

Lecture 5: Difference in Differences

Lingguo Cheng

2025 Fall

Business School, Nanjing University

I. Panel data

- A data set where we can observe the same units over more than one time
- Pooled cross-section data
- Depending on the level of aggregation, the data demand can be quite different for the fixed effect estimation and the difference-in-differences estimation
- Generally speaking, which one is more data-demanding?
 - Individual fixed effect estimation
 - Canonical two-way fixed effect estimation for the DID approach

APC Model

- The effect of age, period, and cohort
- How can we disentangle the effects of age, period, and cohort?
 - For a cross-sectional dataset, it is impossible to isolate age and cohort effects
 - For a panel dataset, we have to tackle the “APC identification problem”
 - Dohmen, T., A. Falk, B. Golsteyn, D. Huffman and U. Sunde (2017). 'Risk attitudes across the life course', *The Economic Journal*, vol. 127(605), pp. F95-F116.
 - Fitzenberger, B., G. Mena, J. Nimczik and U. Sunde (2021). 'Personality Traits Across the Life Cycle: Disentangling Age, Period and Cohort Effects*', *The Economic Journal*, vol. 132(646), pp. 2141-2172.

II. Fixed Effect Estimation

- Setup of Fixed Effect Model

$$y_{it} = x_{it}\beta + c_i + u_{it} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T$$

Where c_i , unobserved heterogeneity; u_{it} , the idiosyncratic error; $c_i + u_{it}$, composite error term

- Fixed effect vs. random effect

- c_i is correlated with x_{it} , fixed effect
- c_i is uncorrelated with x_{it} , random effect model
 - The most important consequence of random effects is that the residual for a given person are correlated across periods

- Two-way fixed effect model

$$y_{it} = x_{it}\beta + c_i + \lambda_t + u_{it} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T$$

Causal Framework of FE

- Union membership vs. wage
- Do workers whose ages set by collective bargaining earn more because of the union membership, or would they earn more anyway (perhaps because they are more experienced or skilled)
 - Y_{it} , the observed earnings of worker i at time t
 - Y_{0it} or Y_{1it} , depending on the union status
 - D_{it} , the union status

- Suppose that (CIA is satisfied)

$$E(Y_{0it} | A_i, X_{it}, t, D_{it}) = E(Y_{0it} | A_i, X_{it}, t)$$

- Where X_{it} is a vector of observed time-varying covariates and A_i is a vector of unobserved but fixed confounder, “Ability”
- It means union status is as good as randomly assigned conditional on A_i and X_{it}

- Suppose that

$$E(Y_{0it} | A_i, X_{it}, t) = \alpha + \lambda_t + A_i' \gamma + X_{it}' \beta$$

- Then (additive and constant causal effect)

$$E(Y_{1it} | A_i, X_{it}, t) = E(Y_{0it} | A_i, X_{it}, t) + \rho$$

Which then implies

$$E(Y_{it} | A_i, X_{it}, t) = \alpha + \lambda_t + A_i' \gamma + X_{it}' \beta + \rho D_{it}$$

Then, the empirical function will be

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}' \beta + \varepsilon_{it}$$

where $\alpha_i \equiv \alpha + A_i' \gamma$, $\varepsilon_{it} \equiv Y_{0it} - E(Y_{0it} | A_i, X_{it}, t)$

Three ways to estimate fixed effect model

- **Method 1: Dummy Variable Estimation**

- Treating the fixed effects, α_i , as parameters to be estimated; the year fixed effect, τ_t , is also treated as a parameter to be estimated. The unobserved individual effects are coefficients on dummies for each individual, while the year effects are coefficients on time dummies.

- **Method 2: within estimator, demeaning (deviation from means) estimator, or fixed effect estimator**

- Method 1 and Method 2 are equivalent mathematically

- **Method 3: Differencing**

Demeaning regression: absorbing the fixed effect

- Fixed effects transformation or within transformation
- (1) Calculate the individual averages

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta}_i + c_i + \bar{u}_i \quad i = 1, 2, \dots, N$$

Where $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$, $\bar{\mathbf{x}}_i = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it}$, $\bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it}$

- (2) Make a demeaning

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

Or $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it} \boldsymbol{\beta} + \ddot{u}_{it}$

- Then the unobserved individual effect is killed

Differencing

- Make a difference between one year and the year before

$$y_{it} - y_{it-1} = (x_{it} - \bar{x}_{it-1})\beta + (u_{it} - u_{it-1})$$

Or

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it-1}$$

- With two periods, differencing is algebraically the same as demeaning, but not otherwise.
- With more than two periods, homoskedastic and serially uncorrelated ϵ_{it} , demeaning is more efficient
- Differenced residuals are inherently serially correlated

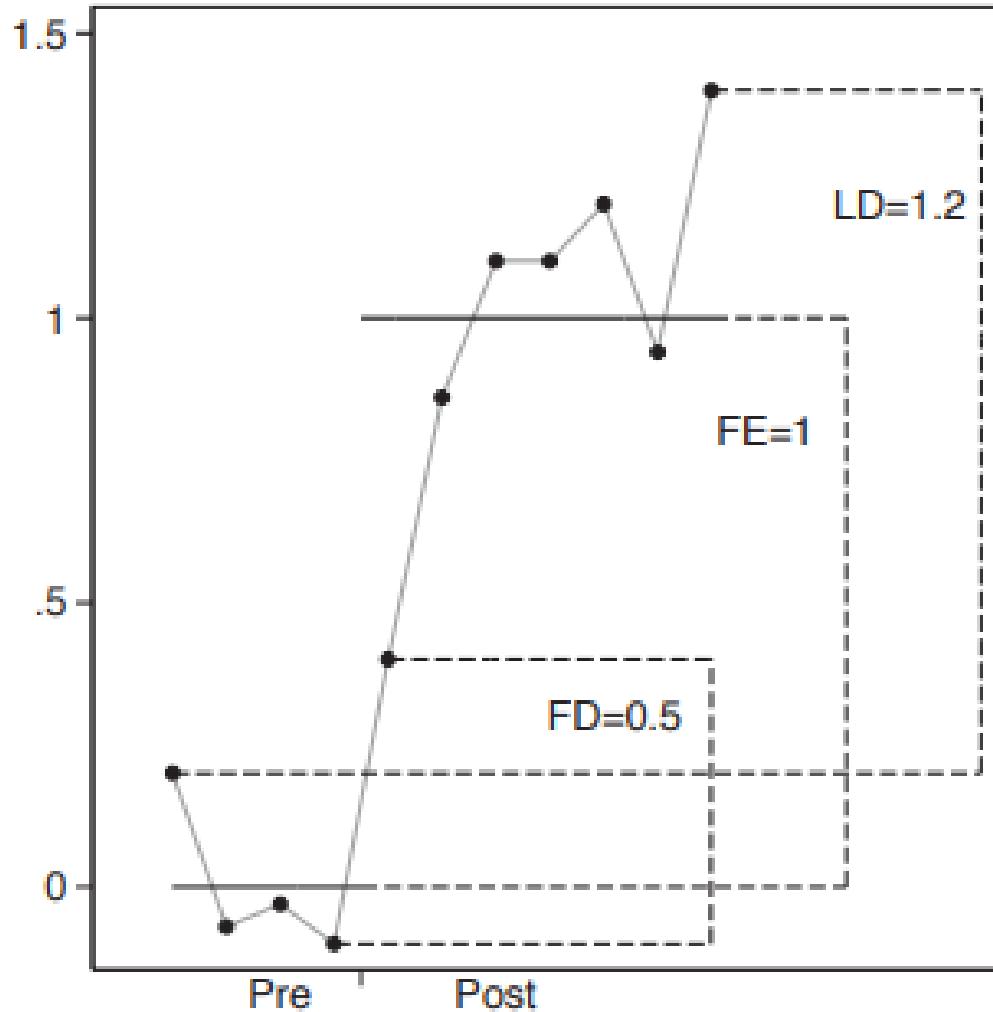


Figure 1: One panel's contributions to FE/FD/LD estimates

Estimated effects of union status on wages, Freeman (1984)

Survey	Cross Section Estimate	Fixed Effects Estimate
May CPS, 1974–75	.19	.09
National Longitudinal Survey of Young Men, 1970–78	.28	.19
Michigan PSID, 1970–79	.23	.14
QES, 1973–77	.14	.16

Notes: Adapted from Freeman (1984). The table reports cross section and panel (fixed effects) estimates of the union relative wage effect. The estimates were calculated using the surveys listed in the left-hand column. The cross section estimates include controls for demographic and human capital variables.

