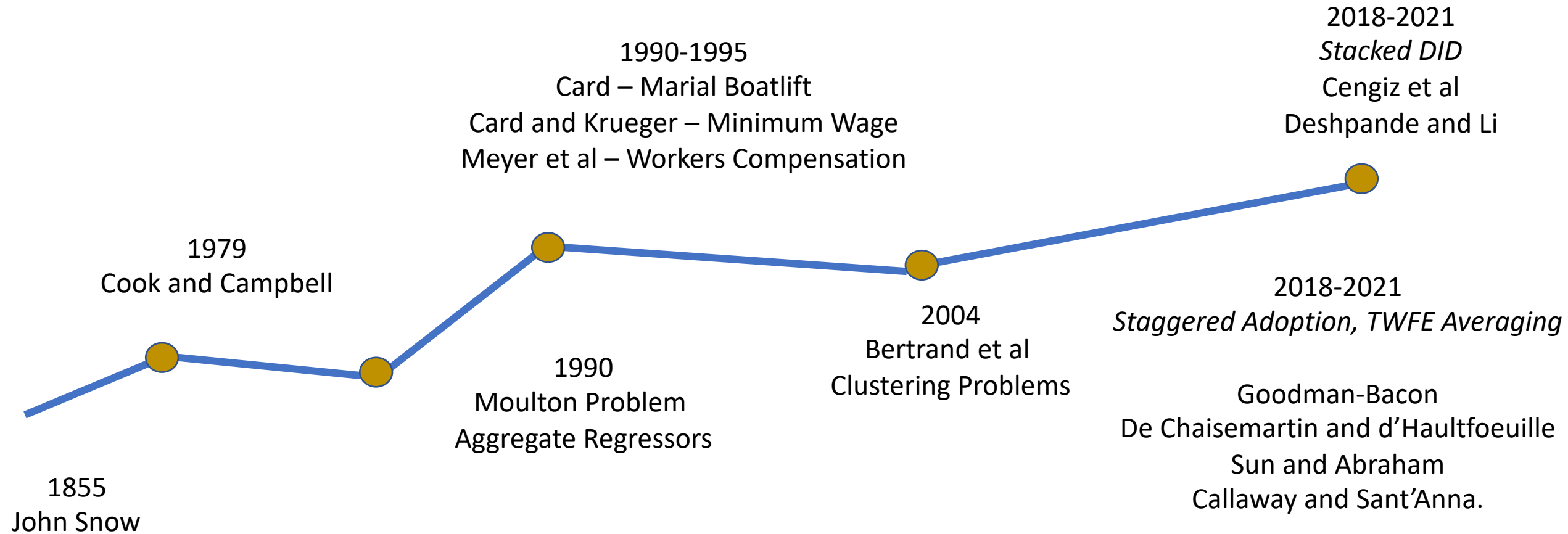# Staggered Adoption Designs
+
Stacked DID and Event Studies

Coady Wing

Workshop in Methods

22 October 2021

One history of difference in differences...

2018-2021
Stacked DID
Cengiz et al
Deshpande and Li

1990-1995
Card – Marial Boatlift
Card and Krueger – Minimum Wage
Meyer et al – Workers Compensation

1979
Cook and Campbell

2004
Bertrand et al
Clustering Problems

2018-2021
Staggered Adoption, TWFE Averaging

1990
Moulton Problem
Aggregate Regressors

Goodman-Bacon
De Chaisemartin and d'Haultfoeuille
Sun and Abraham
Callaway and Sant'Anna.

1855
John Snow

The Plan for Today

Fast review of 2x2 DID

Staggered Adoption DID
   Toy Example
   Analysis Using TWFE and New Problems

Stacked DID

Inference in Stacked DID

Discussion

The Simple 2x2 DID

Simple 2 x 2 DID → Data and Estimation

Table Form

	Pre	Post	Differences
Treat	$Y_{T,Pre}$	$Y_{T,Post}$	$\Delta_T = Y_{T,Post} - Y_{T,Pre}$
Control	$Y_{C,Pre}$	$Y_{C,Post}$	$\Delta_C = Y_{C,Post} - Y_{C,Pre}$
			$\boldsymbol{\beta^{DD} = \Delta_T - \Delta_C}$

Outcomes measured for two groups in two periods.

Treatment status changes between periods for the "treatment group"

No change in treatment status between periods for the "control group".

Regression Form

$$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 (Treat_s \times Post_t) + \epsilon_{st}$$

In the simple 2x2 setting, $\boldsymbol{\beta^{DD} = \beta_3}$.

# *The 2x2 DID nests two simpler research designs*

Pre-Post Design: could be biased by "time trends"

   Unmeasured time varying factors that change along with treatment

   History effects, Maturation effect

Non-equivalent Control Group Design: could be biased because group differences in treatment are confounded by other differences across groups.

