Difference-in-Differences

EDLD 650: Week 2

David D. Liebowitz

Agenda

1. Roadmap and Goals (9:00-10:10)

2. Discussion of Difference-in-Differences (DD) strategy (9:10-10:20)

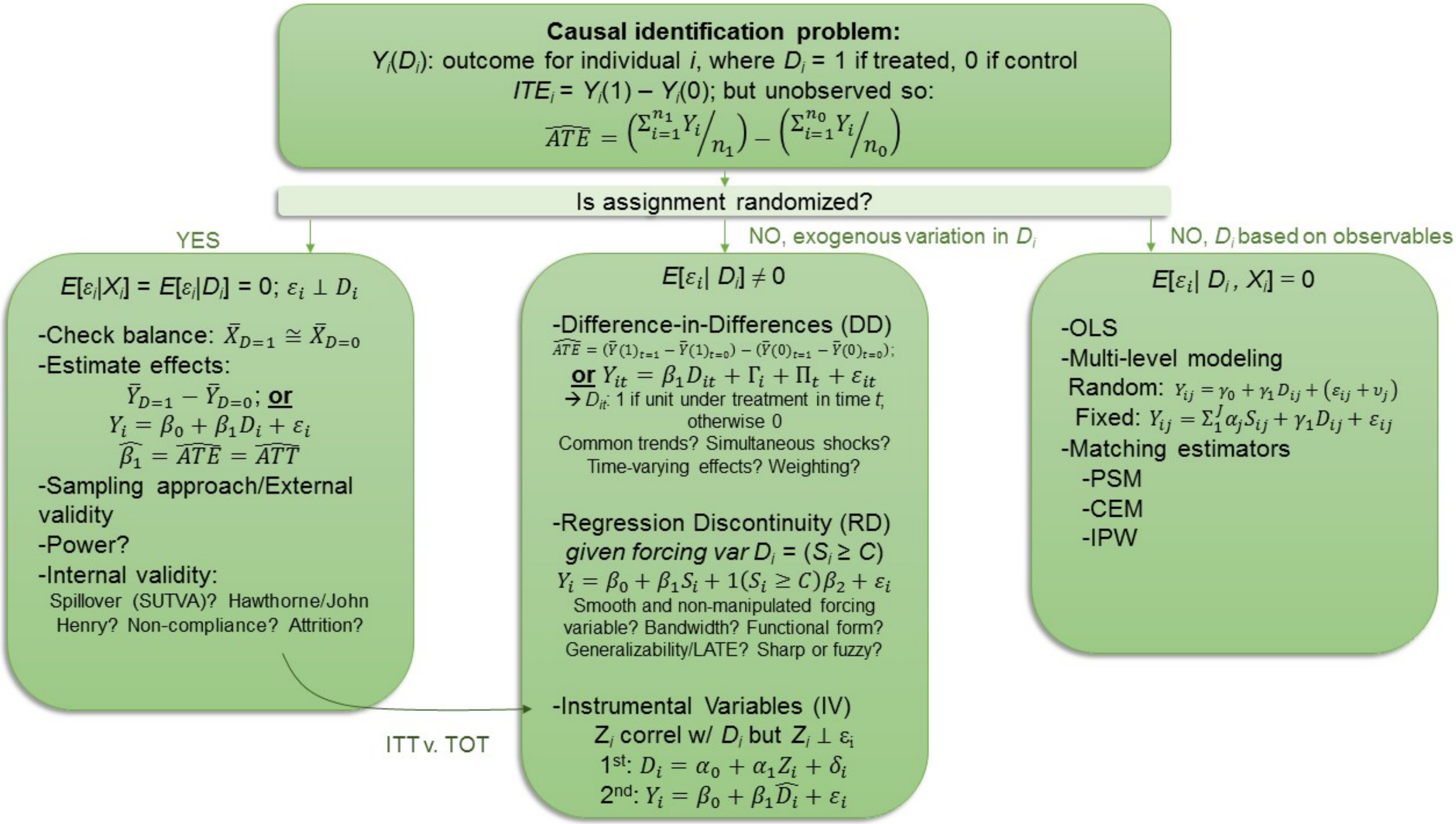
3. Break (10:20-10:30)

4. Estimating DD effects in data (10:30-11:40)

5. Wrap-up (11:40-11:50)

- DARE #1
- Plus/Deltas & Clear/Murky

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:
$\widehat{ATE} = \left(\sum_{i=1}^{n_1} Y_i / n_1\right) - \left(\sum_{i=1}^{n_0} Y_i / n_0\right)$
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 threats to validity in difference-in-differences (DD) identification strategy and multiple approaches to address these threats.

2. Using a cleaned dataset, estimate multiple DD specifications in R and interpret these results

Cold-calling

Purpose

- Formative assessment
- Equitable distribution of class participation
- Shared accountability for deep understanding of complex and technical readings

Norms

- Questions posted by Thursday PM
- Preparation is expected
- These are hard concepts; mistakes are expected
- Judgments on accuracy of responses are about the responses, not the individual
- Questions and response are about learning, not performance

Structure

- All cold calls will be telegraphed
- Questions will come directly from question list
- Random draw (w/ replacement) from class list
- Ample wait time; multiple "at-bats"
- Teaching staff will identify incomplete or incorrect response and seek clarification
- Extension questions on a volunteer basis

Discussion questions

Break

Programming in EDLD 650

What you won't get 😦

- A heavy dose of data management and visualization strategies
- The most efficient code (My coding skills are 💩)

What you will get 😄

- A review of the programming steps you should take as part of the **actual** research process
- *Some* model code for management and visualization
- Programming strategies and packages that can be used to estimate the causal inference techniques we will study
- A community of knowledge programmers who will expand our knowledge base!

Estimating a classic, two-period difference-in-differences (DD) model

Replicating Dynarski (2003)

Recall Dynarski's primary model (Eq. 2):

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

Let's try to fit this!

