A Different Diff-in-Diffs?
Fresh Takes on a Familiar Estimator

Tim Simcoe

Boston University, Questrom School of Business
and NBER

Utah Winter Strategy Conference
January 2022

Slides & Example Code `http://people.bu.edu/tsimcoe/data`

Diff-in-Diffs Setup

- RQ: What is the effect of $T \in \{0, 1\}$ on Y?

Diff-in-Diffs Setup

- RQ: What is the effect of $T \in \{0, 1\}$ on Y?
- Suppose we have Before (Y_B) and After (Y_A) data
- Three intuitive estimators
  1. Treatment vs. Control: $E[Y_A|T = 1] - E[Y_A|T = 0]$
  2. Before vs. After: $E[Y_A|T = 1] - E[Y_B|T = 1]$
  3. Diff-in-Diffs: $E[Y_A - Y_B|T = 1] - E[Y_A - Y_B|T = 0]$

Diff-in-Diffs Setup

- RQ: What is the effect of $T \in \{0, 1\}$ on Y?
- Suppose we have Before (Y_B) and After (Y_A) data
- Three intuitive estimators
  1. Treatment vs. Control: $E[Y_A | T = 1] - E[Y_A | T = 0]$
  2. Before vs. After: $E[Y_A | T = 1] - E[Y_B | T = 1]$
  3. Diff-in-Diffs: $E[Y_A - Y_B | T = 1] - E[Y_A - Y_B | T = 0]$
- DD + parallel trends assumption $\Rightarrow$ causal estimate

$$E[\underbrace{Y_A^1 - Y_B | T = 0}_{\text{Observed}}] = E[\underbrace{Y_A^0 - Y_B | T = 1}_{\text{Counterfactual}}]$$

John Snow (1854)

Cholera Cases per 10K Households

	1849	1854	After - Before
Lambeth (T=1)	85	19	-66
Southwark & Vauxhall	135	147	12
Treated - Control	-50	-128	-78

Card & Kruger (1994)

TABLE 5.2.1
Average employment in fast food restaurants before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (.94)	21.03 (.52)	−.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	.59 (.54)	2.76 (1.36)

Notes: Adapted from Card and Krueger (1994), table 3.

Popular Variations on Diff-in-Diffs

- DD as Linear Regression
  - $E[Y_{it}] = \alpha_0 + \alpha_1 Treated_i + \lambda Post_t + \beta Treated_i * Post_t$

Popular Variations on Diff-in-Diffs

- DD as Linear Regression
  - $E[Y_{it}] = \alpha_0 + \alpha_1 Treated_i + \lambda Post_t + \beta Treated_i * Post_t$
- Two-way Fixed-effects (TWFE) Specification
  - $E[Y_{it}] = \alpha_i + \lambda_t + \beta PostTreated_{it}$

Popular Variations on Diff-in-Diffs

- DD as Linear Regression
  - $E[Y_{it}] = \alpha_0 + \alpha_1 Treated_i + \lambda Post_t + \beta Treated_i * Post_t$
- Two-way Fixed-effects (TWFE) Specification
  - $E[Y_{it}] = \alpha_i + \lambda_t + \beta PostTreated_{it}$
- Event Study Specification
  - $E[Y_{it}] = \alpha_i + \lambda_t + \sum_k \beta_k 1[t - TreatmentYear_i = k]$
  - Plot dynamic treatment effects β_k (normalizing $\beta_{-1} = 0$)
  - "Pre trends" falsification test: $H_0 : \beta_k = 0$ for $k \leq -1$

