Intermediate Data Analysis & Econometrics

Lecture7

Edward Vytlacil

Yale University, Department of Economics

March 1, 2023

Overview for this lecture:

- **Agenda**
  - Difference-in-differences.
    - How to implement,
    - Equivalent regression formulation,
    - Key assumption: common time trends.
  - Application: ‘Agan and Starr 2018 QJE.
- Code for this lecture available here. Data used is available here.

Difference-in-Difference: Setting

Suppose you want to evaluate a policy change, and

- Have a “treatment” and “control" group;
  - policy change effects treatment group;
  - policy change does not effect control group
- Observe both groups both before and after policy change;

Example: Card and Krueger (1994)

- Want to evaluate effect of law increasing minimum wage on employment in fast-food restaurants.
- New Jersey increased minimum wage in April 1992.
- Neighboring Pennsylvania had no change in minimum wage law.
- Treated group are fast food restaurants in New Jersey.
- Control group are fast food restaurants in Pennsylvania.
- Pre-period is before April 1982, post period is after April 1982.

Difference-in-Difference: Setting

Suppose you want to evaluate a policy change, and

- Have a “treatment” and “control" group;
  - policy change effects treatment group;
  - policy change does not effect control group
- Observe both groups both before and after policy change;

Could use **cross-section estimator**:

- Cross-section estimator compares treated to control observations at one point in time (either pre-period or post period).
- For example:
  - could use mean-difference between treated and controls using only post (or only pre) data.
  - could run corresponding regression.
  - could add additional covariates to regression to adjust for additional variables (Cross-Section Regression Control).
- Disadvantage: **Omitted Variable Bias**
  - Even including additional covariates, may still have omitted variable bias.

Difference-in-Difference: Setting

Suppose you want to evaluate a policy change, and

- Have a “treatment” and “control" group;
  - policy change effects treatment group;
  - policy change does not effect control group
- Observe both groups both before and after policy change;

Could use **before-after estimator**:

- Before-after estimator compares treated observations after the policy change to treated observations before the policy change.
- For example:
  - could use mean-difference between treated post- vs before- policy change.
  - could run corresponding regression.
  - could add additional time-varying covariates to regression to adjust for additional variables.
- Disadvantage: **Time-Trends!**
  - Perhaps what the before-estimator is estimating is a time-trend that would have happened even without the policy change.

Difference-in-Difference: Setting

Suppose you want to evaluate a policy change, and

- Have a “treatment” and “control" group;
  - policy change effects treatment group;
  - policy change does not effect control group
- Observe both groups both before and after policy change;

Difference-in-Difference estimator:

- Difference between before-after difference for treated and before-after difference for controls;
  - Intuition: using before-after difference for controls as proxy for what time trend for treated would have been without the policy change.
- Equivalently, difference between cross-section difference for treated vs control in post period, and cross-section difference for treated vs control in pre period;
  - Intuition: Use the cross-section difference in the pre-period as a proxy for what the difference between treated and control would have been had it not been for the policy change.
- Key assumption: time trend for control observations is what the time trend for treated observations would have been without the policy change. (“common trends assumption.”)
- Can represent as a linear regression, which then motivates controlling for additional covariates thought that linear regression representation.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- Have a “treatment” and “control" group;
- Observe both groups both before and after policy change;
- Policy change effects treatment group;
- Policy change has no effect on control group.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{\cdot,Post} = \overline{Y}_{T,Post} - \overline{Y}_{C,Post}$ is **cross section** estimate from post-period.
    - Worry: omitted variable bias.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{\cdot,Post} = \overline{Y}_{T,Post} - \overline{Y}_{C,Post}$ is **cross section** estimate from post-period.
  - Worry: omitted variable bias.
- $\Delta\overline{Y}_{\cdot,Pre} = \overline{Y}_{T,Pre} - \overline{Y}_{C,Pre}$ is cross section estimate from pre-period.
  - Differences between treated and control from before the policy change.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{\cdot,Post} = \overline{Y}_{T,Post} - \overline{Y}_{C,Post}$ is **cross section** estimate from post-period.
  - Worry: omitted variable bias.
- $\Delta\overline{Y}_{\cdot,Pret} = \overline{Y}_{T,Pre} - \overline{Y}_{C,Pre}$ is cross section estimate from pre-period.
  - Differences between treated and control from before the policy change.
- $\text{DiD} = \Delta\overline{Y}_{\cdot,Post} - \Delta\overline{Y}_{\cdot,Pre}$ is **Difference-in-Difference** estimate, difference between the treated and control differences in the post period versus pre period.
- Intuition: DiD adjusting for difference between treated and control that was present before the policy change.

