Introduction to Difference-in-Differences with Variation in Treatment Timing

Much Rosidi

November 27, 2023

Presented at Department of Accounting, Faculty of Economics and Business, Universitas Indonesia

About Me

- PhD student in Public Policy & Administration at University of Kentucky
- Previous education: MA (Applied Economics) from University of Michigan (2014), BAppSc (Accounting) from STAN (2006)
- Not an econometrician, but a user of econometrics. I firmly believe that teaching is the most effective way to master a subject.
- Conducts research primarily in public finance and public financial management using Indonesian data.

Agenda

1. Canonical DID (2 groups 2 periods)
2. DID with variation in treatment timing
3. Recent development of DID
4. STATA simulation

1. Canonical DID (2 Groups 2 Periods)

Canonical DID Model

- DID estimator compares the mean outcomes from period 1 to 2 between a treatment group s (switches from untreated to treated) and a control group n (untreated at both periods).

$$DID = \underbrace{(\bar{Y}_{s,2} - \bar{Y}_{s,1})}_{\text{difference for treated group}} - \underbrace{(\bar{Y}_{n,2} - \bar{Y}_{n,1})}_{\text{difference for control group}}$$

- DID relies on parallel trends assumption: In the absence of treatment, both treated and control groups would have experienced the same trend.

$$E[Y_{s,2}(0) - Y_{s,1}(0)] = E[Y_{n,2}(0) - Y_{n,1}(0)]$$

Canonical DID - Example

$$DID = \underbrace{(\bar{Y}_{s,2} - \bar{Y}_{s,1})}_{\text{difference for treated group}} - \underbrace{(\bar{Y}_{n,2} - \bar{Y}_{n,1})}_{\text{difference for control group}}$$

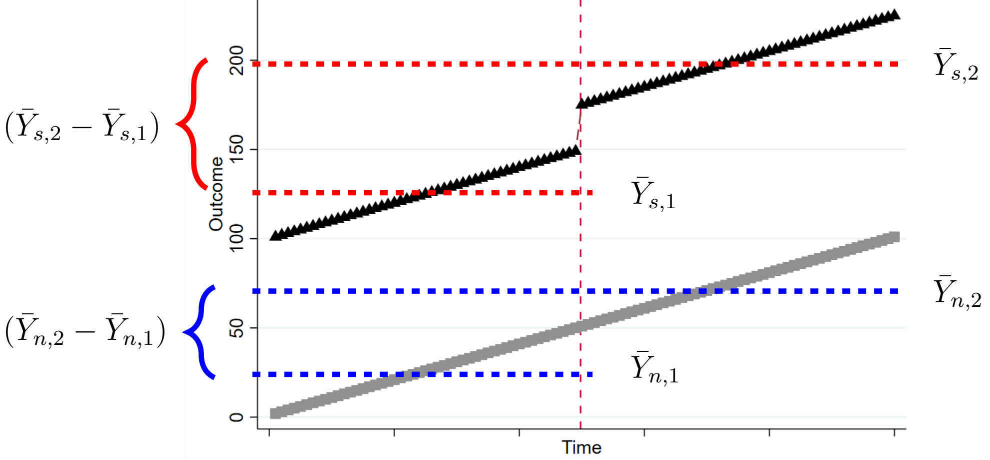
	Treatment (NJ)	Control (PA)	Difference
T=2 (after)	21,03	21,15	-0,12
T=1 (before)	20,44	23,30	-2,86
Difference	0,59	(2,15)	2,74

difference-in-differences

Source: Card and Krueger (1994) in Cunningham (2021)

Visualizing the Setting

$$DID = \underbrace{(\bar{Y}_{s,2} - \bar{Y}_{s,1})}_{\text{difference for treated group}} - \underbrace{(\bar{Y}_{n,2} - \bar{Y}_{n,1})}_{\text{difference for control group}}$$



