Advanced Panel Data: Two-way FE and Diff-in-Diff

Chris Conlon

May 3, 2021

NYU Stern

Staggered TWFE Designs

Staggered TWFE Designs

A very common method is the following:

$$y_{it} = x_{it}\beta + \tau \cdot D_{it} + \eta_i + \lambda_t + \varepsilon_{it}$$

- e.g. Different states i enact a policy $D_{it} = 1$ in different years t.
- We call this "staggered" if not every entity is treated at the same time.
- Surely this is better than:
    - Single treated state (usual DiD)
    - Contemporaneous role out (can't separate τ and λ_t).
- But what does τ or τ_{it} actually measure?

Literature

This is part of a rapidly growing literature

- Callaway and Sant'Anna (2020)
- Goodman-Bacon (2021)
- de Chaisemartin and D'Haultfoeulle (2020
- Sun and Abraham (2020)
- R Vignette

Recap: Regular DiD

$$Y_{it} = D_i \cdot Y_{it}(1) - (1 - D_i) \cdot Y_{it}(0), \quad Y_{i,t-1} = Y_{i,t-1}(0)$$

$$y_{it} = \alpha + \gamma_i \cdot D_i + \lambda_t \cdot \text{Post}_t + \tau \cdot D_i \times \text{Post}_t + \varepsilon_{it}$$

- Nobody treated at $t-1$ and some people treated at t.
- Parallel trends $\mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) | D_{it} = 1] = \mathbb{E}[Y_{it}(0) - Y_{i,t-1}(0) | D_{it} = 0]$
- If parallel trends holds: $\tau = ATT = \mathbb{E}[Y_{it} - Y_{i,t-1} | D_{it} = 1] - \mathbb{E}[Y_{it} - Y_{i,t-1} | D_{it} = 0]$

Simple Question?

When does TWFE deliver the $ATT = \mathbb{E}[\tau_{it} | D_{it} = 1]$?

$$y_{it} = x_{it}\beta + \tau_{it} \cdot D_{it} + \eta_i + \lambda_t + \varepsilon_{it}$$

- Only two periods with $D_{i0} = 0$ for all i.
- Constant treatment effects $\tau_{it} = \tau$.
- Otherwise mostly no...

What are we comparing

$$y_{it} = x_{it}\beta + \tau_{it} \cdot D_{it} + \eta_i + \lambda_t + \varepsilon_{it}$$

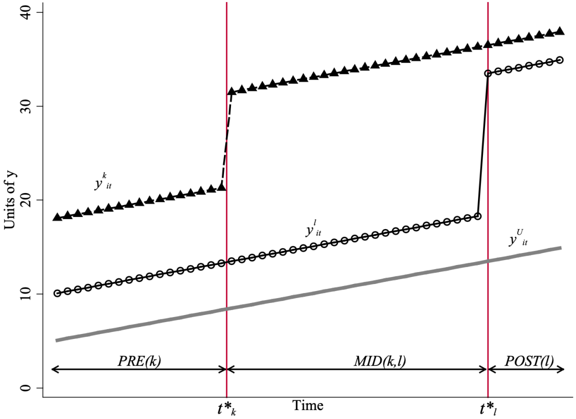
Three different control groups when we compare $Y_{it} | D_{it} = 1$:

1. Units that never get treated
2. Units that will be treated later: $t' > t$
3. But previously treated units enter your control group [bad!]

How τ_{it} varies over time for different cohorts will now affect an estimate of τ.

What do we get? (Goodman-Bacon 2020)

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



Notes: The figure plots outcomes in three groups: a control group, U, which is never treated; an early treatment group, E, which receives a binary treatment at $t_k^* = \frac{34}{100}T$; and a late treatment group, ℓ, which receives the binary treatment at $t_\ell^* = \frac{85}{100}T$. The x-axis notes the three sub-periods: the pre-period for group k, $[1, t_k^* - 1]$, denoted by $PRE(k)$; the middle period when group k is treated and group ℓ is not, $[t_k^*, t_\ell^* - 1]$, denoted by $MID(k, \ell)$; and the post-period for group ℓ, $[t_\ell^*, T]$, denoted by $POST(\ell)$. I set the treatment effect to 10 in group k and 15 in group ℓ.