$$\begin{aligned}\text{DiD} &= \left\{\overline{Y}_{T,Post} - \overline{Y}_{C,Post}\right\} - \Delta\overline{Y}_{\cdot,Pre}\\ &= \overline{Y}_{T,Post} - \left\{\underbrace{\overline{Y}_{C,Post} + \Delta\overline{Y}_{\cdot,Pre}}_{\text{hypothetical value if no treatment effect}}\right\}\end{aligned}$$

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{T,\cdot} = \overline{Y}_{T,Post} - \overline{Y}_{T,Pre}$ is **before-after** estimate.
    - Worry: time trend.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{T,\cdot} = \overline{Y}_{T,Post} - \overline{Y}_{T,Pre}$ is **before-after** estimate.
    - Worry: time trend.
- $\Delta\overline{Y}_{C,\cdot} = \overline{Y}_{C,Post} - \overline{Y}_{C,Pre}$ is before-after for controls.
    - No policy change for controls, so is estimate of time trend.

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- $\Delta\overline{Y}_{T,\cdot} = \overline{Y}_{T,Post} - \overline{Y}_{T,Pre}$ is **before-after** estimate.
  - Worry: time trend.
- $\Delta\overline{Y}_{C,\cdot} = \overline{Y}_{C,Post} - \overline{Y}_{C,Pre}$ is before-after for controls.
  - No policy change for controls, so is estimate of time trend.
- $\text{DiD} = \Delta\overline{Y}_{T,\cdot} - \Delta\overline{Y}_{C,\cdot}$ is **Difference-in-Difference** estimate,
  Difference between the before-after estimate for treated and the before-after estimate for controls.
- Intuition: using before-after for controls to adjust for time trend for treated.

$$\begin{aligned} \text{DiD} &= \left\{ \overline{Y}_{T,Post} - \overline{Y}_{T,Pre} \right\} - \Delta\overline{Y}_{C,\cdot} \\ &= \overline{Y}_{T,Post} - \underbrace{\left\{ \overline{Y}_{T,Pre} + \underbrace{\Delta\overline{Y}_{C,\cdot}}_{\text{time trend if no treatment}} \right\}}_{\text{hypothetical value if no treatment effect}} \end{aligned}$$

Diff-in-Diff Setup

	Pre	Post	Difference
Control (C)	$\overline{Y}_{C,Pre}$	$\overline{Y}_{C,Post}$	$\Delta\overline{Y}_{C,\cdot}$
Treatment (T)	$\overline{Y}_{T,Pre}$	$\overline{Y}_{T,Post}$	$\Delta\overline{Y}_{T,\cdot}$
Difference	$\Delta\overline{Y}_{\cdot,Pre}$	$\Delta\overline{Y}_{\cdot,Post}$	DiD

- Difference in Difference can equivalently be computed either way:

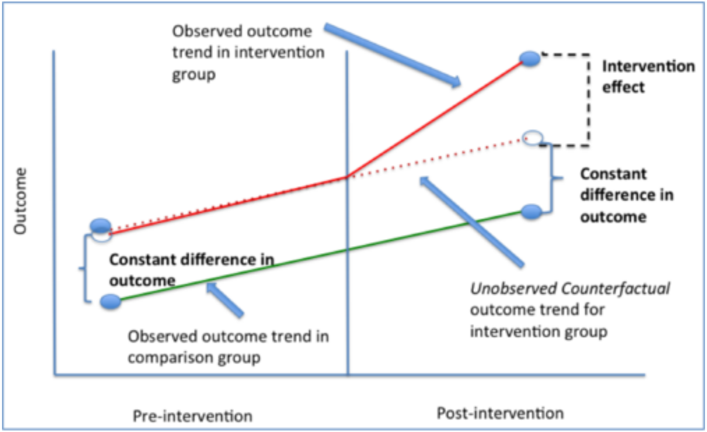
$$\begin{aligned}
\text{DiD} &= \Delta\overline{Y}_{\cdot,Post} - \Delta\overline{Y}_{\cdot,Pre} && \text{difference in cross-section differences post vs pre} \\
&= \overline{Y}_{T,Post} - \left\{ \overline{Y}_{C,Post} + \Delta\overline{Y}_{\cdot,Pre} \right\} \\
&= \overline{Y}_{T,Post} - \left\{ \overline{Y}_{T,Pre} + \Delta\overline{Y}_{C,\cdot} \right\} \\
. &= \Delta\overline{Y}_{T,\cdot} - \Delta\overline{Y}_{C,\cdot} && \text{difference in before-after differences treated vs controls}
\end{aligned}$$