Canonical DID Model

- DID is equal to coefficient β in OLS regression:

$$Y_{it} = \alpha_i + \theta_t + \beta D_{it} + \epsilon_{it}$$

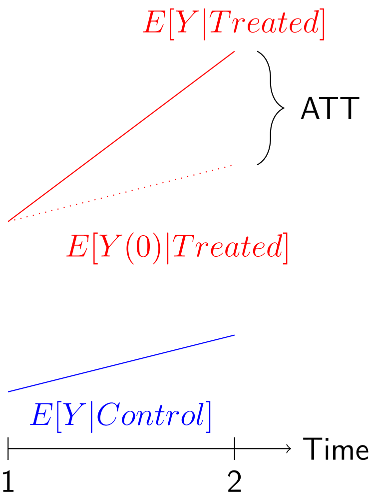
  where $D_{it} = D_i * [t = 2]$

- Under parallel trends assumption, DID is unbiased for the average treatment effect on the treated (ATT)

$$E[DID] = E[Y_{s,2}(1) - Y_{s,2}(0)]$$

  $Y_{s,2}(0)$ is never observed. In potential outcome framework, We can only observe one outcome ($Y_{s,2}(1)$).

Visualizing DID



- The red dotted line is counterfactual* and never observed. DiD "impute" it with control group outcome.
- Parallel (or common) trend assumption is very important in DID, and by definition untestable.

*what would have happened in the absence of a particular treatment.

Proof

Adding zero to the switching equation:

$$
\begin{aligned}
E[DID] &= E[Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1})] \\
&= E[Y_{s,2}(1) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0))] \\
&= E[Y_{s,2}(1) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0))] + \underbrace{E[Y_{s,2}(0) - Y_{s,2}(0)]}_{\text{Zero}} \\
&= \underbrace{E[Y_{s,2}(1) - Y_{s,2}(0)]}_{\text{ATT}} + \underbrace{E[Y_{s,2}(0) - Y_{s,1}(0)] - E[Y_{n,2}(0) - Y_{n,1}(0)]}_{\text{Parallel trend}}
\end{aligned}
$$

For DID estimates to be unbiased, the second term needs to be zero.

Parallel Trend

- Evaluating the parallel trend assumption requires counterfactual which is unobservable.
- Instead, researchers test for parallel pre-trend, estimating placebo pre-treatment DID coefficients. If placebo coefficients are not statistically significant, treatment and control group follow a similar trend in pre-treatment periods.
- But, if they had been similar before, what guarantee they continue trending similarly post treatment absent the treatment?

Exogeneity of Treatment in a DID

- Testing for a parallel pre-trend is crucial, yet it does not guarantee a common trend. The parallel trend assumption can be violated if the treatment is endogenous (Cunningham, 2021).
- Ensuring that treatment is independent of outcome rules out the possibility that a unit is treated due to experiencing negative shocks (Ashenfelter's dip).

Defending the Parallel Trends Assumption

Researchers use these approaches:

- Compelling graph
- Falsification test
- Placebo (parallel pre-trend) test in event study

Compelling Graph - Defending the Parallel Trends

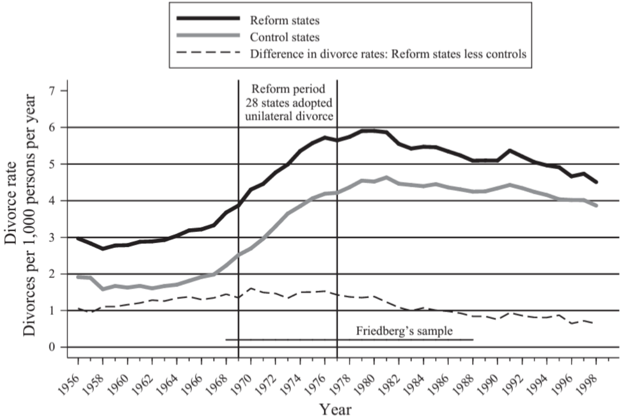


FIGURE 1. AVERAGE DIVORCE RATE: REFORM STATES AND CONTROLS

Source: Wolfers (2006): Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results.