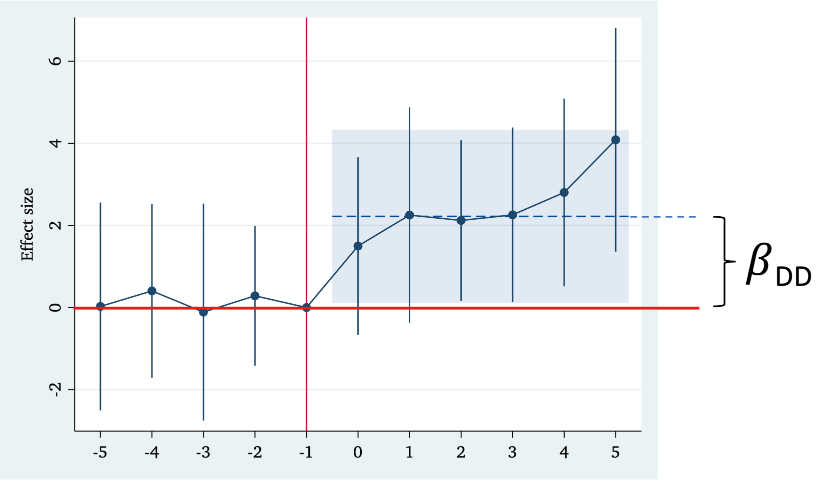
So What is the Problem?

Recent econometrics literature has emphasized two issues with DD:

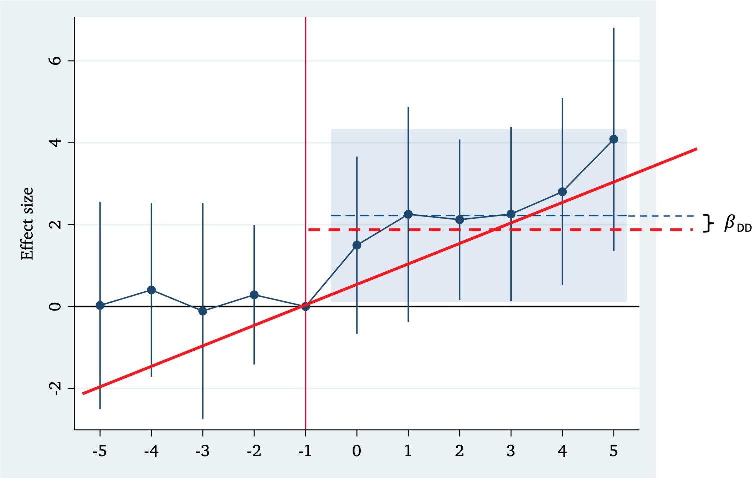
1. Violations of parallel trends assumption
2. Identification under staggered adoption

Takeaway: If you are willing to assume parallel trends, and do not have staggered adoption, previous DD estimators work fine!

The Power of Parallel Trends



The Power of Parallel Trends



What to do about $\nparallel$ trends?

- Keep in mind, parallel trends is a maintained assumption

What to do about $\not\parallel$ trends?

- Keep in mind, parallel trends is a maintained assumption
- Enlarge the standard errors?
  - "An Honest Approach to $\parallel$ Trends" (Rambachan & Roth, WP)
  - User specified close-to-parallel trends $\Rightarrow$ set identification

What to do about $\nparallel$ trends?

- Keep in mind, parallel trends is a maintained assumption
- Enlarge the standard errors?
  - "An Honest Approach to $\parallel$ Trends" (Rambachan & Roth, WP)
  - User specified close-to-parallel trends $\Rightarrow$ set identification
- Relax our rhetoric
  - Failing to reject pre-trends $\neq$ "Proving" parallel trends!!!
  - Authors: Don't over-sell your noisy falsification tests
  - Referees: Don't be Manichean about pre-trend testing

Implications of Staggered Adoption

- Staggered Adoption $\Rightarrow$ units i have different treatment dates

Implications of Staggered Adoption

- Staggered Adoption $\Rightarrow$ units i have different treatment dates
- Consider the DD and TWFE regressions:
  1. $E[Y_{it}] = \alpha_0 + \alpha_1 Treated_i + \lambda Post_t + \beta Treated_i * Post_t$
  2. $E[Y_{it}] = \alpha_i + \lambda_t + \beta PostTreated_{it}$
- Can't estimate (1), because $Post_t$ is undefined for controls...
- but can estimate (2), even without a control group!!

