# Political Methodology II
# Section: SOO Sensitivity + Intro Difference in Differences

Apoorva Lal

February 19, 2021

Stanford University

# Roadmap

- Pset Review
- Review of Selection on Observables
- Regression Adjustment
- Falsification Tests
  - Falsification Tests: Placebo outcome with zero effect
  - Multiple Control Groups
- Sensitivity Analysis
- Difference-in-differences

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

Sensitivity Analysis

Difference-in-differences

# Notes from Previous Pset

- What is an identification proof?

# Notes from Previous Pset

- What is an identification proof?

$$
\begin{aligned}
\overbrace{\text{Diff. in means}}^{\text{Estimator}} &= \overbrace{E[Y_i|D=1]}^{\text{Observed}} - \overbrace{E[Y_i|D=0]}^{\text{Observed}} \\
&= E[Y_{1i}|D=1] - E[Y_{0i}|D=0] \text{ (1)} \\
&= E[Y_{1i}] - E[Y_{0i}] \text{ (2)} \\
&= \underbrace{E[Y_{1i} - Y_{0i}]}_{\text{Estimand}}
\end{aligned}
$$

In an identification proof you either start with the thing you can measure in your data and arrive at the quantity of interest (defined by potential outcomes), or the other way around. Which assumptions correspond to steps (1) and (2) in the proof?

# Notes from Previous Pset

- What is an identification proof?

$$\begin{aligned}
\overbrace{\text{Diff. in means}}^{\text{Estimator}} &= \overbrace{E[Y_i|D=1]}^{\text{Observed}} - \overbrace{E[Y_i|D=0]}^{\text{Observed}} \\
&= E[Y_{1i}|D=1] - E[Y_{0i}|D=0] \text{ (1)} \\
&= E[Y_{1i}] - E[Y_{0i}] \text{ (2)} \\
&= \underbrace{E[Y_{1i} - Y_{0i}]}_{\text{Estimand}}
\end{aligned}$$

In an identification proof you either start with the thing you can measure in your data and arrive at the quantity of interest (defined by potential outcomes), or the other way around. Which assumptions correspond to steps (1) and (2) in the proof? (1) SUTVA, (2) Random assignment: $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i$.

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

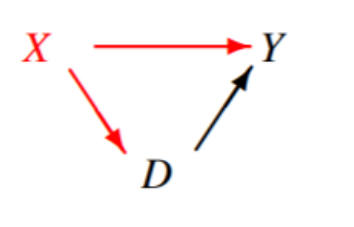
Sensitivity Analysis

Difference-in-differences

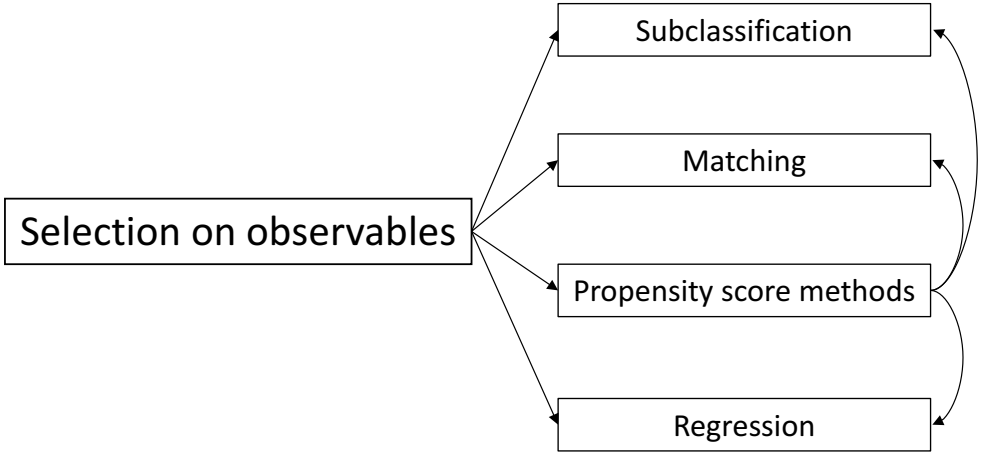
# Idea of Selection on Observables

- Causal inference is all about understanding the treatment assignment mechanism.
- Selection on observables says that once we condition on some observable covariates, treatment assignment was as good as random.
- Requires substantive justification – if there's selection on unobservables (particularly the potential outcomes), matching won't help.
- Estimation methods include subclassification, matching, propensity score methods, regression, and hybrid approaches.