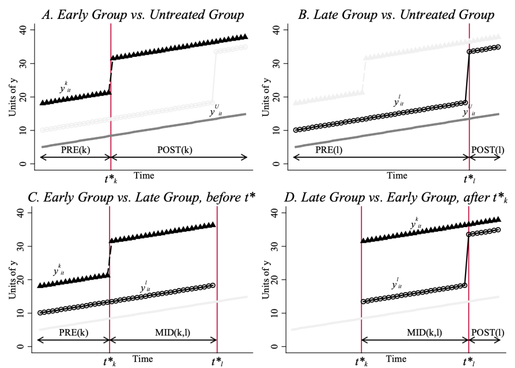
What do we get? (Goodman-Bacon 2020)

One thing we might want to know is whether our TWFE estimate of $\widehat{\tau}$ is at all useful. AKA what did we just estimate?

- Suppose there are two groups G of different treatment timings: {early, late }.
- Think about every possible 2×2 DD estimate.
    - Early group vs. never treated
    - Late group vs. never treated
    - Early group vs. late group before $t*$
    - Late group vs. early group after $t*$.

What do we get? (Goodman-Bacon 2020)

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



Notes: The figure plots the groups and time periods that generate the four simple 2x2 difference-in-difference estimates in the case with an early treatment group, a late treatment group, and an untreated group from Figure 1. Each panel plots the data structure for one 2x2 DD. Panel A compares early treated units to untreated units ($\hat{\beta}^{DD}_{kU}$); panel B compares late treated units to untreated units ($\hat{\beta}^{DD}_{\ell U}$); panel C compares early treated units to late treated units during the late group's pre-period ($\hat{\beta}^{DD,k}_{k\ell}$); panel D compares late treated units to early treated units during the early group's post-period ($\hat{\beta}^{DD,\ell}_{k\ell}$). The treatment times mean that $\overline{D}_k = 0.67$ and $\overline{D}_\ell = 0.16$, so with equal group sizes, the decomposition weights on the 2x2 estimate from each panel are 0.365 for panel A, 0.222 for panel B, 0.278 for panel C, and 0.135 for panel D.

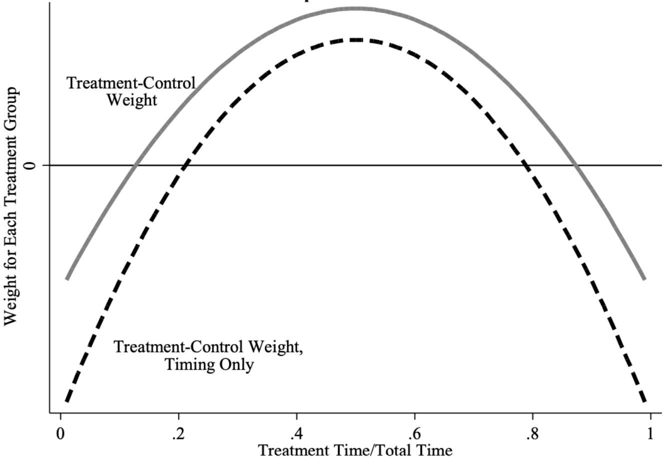
What do we get? (Goodman-Bacon 2020)

$$
\begin{aligned}
\widehat{\tau}^{DD} &= \sum_{k \neq U} s_{kU} \widehat{\tau}_{kU} + \sum_{k \neq U} \sum_{\ell > k} \left[s^k_{k\ell} \widehat{\tau}^k_{k\ell} + s^\ell_{k\ell} \widehat{\tau}^\ell_{k\ell} \right] \\
\widehat{\tau}^{2x2}_{kU} &\equiv \left(\bar{y}_k^{POST(k)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_U^{POST(k)} - \bar{y}_U^{PRE(k)} \right) \\
\widehat{\tau}^{2x2,k}_{k\ell} &\equiv \left(\bar{y}_k^{MID(k,\ell)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_\ell^{MID(k,\ell)} - \bar{y}_\ell^{PRE(k)} \right) \\
\widehat{\tau}^{2x2,\ell}_{k\ell} &\equiv \left(\bar{y}_\ell^{POST(\ell)} - \bar{y}_\ell^{MID(k,\ell)} \right) - \left(\bar{y}_k^{POST(\ell)} - \bar{y}_k^{MID(k,\ell)} \right)
\end{aligned}
$$

Corresponding set of weights $s_{k,\ell}$ which depend on size and variance of each group.
Variance is largely about how concentrated D_{it} is within each group.