Other examples: Twins identification studies

- Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) estimate the returns to schooling using samples of twins, controlling for family-fixed effects.
 - Ashenfelter and Krueger (1994) use cross-sibling reports to construct instruments for schooling differences between twins.
- Li, H., P. W. Liu, J. Zhang and N. Ma (2007). 'Economic returns to communist party membership: Evidence from urban Chinese twins', *The Economic Journal*, vol. 117(523), pp. 1504-1520.

Caveats of FE

- At a minimum, therefore, it's important to avoid overly strong causality claims when interpreting fixed effects estimates (never bad advice for an applied econometrician in any case).
- **xtreg, fe** (be cautious with the reported number of observations)

xtreg2, fe or reghdfe

- Time-invariant covariates will be demeaned or differenced
- Always keep in mind the level your analysis is conducted

$$y_{ipt} = x_{ipt}\beta + c_i + \delta_p + \lambda_t + u_{ipt} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T$$

(delta_p might be redundant)

$$y_{ipt} = x_{ipt}\beta + \delta_p + \lambda_t + u_{ipt} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T$$

(if the endogeneity is due to individual-level fixed effect, then endogeneity is still there)

- Source of variations that your strategy depends on

Difference-in-differences (Intuition)

- The fixed effects strategy requires panel data, that is, repeated observations of subjects at the level your study focuses on (or firms, or whatever the unit of observation might be).
- Often, however, the regressor of interest varies only at a more aggregate or group level, such as state or cohort. (*when using pooled cross-sectional data, the comparability of group composition should be justified*)
- For example, state policies regarding health care benefits for pregnant workers may change over time but are fixed across workers within states. The source of OVB when evaluating these policies must, therefore, be unobserved variables at the state and year level.
- In some cases, group-level omitted variables can be captured by group-level fixed effects, an approach that leads to the differences-in-differences (DD) identification strategy.

The DD pioneer John Snow (1855)

- John Snow (1855) studied cholera epidemics in London in the mid-nineteenth century
 - Establish that cholera is transmitted by contaminated drinking water (as opposed to “bad air,” the prevailing theory at the time).
 - Most medical opinion about cholera transmission at that time was *miasma*, which said diseases were spread by microscopic poisonous particles that infected people by floating through the air. These particles were thought to be inanimate, and because microscopes at that time had incredibly poor resolution, it would be years before microorganisms would be seen.
 - Treatments, therefore, tended to be designed to stop poisonous dirt from spreading through the air. But tried and true methods like quarantining the sick were strangely ineffective at slowing down this plague.
- Imagine, if you were in Snow's shoes, what will and can you do to verify your guess?

Table 70. Compared to what? Different companies.

Company	Outcome
Lambeth	$Y = L + D$
Southwark and Vauxhall	$Y = SV$

Table 71. Compared to what? Before and after.

Company	Time	Outcome
Lambeth	Before	$Y = L$
	After	$Y = L + (T + D)$

Table 72. Compared to what? Difference in each company's differences.

Companies	Time	Outcome	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + T + D$	$T + D$	
				D
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + T$	T	

Table 69. Modified Table XII (Snow 1854).

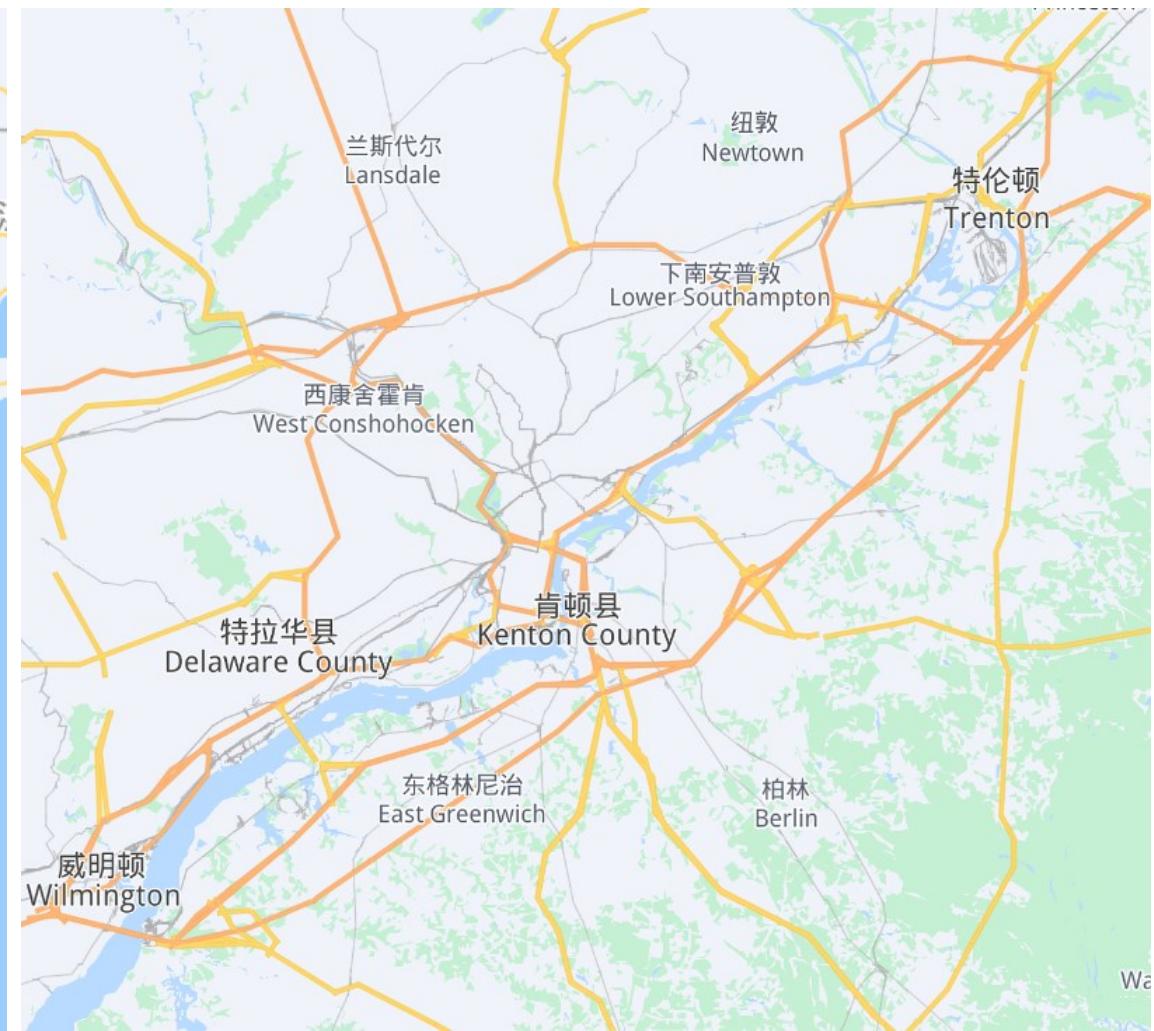
Company name	1849	1854
Southwark and Vauxhall	135	147
Lambeth	85	19

The Parallel Trends Assumption

- We are assuming that there is no time-variant company-specific unobservables.
- That is, without treatment, the average outcomes of the actually treated and nontreated subjects would experience the same change over time.
- This is equivalent to assuming that T is the same for all units. “*the parallel trends assumption*” or “common trend assumption”
- Actually, two factors will lead to the failure of the assumption
 - ❖ The trend is actually not common (group-specific trend)
 - ❖ There might be other time-varying confounders
 - Nothing unobserved in Lambeth households that is changing between these two periods that also determines cholera deaths. (for example, an NGO was distributing masks or kettles in the residence zone of Lambeth)

DD: A formal framework

- Card and Kruger (1994)
- On April 1, 1992, New Jersey raised the state minimum wage from \$4.25 to \$5.05. Card and Krueger collected data on employment at fast-food restaurants in New Jersey in February 1992 and again in November 1992. These restaurants (Burger King, Wendy's, and so on) are big minimum-wage employers.
- Card and Krueger also collected data from the same type of restaurants in eastern Pennsylvania, just across the Delaware River. The minimum wage in Pennsylvania stayed at \$4.25 throughout this period.
- They used their data set to compute differences-in-differences (DD) estimates of the effects of the New Jersey minimum wage increase. That is, they compared the February-to-November change in employment in New Jersey to the change in employment in Pennsylvania over the same period.