# Graphs to visualise conditioning




- **directed acyclic graphs** (DAGs) are a popular (complementary/alternative) framework for causal inference
- especially for observational settings, they stipulate specific 'adjustment sets' to condition on to justify selection on observables
- 'Backdoor criterion': condition on variables that affect both $Y$ and $D$
- See Morgan and Winship and Cunningham for details


Selection on observables
Subclassification
Matching
Propensity score methods
Regression

# Commonalities and Differences

Commonalities among these estimation strategies:

- They're ways of *imputing* the counterfactual potential outcome for treatment units by adjusting for covariates.
- They all impute this counterfactual by using a *weighted average* of observed outcomes for other units.
- Matching and subclassification are nonparametric estimators of the counterfactual, while linear regression is a parametric / semi-parametric estimator.

# Commonalities and Differences

Commonalities among these estimation strategies:

- They're ways of *imputing* the counterfactual potential outcome for treatment units by adjusting for covariates.
- They all impute this counterfactual by using a *weighted average* of observed outcomes for other units.
- Matching and subclassification are nonparametric estimators of the counterfactual, while linear regression is a parametric / semi-parametric estimator.

The main difference is that they use different methods of picking the weights – making different assumptions about the functional form relating the covariates, treatment, and outcomes. Review paper from syllabus


NONPARAMETRIC ESTIMATION OF AVERAGE TREATMENT EFFECTS UNDER EXOGENEITY: A REVIEW*

Guido W. Imbens

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

Sensitivity Analysis

Difference-in-differences

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator.

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator. Using our partialling out formula, we can write $\delta_R$ as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)}$$

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator. Using our partialling out formula, we can write $\delta_R$ as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]}$$

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator. Using our partialling out formula, we can write $\delta_R$ as

$$\delta_R = \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]}$$

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator. Using our partialling out formula, we can write $\delta_R$ as

$$\begin{aligned}\delta_R &= \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]} \\ &= \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[Var[D_i|X]]}\end{aligned}$$

# Regression as a type of propensity weighting (Angrist and Pischke, p.83)

Let's our estimand for the ATE comes from a regression of the form

$$\mathbb{E}[Y_i|D_i, X_i] = \alpha + X_i^T\beta + \delta_R D_i$$

where $X_i$ is a vector of covariates and $D_i$ is the treatment indicator. Using our partialling out formula, we can write $\delta_R$ as

$$\begin{aligned}\delta_R &= \frac{Cov(\tilde{D}_i, Y_i)}{Var(\tilde{D}_i)} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])Y_i]}{\mathbb{E}[\mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]]} \\ &= \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[Var[D_i|X]]} = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]}\end{aligned}$$

# Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

# Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates $X_i$.

# Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates $X_i$. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide?

# Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates $X_i$. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide? Constant treatment effects across strata of $X_i$.

# Regression as a type of propensity weighting

Now let's substitute in the definition of the propensity score function, $p(X_i) = \mathbb{E}[D_i|X_i]$.

$$\delta_R = \frac{\mathbb{E}[(D_i - \mathbb{E}[D_i|X])Y_i]}{\mathbb{E}[\mathbb{E}[D_i|X_i](1 - \mathbb{E}[D_i|X_i])]} = \frac{\mathbb{E}[(D_i - p(X_i))Y_i]}{\mathbb{E}[p(X_i)(1 - p(X_i))]}$$

We can see that this estimand is equal to the weighted propensity score estimand:

$$\delta_R = \mathbb{E}\Big[\frac{p(X_i)(1 - p(X_i))}{\mathbb{E}[p(X_i)(1 - p(X_i))]}\Big(\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big)\Big]$$

where $\frac{p(X_i)(1-p(X_i))}{\mathbb{E}[p(X_i)(1-p(X_i))]}$ is the weight for observations with covariates $X_i$. Compare this to the unweighted propensity score estimand:

$$\delta_{ATE} = \mathbb{E}\Big[\frac{Y_i D_i}{p(X_i)} - \frac{Y_i(1 - D_i)}{1 - p(X_i)}\Big]$$

When will these two coincide? Constant treatment effects across strata of $X_i$. Otherwise, OLS does not estimate ATE/ATT.