Assuming Equal Groups (Goodman-Bacon 2020)

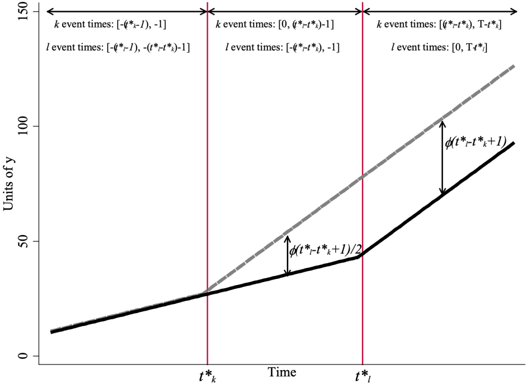
Figure 4. Weighted Common Trends: The Treatment/Control Weights as a Function of the Share of Time Spent Under Treatment



Notes: The figure plots the weights that determine each timing group's importance in the weighted common trends expression in equations (16) and (17).

Time varying TE τ_{it} (Goodman-Bacon 2020)

Figure 3. Difference-in-Differences Estimates with Variation in Timing Are Biased When Treatment Effects Vary Over Time



Notes: The figure plots a stylized example of a timing-only DD set up with a treatment effect that is a trend-break rather than a level shift (see Meer and West 2013). Following section II.A.ii, the trend-break effect equals $\phi \cdot (t - t^* + 1)$. The top of the figure notes which event-times lie in the $PRE(k)$, $MID(k, \ell)$, and $POST(\ell)$ periods for each unit. The figure also notes the average difference between groups in each of these periods. In the $MID(k, \ell)$ period, outcomes differ by $\frac{\phi}{2}(t_\ell^* - t_k^* + 1)$ on average. In the $POST(\ell)$ period, however, outcomes had already been growing in the early group for $t_\ell^* - t_k^*$ periods, and so they differ by $\phi(t_\ell^* - t_k^* + 1)$ on average. The 2x2 DD that compares the later group to the earlier group is biased and, in the linear trend-break case, weakly negative despite a positive and growing treatment effect.

Application: (Goodman-Bacon 2020)

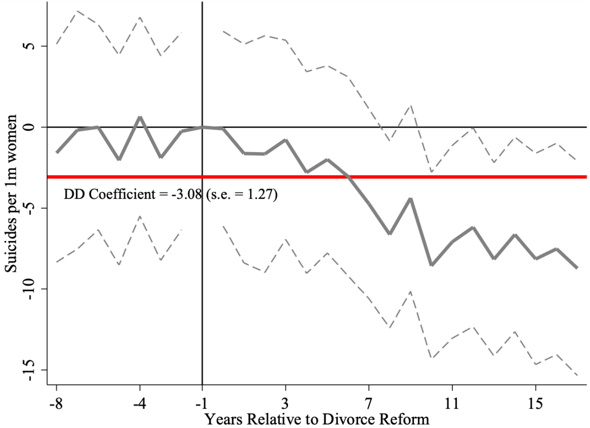
Table 1. The No-Fault Divorce Rollout: Treatment Times, Group Sizes, and Treatment Shares

No-Fault Divorce Year (t_k^*)	Number of States	Share of States (n_k)	Treatment Share ($\bar{D}_k$)
Non-Reform States	5	0.10	.
Pre-1964 Reform States	8	0.16	.
1969	2	0.04	0.85
1970	2	0.04	0.82
1971	7	0.14	0.79
1972	3	0.06	0.76
1973	10	0.20	0.73
1974	3	0.06	0.70
1975	2	0.04	0.67
1976	1	0.02	0.64
1977	3	0.06	0.61
1980	1	0.02	0.52
1984	1	0.02	0.39
1985	1	0.02	0.36

Notes: The table lists the dates of no-fault divorce reforms from Stevenson and Wolfers (2006), the number and share of states that adopt in each year, and the share of periods each treatment timing group spends treated in the estimation sample from 1964-1996

Event Study Plot (Goodman-Bacon 2020)

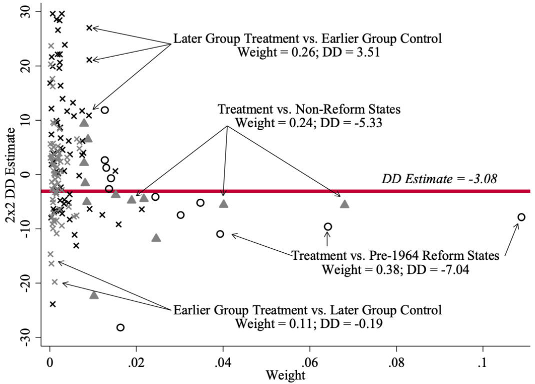
Figure 5. Event-Study and Difference-in-Differences Estimates of the Effect of No-Fault Divorce on Female Suicide: Replication of Stevenson and Wolfers (2006)



Notes: The figure plots event-study estimates from the two-way fixed effects regression equation on page 276 and plotted in figure 1 of Stevenson and Wolfers (2006), along with the DD coefficient. The specification does not include other controls and does not weight by population. Standard errors are robust to heteroskedasticity.