- Y_{1ist} , the potential fast good employment at restaurant i in state s and period t if there is a high state minimum wage;
- Y_{0ist} , the potential fast good employment at restaurant i in state s and period t if there is a low state minimum wage; (counterfactual outcomes)
- The heart of the DD setup is an additive structure for potential outcomes in the no-treatment state. Specifically, we assume that:

$$E(Y_{0ist} | s, t) = \gamma_s + \lambda_t$$

where s denotes state (New Jersey or Pennsylvania) and t denotes period (February, before the minimum wage increase, or November, after the increase).

- This says that in the absence of a minimum wage change employment is given by state effect, and a time effect, which is assumed to be the same in both states (**“common trend hypothesis”**).

- Let D_{st} be a dummy for high-minimum-wage states and periods, assume that,

$$E(Y_{1ist} - Y_{0ist} | s, t) = \delta$$

- Then the observed employment in restaurant i as

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where $E(\varepsilon_{ist} | s, t) = 0$

- From here, we can get

$$\begin{aligned} & E(Y_{ist} | s = NJ, t = Nov) - E(Y_{ist} | s = NJ, t = Feb) \\ &= (\gamma_{NJ} + \lambda_{Nov} + \delta) - (\gamma_{NJ} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb} + \delta \end{aligned}$$

$$\begin{aligned} & E(Y_{ist} | s = PA, t = Nov) - E(Y_{ist} | s = PA, t = Feb) \\ &= (\gamma_{PA} + \lambda_{Nov}) - (\gamma_{PA} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb} \end{aligned}$$

- The population difference-in-differences is the causal effect of interest.

$$\begin{aligned} & \{E(Y_{ist} | s = NJ, t = Nov) - E(Y_{ist} | s = NJ, t = Feb)\} \\ & - E(Y_{ist} | s = PA, t = Nov) - E(Y_{ist} | s = PA, t = Feb) = \delta \end{aligned}$$

November 1992

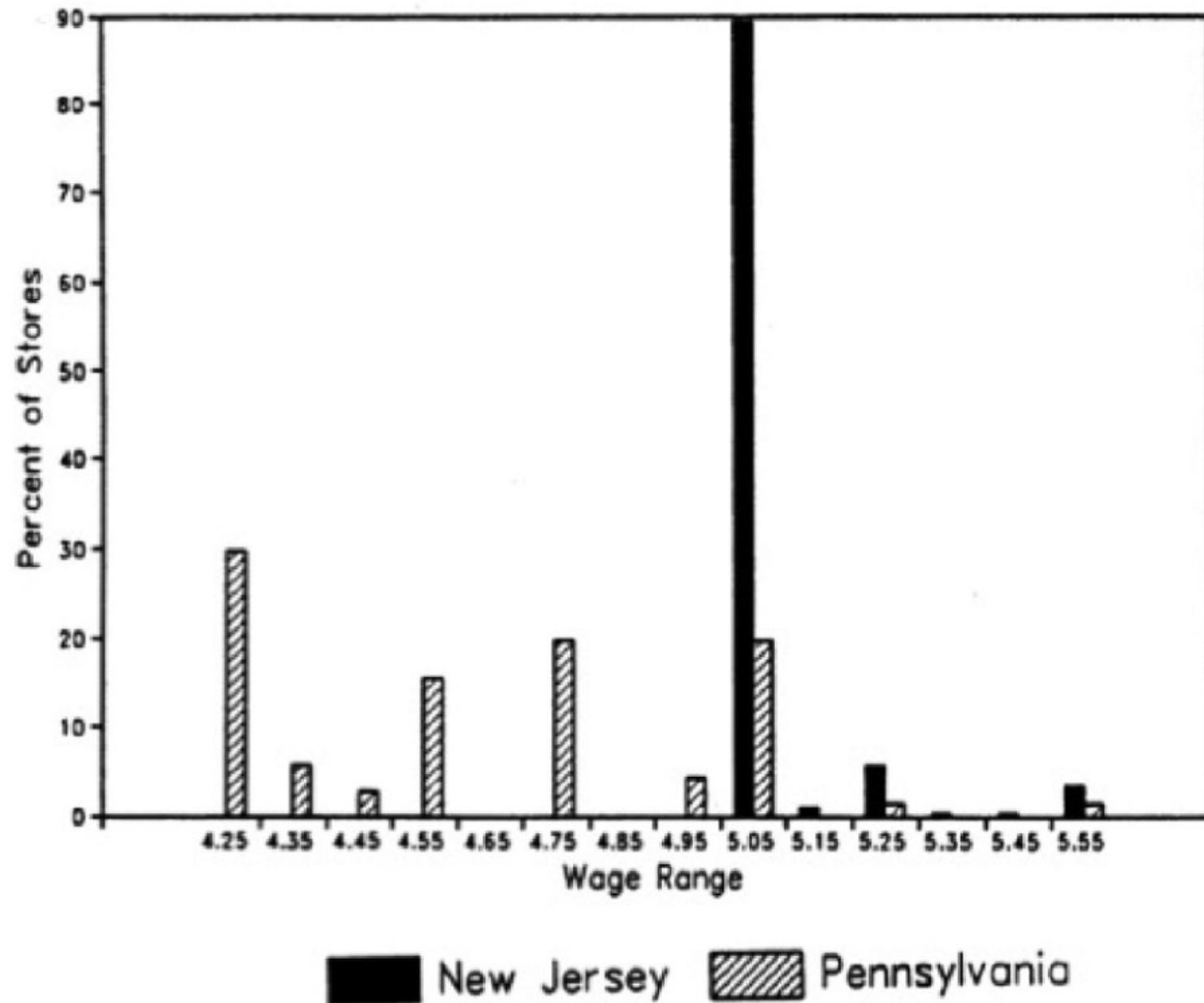


Figure 54. Distribution of wages for NJ and PA in November 1992.

Notice how effective this is at convincing the reader that *the minimum wage in New Jersey was binding*. This piece of data visualization *is not a trivial, or even optional, strategy* to be taken in studies such as this. Beautiful pictures displaying the “first stage” effect of the intervention on the treatment are crucial in the rhetoric of causal inference

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. The table reports average full-time-equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all restaurants with data on employment. Employment at six closed restaurants is set to zero. Employment at four temporarily closed restaurants is treated as missing. Standard errors are reported in parentheses.

- Why we can estimate through table 5.2.1: Repeated observations at group level + group fixed effect
- The Policy shock occurs at the aggregate (group) level

- **Identification assumption**

- The key identifying assumption here is that employment trend would be same in both states in the absence of treatment.
- *Treatment induces a deviation from this common trend*, as illustrated in the figure.
- Although the treatment and control states can differ, this difference is meant to be captured by the state fixed effect, which plays the same role as the unobserved individual effect.