Key Assumption: Common Trends
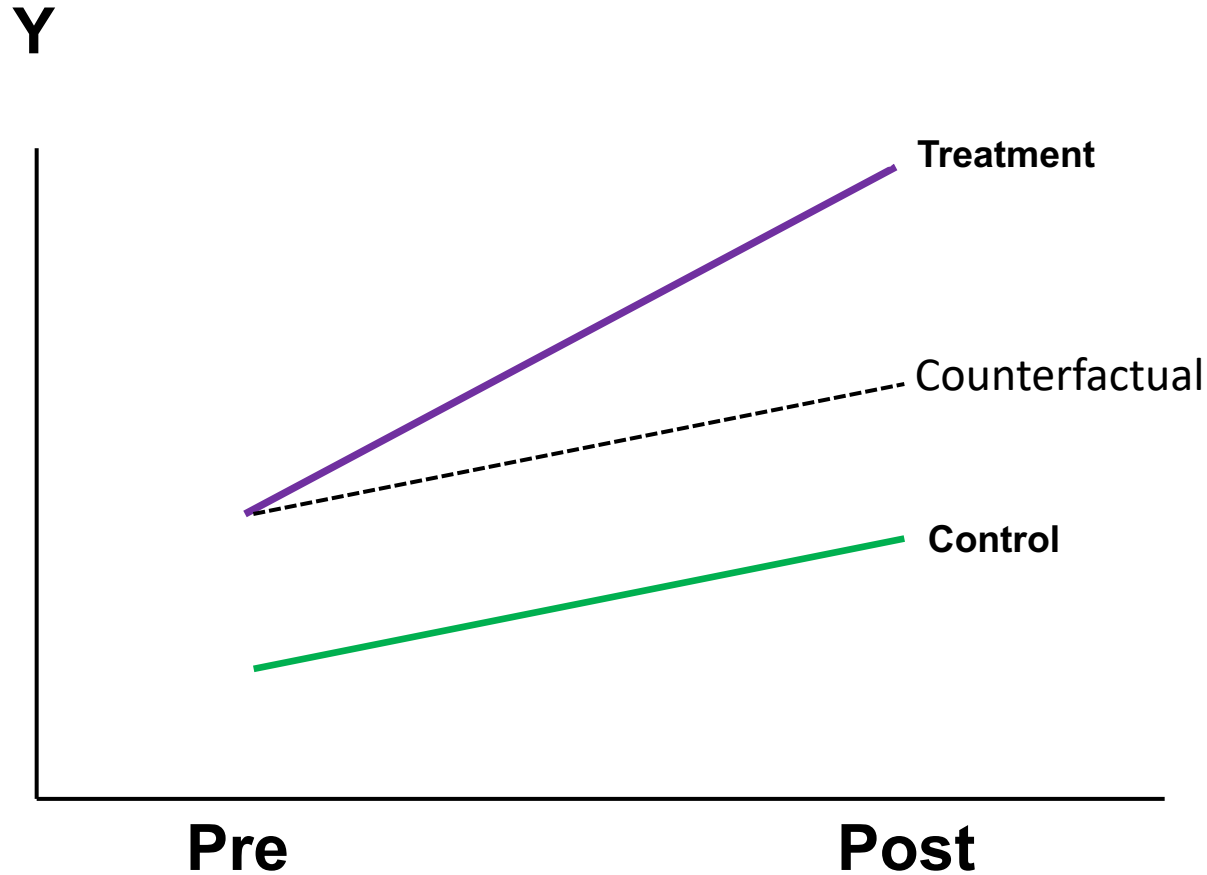
Common trends assumption encapsulates two restrictions:

Time-varying confounders must affect outcomes in both groups in the same way.
Time fixed effects

Group-varying confounders must be time-invariant.
Group fixed effects.

Extensions to Multiple Groups and Multiple Periods

What about cases with more than two groups and more than two periods?

Old and Busted

"Generalized Difference in Differences"

Analyzed using a two-way fixed effect regression model.

Account for clustering using {some method}.

New Hotness

"Staggered Adoption Design"

Analyzed using ...{new method}.

Account for clustering using...{some method}.

Twoway Fixed Effects: The current workhorse

Suppose you have $s = 1 \dots S \ states \ and \ t = 1 \dots T \ periods.$
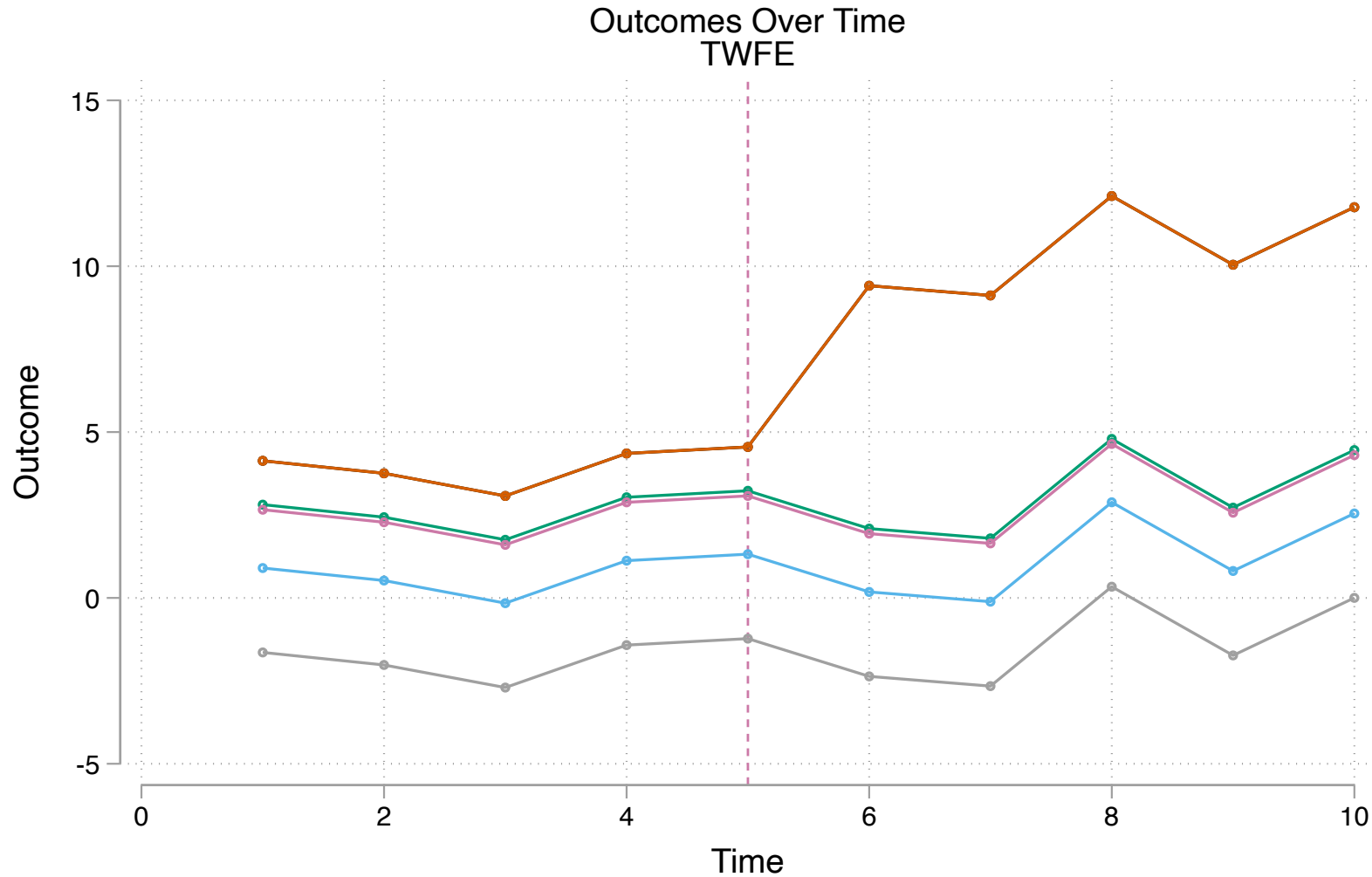
Treatment "turns on" at different times in different states.