Implications of Staggered Adoption

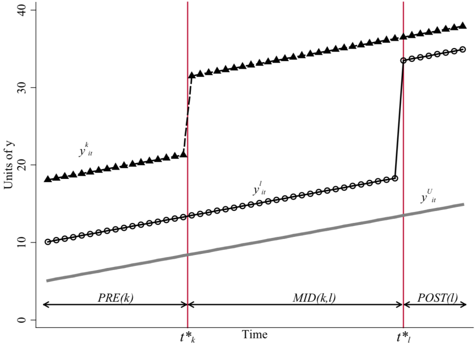
- Staggered Adoption $\Rightarrow$ units i have different treatment dates
- Consider the DD and TWFE regressions:
  1. $E[Y_{it}] = \alpha_0 + \alpha_1 Treated_i + \lambda Post_t + \beta Treated_i * Post_t$
  2. $E[Y_{it}] = \alpha_i + \lambda_t + \beta PostTreated_{it}$
- Can't estimate (1), because $Post_t$ is undefined for controls...
- but can estimate (2), even without a control group!!

Key Point: Differences in treatment timing produce new comparisons, and therefore new sources of identification.

Goodman-Bacon (2019)

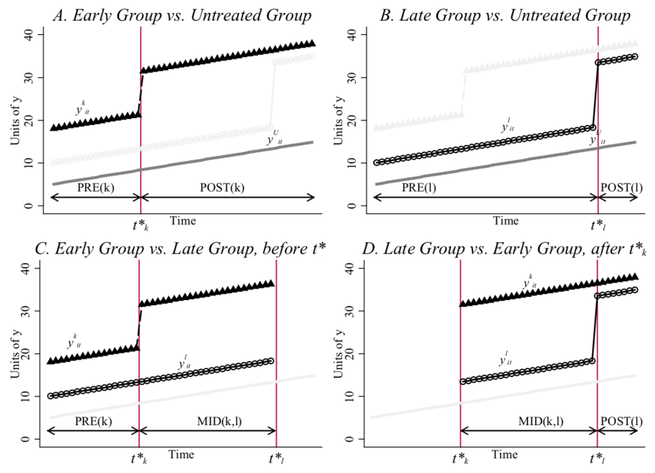
- TWFE β is a weighted average of 2 x 2 DD's
- Three Group Example: Early, Late & Never Treated

Figure 1. Difference-in-Differences with Variation in Treatment Timing: Three Groups



Components of the GB Decomposition

Figure 2. The Four Simple (2x2) Difference-in-Differences Estimates from the Three Group Case



GB Decomposition Theorem

Theorem 1. Difference-in-Differences Decomposition Theorem
Assume that the data contain $k = 1,\dots,K$ groups of units ordered by the time when they receive a binary treatment, $t_k^* \in (1,T]$. There may be one group, U, that never receives treatment. The OLS estimate, $\widehat{\beta}^{DD}$, in a two-way fixed-effects regression (2) is a weighted average of all possible two-by-two DD estimators.

$$\widehat{\beta}^{DD} = \sum_{k \neq U} s_{kU}\, \widehat{\beta}_{kU}^{2x2} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^{k}\, \widehat{\beta}_{k\ell}^{2x2,k} + s_{k\ell}^{\ell}\, \widehat{\beta}_{k\ell}^{2x2,\ell} \right]. \qquad (10a)$$

Where the 2x2 DD estimators are:

$$\widehat{\beta}_{kU}^{2x2} \equiv \left(\overline{y}_k^{POST(k)} - \overline{y}_k^{PRE(k)} \right) - \left(\overline{y}_U^{POST(j)} - \overline{y}_U^{PRE(j)} \right), \qquad (10b)$$

$$\widehat{\beta}_{k\ell}^{2x2,k} \equiv \left(\overline{y}_k^{MID(k,\ell)} - \overline{y}_k^{PRE(k)} \right) - \left(\overline{y}_\ell^{MID(k,\ell)} - \overline{y}_\ell^{PRE(k)} \right), \qquad (10c)$$

$$\widehat{\beta}_{k\ell}^{2x2,\ell} \equiv \left(\overline{y}_\ell^{POST(\ell)} - \overline{y}_\ell^{MID(k,\ell)} \right) - \left(\overline{y}_k^{POST(\ell)} - \overline{y}_k^{MID(k,\ell)} \right). \qquad (10d)$$

Weights, s_k, reflect sample size and treatment-variance for each "timing group" k

