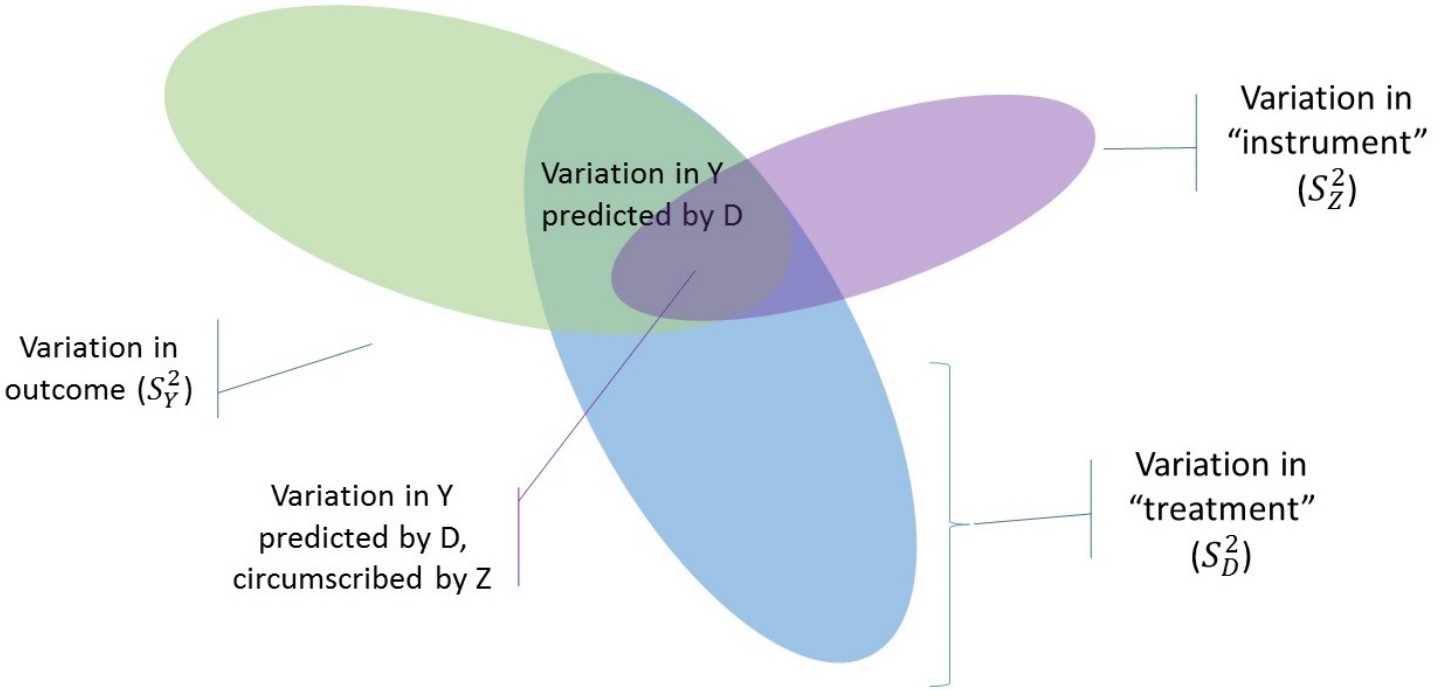
$$p(COLL_i = 1) = \beta_0 + \beta_1 TESTSCORE_i + 1(TESTSCORE_i \geq 60)\beta_2 + \varepsilon_i$$

RD: Issues and Assumptions

1. A Local Average Treatment Effect (LATE)
    - bandwidth selection (bias v. variance tradeoff)
2. Functional-form specification
3. Forcing variable predicts treatment discontinuously
4. No manipulation
5. No bunching

Questions?

Instrumental variables



IV estimate: ratio of area of *overlap of* Y *and* Z to area of *overlap of* D *and* Z. Depends entirely on variation in Z that predicts variation in Y and D:

$$\hat{\beta}_1^{IVE} = \frac{S_{YD}}{S_{DZ}}$$

a Local Average Treatment Effect

2SLS IV set-up

1st stage:

Regress the endogenous treatment (D_i) on instrumental variable (Z_i) using OLS:

$$D_i = \alpha_0 + \alpha_1 Z_i + \nu_i$$

Obtain the *predicted values* of the treatment $(\hat{D}_i)$ from this fit.