Falsification Test - Defending the Parallel Trends

TABLE III
IMPACT OF LOJACK ON CITY AUTO THEFT RATES

Variable	(1)	(2)	(3)	(4)
Years of Lojack availability	−.109	−.157	—	—
	(.013)	(.021)		
Lojack share	—	—	−.242	−.463
			(.031)	(.065)
Unemployment rate	.019	.026	.017	.028
	(.009)	(.010)	(.009)	(.010)
State real per capita income (×1000)	.022	.028	.016	.022
	(.014)	(.015)	(.014)	(.016)
% Black	−.005	−.005	−.002	.001
	(.008)	(.008)	(.009)	(.009)
% Aged 0–17	.106	.115	.102	.118
	(.030)	(.026)	(.030)	(.027)
% Aged 18–24	.003	−.005	−.004	−.027
	(.039)	(.039)	(.039)	(.041)
% Aged 25–44	.028	.059	.008	.056
	(.039)	(.038)	(.039)	(.039)
ln (sworn officers/per capita)	.044	.060	−.001	−.009
	(.130)	(.133)	(.131)	(.137)
Instrument w/years since Lojack began regulatory process?	No	Yes	No	Yes
Adjusted R^2	.883	—	.882	—
Coefficient on Lojack excluding covariates from the specification	−.086	−.113	−.200	−.333
	(.012)	(.018)	(.028)	(.053)

Dependent variable is ln(reported auto thefts per capita). Data cover the period 1981–1994 and include all 57 U. S. central cities with a population greater than 250,000 in 1981. Lojack share is the estimated percent of total vehicles registered that have Lojack installed in the market. Number of observations is equal to 751 in all columns as a result of occasional missing data. In columns (2) and (4) the number of years elapsed since Lojack began the regulatory approval process is used as an instrument for the Lojack variables. All columns include year dummies and city-fixed effects in addition to the variables shown. Unemployment is the annual SMSA unemployment rate. % Black is linearly interpolated between decennial census years. Age categories refer to state age distributions; the omitted category is percent of the population over age 45. White standard errors are in parentheses. The bottom row of the table presents the coefficient on the Lojack variable in specifications that include only year dummies and city-fixed effects as covariates.

TABLE V
IMPACT OF LOJACK ON CRIMES OTHER THAN AUTO THEFT

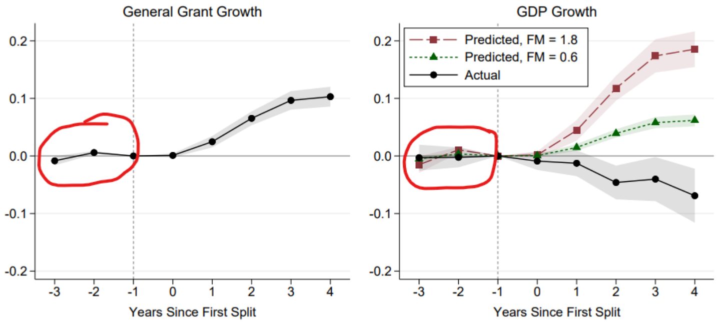
	Substitutable crimes (robbery, burglary, larceny)		Nonsubstitutable crimes (assault, rape, murder)	
Variable	(1)	(2)	(3)	(4)
Years of Lojack availability	−.015	—	−.005	—
	(.009)		(.006)	
Lojack share	—	−.059	—	−.015
		(.015)		(.016)
Unemployment rate	.024	025	−.022	−.022
	(.005)	(.004)	(.006)	(.006)
State real per capita income (×1000)	−.019	−.019	.003	.003
	(.009)	(.009)	(.010)	(.010)
% Black	−.005	−.004	−.001	−.001
	(.004)	(.004)	(.006)	(.006)
% Aged 0–17	−.065	−.064	−.015	−.016
	(.013)	(.013)	(.018)	(.018)
% Aged 18–24	−.037	−.041	−.019	−.020
	(.022)	(.022)	(.029)	(.029)
% Aged 25–44	.099	.102	−.012	−.012
	(.024)	(.024)	(.023)	(.022)
ln(sworn police per capita)	.077	.070	.398	.396
	(.064)	(.063)	(.090)	(.090)
Adjusted R^2	.819	.839	.928	.936
Coefficient on Lojack excluding covariates	.005	−.016	−.016	−.040
	(.006)	(.011)	(.005)	(.008)