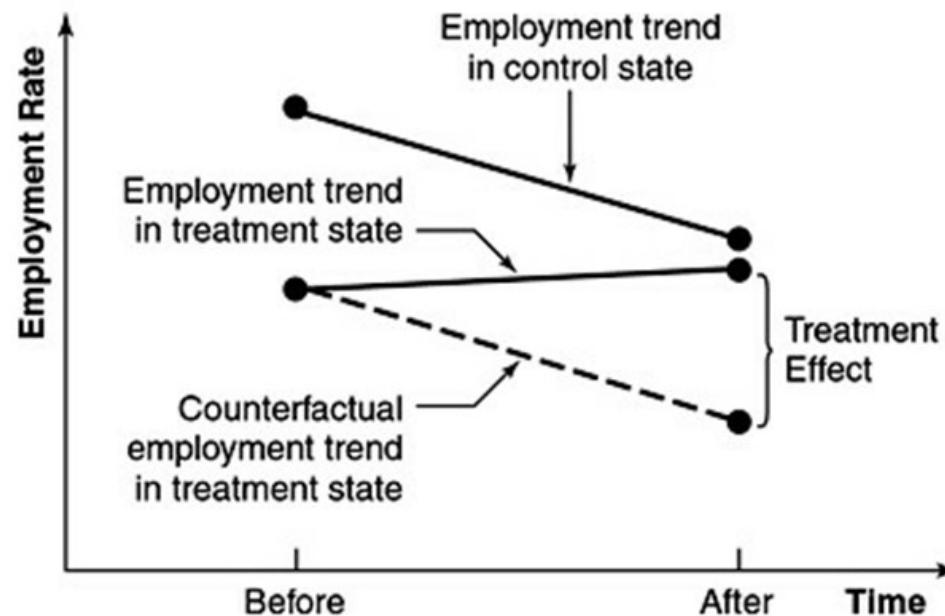
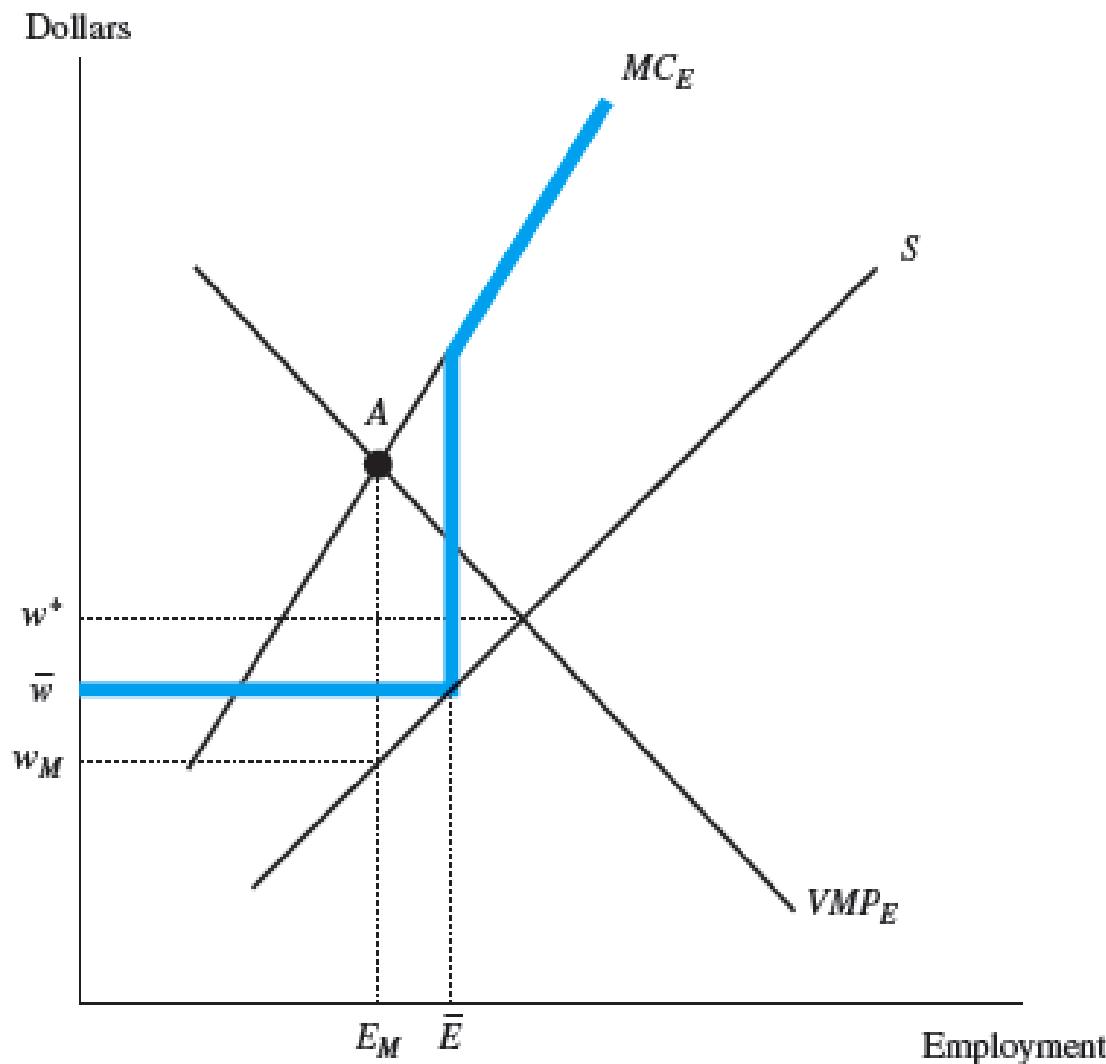


Figure 5.2.1
Causal effects in
the DD model

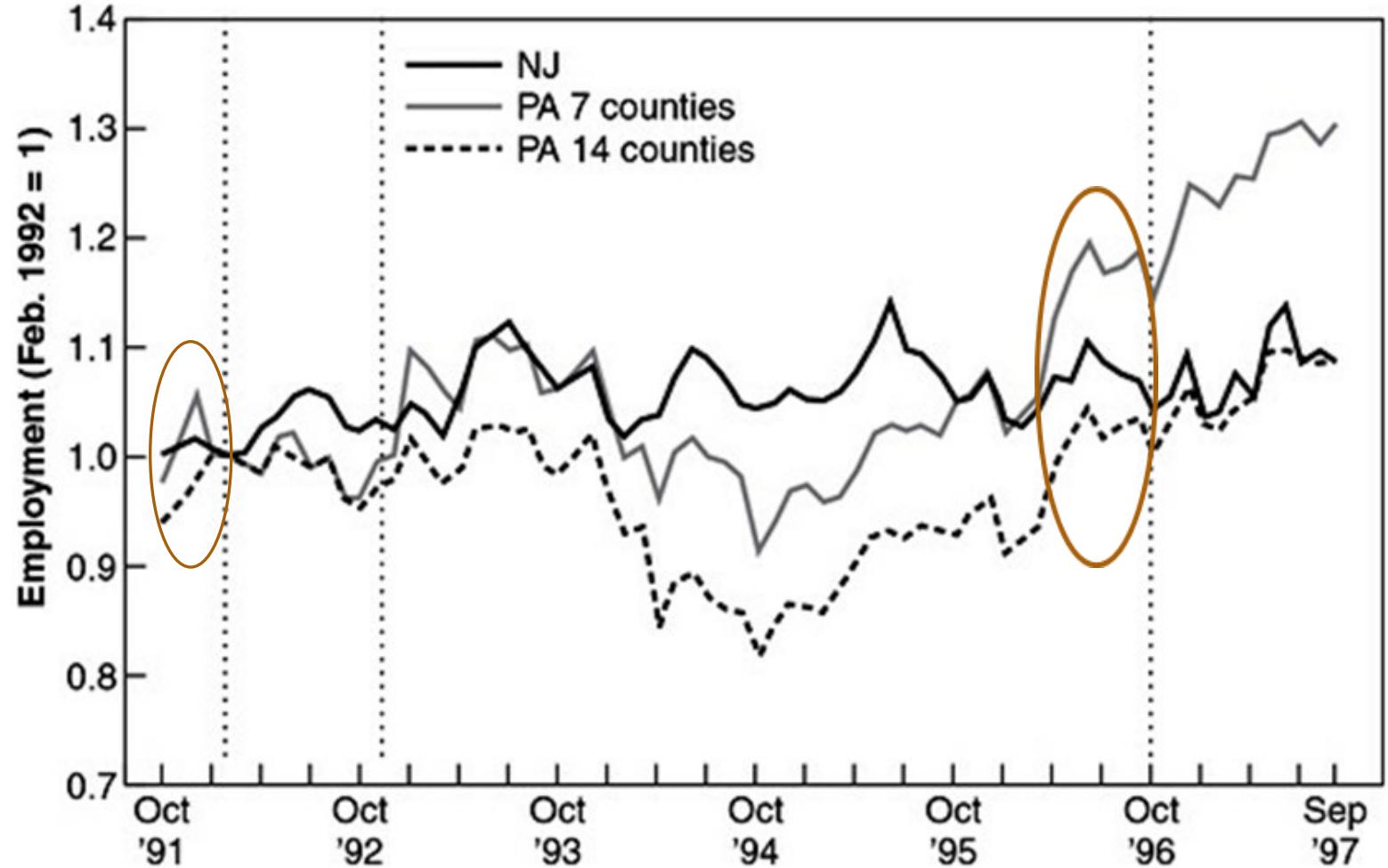
FIGURE 4-20 The Impact of the Minimum Wage on a Nondiscriminating Monopsonist