2nd stage:

Regress the outcome (Y_i) on the predicted values of the treatment $(\hat{D}_i)$:

$$Y_i = \beta_0 + \beta_1 \hat{D}_i + \varepsilon_i$$

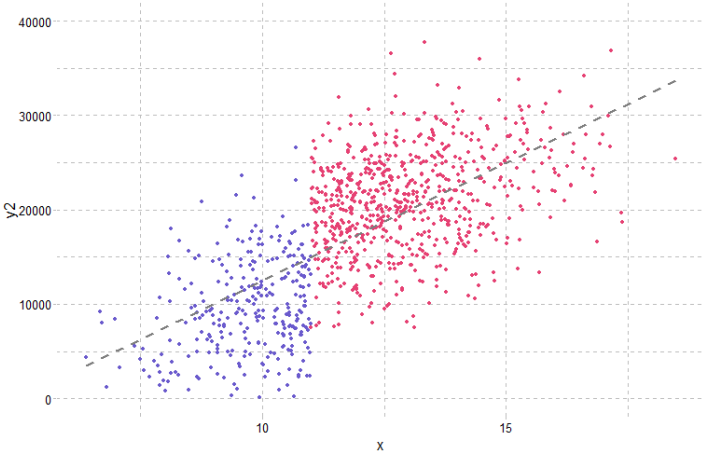
Valid instruments

1. Instrument (Z_i) must be correlated with treatment (D_i), *but*
2. Instrument (Z_i) must be orthogonal $(\perp)$ to all other determinants of the outcome (Y_i)
    - Another way of saying it must be uncorrelated with the residuals (ε_i)
3. Instrument must be related to the outcome *only* through the treatment
    - This is known as the **exclusion restriction**

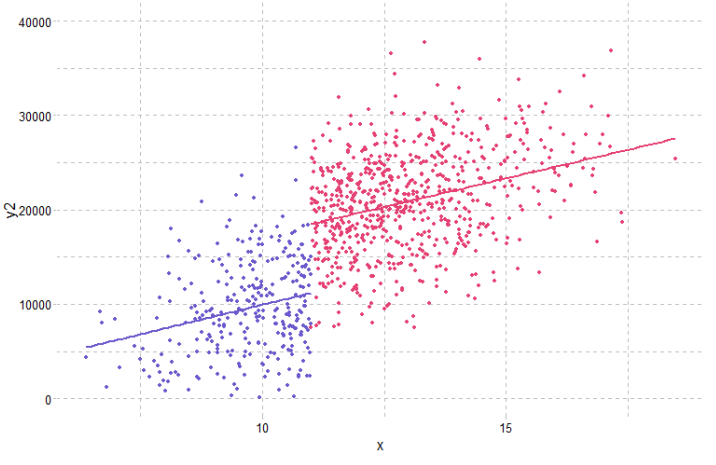
Questions?

Matching

Ignore biased observed relationship



Estimate treatment effect absent bias



Big idea: if we were sure we knew that the only factor driving selection into treatment was individuals' membership in this group:

- We can ignore overall point cloud and refuse to estimate the biased Y|X slope
- Instead, conduct analysis within each subsidiary point clouds
  - Obtain estimates of treatment effect within each point cloud
  - Average to obtain overall *unbiased* estimate of treatment effect of more educational attainment

Matching routines

Phase I:

1. Investigate the selection process explicitly by fitting a "selection model":
   - Could use exact, coarsened exact, etc. family of approaches
   - Or fit a logistic model, with treatment group membership as outcome, and predictors you believe describe the process of selection explicitly:

$$D_i = \frac{1}{1 + e^{-\mathbf{X}_i \theta_i}}$$

2. Use selection model to estimate fitted probability of selection into treatment $(\hat{p})$ for each participant

Phase II:

1. Enforce overlap in sample
2. Check balancing condition has been satisfied
3. Estimate treatment effect in matched (weighted) sample

Questions?