# OLS with heterogeneous treatment effects


In this case your estimat**or** is picking your estim**and**.

# Hybrid Approaches: AIPW Estimator

$$\widehat{\tau}_{DR} = \frac{1}{N}\sum_{i=1}^{n}\left(\frac{D_i(Y_i - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \frac{(1-D_i)(Y_i - \hat{\mu}_0(X_i))}{1-\hat{\pi}(X_i)} + \{\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)\}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left[\overbrace{\hat{\mu}_1(\boldsymbol{X}_i)}^{\text{Regression}} + \overbrace{\frac{D_i(Y_i - \hat{\mu}_1(\boldsymbol{X}_i))}{\hat{\pi}(\boldsymbol{X}_i)}}^{\text{IPW}}\right]}_{\text{estimator for } \mathbb{E}[Y_i(1)]} - \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left[\hat{\mu}_0(\boldsymbol{X}_i) + \frac{(1-D_i)(Y_i - \hat{\mu}_0(\boldsymbol{X}_i))}{1-\hat{\pi}(\boldsymbol{X}_i)}\right]}_{\text{estimator for } \mathbb{E}[Y_i(0)]}$$

Fit $\widehat{\mu}$, $\widehat{\pi}$ using learner of choice.

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

Sensitivity Analysis

Difference-in-differences

# Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility

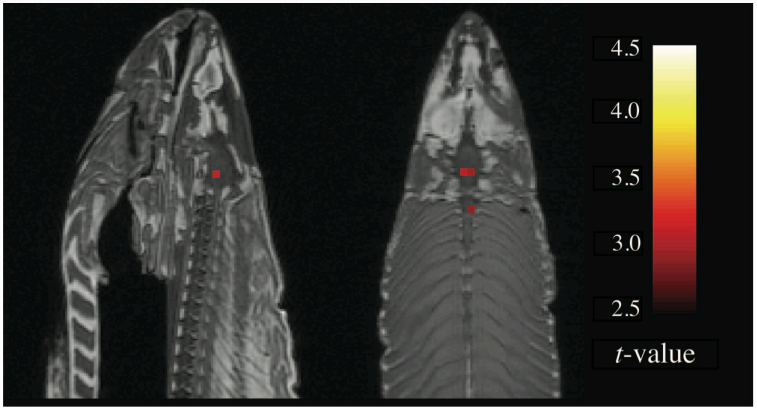
# Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility
- Falsification tests:
  - Estimating a causal effect that is known to equal zero for a **placebo treatment** or **placebo outcome**
  - If we find that the placebo effect is not zero, then the selection on observables assumption is considered less plausible
  - Can leverage multiple control groups to determine plausibility of selection on observables assumption

# Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility
- Falsification tests:
  - Estimating a causal effect that is known to equal zero for a **placebo treatment** or **placebo outcome**
  - If we find that the placebo effect is not zero, then the selection on observables assumption is considered less plausible
  - Can leverage multiple control groups to determine plausibility of selection on observables assumption
- Sensitivity Analysis:
  - How much imbalance in unobservables do we need to eliminate or sufficiently change the estimated treatment effect?

# Post-Mortem Atlantic Salmon




At the time the study was presented, between 25-40% of studies on fMRI being published were NOT using the corrected comparisons. After this group won the Ignobel, that number had dropped to 10%.

Link to study: http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf

# Placebo outcome with zero effect

- Imagine we have data on a "placebo" outcome that is known to be unaffected by the treatment
  - E.g. lags of the outcome variable that are measured before treatment
- For example, assume $D$ is realized at $t = 0$ and is ignorable conditional on a set of $T$ lags of the outcome $Y_1, Y_0 \perp\!\!\!\perp D | Y_{t=-1}, Y_{t=-2}, ..., Y_{t=-T}, X$
- Given a stability assumption we should have ignorability conditional on all lags but one:

$$Y_{t=-1} \perp\!\!\!\perp D | Y_{t=-2}, ..., Y_{t=-T}, X$$

- If we find a non-zero placebo effect for the first lag, then ignorability conditional on all lags seems not very credible (esp. with many lags).