$$Y_{st} = \beta^{FE} D_{st} + a_s + b_t + \epsilon_{st}$$

D_{st} → Dummy variable set to 1 if the policy is in force in state s during period t.

a_s → state fixed effect (time invariant factor)

b_t → time fixed effect (time varying common factor)

How is the TWFE model like a DID?

Time fixed effects: trends are flexible, but exactly the same across groups.

Group fixed effects: groups are different even before treatment. But group differences never change.

But what about messier cases?

Multiple states adopt the treatment.

But they adopt at different times. No clear "pre" and "post".

What if treatment effects are heterogeneous?

Across Units?

Over Time (phase in effects, interactions with calendar time)

Staggered Adoption Design

Panel Data Setting: collection of $s = 1 \dots S$ units each observed at $t = T_1 \dots T_T$ calendar time periods.

Y_{st} is a realized outcome for unit s in period t.

A_s is the policy adoption year in state s.

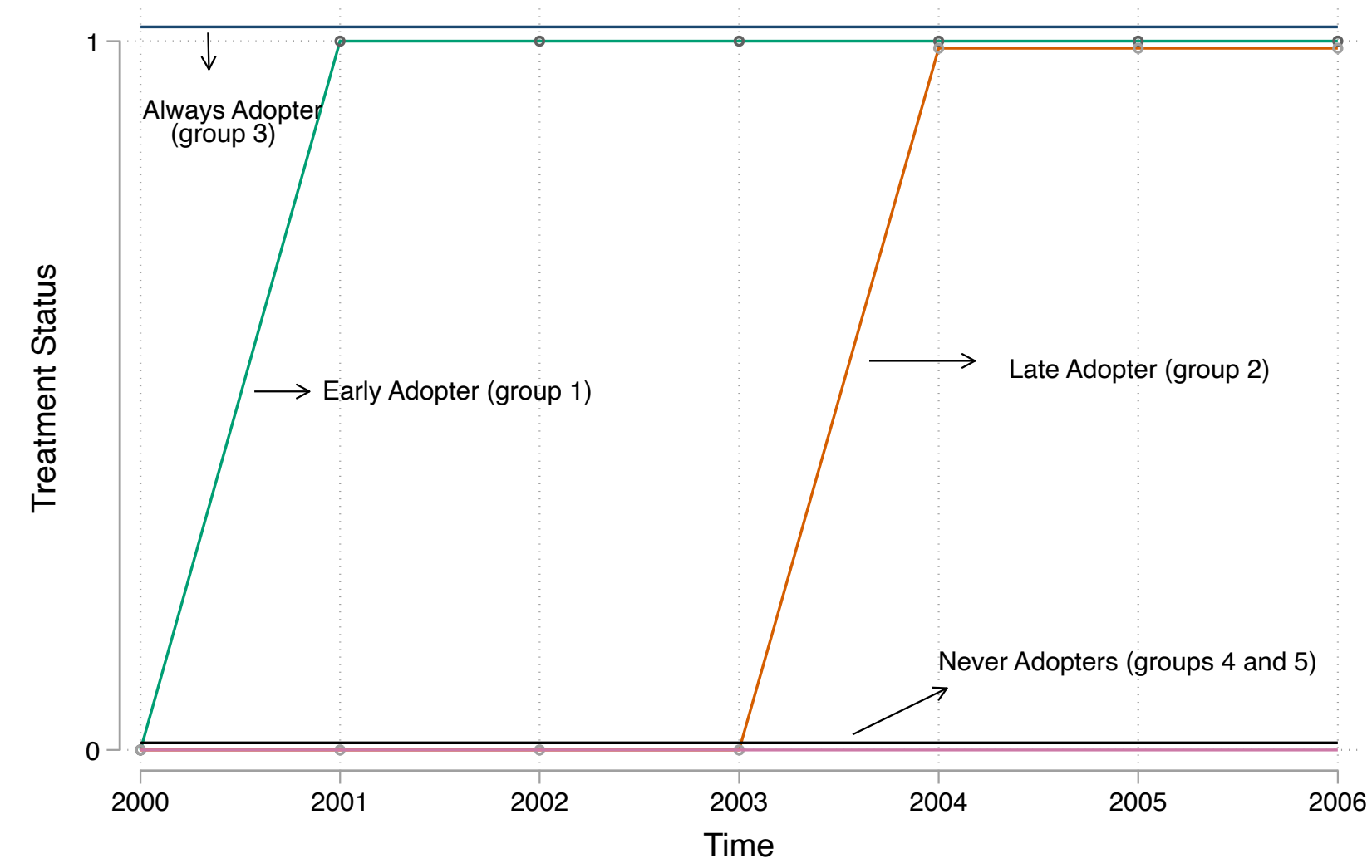 If Indiana adopts a policy in 2005 then $A_{IN} = 2005$
  If New York has never adopted the policy, then think of $A_{NY} = \infty$

$D_{st} = 1(t \geq A_s)$ is an indicator that the policy is "active" in state s in period t.