Dependent variable is the natural log of the crime categories named. Substitutable crimes are those that are presumed to be close substitutes for auto theft, i.e., robbery, burglary, and larceny. Nonsubstitutable crimes are murder, rape, and aggravated assault. In both cases, the sum of the reported crime rates within the various crime categories is used. Data cover the period 1981–1994 and include all 57 U. S. central cities with a population greater than 250,000 in 1981. Number of observations varies between 742 and 767 based on the number of missing observations. All columns include year dummies and city-fixed effects in addition to the variables shown. Lojack share is the estimated percent of total vehicles registered that have Lojack installed in the market. Unemployment is the annual SMSA unemployment rate. % Black is linearly interpolated between decennial census years. Age categories refer to state age distributions; the omitted category is percent of the population over age 45. White-standard errors are in parentheses. The bottom row of the table presents estimates of the Lojack coefficient from specifications including only year dummies and city-fixed effects as covariates.

Source: Ayres and Levitt (1998): Measuring Positive Externalities From Unobservable Victim Precaution: An Empirical Analysis of LoJack.

Parallel Pre-Trend Test - Defending the Parallel Trends

Figure 2: The Effect of District Splits on General Grant and GDP



Notes: This figure plots estimates of the cohort-size-weighted CATT (Equation (2)) and their 95-percent confidence intervals. (The estimates for $h = -3$ omit the 2002 cohort and adjust the weights accordingly, because the data start in 2000.) The left panel shows the impact of the first district split on growth in general grant revenue relative to the year before the split, scaled by GDP in that year. The right panel shows the impact on GDP growth relative to year before the split as predicted by fiscal multiplier values of 0.6 and 1.8 given the one-for-one increase in expenditure due to the increase in general grants. It also plots the impact on actual GDP growth. The confidence intervals are robust to heteroskedasticity and clustering by district.

Source: Cassidy & Velayudhan (2023) : Government Fragmentation and Economic Growth.

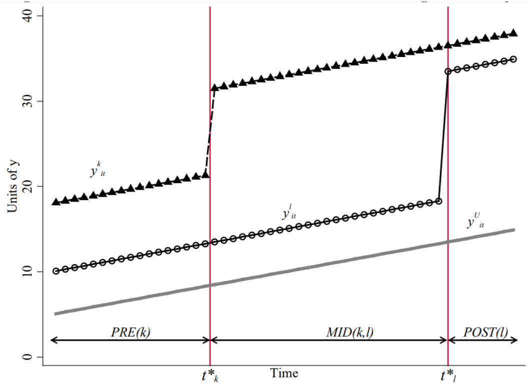
2. DID with Variation in Treatment Timing

Variation in Treatment Timing

- In the canonical DID model: 2 periods (pre and post) and 2 groups (treatment and control)
- Now we focus on the recent literature on multiple periods and variation in adoption timing: ex. provinces/districts adopted policy in different times.

Visualizing the Setting

In 3 group case, $\boxed{D_{it}}$ starts at different times.



Source: Goodman-Bacon (2021): Difference-in-differences with variation in treatment timing.