Key Assumptions

Key Assumptions:

- Treatment and Control groups have ***parallel trends*** in outcome.
- Intervention was not related to differences in expected trends .
- No big compositional shifts in groups from Pre to Post period.
- Introduction of treatment doesn't affect control group's outcomes (e.g. no spillovers, no general equilibrium effects).

Graphical Intuition



Source: Columbia University School of Public Health

https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Diff-in-Diff data needs

You need data on

- The treatment group and control group.
- Covering pre- and post-treatment periods.

Structure

- Aggregate data
- **or** individual-level data, either
  - Pooled cross-sections
  - **or** longitudinal data.

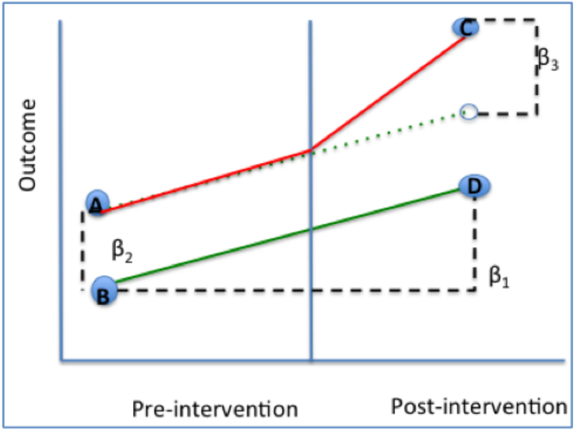
With individual-level data, you can add additional controls to linear regression representation of difference-in-differences.

Diff-in-Diff with regression

$$Y = \beta_0 + \beta_1 D_{\text{post}} + \beta_2 D_{\text{T}} + \beta_3 D_{\text{post}} D_{\text{T}} + e$$

- Y is the outcome of interest
- D_{post} is a time dummy (1 = post intervention),
  β_1 is time trend for controls.
- D_{T} is a treatment group dummy (1 = treatment),
  β_2 is cross-section difference between treated and controls in pre-period.
- D_{post} D_{T} is a time *times* treatment interaction.
- **OLS estimate of β_3 is numerically equivalent to difference-in-difference**.

Diff-in-Diff Regression Visualized



Source: Columbia University School of Public Health

https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Diff-in-Diff with regression

- Advantage of estimating difference-in-difference using linear regression representation: Can add controls!

$$Y = \beta_0 + \beta_1 D_{\text{post}} + \beta_2 D_{\text{T}} + \beta_3 D_{\text{post}} D_{\text{T}} + \beta_4 X + e$$

- For example, in Card and Krueger (1994), add controls for chain-type, whether store is company owned or not, and region dummies.

Application: Agan and Starr 2018.

Consider ‘Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment”
(Agan and Starr 2018 QJE)

Motivation:

- Many Americans have a criminal record, disproportionately African-American men.
- Many firms include a “box” on application forms,
  asking whether applicant has a criminal record, and then often do no consider applicants who check “yes” in response.
- Results in barrier to employment for many individuals, disproportionately African-American men.
- Motivates “Ban-the-Box” (BTB) laws that ban asking about applicant’s criminal history on *initial* job application form.
- BTB laws aim in part to reduce racial disparities in employment.
- Concern: if employers are not allowed to ask about criminal record of applicants, might engage in statistical discrimination based on race.

Example of Triple-Diff: Agan and Star 2018.

Agan and Star (2018) conduct correspondence study:

- submitted approximately 15,000 fictitious online job applications to entry-level positions,
- fictitious applicants:
  - all young males
  - half of resumes randomly assigned a stereotypically white name, and half a stereotypically black name.
  - half of resumes randomly assigned a criminal record;
  - randomized other aspects of resume that indirectly signal a criminal record (GED receipt, gaps in resume).

Example of Triple-Diff: Agan and Star 2018.