# Estimates on Experimental JTPA Data

| | Earnings75 Outcome | | |
|---|---|---|---|
| | mean | se | t-stat |
| Simple Difference | 0.27 | 0.30 | 0.9 |
| OLS (parallel) | 0.15 | 0.22 | 0.7 |
| OLS (separate) | 0.12 | 0.22 | 0.6 |
| Propensity Score Weighting | 0.15 | 0.30 | 0.5 |
| Propensity Score Blocking | 0.10 | 0.03 | 3.4 |
| Propensity Score Regression | 0.16 | 0.30 | 0.5 |
| Propensity Score Matching | 0.23 | 0.37 | 0.6 |
| Matching | 0.14 | 0.28 | 0.50 |
| Weighting and Regression | 0.15 | 0.21 | 0.7 |
| Blocking and Regression | 0.09 | 0.02 | 3.8 |
| Matching and Regression | 0.06 | 0.28 | 0.2 |

# Estimates on CPS Data

| | Earnings75 Outcome | | |
|---|---|---|---|
| | mean | se | t-stat |
| Simple Difference | -12 | 0.68 | -18 |
| OLS (parallel) | -1.2 | 0.36 | -3 |
| OLS (separate) | -1.1 | 0.36 | -3 |
| Propensity Score Weighting | -1.2 | 0.26 | -5 |
| Propensity Score Blocking | -1.4 | 0.32 | -9 |
| Propensity Score Regression | -1.7 | 0.79 | -2 |
| Propensity Score Matching | -1.3 | 0.46 | -3 |
| Matching | -1.3 | 0.41 | -3 |
| Weighting and Regression | -1.2 | 0.24 | -5 |
| Blocking and Regression | -1.3 | 0.25 | -5 |
| Matching and Regression | -1.3 | 0.42 | -3 |

# Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment

# Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.

# Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.

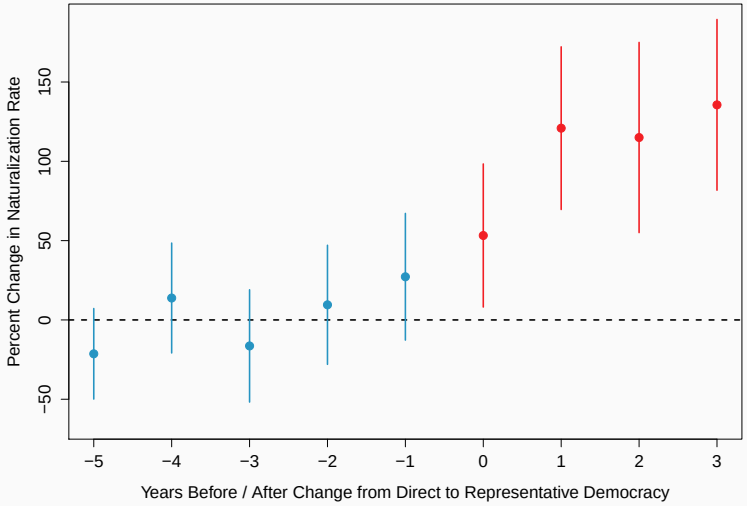
# Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.

# Placebo outcome with zero effect

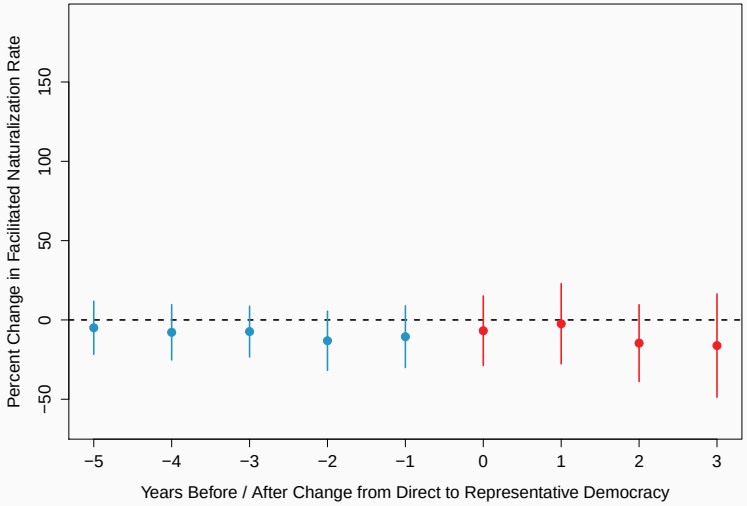
- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

# Effect on Naturalization Rate




Hainmueller and Hangartner (2015, AJPS)