Static Model

Two Way Fixed Effect (TWFE) Estimator:

$$Y_{it} = \alpha_i + \theta_t + \hat{\beta}^{DD} D_{it} + \epsilon_{it}$$

unit fixed effects

time fixed effects

treatment dummy

STATA code: reg outcome treatment i.state i.year, vce(cluster state)

Dynamic TWFE Model (Event Study)

- TWFE event study equation:

$$Y_{it} = \alpha_i + \theta_t + \underbrace{\boxed{\sum_{\ell=-K}^{-2} \gamma_\ell D_{it}^\ell}}_{\text{lags (pre-treatment)}} + \underbrace{\boxed{\sum_{\ell=0}^{L} \gamma_\ell D_{it}^\ell}}_{\text{leads (post-treatment)}} + \epsilon_{it}$$

where $D_{it}^\ell = 1\{t - G_i = \ell\}$ are lags and leads dummies.

- Example with 3 lags and 3 leads:

$\ell = -1$ omitted

$$Y_{it} = \alpha_i + \theta_t + \gamma_{-3} D_{it}^{-3} + \gamma_{-2} D_{it}^{-2} + \gamma_0 D_{it}^0 + \gamma_1 D_{it}^1 + \gamma_2 D_{it}^2 + \gamma_3 D_{it}^3 + \epsilon_{it}$$

A Stylized Example of TWFE Datasets (1)

State	Year	Start	Treat	Time to Treat	Lag 3	Lag 2	Lag 1	Lead 0	Lead 1	Lead 2	Lead 3
State 1	2001	2004	0	-3	1	0	0	0	0	0	0
State 1	2002	2004	0	-2	0	1	0	0	0	0	0
State 1	2003	2004	0	-1	0	0	**0**	0	0	0	0
State 1	2004	2004	1	0	0	0	0	1	0	0	0
State 1	2005	2004	1	1	0	0	0	0	1	0	0
State 1	2006	2004	1	2	0	0	0	0	0	1	0
State 1	2007	2004	1	3	0	0	0	0	0	0	1
State 1	2008	2004	1	4	0	0	0	0	0	0	**1**
State 1	2009	2004	1	5	0	0	0	0	0	0	**1**
State 1	2010	2004	1	6	0	0	0	0	0	0	**1**
State 2	2001	2005	0	-4	**1**	0	0	0	0	0	0
State 2	2002	2005	0	-3	1	0	0	0	0	0	0
State 2	2003	2005	0	-2	0	1	0	0	0	0	0
State 2	2004	2005	0	-1	0	0	**0**	0	0	0	0
State 2	2005	2005	1	0	0	0	0	1	0	0	0
State 2	2006	2005	1	1	0	0	0	0	1	0	0
State 2	2007	2005	1	2	0	0	0	0	0	1	0
State 2	2008	2005	1	3	0	0	0	0	0	0	1
State 2	2009	2005	1	4	0	0	0	0	0	0	**1**
State 2	2010	2005	1	5	0	0	0	0	0	0	**1**

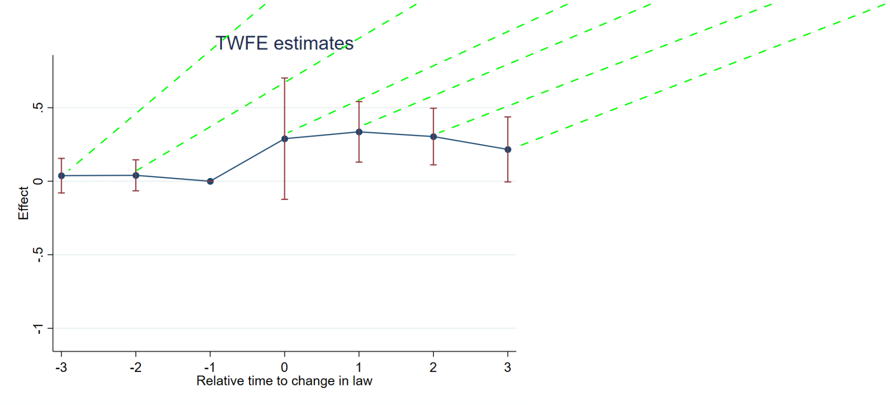
A Stylized Example of TWFE Datasets (2)