Agan and Star (2018) conduct correspondence study:

- submitted approximately 15,000 fictitious online job applications to entry-level positions,
- fictitious applicants:
  - all young males
  - half of resumes randomly assigned a stereotypically white name, and half a stereotypically black name.
  - half of resumes randomly assigned a criminal record;
  - randomized other aspects of resume that indirectly signal a criminal record (GED receipt, gaps in resume).
- Pairs of black and white applications sent to:
  - 4,291 establishments (stores) in 293 chains in New Jersey and New York City
  - Sent to same employers both before ("Pre" period) and after ("Post" period) the effective dates of private-sector BTB laws in New Jersey (March 1, 2015) and New York City (October 27, 2015).
  - Outcome of interest: Callbacks.

Example of Triple-Diff: Agan and Star 2018.

Agan and Star (2018) conduct correspondence study:

- submitted approximately 15,000 fictitious online job applications to entry-level positions,
- fictitious applicants:
  - all young males
  - half of resumes randomly assigned a stereotypically white name, and half a stereotypically black name.
  - half of resumes randomly assigned a criminal record;
  - randomized other aspects of resume that indirectly signal a criminal record (GED receipt, gaps in resume).
- Pairs of black and white applications sent to:
  - 4,291 establishments (stores) in 293 chains in New Jersey and New York City
  - Sent to same employers both before (“Pre” period) and after (“Post” period) the effective dates of private-sector BTB laws in New Jersey (March 1, 2015) and New York City (October 27, 2015).
  - Outcome of interest: Callbacks.
- Methodology is difference-in-difference for effect of banning the box on white, and separately for effect of banning the box on white, and is a difference between those two diff-in-diff (**a triple-diff**) for effect on black-white gap.

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks

-

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Cross-Section Estimation)

- Estimate effect of banning the box on black callbacks (using pre-period data)

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

Cross-section difference in means in pre-period: 0.009.

$\Longrightarrow$ estimate that removing the box decreases calls for blacks by 0.009 (since Box Removers had box in pre-period)

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Cross-Section Estimation)

- Estimate effect of banning the box on black callbacks (using pre-period data)

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

- Cross section difference equivalent to running regression on dummy variable for having a box using only pre-period data.

```
> #Only use Data from Pre Period
> #Equivalent Regression Representation of Cross-Section Estimator
>>lm(response ~ remover, data = subset(df,white==0 & post==0))

Call:
lm(formula = response ~ remover, data = subset(df, white == 0 &
    post == 0))

Coefficients:
(Intercept)      remover
   0.096154     0.008928
```

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Cross-Section Estimation)

- Estimate effect of banning the box on black callbacks (using pre-period data)

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

Cross-section difference in means in pre-period: 0.009.
$\Rightarrow$ estimate that removing the box decreases calls for blacks by 0.009 (since Box Removers had box in pre-period)

- Cross section difference equivalent to running regression of dummy variable for having a box using only pre-period data.
- Concern: Stores that select to have a box can be different from stores that select not to have a box in pre-period, not randomized.
- Using regression function representation, can regression adjust other covariates, but can still have omitted variable bias.

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Before-After Estimation)

- Estimate effect of banning the box on black callbacks (using box-removers, pre/post)

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

Before-After estimate for box-removers: 0.005.

- Advantage: Compares same stores with and without the box.
- Disadvantage: Trend that would happened without policy change?
  is estimated effect the effect of removing the box? or of time trends that would have happened anyway?

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Before-After Estimation)

- Estimate time trend for controls (using no-box stores, pre/post).

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

Can use those unaffected by the policy change to estimate what we think the trend would have been for those effected by the policy change if it had not been for the policy change.

- Estimate trend for no-box stores is 0.016.

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Diff-in-Diff)

- Difference in difference estimate:

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

- Difference in difference estimate is difference in before-after for treated vs control:

$$0.005 - 0.016 = -0.011.$$

- Equivalently, difference in difference estimate is difference in cross section estimates, post versus pre:

$$-0.002 - 0.009 = -0.011.$$

Example of Triple-Diff: Agan and Star 2018.

Consider Effect of Banning the Box for Blacks (Diff-in-Diff)

- Difference in difference estimate:

	Pre-Period	Post-Period	Difference
No Box	0.096	0.112	0.016
Box Remover	0.105	0.110	0.005
Difference	0.009	-0.002	-0.011

```
> #Diff-in-Diff Regression Black Applicants
> lm(response ~ post*remover, data = subset(df,white==0))

Call:
lm(formula = response ~ post * remover, data = subset(df, white == 0))

Coefficients:
 (Intercept)          post       remover  post:remover
    0.096            0.016         0.009        -0.011
```

Example of Triple-Diff: Agan and Star 2018.