Goodman-Bacon Takeaways

1. GB is not a method or solution to biased TWFE estimates
    - The Stata/R "bacondecomp" module is a diagnostic tool

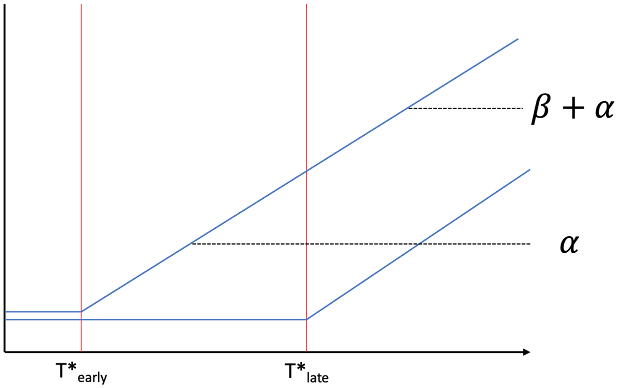
Goodman-Bacon Takeaways

1. GB is not a method or solution to biased TWFE estimates
   - The Stata/R "bacondecomp" module is a diagnostic tool
2. GB highlights key problem with Staggered DD
   - TWFE uses early-treated as control group for late-treated!
   - ...which is also why we can estimate β without controls

Goodman-Bacon Takeaways

1. GB is not a method or solution to biased TWFE estimates
    - The Stata/R “bacondecomp” module is a diagnostic tool
2. GB highlights key problem with Staggered DD
    - TWFE uses early-treated as control group for late-treated!
    - ...which is also why we can estimate β without controls
3. Points towards excluding “forbidden comparisons”
    - Callaway and Sant’Anna (aggregation)
    - Borusyak, Jaravel and Spiess (imputation)
    - Matching with pseudo-treatment

Good News! Staggered DD Estimates are (Probably) Conservative



$DD^{late} =$

$$(\alpha - 0) - (\beta + \alpha - \alpha) = \alpha - \beta < \alpha$$

Callaway and Sant'Anna (JOE 2021)

- Let G_i be treatment cohort of unit i ($G_i = \infty$ for Controls)
- C&S define group-time average treatment effects

$$DD(G,T) = E[Y_T - Y_{T_0} | T_0 < G_i < T] - E[Y_T - Y_{T_0} | T < G_i]$$

Callaway and Sant'Anna (JOE 2021)

- Let G_i be treatment cohort of unit i ($G_i = \infty$ for Controls)
- C&S define group-time average treatment effects

$$DD(G,T) = E[Y_T - Y_{T_0} | T_0 < G_i < T] - E[Y_T - Y_{T_0} | T < G_i]$$

- Construct $\widehat{\beta^{DD}}$ as weighted average of $DD(G,T)$'s
  - Basic idea: aggregation of "clean" 2 x 2's
  - Researcher chooses weights $\Rightarrow$ many possibilities
  - Loop over (G, T_0, T) $\Rightarrow$ slow on large panels
  - Paper discusses issues with inference (SEs)
- Wooldridge (2021) shows how to recover $DD(G,T)$'s from OLS with many interacted fixed effects

Borusyak, Jaravel and Spiess (2021, WP)

1. Estimate TWFE model using untreated observations
   - $Y_{it}^0 = \alpha_i + \lambda_t + X_{it}\theta$

Borusyak, Jaravel and Spiess (2021, WP)

1. Estimate TWFE model using untreated observations
   - $Y_{it}^0 = \alpha_i + \lambda_t + X_{it}\theta$
2. Calculate (imputed) treatment effect for treated observations
   - $DD_{it} = Y_{it} - \widehat{Y_{it}^0}$

Borusyak, Jaravel and Spiess (2021, WP)

1. Estimate TWFE model using untreated observations
   - $Y_{it}^0 = \alpha_i + \lambda_t + X_{it}\theta$
2. Calculate (imputed) treatment effect for treated observations
   - $DD_{it} = Y_{it} - \widehat{Y_{it}^0}$
3. Construct $\widehat{\beta^{DD}}$ as weighted average of DD_{it}'s
   - Weighting choices $\Rightarrow$ researcher DOF
   - Faster than C&S, but need to store α_i