State	Year	Start	Treat	Time to Treat	Lag 3	Lag 2	Lag 1	Lead 0	Lead 1	Lead 2	Lead 3
State 3	2001	.	0	.	0	0	0	0	0	0	0
State 3	2002	.	0	.	0	0	0	0	0	0	0
State 3	2003	.	0	.	0	0	0	0	0	0	0
State 3	2004	.	0	.	0	0	0	0	0	0	0
State 3	2005	.	0	.	0	0	0	0	0	0	0
State 3	2006	.	0	.	0	0	0	0	0	0	0
State 3	2007	.	0	.	0	0	0	0	0	0	0
State 3	2008	.	0	.	0	0	0	0	0	0	0
State 3	2009	.	0	.	0	0	0	0	0	0	0
State 3	2010	.	0	.	0	0	0	0	0	0	0
State 4	2001	2006	0	-5	**1**	0	0	0	0	0	0
State 4	2002	2006	0	-4	**1**	0	0	0	0	0	0
State 4	2003	2006	0	-3	1	0	0	0	0	0	0
State 4	2004	2006	0	-2	0	1	0	0	0	0	0
State 4	2005	2006	0	-1	0	0	**0**	0	0	0	0
State 4	2006	2006	1	0	0	0	0	1	0	0	0
State 4	2007	2006	1	1	0	0	0	0	1	0	0
State 4	2008	2006	1	2	0	0	0	0	0	1	0
State 4	2009	2006	1	3	0	0	0	0	0	0	1
State 4	2010	2006	1	4	0	0	0	0	0	0	**1**

Understanding TWFE Plot

reg outcome lag3 lag2 lead0 lead1 lead2 lead3 i.state i.year, vce(cluster state)

$$Y_{it} = \alpha_i + \theta_t + \gamma_{-3} D_{it}^{-3} + \gamma_{-2} D_{it}^{-2} + \gamma_0 D_{it}^0 + \gamma_1 D_{it}^1 + \gamma_2 D_{it}^2 + \gamma_3 D_{it}^3 + \epsilon_{it}$$



TWFE Difference-in-Differences

Two Way Fixed Effect (TWFE) Estimator:

$$Y_{it} = \underbrace{\alpha_i}_{\boxed{\text{unit fixed effects}}} + \underbrace{\theta_t}_{\boxed{\text{time fixed effects}}} + \hat{\beta}^{DD} \underbrace{D_{it}}_{\boxed{\text{treatment dummy}}} + \epsilon_{it}$$

What is $\hat{\beta}^{DD}$ in different treatment timing setting?

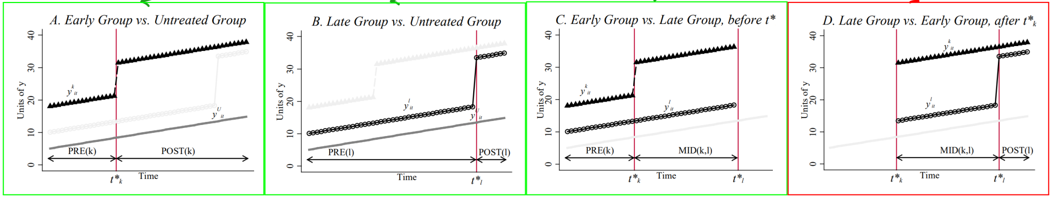
Hint: Goodman-Bacon (2021) & others have the answer.

Decomposition

For 3 groups case:

$$Y_{it} = \alpha_i + \theta_t + \hat{\beta}^{DD} D_{it} + \epsilon_{it}$$

$$\beta_{DD} = s_{kU} \boxed{\hat{\beta}^{2x2}_{kU}} + s_{\ell U} \boxed{\hat{\beta}^{2x2}_{\ell U}} + s^k_{k\ell} \boxed{\hat{\beta}^{2x2,k}_{k\ell}} + s^k_{k\ell} \boxed{\hat{\beta}^{2x2,\ell}_{k\ell}}$$



"Forbidden comparison"

Source: Goodman-Bacon (2021): Difference-in-differences with variation in treatment timing.