The minimum wage may increase both wages and employment when imposed on a monopsonist. A minimum wage set at \bar{w} increases employment to \bar{E} .



To test the “Common Trend Assumption”

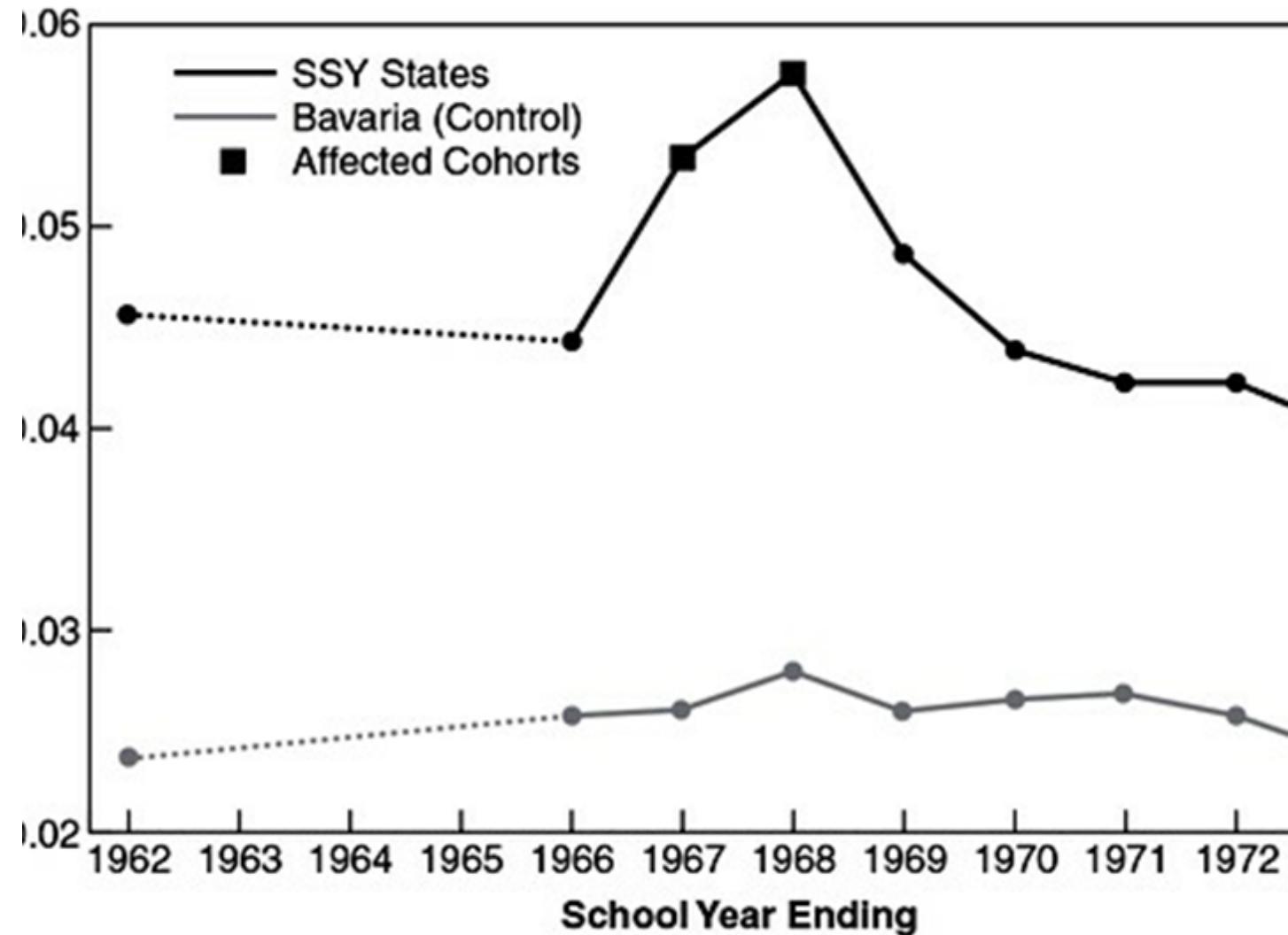
- The common trend assumption can be investigated using data on multiple periods: To plot the data is always a good idea.
- #1 A poor example: ([David Card and Alan B Krueger, 1994](#))



However, to check out the pre-treat trend is neither necessary nor sufficient for the parallel trend hypothesis.

#2. A good example

Figure 5.2.3 Average grade repetition rates in second grade for treatment and control schools in Germany (from Pischke, 2007). The data span a period before and after a change in term length for students outside Bavaria (SSY states).



Regression DD (2*2)

- Regression function:

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$

- Let NJ_s be a dummy for restaurants in New Jersey and d_t be a time-dummy that switches on for observations obtained in November. Actually, we can write this in a more general and equivalent form as $D_{st} = NJ_s * d_t$.

$$\alpha = E(Y_{ist} | s = PA, t = Feb) = \gamma_{PA} + \lambda_{Feb}$$

$$\gamma = E(Y_{ist} | s = NJ, t = Feb) - E(Y_{ist} | s = PA, t = Feb) = \gamma_{NJ} - \gamma_{PA}$$

$$\lambda = E(Y_{ist} | s = PA, t = Nov) - E(Y_{ist} | s = PA, t = Feb)$$

$$= (\gamma_{PA} + \lambda_{Nov}) - (\gamma_{PA} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb}$$