  Policy adoption is an absorbing state. (State's do not abandon the policy once they put it in place.)

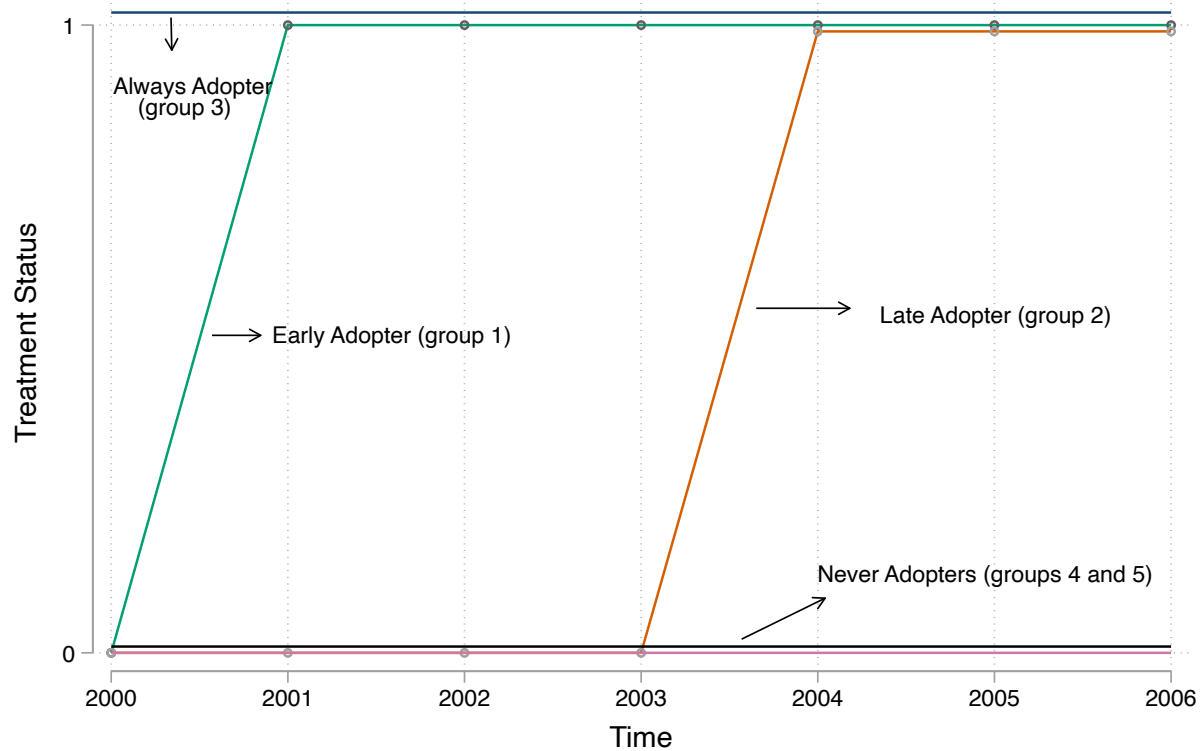*Staggered Adoption Design Contains Many 2x2 DIDs*

A simple staggered adoption design

Group ID (State)	Adoption Year	Calendar Years of Data
1	2001	
2	2004	
3	1999	2000-2006
4	NA	
5	NA	

How many simple 2x2 DIDs are in this staggered adoption design?

Comparison	2x2 DID
E vs N	$\beta_{1,5} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{5,2001} - Y_{5,2000}\right)$
L vs N	$\beta_{2,5} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{5,2004} - Y_{5,2003}\right)$
E vs L	$\beta_{1,2} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{2,2001} - Y_{2,2000}\right)$
L vs E	$\beta_{2,1} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{1,2004} - Y_{5,2003}\right)$
E vs A	$\beta_{1,3} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{3,2001} - Y_{3,2000}\right)$
L vs A	$\beta_{2,3} = \left(Y_{1,2004} - Y_{1,2003}\right) - \left(Y_{3,2004} - Y_{3,2003}\right)$

Counting the pairwise DIDs

1. Find all of the "switchers".

2. For each switcher: compute a simple pre-post change using the last pre-period and the first post-period.

3. Now swap in every possible control case where treatment status doesn't change over those two periods.

Note: if we are slightly fancier, we could take averages over multiple periods instead of simple two-period comparisons.

What happens to all of these underlying 2x2 DIDs? How do we turn them into a single effect estimate?

Goodman-Bacon (2021) asks how the TWFE regression coefficient is connected to underlying 2x2 DID comparisons in a staggered adoption designs.

$$Y_{st} = \beta^{FE} D_{st} + a_s + b_t + \epsilon_{st}$$

Turns out that β^{FE} is a (somewhat complicated) variance weighted average of all pairwise 2x2 DIDs.

A pairwise DID gets more weight if the policy change occurs close to the middle of the study window.

A pairwise DID gets more weight if it includes more observations.

CITATION: Goodman-Bacon, Andrew. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* (2021).

Strange weighted averaging is not such a big issue…

But time-varying treatment effects can create big problems.

Why might treatment effects vary over time?

Phase in effects: it might take some time for an intervention to produce measurable changes in behavior.

> Knowledge of a new tax might spread gradually. People might not change consumption or production behavior until they figure it out.

> Marijuana might be legalized in one year, but it will take a few years for dispensaries to open.

Calendar Period x Treatment Effect Interactions: the same policy may have a different effect on behavior in some calendar periods than others.

> Expanded Unemployment Insurance benefits might have negligible effects on labor supply when the economy is depressed. But they could have larger effects on labor supply when the economy is booming.