Issues with Dynamic TWFE

Sun and Abraham (2021)

$$Y_{it} = \alpha_i + \theta_t + \sum_{\ell=-K}^{-2} \gamma_\ell D_{it}^\ell + \sum_{\ell=0}^{L} \gamma_\ell D_{it}^\ell + \epsilon_{it}$$

where $D_{it}^k = 1\{t - G_i = \ell\}$ are lags and leads dummies.

- Like in static setting, γ_ℓ may contains negative result in the presence of heterogeneous treatment effects.
- The coefficient on a lag and lead may be contaminated by effects from other periods in the presence of heterogeneous treatment effects.

Problem and Solution

- Negative result: DID Estimates with variation in timing are biased in the presence of heterogeneous treatment effect, e.i. when treatment effects vary over time.
- Negative results come from forbidden comparison: late vs early.
- In event study: Contamination effect from other leads and lags.
- New estimators aim to avoid "forbidden comparisons" by employing selective control groups, such as those that were never treated, not yet treated, or last treated.
- New estimators: Perform better under heterogeneous treatment effects.

3. Recent Development of DID

New Estimators

- Several new estimators have been proposed to address the negative weight/forbidden comparison issue.
- A common characteristic among these estimators is their reliance on a "clean comparison" approach (group that were never treated, not yet treated, or last treated).

Callaway and Sant'Anna (2021) Estimator

- Needs to choose comparison groups either never treated or not-yet treated.
- Parameter of interest: ATT in period t for units first treated at period g

$$ATT(g,t) = E[Y_t(g) - Y_t(0)|G = g]$$

- Group-time average treatment effects, based on a "never treated"

$$ATT(g,t) = E[Y_t - Y_{g-1}|G = g] - E[Y_t - Y_{g-1}(0)|C = 1]$$

  where $C = 1$ is never-treated units
- Group-time average treatment effects, based on a "not-yet treated"

$$ATT(g,t) = E[Y_t - Y_{g-1}|G = g] - E[Y_t - Y_{g-1}(0)|D_t = 0, G \neq g]$$

Chaisemartin & D'Haultfuille (2022) Estimator

- Uses comparison groups not-yet treated.

$$DID_{g,l} = Y_{g,F_g+l} - Y_{g,F_g-1} - \sum_{g':D_{g',1}=0,F_{g'}>F_g+l} \frac{N_{g',F_g+l}}{N^u_{F_g+l}}(Y_{g',F_g+l} - Y_{g',F_g-1})$$

- Aggregated into each lead

$$DID_{+,l} = \sum_{g:D_{g,1}=0,F_g \leq T_u-l} \frac{\beta^{F_g+l} N_{g,F_g+l}}{N^1_l} DID_{g,l}$$

4. STATA Simulation

Replication Project

- **Goal**: Replicate 3 graphs from Figure 3 in Chaisemartin & D'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey."
- **Data source**: Initially sourced from Wolfers (2006): "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results."
- **Approach**: We will proceed step-by-step, conducting estimations using TWFE, CS, and dCDH estimators.

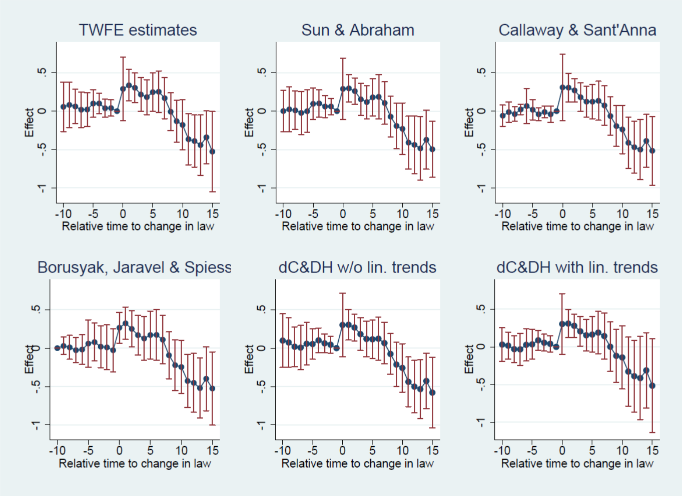
Replication Project

Year	Prior	1969	1970	1971	1972	1973	1974	1975	1976	1977	1985	After	Total
# States treated	2	1	2	6	3	10	2	2	1	1	1	20	51