$$\delta = \{E(Y_{ist} | s = NJ, t = Nov) - E(Y_{ist} | s = NJ, t = Feb)\}$$

$$-\{E(Y_{ist} | s = PA, t = Nov) - E(Y_{ist} | s = PA, t = Feb)\}$$

Labor Supply

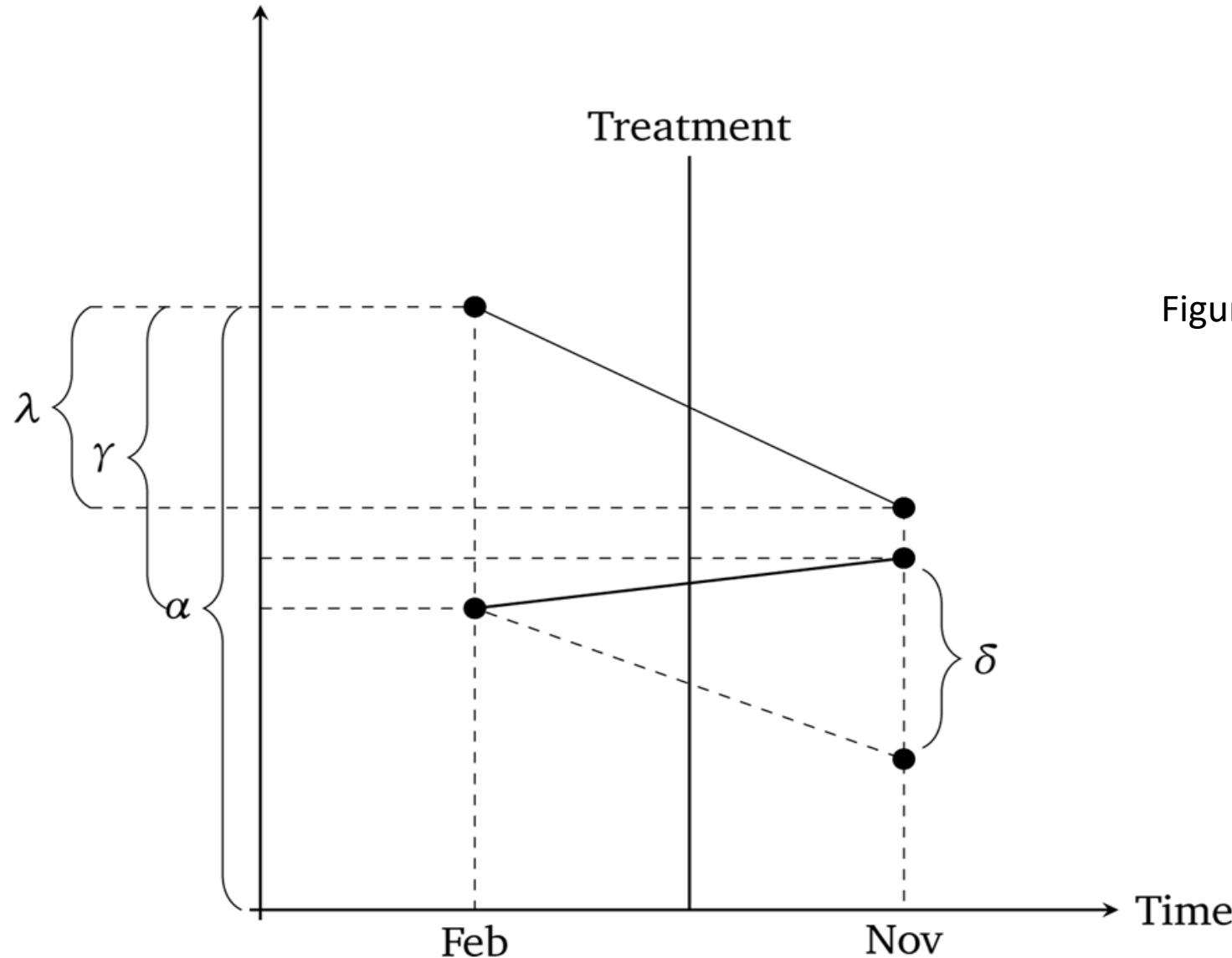
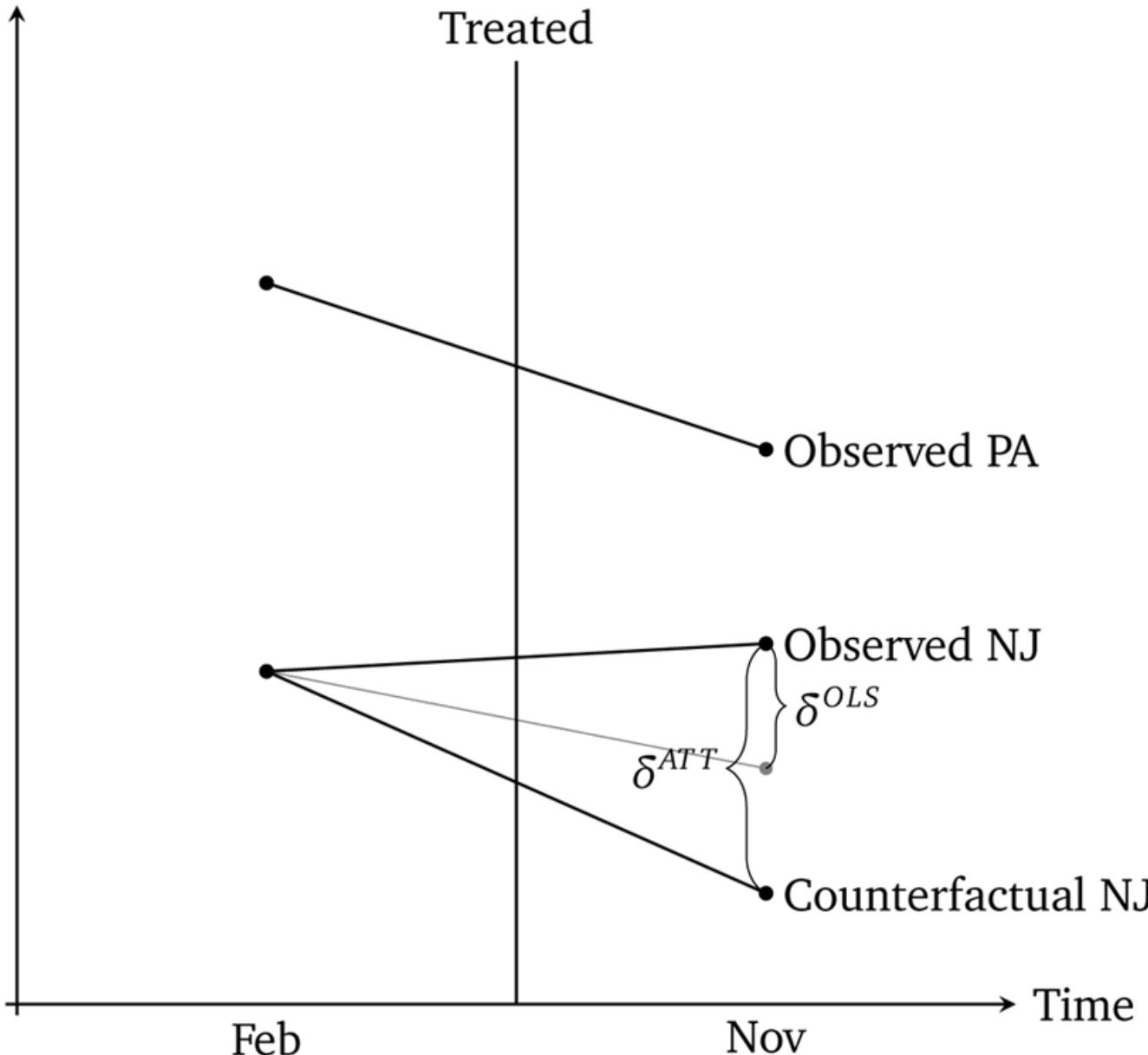


Figure 55. DD regression diagram

Labor Supply



Once again, the parallel trend hypothesis is the most crucial thing in a DD setup.

Notice that

$$E(Y_{ist} \mid s = PA, t = Feb) = \alpha$$

$$E(Y_{ist} \mid s = NJ, t = Feb) = \alpha + \gamma$$

$$E(Y_{ist} \mid s = PA, t = Nov) = \alpha + \lambda$$

$$E(Y_{ist} \mid s = NJ, t = Nov) = \alpha + \gamma + \lambda + \delta$$

$$E(Y_{ist} \mid s, t) = \gamma_s + \lambda_t$$

$$\alpha = E(Y_{ist} \mid s = PA, t = Feb) = \gamma_{PA} + \lambda_{Feb}$$

Extension 1: incorporate X_{ist}

- Keep in your mind all the time to which level of the variable of interest is aggregated
- It's easy to add additional covariates in this framework. In other words, we can model counterfactual employment in the absence of a change in the minimum wage as

$$E(Y_{0ist} | s, t, X_{ist}) = \gamma_s + \lambda_t + X_{ist}' \beta$$

$$E(Y_{1ist} | D = 1; s, t, X_{st}) = \gamma_s + \lambda_t + \delta + X_{ist}' \beta$$

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + X_{ist}' \beta + \varepsilon_{ist}$$

- Variables in X_{ist} could be individual level variables or *time varying variables at the state level*. Including individual level variables may not only help to control for confounding trends, but may also reduce the standard errors of the estimate.
- What about the time constant variable at the state level? Duflo (2003) gives a good example

Addition of state-specific time trend

- An alternative check on the DD identification strategy adds state-specific time trends to the regressors in X_{ist} .

$$Y_{ist} = \gamma_{0s} + \gamma_{1s} t + \lambda_t + \beta D_{s,t} + X_{ist} \delta + \varepsilon_{ist}$$

- Where γ_{0s} is a state-specific intercept, and γ_{1s} is a state-specific trend coefficient multiplying the time trend variable, t . This allows treatment and control states to follow different trends in a limited but potentially revealing way.
- How to incorporate a linear time trend in panel data analysis using Stata?