Return to the staaaered adoption design...
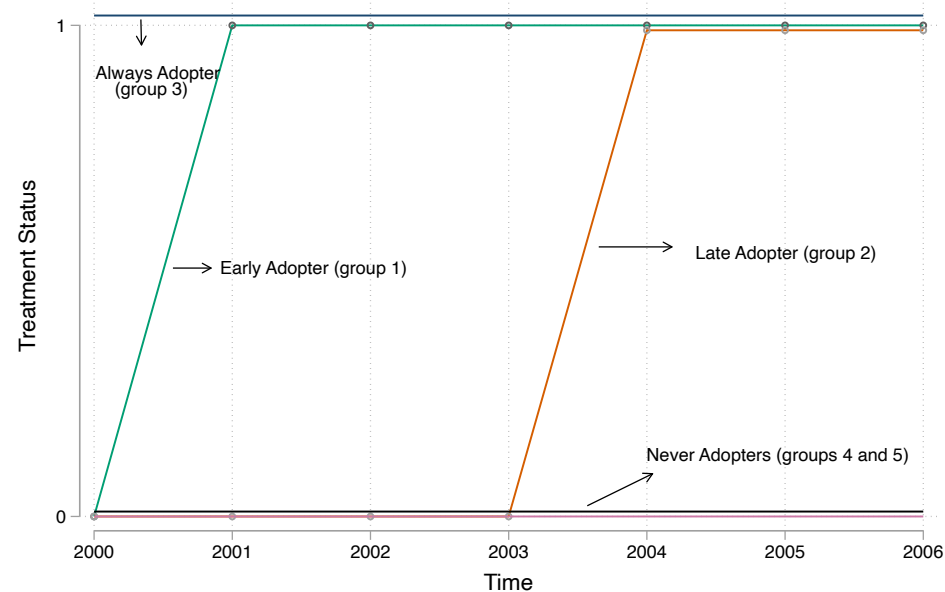
Group ID (State)	Adoption Year	Calendar Years of Data
1	2001	
2	2004	
3	1999	2000-2006
4	NA	
5	NA	

Population Model

$$Y_{st} = \beta_{st}D_{st} + a_s + b_t + \epsilon_{st}$$

Assume:

$$a_s = Group\ ID$$
$$b_t = Calendar\ Year$$
$$\epsilon_{st} = 0$$

Compare Two Versions of the treatment effect

Constant Treatment Effect: $\beta_{st} = \beta = 4$ for all units and periods.

Phase In Effect: $\beta_{st} = \left(1(t \geq A_s) \times \frac{(t - A_s)}{4}\right) \times 4$, where A_s is the adoption data in state s.

Try out the Constant Effects Model

Constant Treatment Effects Model: $a_s = s, b_t = t, \epsilon_{st} = 0, \beta_{st} = 4$

Group ID	2000	2001	2002	2003	2004	2005	2006
1	1 + 2000 = 2001	1+2001 + 4 = 2006	2007	2008	2009	2010	2011
2	2 + 2000 = 2002	2003	2004	2005	2010	2011	2012
3	3 + 2000 + 4 = 2007	2008	2009	2010	2011	2012	2013
4	4 + 2000 = 2004	2005	2006	2007	2008	2009	2010
5	5 + 2000 = 2005	2006	2007	2008	2009	2010	2011

Comparison	2x2 DID	
E vs N	$\beta_{1,5} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{5,2001} - Y_{5,2000}\right)$	(2006 - 2001)-(2006-2005) = 4
L vs N	$\beta_{2,5} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{5,2004} - Y_{5,2003}\right)$	(2010-2005)-(2009-2008) = 4
E vs L	$\beta_{1,2} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{2,2001} - Y_{2,2000}\right)$	(2006-2001)-(2003-2002) = 4
L vs E	$\beta_{2,1} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{1,2004} - Y_{5,2003}\right)$	(2010-2005)-(2009-2008) = 4
E vs A	$\beta_{1,3} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{3,2001} - Y_{3,2000}\right)$	(2006-2001)-(2008-2007) = 4
L vs A	$\beta_{2,3} = \left(Y_{1,2004} - Y_{1,2003}\right) - \left(Y_{3,2004} - Y_{3,2003}\right)$	(2010-2005)-(2011-2010) = 4

Try out the Phase in Model

Constant Treatment Effects Model: $a_s = s, b_t = t, \epsilon_{st} = 0, \beta_{st} = \left(1(t \geq A_s) \times \frac{(t-A_s)}{4}\right) \times 4$

Group ID	2000	2001	2002	2003	2004	2005	2006
1	1 + 2000 = 2001	1+2001 + 0x4 = 2002	1+2002 + (¼)4 = 2004	2006	2008	2010	2012
2	2002	2003	2004	2005	2006	2008	2010
3	2003 + ¼ x 4 = 2004	2004+2/4 x 4 = 2006	2005 + 3 = 2008	2010	2012	2014	2016
4	2004	2005	2006	2007	2008	2009	2010
5	2005	2006	2007	2008	2009	2010	2011

Comparison	2x2 DID	
E vs N	$\beta_{1,5} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{5,2001} - Y_{5,2000}\right)$	(2002-2001)-(2006-2005) = 0
L vs N	$\beta_{2,5} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{5,2004} - Y_{5,2003}\right)$	(2006-2005)-(2009-2003) = 0
E vs L	$\beta_{1,2} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{2,2001} - Y_{2,2000}\right)$	(2002-2001)-(2003-2002) = 0
L vs E	$\beta_{2,1} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{1,2004} - Y_{5,2003}\right)$	(2006-2005)-(2008-2006) = -1
E vs A	$\beta_{1,3} = \left(Y_{1,2001} - Y_{1,2000}\right) - \left(Y_{3,2001} - Y_{3,2000}\right)$	(2002-2001)-(2006-2004) = -1
L vs A	$\beta_{2,3} = \left(Y_{2,2004} - Y_{2,2003}\right) - \left(Y_{3,2004} - Y_{3,2003}\right)$	(2006-2005)-(2012-2010) = -1

Time varying treatment effects cause bias when you use a TWFE regression to analyze data from a staggered adoption design

Some of the 2x2 DIDs are using "already treated" groups as control group.