Decomposition Plot (Goodman-Bacon 2020)

Figure 6. Difference-in-Differences Decomposition for Unilateral Divorce and Female Suicide



Notes: The figure plots each 2x2 DD components from the decomposition theorem against their weight for the unilateral divorce analysis. The open circles are terms in which one timing group acts as the treatment group and the pre-1964 reform states act as the control group. The closed triangles are terms in which one timing group acts as the treatment group and the non-reform states act as the control group. The *x*'s are the timing-only terms. The figure notes the average DD estimate and total weight on each type of comparison. The two-way fixed effects estimate, -3.08, equals the average of the *y*-axis values weighted by their *x*-axis value.

Application: Estimates (Goodman-Bacon 2020)

Table 2. DD Estimates of the Effect of Unilateral Divorce Analysis on Female Suicide: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	No Untreated States	WLS	Propensity Score Weighting	Controls	Unit-Specific Trends	Group-Specific Pre-Trends	Region-by-Year Fixed Effects
Unilateral Divorce	-3.08	2.42	-0.35	1.04	-2.52	0.59	-6.52	-1.16
	[1.27]	[1.81]	[1.97]	[1.78]	[1.09]	[1.35]	[2.98]	[1.37]
Difference from baseline specification		5.50	2.73	4.12	0.56	3.67	-3.44	1.92
Share due to:								
2x2 DDs		0	0.52	1	0.22	0.90	1	0.37
Weights		1	0.39	0	0.05	0.47	0	0.76
Interaction		0	0.09	0	<0.01	-0.36	0	-0.13
Within Term		0	0	0	0.73	0	0	0

Notes: The table presents DD estimates from the alternative specifications discussed in section III. Column (1) is the two-way fixed effects estimate from equation (2). Column (2) drops the pre-1964 reform and non-reform states. Column (3) weights by state adult populations in 1964. Column (4) weights by the inverse propensity score estimated from a probit model that contains the sex ratio, per-capita income, the general fertility rate, and the infant mortality rate all measured in 1960. Column (5) controls for per-capita income, female homicide rates, and per-capita welfare caseloads. Column (6) includes state-specific linear time trends. Column (7) comes from a two-step procedure that first estimates group-specific trends from 1964-1968, subtracts them from the suicide rate, and estimates equation (2) on the transformed outcome variable. Column (8) includes region-by-year fixed effects. Below the standard errors I show the difference between each estimate and the baseline result, an the last three rows show the share of this difference that comes from changes in the 2x2 DD's, the weights, or their interaction as shown in equation (18).

Recap

- The Goodman-Bacon (2020) paper tells us what we are measuring with TWFE.
- Weird things can happen like negative weights.
- But it doesn't really tell us what we should do.
- Other than be careful and plot the event study plot always.

Callaway and Sant'Anna (2020)

Some Theory

Definitions

- G_i when is i treated (group/cohort).
- C_i set of never treated individuals
- $Y_{it} = [G_i > t]Y_{it}(0) + [G_i \leq t]Y_{it}(G_i)$
- Idea: multiple PO for $Y_{it}(z)$

Assumptions

- Irreversibility: $D_{it} \geq D_{i,t-1}$.
- Modified Parallel trends for $t \geq g$ and $s \geq t$:

$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid G = g\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid C = 1\right]$$
$$\mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid G = g\right] = \mathbb{E}\left[Y_t(0) - Y_{t-1}(0) \mid D_s = 0, G \neq g\right]$$

Can also do parallel trends conditional on covariates.

Grouped TE

All observations treated at time g are grouped into a cohort

$$ATT(t, g) = \mathbb{E}\left[Y_t(g) - Y_t(0) \mid G = g\right]$$

We can estimate a time-varying and cohort-specific ATT using never-treated or not-yet-treated.

$$\mathrm{ATT}(g, t) = E\left[Y_t - Y_{g-1} \mid G = g\right] - E\left[Y_t - Y_{g-1} \mid C = 1\right]$$
$$\mathrm{ATT}(g, t) = E\left[Y_t - Y_{g-1} \mid G = g\right] - E\left[Y_t - Y_{g-1} \mid D_t = 0, G \neq g\right]$$

In may be better to simply estimate and report these objects.
(Assuming modified parallel trends hold).
Unless you have lots of groups – I would stop here.

Combining Grouped TE

We could average the $\mathrm{ATT}(g,t)$ and report those (if we have too many groups).

$$\theta_S(g) = \frac{1}{\mathcal{T} - g + 1} \sum_{t=2}^{\mathcal{T}} 1\{g \leq t\} ATT(g,t)$$

$$\theta_S^O := \sum_{g=2}^{\mathcal{T}} \theta_S(g) P(G = g)$$

This is the time-average for each group and the overall average resepectively.