# Effect on Facilitated Naturalization Rate




Hainmueller and Hangartner (2015, AJPS)

# Using Multiple Control Groups

- Imagine three groups:

$$D = \begin{cases} 1 & \text{Treatment Group} \\ 0 & \text{Control Group A} \\ -1 & \text{Control Group B} \end{cases}$$

  - E.g. treated, control participants, and eligible non-participants

- Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U$ is unobserved
- If the two control groups are expected to vary on $U$, we can bracket the treatment effect by comparing treated vs. control A and treated vs. control B.
- If the effect estimates are similar, we have some evidence that $U$ might be ignorable and therefore adjusting for $X$ is sufficient since $Y_1, Y_0 \perp\!\!\!\perp D|X$ holds
- Can also compare control A to control B and we expect a zero effect. If not, at least one control group is invalid.

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

Sensitivity Analysis

Difference-in-differences

# Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder $U$ have to be to eliminate or sufficiently change the estimated treatment effect?

# Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder $U$ have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in **Imbens (2003, AER)**:
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and $U$ and $X$ are independent.

# Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder $U$ have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in **Imbens (2003, AER)**:
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and $U$ and $X$ are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so $\gamma$ indicates strength of relationship between $U$ and $D|X$

# Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder $U$ have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in **Imbens (2003, AER)**:
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and $U$ and $X$ are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so $\gamma$ indicates strength of relationship between $U$ and $D|X$
  - $Y$ is conditionally normal with constant treatment effect $\alpha$ so $Y|X, U \sim N(\alpha D + X + \delta U, \sigma^2)$ and $\delta$ indicates strength of relationship between $U$ and $Y|X$

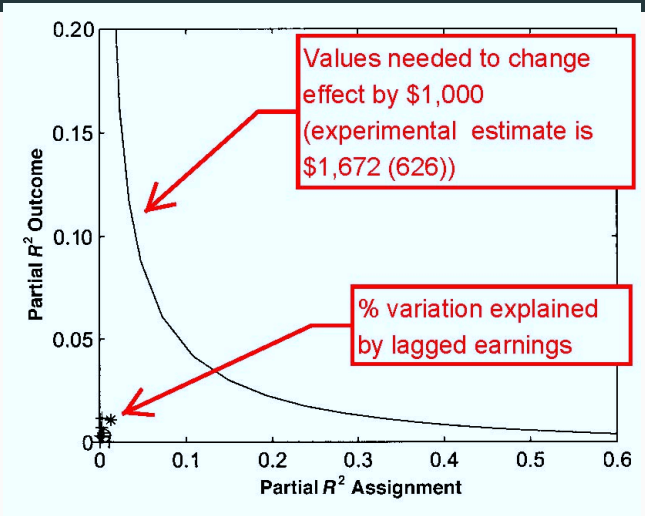
# Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder $U$ have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in **Imbens (2003, AER)**:
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and $U$ and $X$ are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so $\gamma$ indicates strength of relationship between $U$ and $D|X$
  - $Y$ is conditionally normal with constant treatment effect $\alpha$ so $Y|X, U \sim N(\alpha D + X + \delta U, \sigma^2)$ and $\delta$ indicates strength of relationship between $U$ and $Y|X$
  - Choose alternative values for $(\gamma, \delta)$ and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, , \sigma^2, \theta, \gamma, \delta)$ for fixed $(\gamma, \delta)$. If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