This would be fine if the treatment effect is constant over time.

But if treatment effects change over time, then the already treated groups are not following the same trend as the treated group.

In essence: some of the 2x2 DIDs will violate the common trends assumption because of the phase in effect.

What should we do about this issue?

Thing 1: Notice that you have a staggered adoption design, and recognize that regression models will be pooling/averaging effects from multiple sub-experiments.

Thing 2: Consider whether it is plausible to assume that effects are constant across groups and over time. Be particularly concerned about the possibility that effects change over time.

Thing 3: Find some analytic strategy that avoids "problematic comparisons" in which already treated groups serve as controls, or at least measure how important these comparisons are in the overall average.

Stacked DID
and
Event Study

(One of several possible approaches)

What is a stacked DID or stacked event study?

A stacked DID or stacked event study is a way to analyze data from a staggered adoption design.

The idea originally appears in two applied papers:

- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics* 134, no. 3 (2019): 1405-1454.

- Deshpande, Manasi, and Yue Li. "Who is screened out? Application costs and the targeting of disability programs." *American Economic Journal: Economic Policy* 11, no. 4 (2019): 213-48.

The scheme is not perfect. It doesn't provide the perfect way to "weight and sum" event specific treatment effects.

But it ensures that you don't have any problematic comparisons and so is robust to the biases those comparisons create

How do you stack your DID or Event Study?

Checklist: Start with a staggered adoption design and some data.

1. Define an Event Window

2. Enumerate Sub-Experiments

3. Define Inclusion Criteria

4. Stack The Data

5. Specify an Estimating Equation

Choose an Event Window for your study

Notation

κ_a is the length of the pre-event window. (The number of years before the policy change that you will require for the analysis.)

κ_b is the length of the post-event window. (The number of years after the policy change that you will require for the analysis.)

Example: Symmetric 4 year window: $\kappa_a = \kappa_b = 4$.

This means that for every "policy event" in your study, you must have at least 4 years of pre-treatment data and 4 years of post treatment data. Common window for the whole analysis.

Choosing κ_a and κ_b is a research design decision. A long window might seem nice, but it could mean that there are many policy changes that you can't study.

Take an inventory of the sub-experiments

T_1 be the earliest calendar period in your data.

T_T is the latest calendar year in your data.

Define Ω_A to be the collection of all policy adoption years that comply with the event window.

$$\Omega_A = \{A_s | T_1 + \kappa_a \leq A_s \leq T_T - \kappa_b\}$$

These are the policy changes that take place:
    at least κ_a periods after the earliest calendar date
    at least κ_b periods before the latest calendar data

Note: some policy events may be "discarded" because they happen too recently or too early to be studied in the stacked DID framework.

Keep track of the compliant policy events…

Let $d = 1 \ldots D$ index the collection of **sub-experiments** in Ω_A.

ω_d represents the policy adoption date of the $\mathrm{d^{th}}$ $\mathrm{sub-experiment}$.

Example:

ω_1 is the adoption date of the first sub-experiment. This is the earliest policy event that is compliant with the event window.

ω_2 is the adoption date of the second event, etc.

Use inclusion criteria to build a data set for each sub-experiment

<u>Inclusion Condition 1</u> (Valid Treated Units)

    Within sub-experiment d all "treated units" have the same adoption date.

    Implication: a unit may serve as a treated unit in at most 1 sub-experiment.

<u>Inclusion Condition 2</u> (Clean Controls)

    Only units with $A_s > \omega_d + \kappa_b$ are eligible to be included as controls in sub-experiment d.
    Implications:

        Clean controls may consist of never treated units or units that are treated far enough in the future to provide a clean post-period.

        Implication: A unit may serve as a clean control in multiple sub-experiment.

<u>Inclusion Condition 3</u> (Valid Time Periods)

    All observations in sub-experiment d must come from time periods that fall inside the sub-experiment's event window.

    That is sub-experiment d only includes observations that satisfy $\omega_d - \kappa_a \leq t \leq \omega_d + \kappa_b$

Inclusion Criteria in Practice: pseudo code

for d in Ω_a {

    Open the full data set

    $T_{s(d)} = 1(A_s = \omega_d)$ → Make a dummy set to 1 if s is a treated unit for sub-experiment d

    $C_{s(d)} = 1(A_s > \omega_d + \kappa_b)$ → Make a dummy set to if s is a clean control for sub-experiment d

    $M_{t(d)} = 1(\omega_d - \kappa_a \leq t \leq \omega_d + \kappa_b)$ → Make a dummy set to 1 if the calendar date t belongs in sub-experiment d.

    $I_{st(d)} = M_{td}(T_{sd} + C_{sd})$ → Make a dummy for inclusion; set to 1 if observation (s,t) belongs in sub-experiment d.

    keep if $I_{st(d)} == 1$

    save sub_experiment_d

}

Make the stacked data set.

for d in Ω_a {

        append using sub_experiment_d

}

Specify an estimating equation

Simplest Form is a regular old DID fitted to the stacked data.

$T_{sd} \rightarrow$ indicator that unit s is a treated unit in sub-experiment d.

$P_{td} \rightarrow$ indicator that period t is in the post period in sub-experiment d.

$$Y_{std} = \beta_0 + \beta_1 T_{sd} + \beta_2 P_{td} + \beta_3 (T_{sd} \times P_{td}) + \epsilon_{std}$$

Equivalent:

$$Y_{std} = \beta_3 (T_{sd} \times P_{td}) + \theta_{sd} + \gamma_{td} + \epsilon_{std}$$

The stacked DID specification: pros and cons

The stacked DID specification will average all of the time-varying effects into a single averaged effect.