- Period:1956-1998 (43 years)
- Total states 51: 2 always treated (dropped) and 20 never treated (treated later in 2000). 29 states were treated within the time window (see detail in the table).
- **List to do**:
  - Run do file step by step!
  - Attempt manual calculation of ATT(g,t) described by CS(2021) to gain deeper insight into the underlying processes!

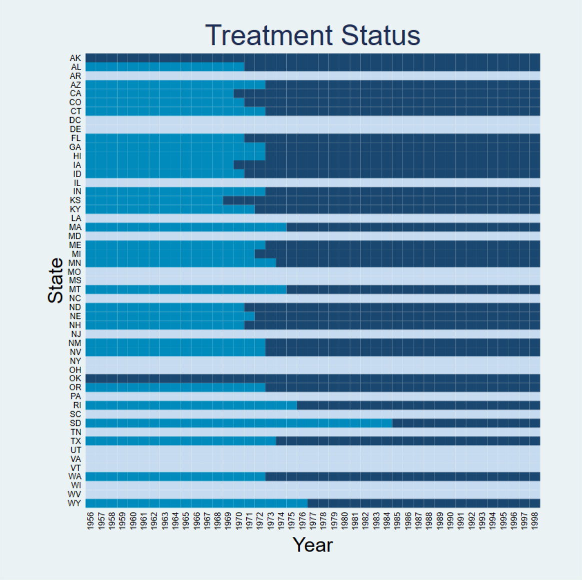
Graphs to be Replicated

Figure 3. Effects of Unilateral Divorce Laws, using the data in Wolfers (2006a)

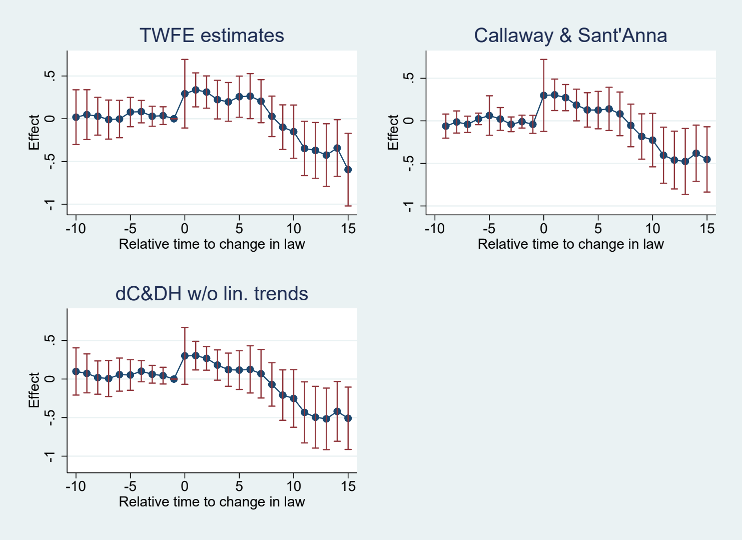


Source: Chaisemartin & D'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey."

Treatment Map



STATA Simulation: Replication Results



Conclusion

- While many scholars are aware of TWFE's limitations, it remains commonly used.
- Mastering every estimator is overwhelming, but our replication reveals their close alignment.
- Many authors presented results new estimator(s) as a robustness check.
- In general, concerns related to Two-Way Fixed Effects (TWFE) tend to diminish when the dataset includes a substantial number of units that have never received treatment.

THANK YOU!

For comments, feedback, or inquiries, please email me at: muchrosidi@gmail.com or much.rosidi@uky.edu.

Visit my website: muchrosidi.com