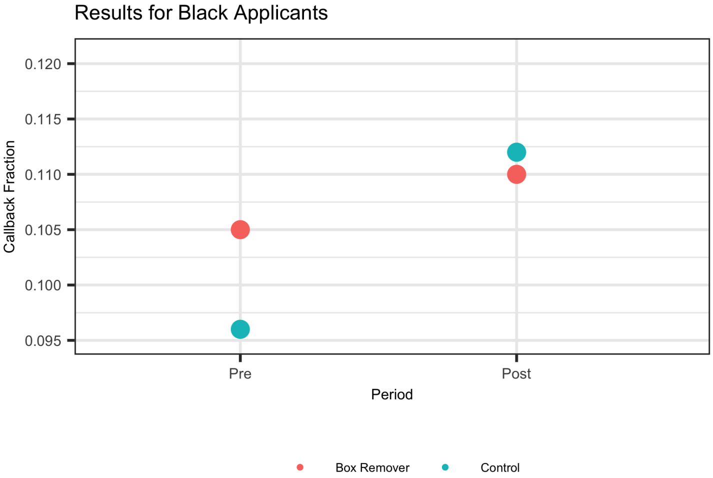
Consider Effect of Banning the Box for Whites (Diff-in-Diff)

- Difference in difference estimate:

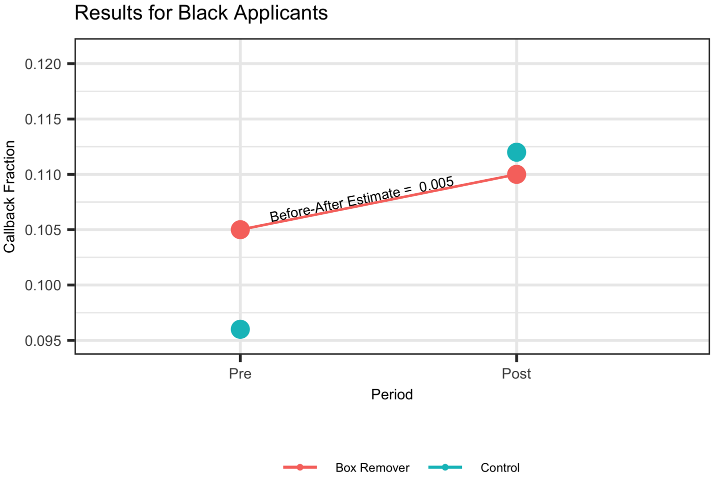
	Pre-Period	Post-Period	Difference
No Box	0.123	0.133	0.010
Box Remover	0.113	0.150	0.037
Difference	-0.010	0.017	0.027

```
> #Diff-in-Diff Regression White Applicants
> lm(response ~ post*remover, data = subset(df,white==1))

Call:
lm(formula = response ~ post * remover, data = subset(df, white == 1))

Coefficients:
 (Intercept)          post       remover   post:remover
    0.123            0.010       -0.010          0.027
```