Reading in the data

```
dynarski <- read_dta(here("data/ch8_dynarski.dta"))

head(dynarski)
```

```
#> # A tibble: 6 x 8
#>      id  hhid   wt88  coll hgc23 yearsr              fatherdec offer
#>   <dbl> <dbl>  <dbl> <dbl> <dbl>  <dbl>              <dbl+lbl> <dbl>
#> 1     9     9 691916     1    13     81 0 [Father not deceased]     1
#> 2    14    13 784204     1    16     81 0 [Father not deceased]     1
#> 3    15    15 811032     1    16     82 0 [Father not deceased]     0
#> 4    21    20 644853     1    16     79 0 [Father not deceased]     1
#> 5    22    22 728189     1    16     80 0 [Father not deceased]     1
#> 6    24    23 776590     0    12     79 0 [Father not deceased]     1
```

Viewing the data

Show 7 entries Search:

	id	coll	hgc23	yearsr	fatherdec	offer
1	9	1	13	81	0	1
2	14	1	16	81	0	1
3	15	1	16	82	0	0
4	21	1	16	79	0	1
5	22	1	16	80	0	1
6	24	0	12	79	0	1
7	26	1	14	80	0	1

Showing 1 to 7 of 3,986 entries

Previous 1 2 3 4 5 … 570 Next

Understanding the data (1)

```
x ← select(dynarski, coll, hgc23, fatherdec, offer)

summary(x)
```

```
#>       coll            hgc23         fatherdec           offer      
#>  Min.   :0.0000   Min.   :10.00   Min.   :0.00000   Min.   :0.000  
#>  1st Qu.:0.0000   1st Qu.:12.00   1st Qu.:0.00000   1st Qu.:0.000  
#>  Median :0.0000   Median :12.00   Median :0.00000   Median :1.000  
#>  Mean   :0.4579   Mean   :13.14   Mean   :0.04792   Mean   :0.723  
#>  3rd Qu.:1.0000   3rd Qu.:14.00   3rd Qu.:0.00000   3rd Qu.:1.000  
#>  Max.   :1.0000   Max.   :19.00   Max.   :1.00000   Max.   :1.000  
```

```
sum(is.na(coll))
```

```
#> [1] 0
```

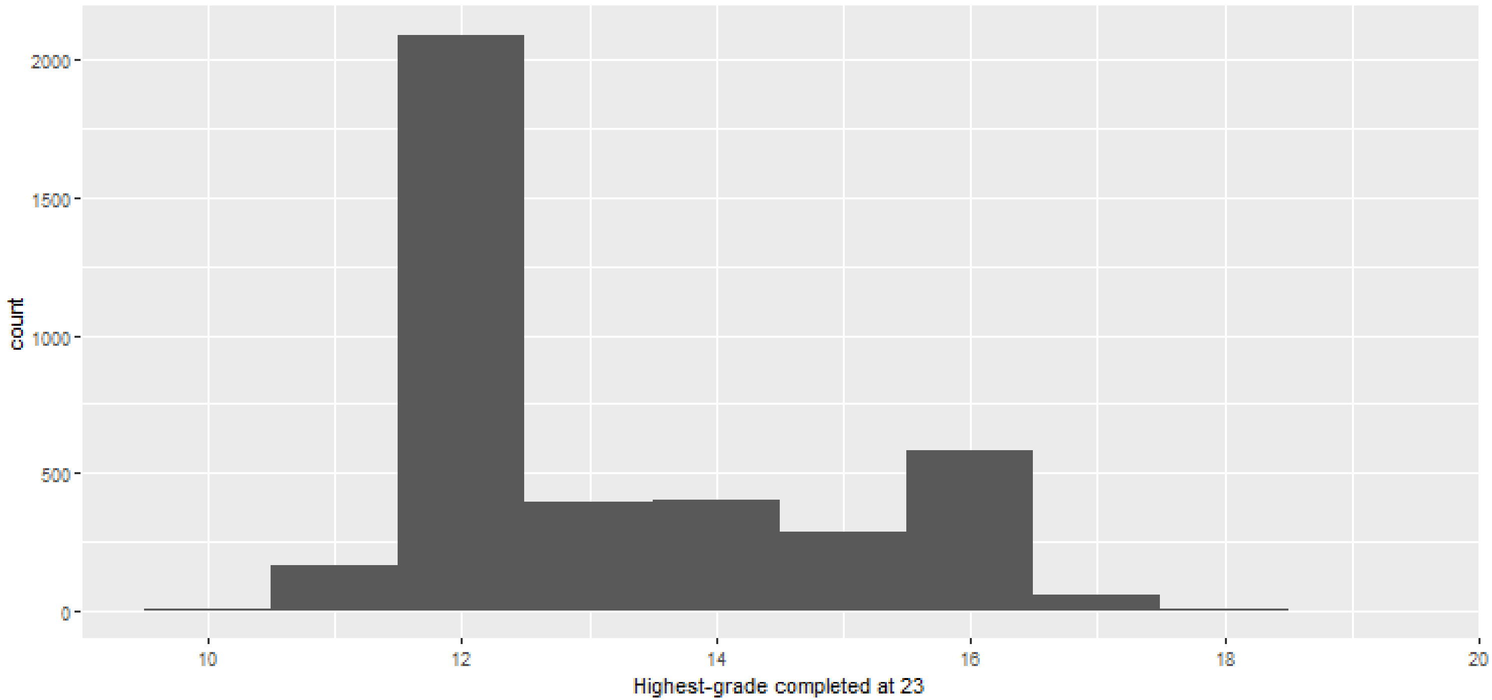
Understanding the data (2)

```
college ← table(dynarski$fac_fatherdec, dynarski$fac_coll)
college
```

```
#>
#>                          No College College
#>   Father not deceased          2059    1736
#>   Father deceased               102      89
```

Plot outcome data

```
hg ← ggplot(dynarski, aes(hgc23)) + geom_histogram(binwidth=1)
  hg + scale_x_continuous(name="Highest-grade completed at 23",
                          breaks=c(10, 12, 14, 16, 18, 20))
```