Stacked Difference-in-Differences

- Deshpande & Li (2019), Cengiz et al (2019)

1. Choose a time-window (t_{pre}, t_{post})
2. For each treatment cohort G create a new dataset containing
   - Periods $G - t_{pre}$ to $G + t_{post}$ for treated cohort
   - All units not treated over same time period
3. Stack datasets (indexed by c) into one large panel
4. Estimate $E[Y_{cgpit}] = \alpha_{cg} + \lambda_{cp} + \beta PostTreated_{it}$
   - where α_{cg}, λ_{cp} are dataset-by-group and -period effects

DD as Matching

1. For each treated i, pick a similar control $\Rightarrow$ 1:1 match
   - Coarsened matching (CEM) or propensity score
2. Estimate DD or TWFE model. Done.

- Exact matching yields $N_{Treated}$ "clean" DD's

DD as Matching

1. For each treated i, pick a similar control $\Rightarrow$ 1:1 match
   - Coarsened matching (CEM) or propensity score
2. Estimate DD or TWFE model. Done.

- Exact matching yields $N_{Treated}$ "clean" DD's
- Should you match on Y_0 and/or pre-trends?

DD as Matching

1. For each treated i, pick a similar control $\Rightarrow$ 1:1 match
   - Coarsened matching (CEM) or propensity score
2. Estimate DD or TWFE model. Done.

- Exact matching yields $N_{Treated}$ "clean" DD's
- Should you match on Y_0 and/or pre-trends?
- It depends. Chabe-Ferret (JOE 2015) $\Rightarrow$ FE and matching are not complementary

I have a DD Paper. What Should I do?

- Simultaneous Adoption $\Rightarrow$ "old" DD or TWFE
  - Can add matching / weighting to address selection
- Sequential Adoption
  - Large N_i / many controls $\Rightarrow$ exact matching
  - Large N_i / mostly treated $\Rightarrow$ BJS
  - Small N_i / mostly treated $\Rightarrow$ C&S or Stacked DD

Leading use case for new estimators: Impact of policy adopted by most states at different times (e.g. smoking bans)

Rysman & Simcoe (MS 2008)

What is the impact of technology standardization on "patent value"?

- Sample of patents "declared essential" ($T_i = 1$) to SSOs
- Controls ($T_i = 0$) matched on vintage and tech-class
- Panel Data: $i =$ Patent, $t =$ Year
  - Outcome $Y_{it} =$ Citation count
  - Disclosure years $\Rightarrow$ staggered treatment
- Data & Code: `http://people.bu.edu/tsimcoe/data`

Baseline 2 x 2 DD

. reg cites PostTreat TreatGroup PostPeriod, cluster(pat)

```
Linear regression                               Number of obs     =     67,367
                                                F(3, 6139)        =      67.96
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0222
                                                Root MSE          =     5.4855

                             (Std. Err. adjusted for 6,140 clusters in pat_id)
```

cites	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
PostTreat	.0938327	.1670108	0.56	0.574	-.2335671	.4212324
TreatGroup	1.578988	.1493449	10.57	0.000	1.286219	1.871756
PostPeriod	.2001718	.0985406	2.03	0.042	.0069977	.393346
_cons	1.717133	.0753659	22.78	0.000	1.569389	1.864876

Baseline DD with Year Effects

. reg cites PostTreat TreatGroup i.year, cluster(pat)

```
Linear regression                               Number of obs     =     67,367
                                                F(32, 6139)       =      29.68
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0400
                                                Root MSE          =     5.4365

                            (Std. Err. adjusted for 6,140 clusters in pat_id)
------------------------------------------------------------------------------
             |               Robust
       cites |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   PostTreat |   .5879134   .1546432     3.80   0.000     .2847585    .8910683
  TreatGroup |   1.302752   .1480573     8.80   0.000     1.012508    1.592997
             |
        year |
       1981  |  -.1254601   .1565342    -0.80   0.423    -.4323219    .1814017
       1982  |  -.1384076   .1424279    -0.97   0.331    -.4176162     .140801
```