Combining Grouped TE

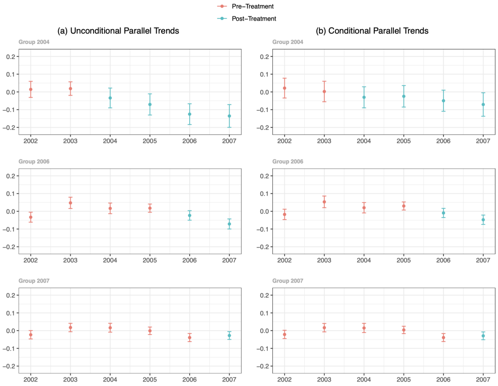
We could also do the group-averaged event study

$$\theta_D(e) := \sum_{g=2}^{\mathcal{T}} 1\{g + e \leq \mathcal{T}\} ATT(g, g+e) P(G = g \mid G + e \leq \mathcal{T})$$

This combines cohorts and plots the average k periods before and after treatment.

Application: Estimates

Figure 1: Minimum Wage Group-Time Average Treatment Effects



Notes: The effect of the minimum wage on teen employment estimated under the unconditional parallel trends assumption (Panel (a)) and the conditional parallel trends assumption (Panel (b)). Red lines give point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the county level. Under the null hypothesis of the parallel trends assumption holding in all periods, these should be equal to 0. Blue lines provide point estimates and uniform 95% confidence bands for the treatment effect of increasing the minimum wage allowing for clustering at the county level. The top row includes states that increased their minimum wage in 2004, the middle row includes states that increased their minimum wage in 2006, and the bottom row includes states that increased their minimum wage in 2007. The estimates in Panel (b) use the the doubly robust estimator discussed in the text.

Application: Estimates

Table 3: Minimum Wage Aggregated Treatment Effect Estimates

(a) Unconditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.037
					(0.006)
Simple Weighted Average					-0.052
					(0.006)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.091	-0.047	-0.028		-0.039
	(0.019)	(0.008)	(0.007)		(0.007)
Event Study	e=0	e=1	e=2	e=3	
	-0.027	-0.071	-0.125	-0.136	-0.090
	(0.006)	(0.009)	(0.021)	(0.023)	(0.013)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.034	-0.071	-0.055	-0.050	-0.052
	(0.019)	(0.02)	(0.009)	(0.006)	(0.013)
Event Study w/ Balanced Groups	e=0	e=1			
	-0.027	-0.071			-0.049
	(0.009)	(0.009)			(0.008)

Application: Estimates

(b) Conditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.008
					(0.006)
Simple Weighted Average					-0.033
					(0.007)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.044	-0.029	-0.029		-0.031
	(0.020)	(0.008)	(0.008)		(0.007)
Event Study	e=0	e=1	e=2	e=3	
	-0.024	-0.041	-0.050	-0.071	-0.046
	(0.006)	(0.009)	(0.022)	(0.026)	(0.013)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.030	-0.025	-0.030	-0.049	-0.033
	(0.022)	(0.021)	(0.009)	(0.007)	(0.012)
Event Study w/ Balanced Groups	e=0	e=1			
	-0.016	-0.041			-0.028
	(0.010)	(0.009)			(0.008)

Application: Estimates

Notes: The table reports aggregated treatment effect parameters under the unconditional and conditional parallel trends assumptions and with clustering at the county level. The row 'TWFE' reports the coefficient on a post-treatment dummy variable from a two-way fixed effects regression. The row 'Simple Weighted Average' reports the weighted average (by group size) of all available group-time average treatment effects as in Equation (3.10). The row 'Group-Specific Effects' summarizes average treatment effects by the timing of the minimum wage increase; here, g indexes the year that a county is first treated. The row 'Event Study' reports average treatment effects by the length of exposure to the minimum wage increase; here, e indexes the length of exposure to the treatment. The row 'Calendar Time Effects' reports average treatment effects by year; here, t indexes the year. The row 'Event Study w/ Balanced Groups' reports average treatment effects by length of exposure using a fixed set of groups at all lengths of exposure; here, e indexes the length of exposure and the sample consists of counties that have at least two years of exposure to minimum wage increases. The column 'Single Parameters' represents a further aggregation of each type of parameter, as discussed in the text. The estimates in Panel (b) use the the doubly robust estimator discussed in the text.

Explore `did` in `R`

https://bcallaway11.github.io/did

Thanks!