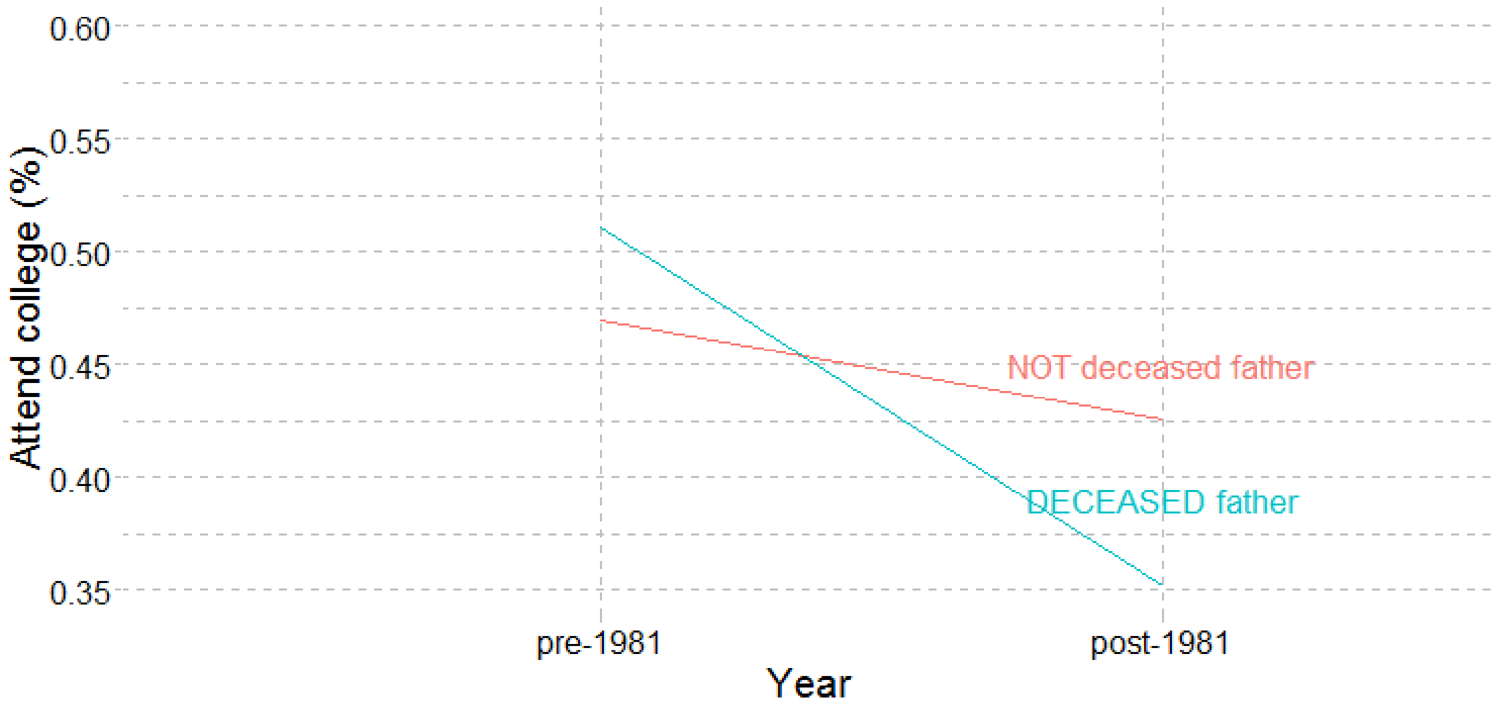
Summary statistics table

Table 1. Descriptive Statistics

Statistic	N	Mean	St. Dev.
Attend college at 23	3,986	0.46	0.50
Years schooling at 23	3,986	13.14	1.63
Father deceased	3,986	0.05	0.21
Offer	3,986	0.72	0.45

Notes: This table presents unweighted means and standard deviations from the NLSY poverty and random samples used in the Dynarski (2003) paper.