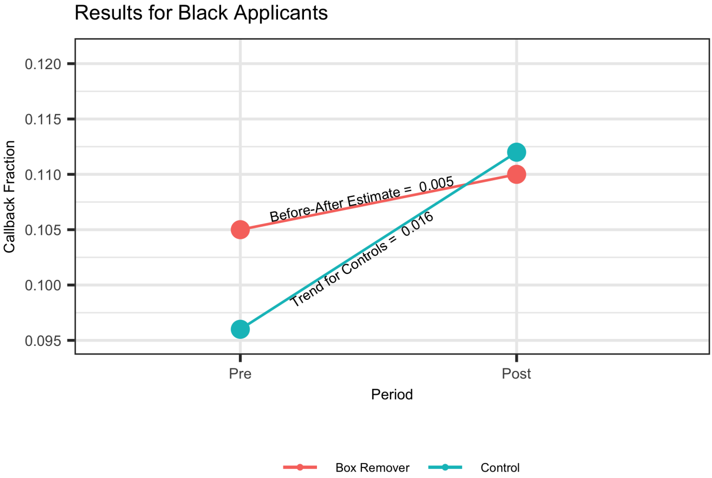
Graphical Intuition



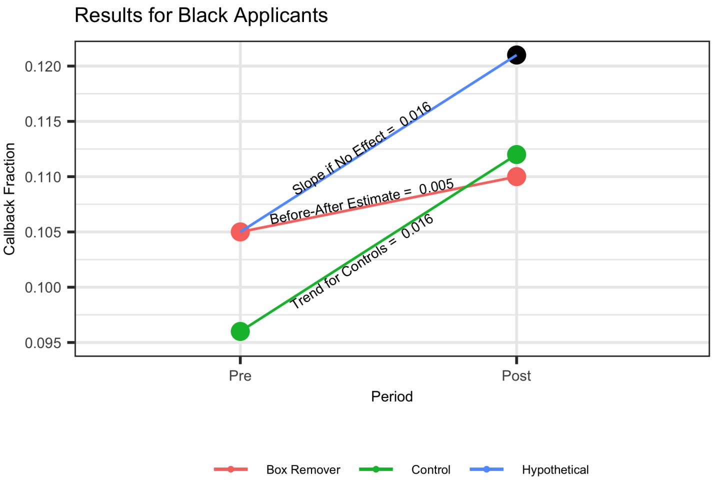
Graphical Intuition



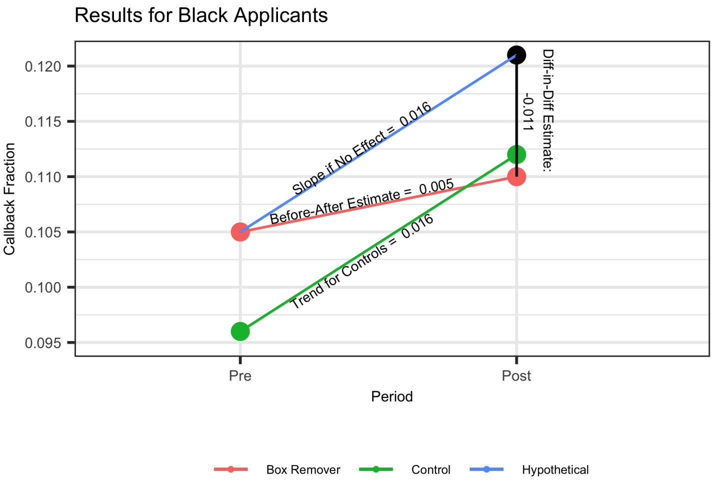
Graphical Intuition



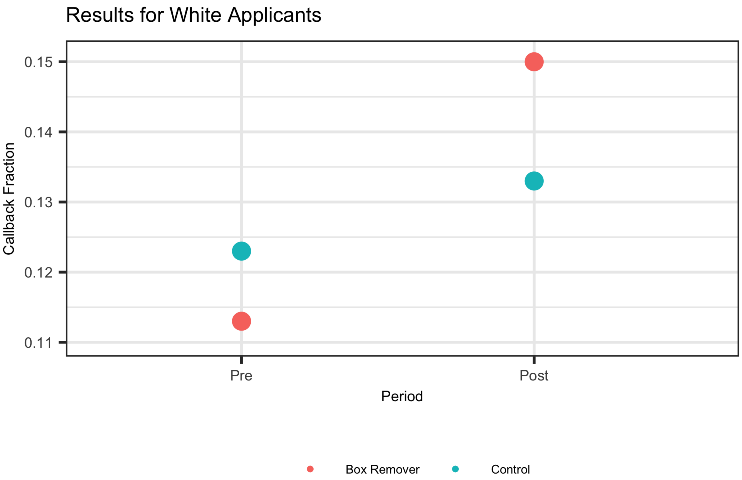
Graphical Intuition



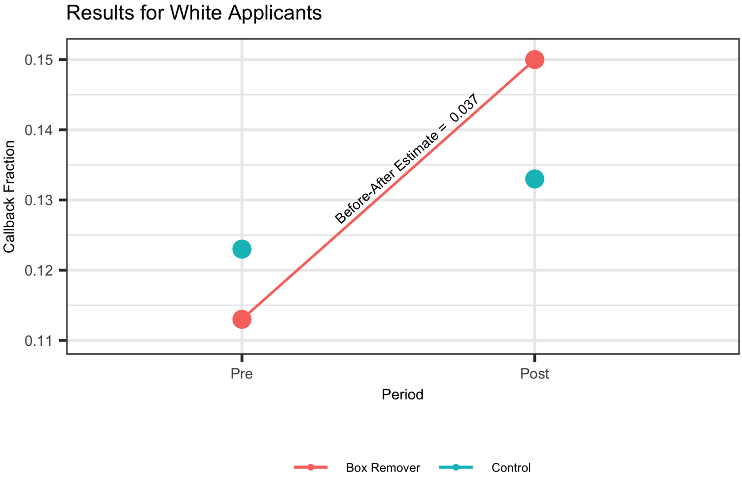
Graphical Intuition



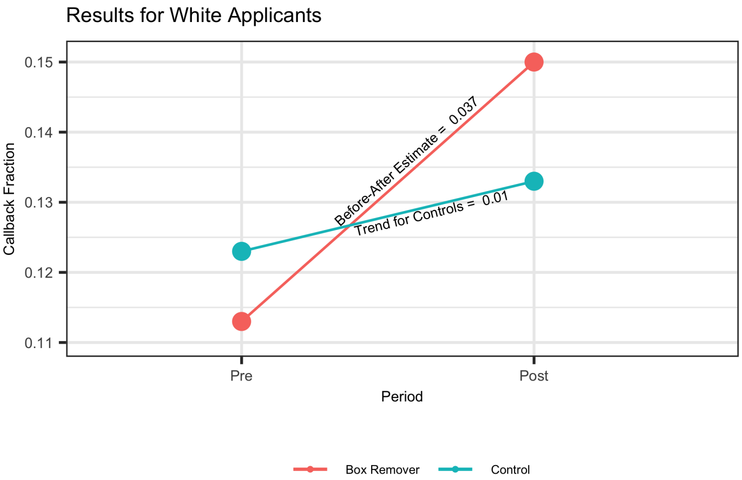
Graphical Intuition



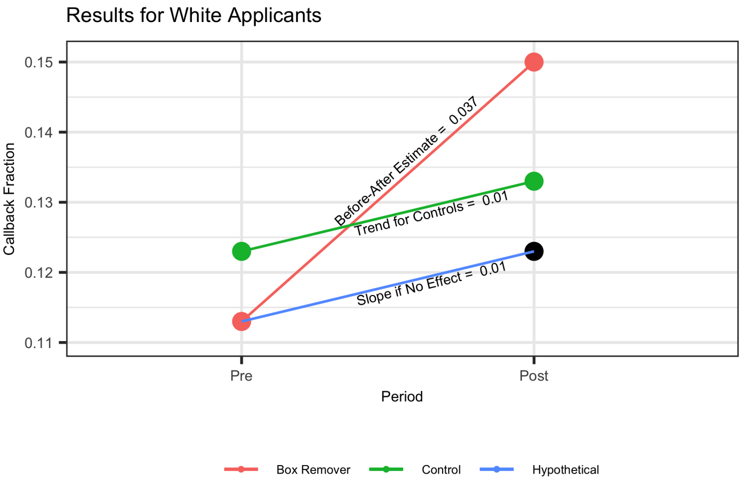
Graphical Intuition