`gen t=year-2000 // 2000 is the baseline year, that is the first`

or

`sort id year`

`bysort id: gen t=_n`

or just the calendar year itself

Or, more generally

- State-specific linear trend

$$Y_{ist} = \beta D_{s,t} + X_{ist} \delta + \gamma_s + \lambda_t + \gamma_s t + \varepsilon_{ist}$$

- State-specific more general time trend

$$Y_{ist} = \beta D_{s,t} + X_{ist} \delta + \gamma_s + \lambda_t + \gamma_s * \lambda_t + \varepsilon_{ispt} \quad \text{or}$$

$$Y_{ist} = \beta D_{s,t} + X_{ist} \delta + \gamma_s + \lambda_t + \phi_{st} + \varepsilon_{ispt}$$

(ideal in theory, but infeasible)

- Province-specific more general time trend

$$Y_{ispt} = \beta D_{s,t} + X_{ispt} \delta + \gamma_s + \lambda_t + \gamma_p t + \varepsilon_{ispt}$$

Where p index province, s, say, prefecture; as an alternative to state-prefecture trend, imperfect but sometimes acceptable (s is affiliated to p, then number of p is smaller than that of s)

Or more complex

- Controlling the confounding program, P_{st}

$$Y_{ist} = \beta D_{s,t} + \alpha P_{st} + X_{ist}\delta + \gamma_s + \lambda_t + \varepsilon_{ist}$$

- To conserve the degrees of freedom, Zs can be an time-invariant continuous covariate at s level

$$Y_{ist} = \beta D_{s,t} + X_{ist}\delta + \gamma_s + \lambda_t + Z_s t + \varepsilon_{ist}$$

- An ultimate control form: combining together but avoid duplicate or redundant controls.

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	-.186 (.064)	-.185 (.051)	-.104 (.039)	.0002 (.020)
Log development expenditure per capita		.240 (.128)	.184 (.119)	.241 (.106)
Log installed electricity capacity per capita		.089 (.061)	.082 (.054)	.023 (.033)
Log state population		.720 (.96)	0.310 (1.192)	-1.419 (2.326)
Congress majority			-.0009 (.01)	.020 (.010)
Hard left majority			-.050 (.017)	-.007 (.009)
Janata majority			.008 (.026)	-.020 (.033)
Regional majority			.006 (.009)	.026 (.023)
State-specific trends	No	No	No	Yes
Adjusted R^2	.93	.93	.94	.95

Notes: Adapted from Besley and Burgess (2004), table IV. The table reports regression DD estimates of the effects of labor regulation on productivity. The dependent variable is log manufacturing output per capita. All models include state and year effects. Robust standard errors clustered at the state level are reported in parentheses. State amendments to the Industrial Disputes Act are coded 1 = pro-worker, 0 = neutral, -1 = pro-employer and then cumulated over the period to generate the labor regulation measure. Log of installed electrical capacity is measured in kilowatts, and log development expenditure is real per capita state spending on social and economic services. Congress, hard left, Janata, and regional majority are counts of the number of years for which these political groupings held a majority of the seats in the state legislatures. The data are for the sixteen main states for the period 1958–92. There are 552 observations.

Besley and Burgess (2004)

Study the effects of labor regulation on businesses in Indian states

TABLE 5.2.3
Estimated effects of labor regulation on the performance of firms in Indian states

Apparently, labor regulation in India increased in states where output was declining anyway. Control for this trend therefore drives the estimated regulation effect to zero.

Extension 2: Non-binary treatment

- Example 1: Card (1992)
- Look at all state minimum wages in the united states.
- Some of these are a little higher than the federal minimum (which covers everyone regardless of where they live), some are a lot higher, and some are the same. The minimum wage is therefore a variable with differing "treatment intensity" across states and over time.
- Moreover, in addition to statutory variation in state minima, the local importance of a minimum wage varies with average state wage levels. For example, the early-1990s federal minimum of \$4.25 was probably irrelevant in Connecticut - with high average wages - but a big deal in Mississippi.

- Card (1992) exploits regional variation in the impact of the federal minimum wage. His approach is motivated by an equation like

$$Y_{ist} = \gamma_s + \lambda_t + \beta(FA_s \cdot dt) + \varepsilon_{ist}$$

- where the variable FAs is a measure of the fraction of teenagers likely to be affected by a minimum wage increase in each state, more specifically, the fraction of workers who are paid less than \$3.80 just before the increase of the minimum wages.
- dt is a dummy for observations after April 1990, when the federal minimum increased from \$3.35 to \$3.80. The FAs variable measures the baseline (pre-increase) proportion of each state's teen labor force earning less than \$3.80.
- Since there are still only two time periods in the Card (1992) setup, the equation can be differenced over time to obtain

$$\Delta Y_{st} = \lambda^* + \beta FA_s + \Delta \varepsilon_{st}$$

Explanatory Variable	Change in Mean Log Wage		Change in Teen Employment-Population Ratio	
	(1)	(2)	(3)	(4)
1. Fraction of affected teens (FA_s)	.15 (.03)	.14 (.04)	.02 (.03)	-.01 (.03)
2. Change in overall emp./pop. ratio	—	.46 (.60)	—	1.24 (.60)
3. R^2	.30	.31	.01	.09

Notes: Adapted from Card (1992). The table reports estimates from a regression of the change in average teen employment by state on the fraction of teens affected by a change in the federal minimum wage in each state. Data are from the 1989 and 1990 CPS. Regressions are weighted by the CPS sample size for each state.

TABLE 5.2.2
Regression DD estimates of minimum wage effects on teens, 1989 to 1990

Not just restricted to calendar year

- In the typical DD setting, T is by default, set to “Calendar year”
- But many times, other dimensions can be possible. One common example is the cohort.
- When the cohort takes the place of the time dimension, DD can be adopted even with a cross-sectional dataset

Example 2: Qian (2008)

- Empirical Strategy

$$\begin{aligned} sex_{ic} = & (tea_i \times post_c) \beta + (orchard_i \times post_c) \delta + (cashcrop_i \times post_c) \rho + Han_{ic} \zeta \\ & + \alpha + \psi_i + \gamma_c + \varepsilon_{ic} \end{aligned}$$

$post_c$ is a dummy variable that indicates if an individual is born after 1979, that is

$$post_c = \begin{cases} 1 & \text{if individuals are born after 1979} \\ 0 & \text{otherwise} \end{cases}$$

Sex_{ic} , the fraction of male in country i , cohort c is a function of the interaction terms between tea_i , the amount of tea planted for each county i , and $post_c$; the interaction terms between $orchard_i$, the amount of orchard planted for each county i , and $post_c$; the interaction terms between $cashcrop_i$, the amount of cash crops planted for each county i , and $post_c$; Han_{ic} , the fraction that is ethnically Han; φ_i , county fixed effect; γ_c , cohort fixed effect. The reference group is composed of individuals born during 1970-1979. It and all its interaction terms are dropped.