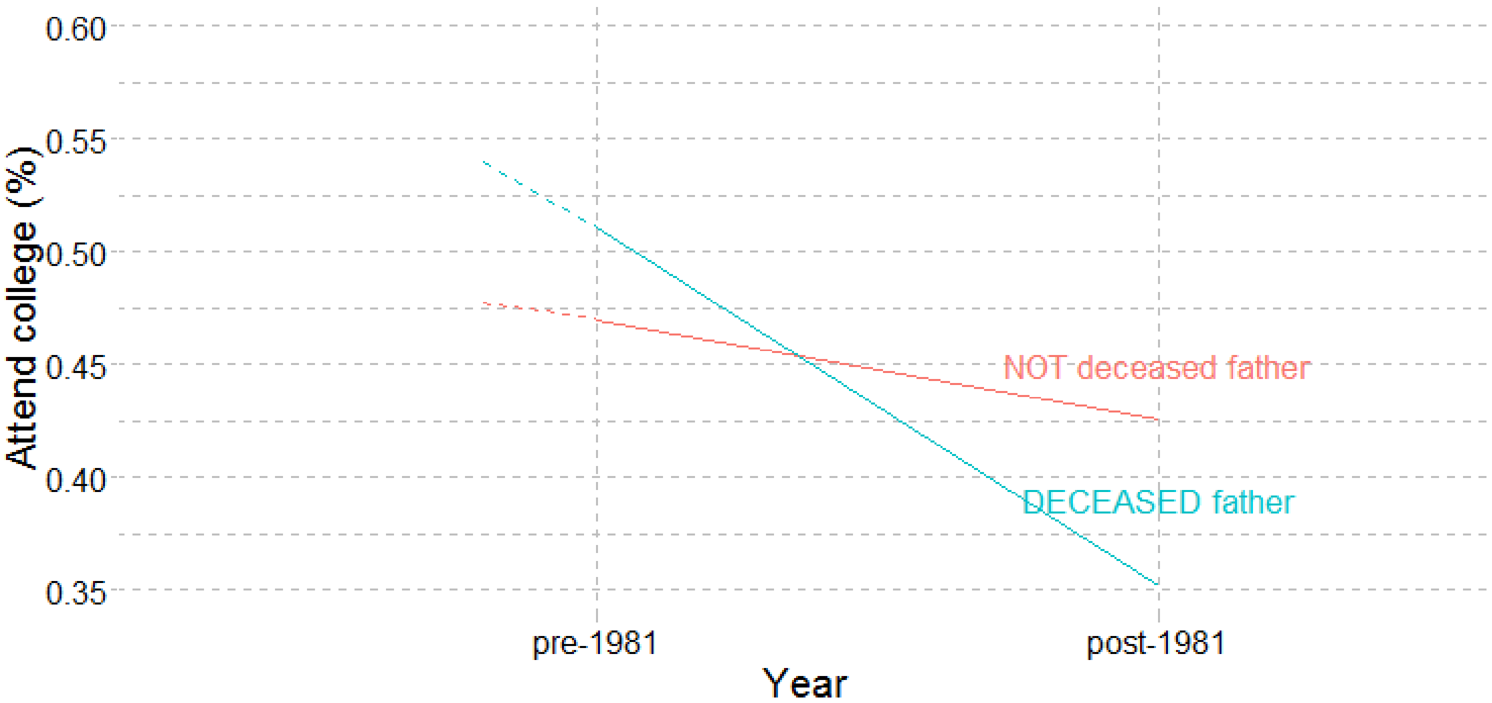
Graphical DD



What is treatment effect?

What is the core identifying assumption underlying the DD framework? How do we know whether we've satisfied it?

Graphical DD



What would you think if you "knew" this was the pattern?

Estimate classic two-period DD

Dynarski's original model:

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

Murnane and Willet have renamed the variable to make clear that a value of 1 means individuals are eligible for aid, so:

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

Estimate classic two-period DD

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

```
lm(coll ~ fatherdec*offer + fatherdec + offer, data=dynarski)
```

```
#>
#> Call:
#> lm(formula = coll ~ fatherdec * offer + fatherdec + offer, data = dynarsk
#>
#> Coefficients:
#>     (Intercept)        fatherdec              offer  fatherdec:offer
#>         0.42571         -0.07386            0.04387          0.11523
```

This doesn't quiet match, let's add the weights in...

Estimate classic two-period DD

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

```
lm(coll ~ fatherdec*offer + fatherdec + offer, data=dynarski,
          weights=dynarski$wt88)
```

```
#>
#> Call:
#> lm(formula = coll ~ fatherdec * offer + fatherdec + offer, data = dynarsk
#>     weights = dynarski$wt88)
#>
#> Coefficients:
#>     (Intercept)        fatherdec              offer  fatherdec:offer
#>         0.47569         -0.12348            0.02601          0.18223
```

Pretty underwhelming output?

Under the hood

```
est_dynarski <- lm(coll ~ fatherdec*offer + fatherdec + offer,
                   data=dynarski, weights=dynarski$wt88)
est_dynarski %>% names()
```

```
#>  [1] "coefficients"  "residuals"     "fitted.values" "effects"
#>  [5] "weights"       "rank"          "assign"        "qr"
#>  [9] "df.residual"   "xlevels"       "call"          "terms"
#> [13] "model"
```

```
est_dynarski %>% tidy()
```

```
#> # A tibble: 4 x 5
#>   term            estimate std.error statistic   p.value
#>   <chr>              <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)       0.476     0.0150     31.8  7.12e-198
#> 2 fatherdec        -0.123     0.0752     -1.64 1.01e-  1
#> 3 offer             0.0260    0.0178      1.46 1.43e-  1
#> 4 fatherdec:offer   0.182     0.0893      2.04 4.14e-  2
```

Further under the hood

```
summary(est_dynarski)
```

```
#> 
#> Call:
#> lm(formula = coll ~ fatherdec * offer + fatherdec + offer, data = dynarsk
#>     weights = dynarski$wt88)
#> 
#> Weighted Residuals:
#>    Min     1Q Median     3Q    Max 
#> -490.9 -230.3 -138.6  247.7  554.0 
#> 
#> Coefficients:
#>                 Estimate Std. Error t value Pr(>|t|)    
#> (Intercept)      0.47569    0.01496  31.793   <2e-16 ***
#> fatherdec       -0.12348    0.07520  -1.642   0.1007    
#> offer            0.02601    0.01777   1.463   0.1435    
#> fatherdec:offer  0.18223    0.08931   2.041   0.0414 *  
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> 
#> Residual standard error: 285.7 on 3982 degrees of freedom
```

Making a no-fuss table

```
stargazer(est_dynarski, type='html', single.row = T)
```

	Dependent variable:
	coll
fatherdec	-0.123 (0.075)
offer	0.026 (0.018)
fatherdec:offer	0.182** (0.089)
Constant	0.476*** (0.015)
Observations	3,986
R^2	0.002
Adjusted R^2	0.001
Residual Std. Error	285.711 (df = 3982)

Central DD asssumptions

In order to fully trust that the estimates produced by a DD analysis are unbiased by endogeneity, we need to make (and defend) the following two 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?

We'll look at how to address some of these in the next section of the lecture, and you'll read more about how to do so in the readings and DARE for next week!

DD in panel data

A. The two-way fixed effect (TWFE) estimator for staggered implementation

B. Appropriate statistical inference

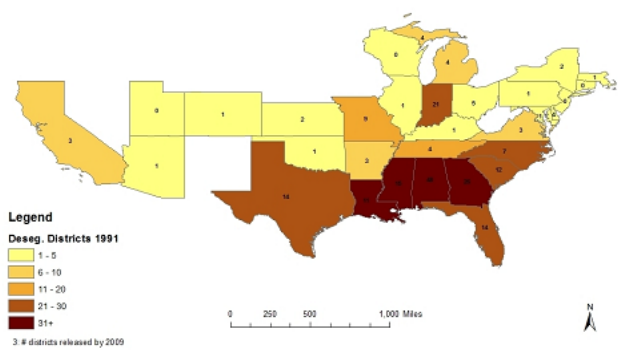
C. Assessing the parallel trends assumption (PTA)

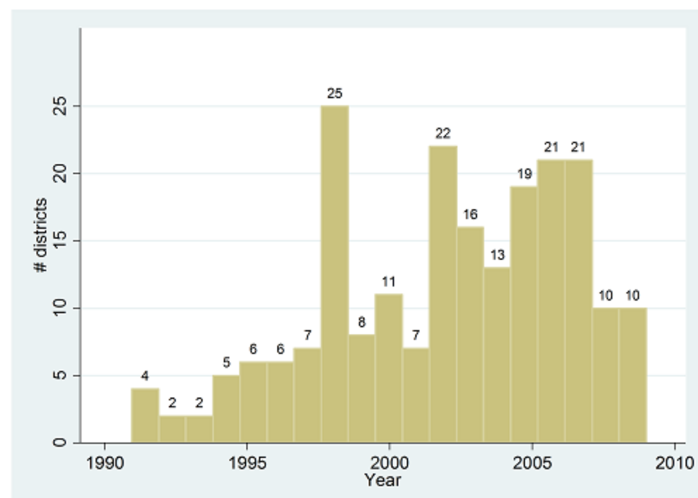
D. The event-study approach

End of desegregation

- In 1991, 480 school districts were under court desegregation order
- In following two decades, nearly half (215) were released and returned to neighborhood assignment patterns
- Timing of release was arguably **exogenous** and **quasi-random**.