# Formal Sensitivity Tests

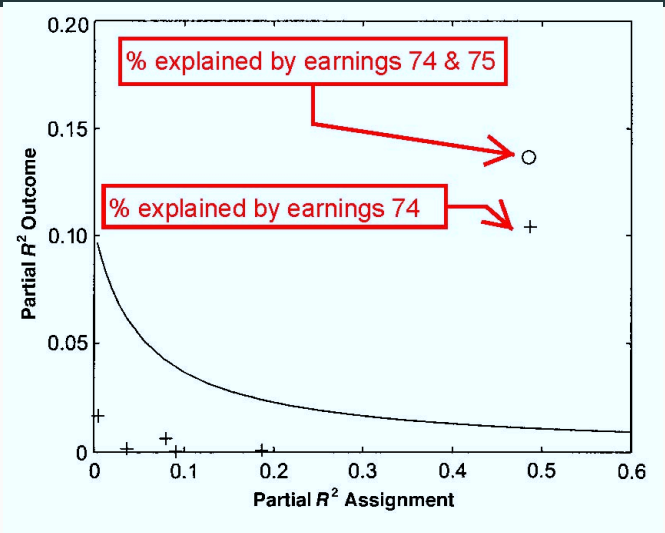
- Choose $(\gamma, \delta)$ and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, , \sigma^2, \theta, \gamma, \delta)$ for fixed $(\gamma, \delta)$.
- $\gamma$ and $\delta$ are difficult to interpret, we instead use:
  - $R^2_{Y,par}(\delta)$: % residual variation in outcome explained by unobserved covariate $U$ (above variation explained by $X$)
  - $R^2_{D,par}(\gamma)$: % residual variation in treatment assignment explained by unobserved covariate $U$ (above variation explained by $X$)
  - Try a range of values. Magnitudes should be compared to explanatory power of $X$ for $Y$ and $D$ respectively.
- Example: JTPA: Experimental data and PSID controls

# Results in Experimental Data




Imbens (2003)

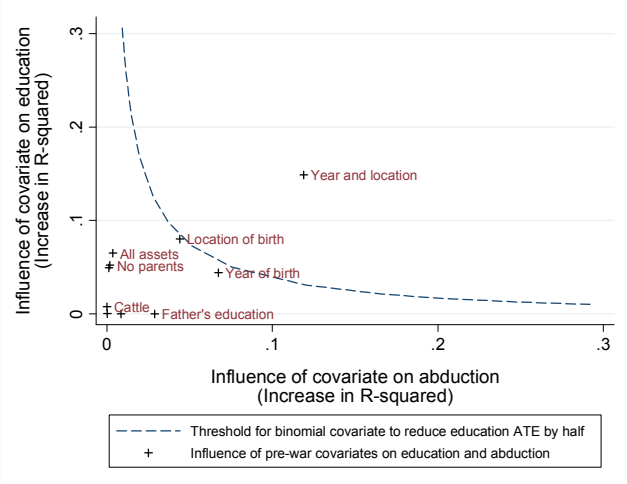
# Results with PSID Controls




Imbens (2003)

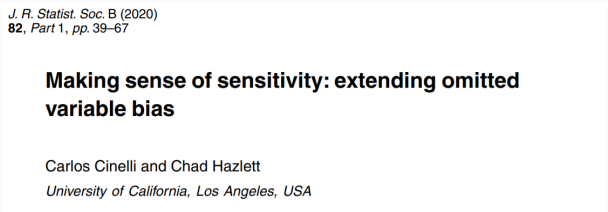
# Effect of Abduction on Education

**Figure 5: Impact of relaxing the assumption of unconfoundedness**




Blattman and Annan (2010, ReStat)

# Alternate Approaches; Active area of research



*J. R. Statist. Soc.* B (2020)
**82**, *Part* 1, *pp.* 39–67

**Making sense of sensitivity: extending omitted variable bias**

Carlos Cinelli and Chad Hazlett
*University of California, Los Angeles, USA*

http://arelbundock.com/posts/robustness_values/

**Sensitivity Analysis via the Proportion of Unmeasured Confounding**

Matteo Bonvini* Edward H. Kennedy†

December 18, 2020

# Roadmap

Pset Review

Review of Selection on Observables

Regression Adjustment

Falsification Tests

- Falsification Tests: Placebo outcome with zero effect
- Multiple Control Groups

Sensitivity Analysis

Difference-in-differences

# DiD: Two Groups and Two Periods

**Estimand (ATT)**

$\tau_{ATT} = E[Y_1(1) - Y_0(1)|D = 1]$

| | **Post-Period (T=1)** | **Pre-Period (T=0)** |
|---|---|---|
| Treated D=1 | $E[Y_1(1)\|D = 1]$ | $E[Y_0(0)\|D = 1]$ |
| Control D=0 | $E[Y_0(1)\|D = 0]$ | $E[Y_0(0)\|D = 0]$ |

**Problem**

*Missing potential outcome:* $E[Y_0(1)|D = 1]$*, ie. what is the average post-period outcome for the treated in the absence of the treatment?*

# Identification with Difference-in-Differences

**Identification Assumption (parallel trends)**

$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$

**Identification Result**

Given parallel trends the ATT is identified as:

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D = 1] &= \Big\{E[Y(1)|D = 1] - E[Y(1)|D = 0]\Big\} \\ &- \Big\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\Big\} \end{aligned}$$

Implicit functional form assumption: Parallel trends in levels != Parallel trends in logs. (cf Jensen's Inequality)

# Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$$

What is the missing data here?

# Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$$

What is the missing data here? The missing data is $E[Y_0(1)|D = 1]$ (the control counterfactual in the post-period for the treated group). To obtain an estimate of this, we assume parallel trends:

$$\begin{aligned} E[Y_0(1) - Y_0(0)|D = 1] &= E[Y_0(1) - Y_0(0)|D = 0] \\ \implies E[Y_0(1)|D = 1] &= E[Y_0(0)|D = 1] + E[Y_0(1) - Y_0(0)|D = 0] \end{aligned}$$

# Identification with Difference-in-Differences

Start with the estimand we want:

$$\tau_{ATT} = E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1]$$

What is the missing data here? The missing data is $E[Y_0(1)|D = 1]$ (the control counterfactual in the post-period for the treated group). To obtain an estimate of this, we assume parallel trends:

$$\begin{aligned} E[Y_0(1) - Y_0(0)|D = 1] &= E[Y_0(1) - Y_0(0)|D = 0] \\ \implies E[Y_0(1)|D = 1] &= E[Y_0(0)|D = 1] + E[Y_0(1) - Y_0(0)|D = 0] \end{aligned}$$

Substitute the assumption in for the missing data and rearrange:

$$\begin{aligned} \tau_{ATT} &= E[Y_1(1)|D = 1] - E[Y_0(1)|D = 1] \\ &= E[Y_1(1)|D = 1] - (E[Y_0(0)|D = 1] + E[Y_0(1)|D = 0] - E[Y_0(0)|D = 0]) \\ &= (E[Y(1)|D = 1] - E[Y(0)|D = 1]) - (E[Y(1)|D = 0] - E[Y(0)|D = 0]) \end{aligned}$$

This is just the difference of pre-post differences between the treated and control groups.

**Estimand (Sample Means: Panel)**

$$\left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(1) - \frac{1}{N_0}\sum_{D_i=0} Y_i(1)\right\} - \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(0) - \frac{1}{N_0}\sum_{D_i=0} Y_i(0)\right\}$$
$$= \left\{\frac{1}{N_1}\sum_{D_i=1} \{Y_i(1) - Y_i(0)\} - \frac{1}{N_0}\sum_{D_i=0} \{Y_i(1) - Y_i(0)\}\right\},$$

where $N_1$ and $N_0$ are the number of treated and control units respectively.

# Regression Estimator

This implies a fully saturated regression model with a two-way interaction:

$$E[Y_{igt}|g,t] = \beta_0 + \beta_1 \text{Treated}_g + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_g \times \text{Post}_t)$$

The subscripts are conventions to indicate the level of variation ($i$ = Individual, $g$ = Group, $t$ = Time).

# Difference-in-Differences estimator

This implies a fully saturated regression model with a two-way interaction:

$$E[Y_{igt}|g,t] = \beta_0 + \beta_1 \text{Treated}_g + \beta_2 \text{Post}_t + \beta_3 (\text{Treated}_g \times \text{Post}_t)$$

The subscripts are conventions to indicate the level of variation ($g$ = Group, $t$ = Time). This translates to our two-by-two table as follows:

| | **Post-Period (T=1)** | **Pre-Period (T=0)** | **Pre/Post Diff.** |
|---|---|---|---|
| Treated D=1 | $\beta_0 + \beta_1 + \beta_2 + \beta_3$ | $\beta_0 + \beta_1$ | $\beta_2 + \beta_3$ |
| Control D=0 | $\beta_0 + \beta_2$ | $\beta_0$ | $\beta_2$ |
| Treated/Control Diff. | $\beta_1 + \beta_3$ | $\beta_1$ | $\boldsymbol{\beta_3}$ |

$\boldsymbol{\beta_3}$ is the diff-in-diff estimate!

HOW MUCH SHOULD WE TRUST DIFFERENCES-IN-DIFFERENCES ESTIMATES?*

MARIANNE BERTRAND
ESTHER DUFLO
SENDHIL MULLAINATHAN

Standard advice: Cluster at least the unit level. More on this later.