It will be free from those strange new biases.

But it still won't actually "reveal" the time varying effects.

Building Up To The Stacked Event Study Form

Define $YSE_{td} = t - \omega_d$ to be the "time since event" variable in sub-experiment d.

By construction, $YSE_{td} = -\kappa_a, \dots, 1, \dots 0, 1, \dots, \kappa_b$ in every sub-experiment.

Sub-Experiment Specific Event Study Form

Imagine we studied each sub-experiment separately using an event study approach. We'd fit models like:

$$Y_{st}^d = \sum_{j=-\kappa_a}^{\kappa_b} \beta_j^d \times 1(TSE_{td} = j) + \sum_{m=-\kappa_a}^{\kappa_b} \delta_j^d \left(T_{sd} \times 1(TSE_{td} = j) \right) + \theta_s^d + \epsilon_{st}^d$$

Things to notice:

(1) Different set of event study coefficients in each sub-experiment. (That's what we may want to average together in the stacked form.)

(2) Different set of state fixed effects in each sub-experiment. (Sometimes the same s will be in more than one of these regressions and that state will have a different fixed effect in each.)

(3) The event study main effects are mathematically equivalent to calendar time fixed effects for the clean control units in the sub-experiment. These two are allowed to differ across sub-experiments.
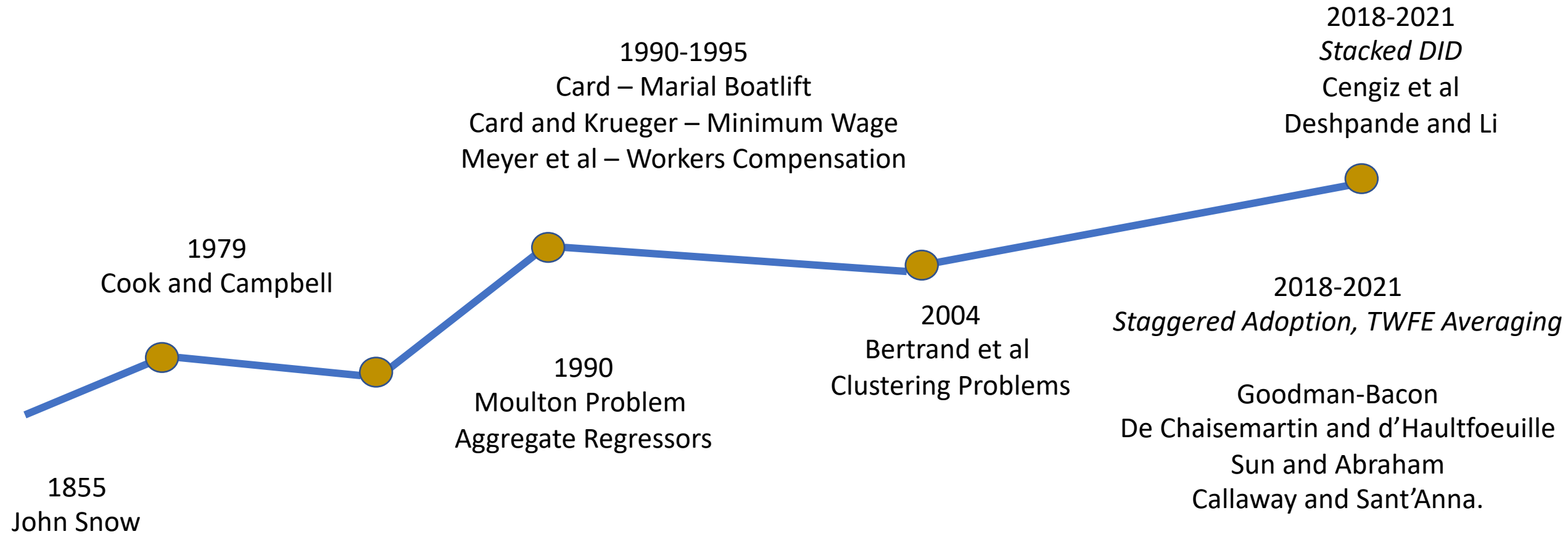
Stacked Event Study Form

$$Y_{std} = \sum_{j=-\kappa_a}^{\kappa_b} \beta_j \times 1(TSE_{td} = j) + \sum_{m=-\kappa_a}^{\kappa_b} \delta_j\big(T_{sd} \times 1(TSE_{td} = j)\big) + \theta_{sd} + \epsilon_{st}^d$$

Notes

(1) Fit this model to the stacked data.

(2) Cross-equation restrictions: there is one set of event study coefficients for each sub-experiment. This is the averaging\pooling idea.

(3) Full set of state x sub-experiment fixed effects.

(4) We could augment the model to replace the event study main effects with a full set of time x sub-experiment fixed effects. (Not clear if this is helpful or not.)

Clustering in the Stacked DID

One history of difference in differences…

1990-1995
Card – Marial Boatlift
Card and Krueger – Minimum Wage
Meyer et al – Workers Compensation

2018-2021
Stacked DID
Cengiz et al
Deshpande and Li

1979
Cook and Campbell

2004
Bertrand et al
Clustering Problems

2018-2021
Staggered Adoption, TWFE Averaging

1990
Moulton Problem
Aggregate Regressors

Goodman-Bacon
De Chaisemartin and d'Haultfoeuille
Sun and Abraham
Callaway and Sant'Anna.

1855
John Snow

The stacked design will often include multiple copies of the same observation.

Never treated units will be used as clean controls in every sub-experiment and the sub-experiments will often "overlap" in calendar time.

Late treated units may also be included multiple times if they are used as controls for some early events.

Duplicate observations mean that the observations in the stacked data set can't really be considered "independent"

How should we account for that?