- This provides strong support to the claim that the districts which were not (or *not yet*) released from court orders were on **parallel trends** in their outcomes with districts that were released and, thus, serve as **valid counterfactuals**.[1]





[1] Liebowitz (2018)

End of desegregation data

Show 9 entries Search:

	leaid	year	STATE	unitary	sd_dropout_prop_b	yrdiss
1	0100030	1990	01	0	0.16343490779399	2002
2	0100030	2000	01	0	0.18518517911434	2002
3	0100030	2010	01	1	0.10169491171836	2002
4	0100090	1990	01	0	0.21333333849906	
5	0100090	2000	01	0	0.15965346992015	
6	0100090	2010	01	0	0.10317460447549	
7	1000230	1990	10	0	0.09617373347282	1996
8	1000230	2000	10	1	0.15832762420177	1996
9	1000230	2010	10	1	0.09976246953010	1996

Showing 1 to 9 of 9 entries

Previous 1 Next

Estimate DD in panel data (1)

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

Take a minute to write down what this model does in words. Use the terms **mean effect**, **time series**, **fixed effects** and **causal parameter of interest**. Share with your neighbor.

> The model takes advantage of **time series** data in which the black dropout rate in each district is observed at three points in time. The model regresses the black dropout rate in a **fixed effect** model in which observations are clustered in two dimensions: within district (Γ_j) and also within time (Π_t). Note: Γ_j represents a vector of dummy indicators that take the value of one if observation *j* is equal to district *j* and zero otherwise. Π_t represents a vector of dummy indicators that take the value of one if observation *j* is in time *t* (1990, 2000 or 2010). β_1 estimates the **mean effect** of being observed after being declared unitary and is the **causal parameter of interest** reflecting the effect of being released from a desegregation order $UNITARY_{jt}$ on the black dropout rate.

In this case, the estimates rely on **repeated cross-sectional** panel data. We could also implement the same framework in **longitudinal** panel data.

Estimate DD in panel data (2)

We are going to shift to using the `fixest` package; an incredibly versatile and robust tool for regression analysis in R from Laurent Berge.

```
ols_unitary1 <- feols(sd_dropout_prop_b ~ unitary | year + leaid,
                      data=desegregation,
                      vcov = "iid", weights=desegregation$sd_t_1619_b)
summary(ols_unitary1)
```

```
#> OLS estimation, Dep. Var.: sd_dropout_prop_b
#> Observations: 1,403
#> Fixed-effects: year: 3,  leaid: 476
#> Standard-errors: IID
#>         Estimate Std. Error t value   Pr(>|t|)
#> unitary 0.013001   0.003121 4.16527 3.4012e-05 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> RMSE: 1.17356     Adj. R2: 0.558947
#>                 Within R2: 0.01843
```

Can you interpret this output? (*ignore the line beginning vcov for now*)

Addressing serial correlation

The worry: within-unit correlation of outcomes (e.g., within-state, across state-years) results in correlated (and therefore too small) standard errors. As a result out **statistical inference** will be incorrect.

The solution: **cluster-robust standard errors**[1]. Clustering standard errors by the k^{th} regressor inflates iid OLS standard errors by:

$$\tau_k \simeq 1 + \rho_{x_k}\rho_\mu(\bar{N}_g - 1)$$

where ρ_{x_k} is the within-cluster correlation of regressor x_{igk}, ρ_μ is the within-cluster error correlation and $\bar{N}_g$ is the average cluster size.

τ_k is **asymptotically** correct as number of clusters increase. Current consensus: this estimate of τ_k is accurate with **~50 clusters**. Fewer than 40, and this approach can dramatically under-estimate SEs (consider bootstrapping).

Best practice: cluster at the unit of treatment (or consider two-way clustering).[2]

[1] Read all about cluster-robust standard errors in Cameron & Miller's (2015) accessible practitioner's guide to standard errors.
[2] Bertrand, Mullainathan & Duflo (2004) and Abadie et al. (2017).

Clustered standard errors (1)

```
ols_unitary2 <- feols(sd_dropout_prop_b ~ unitary | leaid + year,
                     data=desegregation,
                     weights=desegregation$sd_t_1619_b)

summary(ols_unitary2)
```

```
#> OLS estimation, Dep. Var.: sd_dropout_prop_b
#> Observations: 1,403
#> Fixed-effects: leaid: 476,  year: 3
#> Standard-errors: Clustered (leaid)
#>         Estimate Std. Error t value  Pr(>|t|)
#> unitary 0.013001   0.004831 2.69102 0.0073747 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> RMSE: 1.17356     Adj. R2: 0.558947
#>                 Within R2: 0.01843
```

Default behavior in `fixest` is to cluster standard errors on the first fixed effect.

Clustered standard errors (2)

```
ols_unitary3 <- feols(sd_dropout_prop_b ~ unitary | leaid + year,
                     data=desegregation,
                     vcov = ~ leaid^year,
                     weights=desegregation$sd_t_1619_b)

summary(ols_unitary3)
```

```
#> OLS estimation, Dep. Var.: sd_dropout_prop_b
#> Observations: 1,403
#> Fixed-effects: leaid: 476,  year: 3
#> Standard-errors: Clustered (leaid^year)
#>          Estimate Std. Error t value  Pr(>|t|)
#> unitary 0.013001   0.004774 2.72341 0.0065413 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> RMSE: 1.17356     Adj. R2: 0.558947
#>                 Within R2: 0.01843
```

We are going to cluster our standard errors **at the level of assignment to treatment**: the district-year .