Two-Way Fixed Effects

. xtreg cites PostTreat i.year, fe i(pat) robust

```
Fixed-effects (within) regression               Number of obs     =     67,367
Group variable: pat_id                          Number of groups  =      6,140

R-sq:                                           Obs per group:
     within  = 0.0270                                         min =          1
     between = 0.0222                                         avg =       11.0
     overall = 0.0228                                         max =         26

                                                F(31,6139)        =      19.49
corr(u_i, Xb)  = 0.0233                         Prob > F          =     0.0000

                             (Std. Err. adjusted for 6,140 clusters in pat_id)
```

cites	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
PostTreat	.9803645	.1001289	9.79	0.000	.7840767	1.176652
year						
1981	.1006381	.3203953	0.31	0.753	-.5274491	.7287252
1982	-.0227574	.4543496	-0.05	0.960	-.9134419	.8679272

Matching + DD

. xtreg cites PostTreat PostPeriod i.year, fe i(pat) robust

```
Fixed-effects (within) regression               Number of obs     =     67,367
Group variable: pat_id                          Number of groups  =      6,140

R-sq:                                           Obs per group:
     within  = 0.0275                                         min =          1
     between = 0.0258                                         avg =       11.0
     overall = 0.0220                                         max =         26

                                                F(32,6139)        =      19.82
corr(u_i, Xb)  = 0.0250                         Prob > F          =     0.0000

                                (Std. Err. adjusted for 6,140 clusters in pat_id)
```

cites	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
PostTreat	.7555577	.1122421	6.73	0.000	.5355238	.9755917
PostPeriod	.3430708	.0834113	4.11	0.000	.1795553	.5065863
year						
1981	.0982575	.3209077	0.31	0.759	-.530834	.727349
1982	-.0426922	.4568111	-0.09	0.926	-.9382021	.8528178

Borusyak, Jaravel & Spiess (Imputation)

. net install did_imputation . replace TreatYr = . if (TreatGroup==0)
. did_imputation cites pat_id year TreatYr, autosample

```
Warning: part of the sample was dropped for the following coefficients because

                                              Number of obs     =     64,757
```

cites	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tau	.7901201	.1123972	7.03	0.000	.5698256	1.010415

Callaway & Sant'Anna (GT Effects)

. net install csdid

. replace TreatYr = 0 if (TreatGroup==0)

. csdid cites, ivar(pat_id) time(year) gvar(TreatYr)

. estat simple

```
Units always treated found. These will be excluded
Panel is not balanced
Will use observations with Pair balanced (observed at t0 and t1)
.....................xxxxxxxxxx...................
....xxxxxxxxx.......................xxxx..........
.................xxx............................xx
............................xx....................
..........xxx.....................................
..................................................
..................................................
........................................x.........
....................xxx...........................
xx............................x...................
..........xxx...........................xxxxxx....
....................xx............................
xxxxxxxx......................xxxxxx..............
..........xxxxxxxxxxx...................xxxxxxxx..
```

References & Extra Resources

References

- Borusyak, K., X. Jaravel & J. Spiess (2021) "Revisiting Event Study Designs: Robust and Efficient Estimation" *Working Paper*.
- Callaway, B. & P. Sant'Anna (2021) "Difference-in-Differences with multiple time periods" *Journal of Econometrics*, 225(2): 200–230.
- Cengiz, D., A. Dube, A. Lindner, & B. Zipperer (2019) "The effect of minimum wages on low-wage jobs." *Quarterly Journal of Economics*, 134(4): 1405–1454.
- Deshpande, M. & Y. Li (2019) "Who is screened out? Application costs and the targeting of disability programs." *AEJ: Economic Policy*, 11(4): 213–248.
- Goodman-Bacon, A. (2021) "Difference-in-differences with variation in treatment timing" *Journal of Econometrics*, 225(2): 254–277.
- Rambachan A. & J. Roth (2021) "An Honest Approach to Parallel Trends" *Working Paper*.
- Wooldridge, J. (2021) "Two-Way Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators" *Working Paper*.

Resources

- Stata/R Packages: bacondecomp, csdid, did_imputation
- `https://taylorjwright.github.io/did-reading-group/`