THE EFFECT OF MINIMUM WAGES ON LOW-WAGE JOBS*

DORUK CENGIZ
ARINDRAJIT DUBE
ATTILA LINDNER
BEN ZIPPERER

We estimate the effect of minimum wages on low-wage jobs using 138 prominent state-level minimum wage changes between 1979 and 2016 in the United States using a difference-in-differences approach. We first estimate the effect of the minimum wage increase on employment changes by wage bins throughout the hourly wage distribution. We then focus on the bottom part of the wage distribution and compare the number of excess jobs paying at or slightly above the new minimum wage to the missing jobs paying below it to infer the employment effect. We find that the overall number of low-wage jobs remained essentially unchanged over the five years following the increase. At the same time, the direct effect of the minimum wage on average earnings was amplified by modest wage spillovers at the bottom of the wage distribution. Our estimates by detailed demographic groups show that the lack of job loss is not explained by labor-labor substitution at the bottom of the wage distribution. We also find no evidence of disemployment when we consider higher levels of minimum wages. However, we do find some evidence of reduced employment in tradeable sectors. We also show how decomposing the overall employment effect by wage bins allows a transparent way of assessing the plausibility of estimates. *JEL* Codes: J23, J38, J88.

Clustered at the unit x sub-experiment level in stacked event studies.

Who Is Screened Out? Application Costs and the Targeting of Disability Programs[†]

By MANASI DESHPANDE AND YUE LI[*]

We study the effect of application costs on the targeting of disability programs. We identify these effects using the closings of Social Security Administration field offices, which provide assistance with filing disability applications. Closings lead to a persistent 16 percent decline in the number of disability recipients in surrounding areas, with the largest effects for applicants with moderately severe conditions and low education levels. Disability applications fall by only 10 percent, implying that the closings reduce targeting efficiency based on current eligibility standards. Increased congestion at neighboring offices appears more important as a channel than higher travel or information costs. (JEL H55, I13, I18, J14)

Clustered at the unit level in stacked event studies.

Statistical Inference in Stacked Event Studies

Coady Wing, Seth Freedman, Laura Montenovo, Patrick Carlin

Data Generating Process

We generate data from a standard two way fixed effects model with clustering induced by state-specific AR(1) serial correlation.

$$Y_{st} = \beta D_{st} + a_s + b_t + \epsilon_{st}$$
$$\epsilon_{st} = \rho_s \times \epsilon_{s,t-1} + u_{st}$$

Today:

$\beta = 10$ → constant treatment effect

$a_s \sim N(0,20)$ unit fixed effect which is correlated with treatment adoption

$b_t \sim N(0,5)$ →time fixed effect.

$u_{st} \sim N(0,20)$→unit x time random effect. (error term)

$\rho_s \sim Bin(k, 100)/100$ (every state gets a ρ_s between 0 and 1, with mean k).

Monte Carlo Experiments

50 States x 19 years

100 States x 19 years

In each simulation:

   40% of states will be treated; 60% never treated.

   Treated states are split evenly across 10 timing groups.

   Evenly distribute the 10 timing groups to 10 treatment adoption dates, chosen to satisfy five pre-periods and five post-periods.

Across simulations, we vary the mean of ρ_s to control the level of clustering/serial correlation.

We use mean $\rho = 0, \ .4, \ and \ .8$

Preliminary Results: Type I error rates across methods

	50 States			100 States		
Serial Correllation→	0	0.4	0.8	0	0.4	0.8
Twoway FE with Robust SEs	0.05	0.16	0.37	0.05	0.18	0.34
Twoway FE with Unit Clustered SEs	0.05	0.05	0.05	0.05	0.04	0.06
SDD with Robust SEs	0.06	0.19	0.35	0.07	0.19	0.34
SDD with Unit Clusters SEs	0.06	0.05	0.06	0.06	0.05	0.06
SDD with UnitXSub Experiment Clustered SEs	0.08	0.07	0.08	0.07	0.06	0.09
SDD with Sub Experiment Clustered SEs	0.05	0.05	0.06	0.05	0.05	0.07

Discussion

As expected, TWFE fares very badly if you don't adjust for clustering but does fine if you cluster standard errors at the state level.

Stacked DID with clustering at the Unit x Design level over rejects the null hypothesis. This makes sense because it does not account for duplication.

Stacked DID with clustering at the Unit Level works pretty well.

That's our recommended approach, for now.

Stacking, Dependence, and Sample Size

Compared to the "regular" TWFE estimator, the stacked regressions will have a larger sample size.

Additional Simulations To Compare:

Stacked estimator with no duplication

vs

Stacked estimator with duplication

Holding total Sample Size Fixed.

Create a no-duplication sample: randomly assign each clean control to a single sub-experiment, this reduces the total number of controls available.

Create a duplication sample using the same number of controls, but allow them to be reapeated.

Experiments to Compare: Dependence vs Stacking, holding N constant

	50 States			100 States		
Serial Correllation	0	0.4	0.8	0	0.4	0.8
Twoway FE	0.05	0.05	0.05	0.05	0.04	0.06
SDD with All Never-Treated	0.06	0.05	0.06	0.06	0.05	0.06
SDD with Repeated Never-Treated	0.09	0.08	0.09	0.07	0.07	0.07
SDD with Partitioned Never-Treated	0.05	0.05	0.05	0.06	0.04	0.06

Results suggest that stacked DID with partitioned Never Treated Units (No duplication) does better than Stacked DID with Repeated Never Treated (Duplication).

Take Home Messages

Staggered Adoption Designs create conditions for bias when treatment effects are time varying, and pose and aggregation/averaging puzzle.

Important Things

Use an estimation strategy based only on "clean controls", such as stacked DID.

Cost: trim treated units that don't have "clean controls" and time periods that fall outside the event window.

Use event study approaches to allow effects to vary over time.

In the stacked estimator, cluster standard errors at the Unit level to account for duplication.

Consider further methods to control aggregation. See work by Callaway and Sant'Anna (2020)

Callaway, Brantly, and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods." *Journal of Econometrics* (2020).