Addressing serial correlation

A taxonomy of models estimating the end of school desegregation on the black dropout rate, by std. error clustering approach

	Unclustered	Clustered (Unit)	Clustered (Unit*Period)
unitary	0.013***	0.013**	0.013**
	(0.003)	(0.005)	(0.005)
Num.Obs.	1403	1403	1403
R2	0.709	0.709	0.709
Std.Errors	IID	by: leaid	by: leaid^year
FE: year	X	X	X
FE: leaid	X	X	X
Notes: The table displays coefficients from Equation X with standard errors in parentheses.			
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001			

Doesn't make too much of a difference here...

Addressing parallel trends

A parametric approach

$$
\begin{aligned}
\text{DROPOUT\_BLACK}_{jt} = & \beta_1 \text{UNITARY}_{jt} + \beta_2 (\text{UNITARY} \times \text{YEAR})_{jt} + \\
& \beta_3 \text{RUN\_TIME}_{jt} + \Gamma_j + \Pi_t + \epsilon_j
\end{aligned}
$$

What is this RUN_TIME_{jt} and how do we code it?

```
desegregation <- desegregation %>%
  mutate(run_time = case_when(
   !is.na(yrdiss) ~ (year - yrdiss),
   is.na(yrdiss) ~ -1 ## <-- this is funky, let's talk about it
  ))
summary(desegregation$run_time)
```

```
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>  -19.00   -1.00   -1.00   -1.51   -1.00   19.00
```

Look at RUN_TIME in the data

Show 9 entries Search:

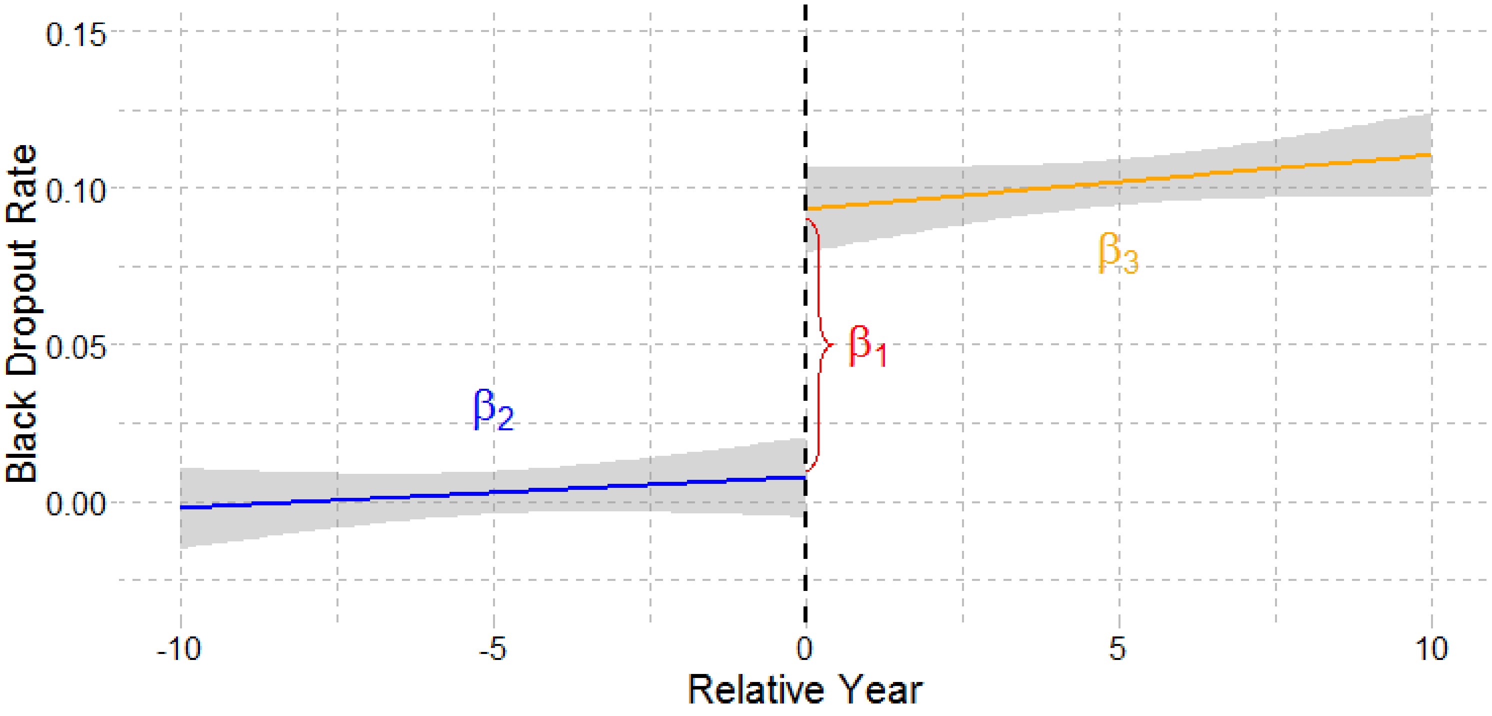
	leaid	year	STATE	unitary	sd_dropout_prop_b	yrdiss	run_time
1	0100030	1990	01	0	0.163434907793999	2002	-12
2	0100030	2000	01	0	0.185185179114342	2002	-2
3	0100030	2010	01	1	0.101694911718369	2002	8
4	0100090	1990	01	0	0.213333338499069		-1
5	0100090	2000	01	0	0.159653469920158		-1
6	0100090	2010	01	0	0.103174604475498		-1
7	1000230	1990	10	0	0.0961737334728241	1996	-6
8	1000230	2000	10	1	0.158327624201775	1996	4
9	1000230	2010	10	1	0.0997624695301056	1996	14

Showing 1 to 9 of 9 entries

Previous 1 Next

Map coefficients to graph

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



Remember: given the structure of our model, these parameters are estimated *relative to untreated and not-yet-treated districts.*