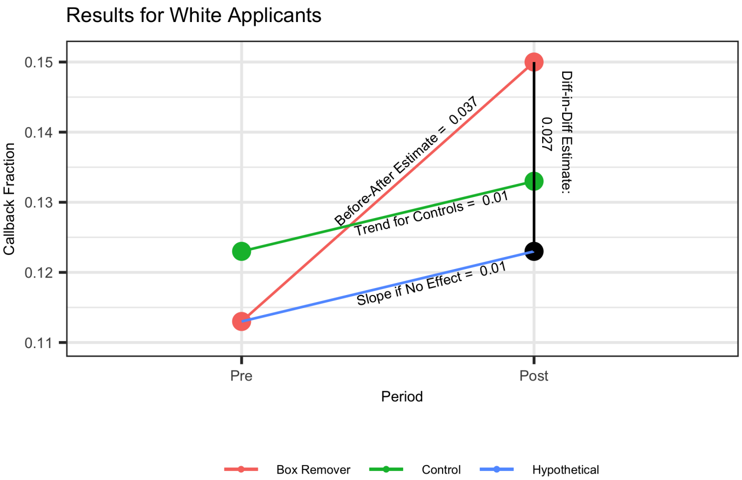
Graphical Intuition



Graphical Intuition



Graphical Intuition



Example of Diff-in-Diff: Agan and Star 2018.

- Diff-in-Diff Estimate for Blacks: -0.011.
- Diff-in-Diff Estimate for Whites: 0.027.
- "Triple Difference" for effect of banning the box on black-white gap:

$$0.027 - \left(-0.011\right) = 0.038,$$

banning the box *increased* the gap in call-backs by 0.038, exacerbating the racial difference.