Putting it all together

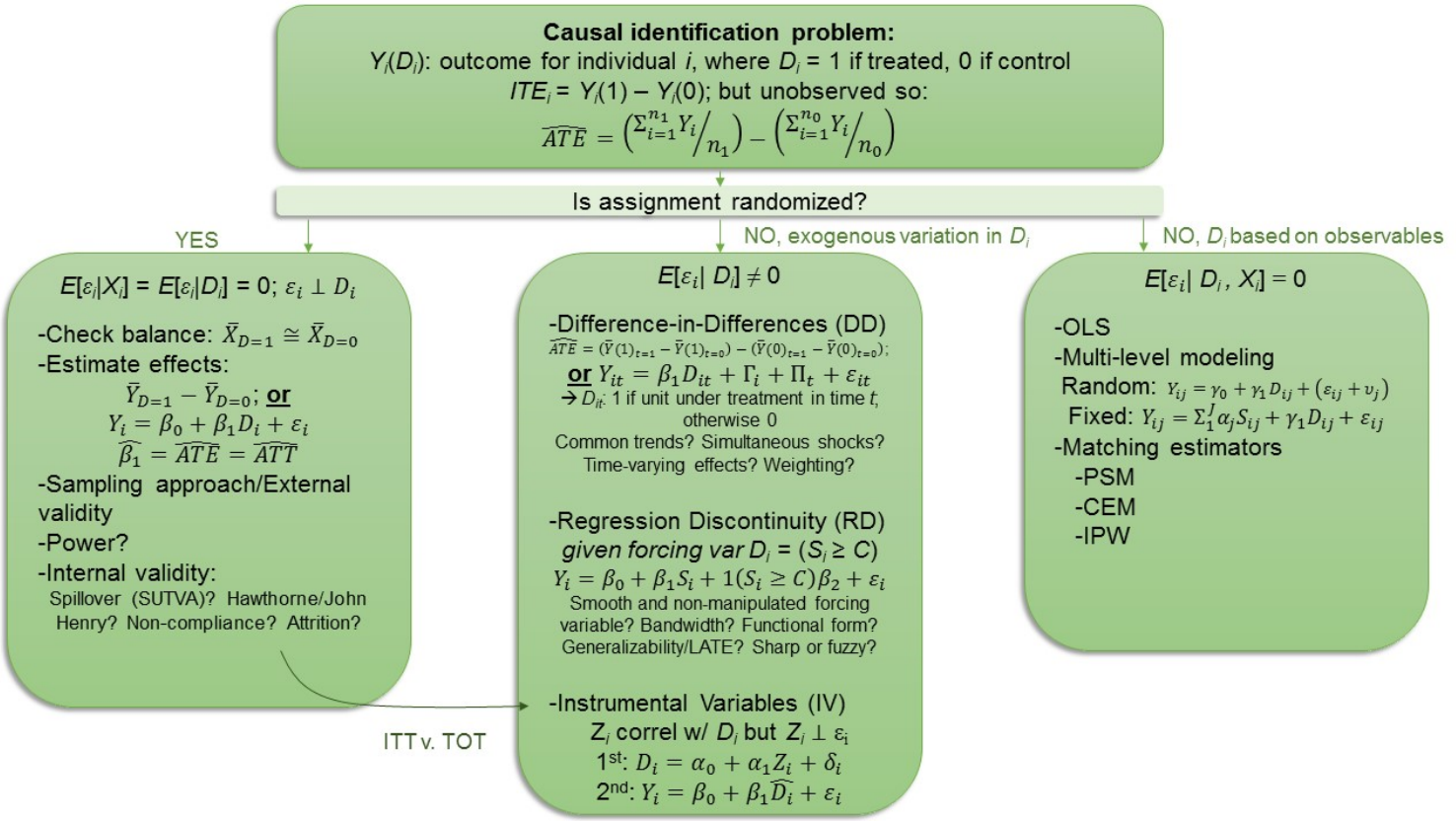
Is there a hierarchy of causal inference strategies? NO! Each research design is the best design and has the highest degree of internal validity, *given the data-generating process*.

A regression discontinuity design with poorly met assumptions (*e.g., manipulation or even insufficient observations on one side of discontinuity that prevent modeling appropriate functional form*) is not a "better" design than a well-justified matching study. An RCT with great internal validity might not teach us as much as a generalizable RD.

Always conduct research designed to make causal inference with a mix of "humble and hotshot" attitude

Continue your education at UO: **EC523, EC524, EC525, EC607, SOC613**, etc.
Beyond UO: **Mixtape Sessions, MethodsU, ICPSR, IES**, etc.

Can you explain this figure?



Logistics and wrap-up

Goals

1. Describe conceptual approach to matching analysis
2. Assess validity of matching approach and what selection on observable assumptions implies
3. Conduct matching analysis in simplified data using both coarsened-exact matching (CEM) and propensity-score matching (PSM)
4. Synthesize strategies for causal inference and articulate value of each strategy given a particular data-generating process

To-Dos

Week 10: Presentate! 😄

Order (randomly generated):

1. Yitong
2. Xiaoqi
3. Seulbi
4. Havi
5. Eunji
6. Yessy
7. Tony
8. Janette
9. Brittany

To-dos

Readings:

- **Optional**: *MM* Ch. 13 and 14

Final Research Project

- Presentation, March 11
- Paper, March 20 (submit March 13 for feedback

Feedback

Student Experience Survey