Parallel trends?

```
ols_unitary_run <- feols(sd_dropout_prop_b ~ unitary +
      unitary:run_time + run_time |
      year + leaid, data=desegregation,
      vcov = ~leaid^year, weights=desegregation$sd_t_1619_b)
summary(ols_unitary_run)
```

```
#> OLS estimation, Dep. Var.: sd_dropout_prop_b
#> Observations: 1,403
#> Fixed-effects: year: 3,  leaid: 476
#> Standard-errors: Clustered (leaid^year)
#>                   Estimate Std. Error  t value Pr(>|t|)
#> unitary           0.008785   0.006112  1.43720 0.150884
#> run_time          0.001120   0.000588  1.90396 0.057119 .
#> unitary:run_time -0.001446   0.000689 -2.09977 0.035928 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> RMSE: 1.16841     Adj. R2: 0.561866
#>                 Within R2: 0.027039
```

How would this graph look different than the one on previous slide?

A complete table!

Table 2. Effects of end of school desegregation on black dropout rate

	1	2	3
Unitary status	0.013**	0.013**	0.009
	(0.005)	(0.005)	(0.006)
Pre-trend			0.001+
			(0.001)
Unitary x Relative-Year			-0.001*
			(0.001)
Covariates?		X	X
Num.Obs.	1403	1403	1403
R2	0.709	0.710	0.712

Notes: `$^{+}p < 0.1, ^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$`. The table displays coefficients from Equation X and district-by-year clustered standard errors in parentheses. All models include fixed effects for year and district. Models 2 and 3 adjust for the proportion of 16-19 year-olds residing in the district in 1990 who were Black, interacted with year.

A flexible approach

What if, instead of assigning a particular functional form to our treatment effects over time (either mean, linear or higher-order polynomial), we specified an entirely flexible model?

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

Could also write as:

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

Think for a moment what this model does?

> The model adjusts its estimates of the mean rate of Black dropout in district *j* by the mean rate of Black dropout in year *t* across all districts. Then, it estimates what effect does being *t* years pre- or post-unitary have. The comparison in each of these β s is to being never or not yet *UNITARY*.

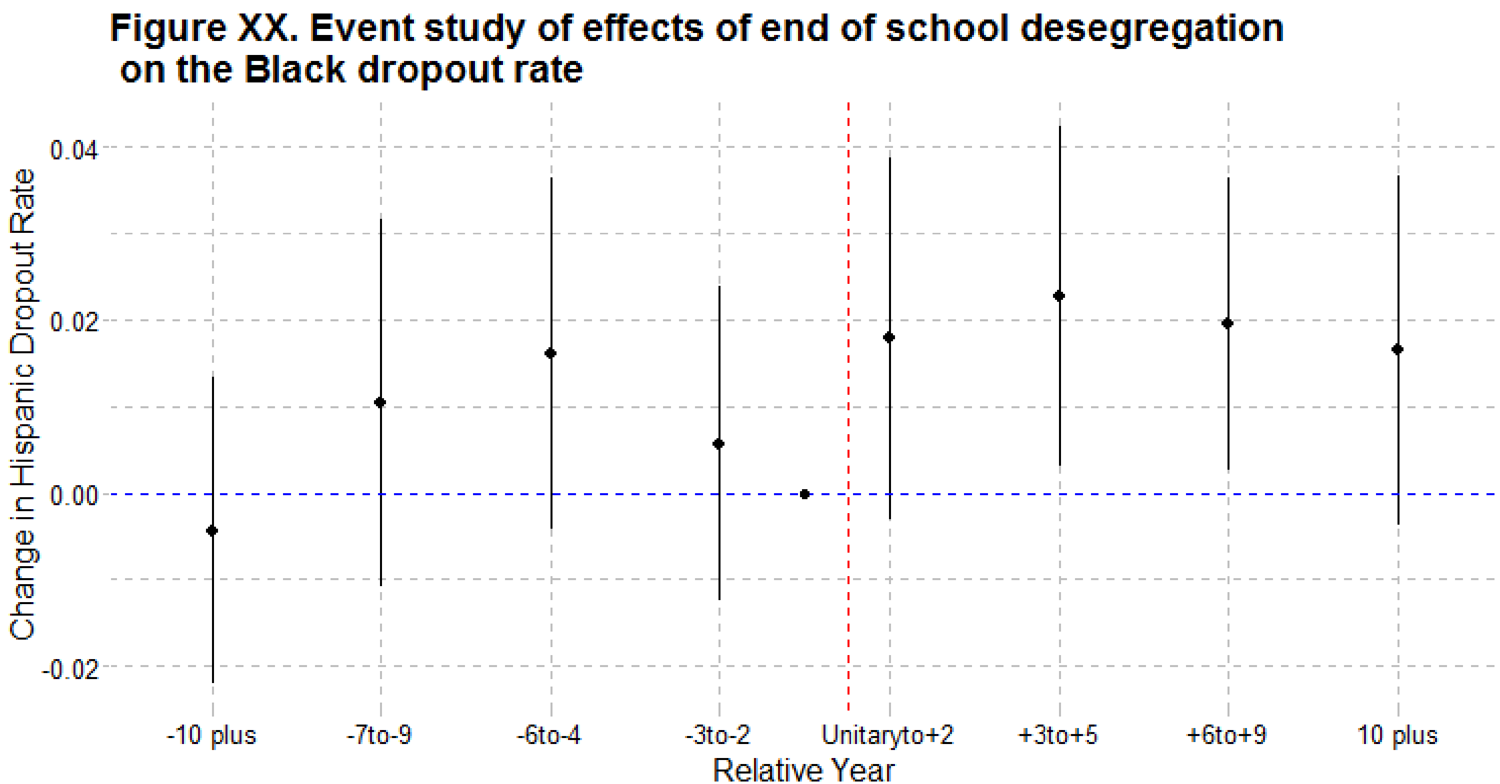
Event study

This would permit a **fully flexible specification**, permitting us to both evaluate **violations of the PTA** and assess potential **dynamic effects** of the treatment:

```
...
#> Observations: 1,403
#> Fixed-effects: year: 3,  leaid: 476
#> Standard-errors: Clustered (leaid^year)
#>              Estimate Std. Error   t value Pr(>|t|)
#> r_10minus   -0.004374   0.009016 -0.485116 0.627670
#> r_7to9minus  0.010307   0.010810  0.953441 0.340531
#> r_6to4minus  0.016110   0.010346  1.557172 0.119655
#> r_3to2minus  0.005682   0.009276  0.612519 0.540294
#> r_0to2plus   0.017822   0.010649  1.673610 0.094430 .
#> r_3to5plus   0.022691   0.009976  2.274513 0.023086 *
#> r_6to9plus   0.019587   0.008598  2.278103 0.022870 *
#> r_10plus     0.016468   0.010330  1.594247 0.111106
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
```