Example of Diff-in-Diff: Agan and Star 2018.

- Diff-in-Diff Estimate for Blacks: -0.011.
- Diff-in-Diff Estimate for Whites: 0.027.
- "Triple Difference" for effect of banning the box on black-white gap:

$$0.027 - (-0.011) = 0.038,$$

  banning the box *increased* the gap in call-backs by 0.038, exacerbating the racial difference.
- Equivalently, estimate fully-saturated regression model:

$$\begin{aligned} Y = \beta_0 + \beta_1 D_{\text{post}} + \beta_2 D_{\text{T}} + \beta_3 \text{White} + \beta_5 D_{\text{post}} \times \text{White} \\ + \beta_6 D_{\text{T}} \times \text{White} + \beta_7 D_{\text{post}} \times D_{\text{T}} + \beta_8 D_{\text{post}} \times D_{\text{T}} \times \text{White} + e \end{aligned}$$

  then
  - $\hat{\beta}_7$ is Diff-in-Diff estimate for blacks,
  - $\hat{\beta}_7 + \hat{\beta}_8$ is Diff-in-Diff estimate for whites,
  - $\hat{\beta}_8$ is triple difference
- Threats to validity?

Example of Diff-in-Diff: Agan and Star 2018.

	Dependent variable:			
	Callback Rate			
	(1)	(2)	(3)	(4)
White			0.0269**	0.0269**
			(0.0111)	(0.0111)
BoxRemover	−0.0004	−0.0004	0.0089	0.0089
	(0.0288)	(0.0288)	(0.0262)	(0.0262)
Post	0.0126	0.0126	0.0154	0.0154
	(0.0130)	(0.0130)	(0.0150)	(0.0150)
BoxRemover x white			−0.0188	−0.0188
			(0.0142)	(0.0142)
Post x white			−0.0054	−0.0054
			(0.0123)	(0.0124)
BoxRemover x post	0.0082	0.0082	−0.0104	−0.0105
	(0.0182)	(0.0182)	(0.0190)	(0.0190)
BoxRemover x post x white			0.0373**	0.0374**
			(0.0179)	(0.0179)
Controls?	No	Yes	No	Yes

Example of Diff-in-Diff: Agan and Star 2018.

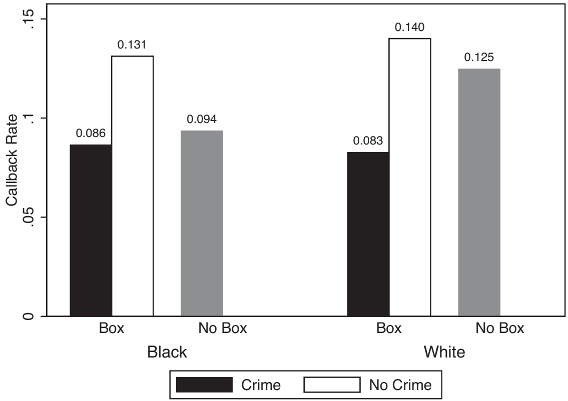


FIGURE I

Callback Rates by Race, Crime, and Box: Preperiod Applications Only

Example of Diff-in-Diff: Agan and Star 2018.

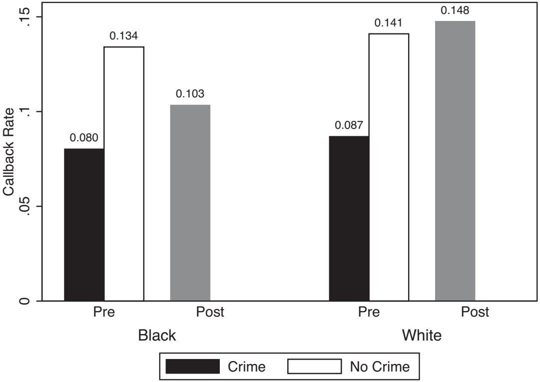


FIGURE II

Callback Rates by Race, Criminal Record, and Period: Balanced Box Removers Only

Example of Diff-in-Diff: Agan and Star 2018.

Conclusions of Agan and Starr:

- In pre-BTB period, strong evidence of discrimination in stores with no box, but not for stores with a box.
- BTB law helped whites, hurt blacks, especially blacks with no criminal background.
- Evidence of exaggerated statistical discrimination.
- Employers not using GED as signal of criminal background, despite GED being a stronger signal than race, and despite discrimination on GED being legal (unlike discrimination on race).