What has happened to our standard errors? (think about **bias v. variance tradeoff**)

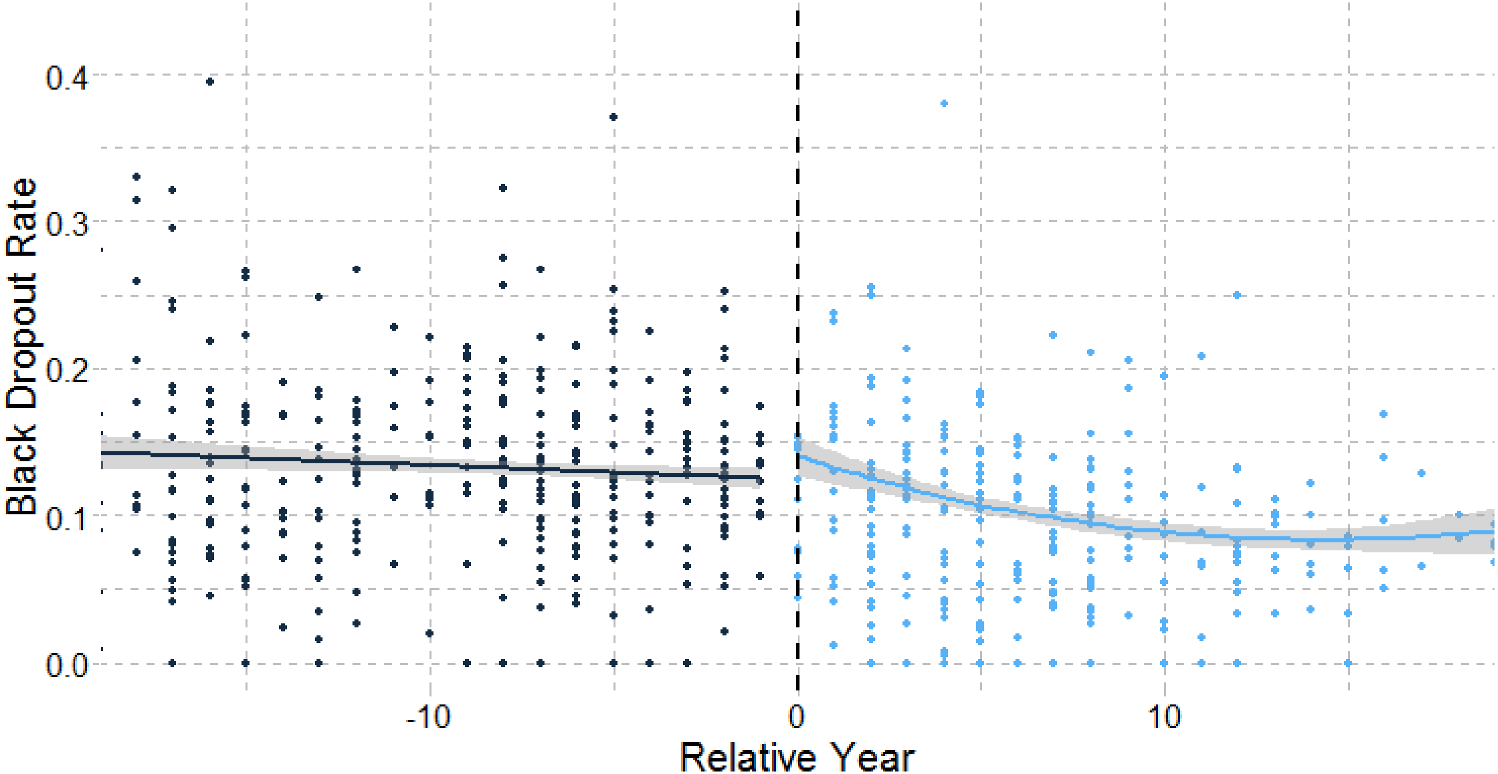
Event study visualized



> The end of desegregation efforts had a causal effect on the Black dropout rate, resulting in a discontinuous and persistent increase of between 1 and 2 percentage points (*caveats, caveats*).

C-ITS

An aside on the related Comparative-Interrupted Time Series approach:



C-ITS considered

Strengths

- Takes advantage of full range of data
- Compared to mean-effect-only DD, allows differentiation of discontinuous jump vs. post-trend
- Permits modeling of fully flexible functional form (can include quadratic, cubic, quartic relationships, interactions and more!)
- Data-responsive approach

Weaknesses

- Encourages over-fitting
- Functional-form dependent
- Risks generating unstable models

Note that a fully-saturated C-ITS model (i.e., a model that estimates a coefficient on an indicator for each time period) is identical to an event study.

Wrap-up

Goals

1. Describe threats to validity in difference-in-differences (DD) identification strategy and multiple approaches to address these threats.

2. Using a cleaned dataset, estimate multiple DD specifications in R and interpret these results

To-Dos

Reading: Liebowitz, Porter & Bragg (2019)

- Critical to read the paper and answer a small set of questions as preparation for DARE
- *Further*: MHE: Ch. 5, 'Metrics: Ch. 5, Mixtape:

DARE #1

- Due 9:00am, January 21 (different due date bc MLK Jr Day)
- Let's look at assignment
- Submit code and memo in response to questions
- Indicate partners (or not)
- I am available for support!

Research Project Proposal due 9am, 1/27

- Talk to me!

Feedback

Plus/Deltas

Front side of index card

Clear/Murky

On back