TABLE III
**OLS AND 2SLS ESTIMATES OF THE EFFECT OF PLANTING TEA AND ORCHARDS ON SEX
 RATIOS CONTROLLING FOR COUNTY LEVEL LINEAR COHORT TRENDS**

	Dependent variables					
	Fraction of males			Tea × post	Fraction of males	
	(1) OLS	(2) OLS	(3) OLS	(4) 1st	(5) IV	(6) IV
Tea × post	-0.012 (0.007)	-0.013 (0.006)	-0.012 (0.005)		-0.072 (0.031)	-0.011 (0.007)
Orchard × post	0.005 (0.002)					
Slope × post	-0.002 (0.002)			0.26 (0.057)		
Linear trend	No	No	Yes	Yes	No	Yes
Observations	28,349	37,756	37,756	37,756	37,756	37,756

Notes. Coefficients of the interactions between dummies indicating whether a cohort was born post-reform and the amount of tea planted in the county of birth. All regressions include county and birth year fixed effects and controls for Han, and cashcrop × post. All standard errors are clustered at the county level. In column (1), the sample includes all individuals born during 1970–1986. In columns (2)–(6), the sample includes all individuals born during 1962–1990. Post = 1 if birthyear > 1979. Data for land area sown are from the 1997 China Agricultural Census.

Example 3: Duflo, Esther (2001)

- The questions of whether investments in infrastructure can cause an increase in educational attainment, and whether an increase in educational attainment causes an increase in earnings are basic concerns for development economists.
- This paper exploits a dramatic change in policy to evaluate the effect building schools has on education and earnings in Indonesia.
- Indonesian children normally attend primary school between the ages of 7 and 12. All children born in 1962 or before were 12 or older in 1974, when the first INPRES schools were constructed. Thus, they did not benefit from the program, since they should have left primary school before the first INPRES schools were opened.
- For younger children, the exposure is an increasing function of their date of birth. Hence, the effect of the program should be close to 0 for children 12 or older in 1974 and increasing for younger children
- The program intensity was related to enrollment rates in 1972, which differed widely across regions, region of birth is a second dimension of variation in the intensity of the program.

Identification strategy

$$S_{ijk} = c + \alpha_j + \beta_k + P_j T_i \gamma + (C_j T_i) \delta + \varepsilon_{ijk}$$

where S_{ijk} is the education of individual i born in region j in year k , c is a constant, α_j is a district of birth fixed effect, β_k is a cohort of birth fixed effect, T_i is a dummy indicating whether the individual belongs to the "young" cohort in the subsample, P_j denotes the intensity of the program in the region of birth, and C_j is a vector of region-specific variables.

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
<i>Panel A: Experiment of Interest</i>						
Aged 2 to 6 in 1974	8.49 (0.043)	9.76 (0.037)	-1.27 (0.057)	6.61 (0.0078)	6.73 (0.0064)	-0.12 (0.010)
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Difference	0.47 (0.070)	0.36 (0.038)	0.12 (0.089)	-0.26 (0.011)	-0.29 (0.0096)	0.026 (0.015)
<i>Panel B: Control Experiment</i>						
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Aged 18 to 24 in 1974	7.70 (0.059)	9.12 (0.044)	-1.42 (0.072)	6.92 (0.0097)	7.08 (0.0076)	-0.16 (0.012)
Difference	0.32 (0.080)	0.28 (0.061)	0.034 (0.098)	0.056 (0.013)	0.063 (0.010)	0.0070 (0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN COHORT DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN THE REGION OF BIRTH

	Observations	Dependent variable						
		Years of education			Log(hourly wage)			
		(1)	(2)	(3)	(4)	(5)	(6)	
<i>Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974</i>								
<i>(Youngest cohort: Individuals ages 2 to 6 in 1974)</i>								
Whole sample	78,470	0.124 (0.0250)	0.15 (0.0260)	0.188 (0.0289)				
Sample of wage earners	31,061	0.196 (0.0424)	0.199 (0.0429)	0.259 (0.0499)	0.0147 (0.00729)	0.0172 (0.00737)	0.0270 (0.00850)	
<i>Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974</i>								
<i>(Youngest cohort: Individuals ages 12 to 17 in 1974)</i>								
Whole sample	78,488	0.0093 (0.0260)	0.0176 (0.0271)	0.0075 (0.0297)				
Sample of wage earners	30,225	0.012 (0.0474)	0.024 (0.0481)	0.079 (0.0555)	0.0031 (0.00798)	0.00399 (0.00809)	0.0144 (0.00915)	
<i>Control variables:</i>								
Year of birth*enrollment rate in 1971		No	Yes	Yes	No	Yes	Yes	
Year of birth*water and sanitation program		No	No	Yes	No	No	Yes	

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). The number of observations listed applies to the specification in columns (1) and (4). Standard errors are in parentheses.

Extension 3: From 2*2 to N*T

- A great Leap but seemingly plain

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$



$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where $E(\varepsilon_{ist} | s, t) = 0$

$$Y_{ist} = \gamma_s + \lambda_t + \delta(\gamma_s \cdot d_t) + X'_{st} \beta + \varepsilon_{ist}$$



$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + X'_{st} \beta + \varepsilon_{ist}$$

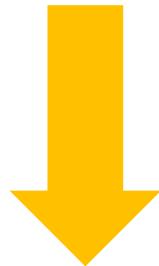


- D_{st} , is a policy dummy that is defined to be unity for groups and time periods subject to the policy.

$$D_{st} = \begin{cases} 1, & \text{if } t \geq \tau \text{ for state } s, \tau \text{ is the time that policy implements} \\ 0, & \text{otherwise} \end{cases}$$

- A seeming irrelevant but essential transformation: a great leap from 2*2 to N*T
- This imposes the restriction that the policy has the same effect in every year (Jeffrey M. Wooldridge, 2010).

ind	year	event_year	year_till_event	Yit	a_sh	bt	as_bt	Xit
BJ	2020				0	0	0	
BJ	2022				0	1	0	
SH	2020	2021	-1		1	0	0	
SH	2022	2021	1		1	1	1	



ind	year	event_year	year_till_eve	Yit	a_sh	bt	as_bt	dst	Xit
BJ	2020				0	0	0	0	
BJ	2022				0	1	0	0	
SH	2020	2021	-1		1	0	0	0	
SH	2022	2021	1		1	1	1	1	

ind	year	ar	nt	Yit	event_ye		year_till_eve		a_sh*b_2022	a_sh*b_2023	a_gz*b_2022	a_gz*b_2023	Dst
					a_sh	a_gz	b_2022	b_2023					
BJ	2020				0	0	0	0	0	0	0	0	0
BJ	2022				0	0	1	0	0	0	0	0	0
BJ	2023				0	0	0	1	0	0	0	0	0
SH	2020	2021	-1		1	0	0	0	0	0	0	0	0
SH	2022	2021	1		1	0	1	0	1	0	0	0	1
SH	2023	2021	2		1	0	0	1	0	1	0	0	1
GZ	2020	2021	-1		0	1	0	0	0	0	0	0	0
GZ	2022	2021	1		0	1	1	0	0	0	1	0	1
GZ	2022	2021	2		0	1	0	1	0	0	0	1	1

With more periods, we can fly

- As a start, let's consider a policy that occurs all at t_0 (e.g. single timing rolled out to treated units)
- More time periods helps in several ways:
 - If we have multiple periods *before* the policy implementation, we can partially test the underlying assumptions. Sometimes referred to as “pre-trends”
 - If we have multiple periods *after* the policy implementation, we can examine the timing of the effect
 - Is it an immediate effect? Does it die off? Is it persistent?
 - If you pool all time periods together into one “post” variable, this estimates the average effect.
 - Or alternatively, you implicitly assume a constant treatment effect over time
 - If the sample is not balanced, can have unintended effects!

Extension 4: Heterogeneous treatment effect and event study (2->T)

- **Heterogeneous treatment effect: q lags (or q after) – posttreatment effect**

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=0}^q \delta_j D_{st}(s, t = \tau + j) + X'_{st} \beta + \varepsilon_{ist}$$

- Moreover, the $\delta_j, \forall j > 0$ may not be identical. For example, the effect of the treatment could accumulate over time, so that increases in j ; or fade away.

- **Heterogeneous treatment effect: m leads (or m before) – anticipatory effect**

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=-m}^{-1} \delta_j D_{st}(s, t = \tau + j) + X'_{st} \beta + \varepsilon_{ist}$$

- A test of the differences assumption is $\delta_j = 0 \forall j < 0$, i.e. the coefficients on all leads of the treatment should be zero. This is thought of direct test of the parallel trend hypothesis.

Granger causality and event study

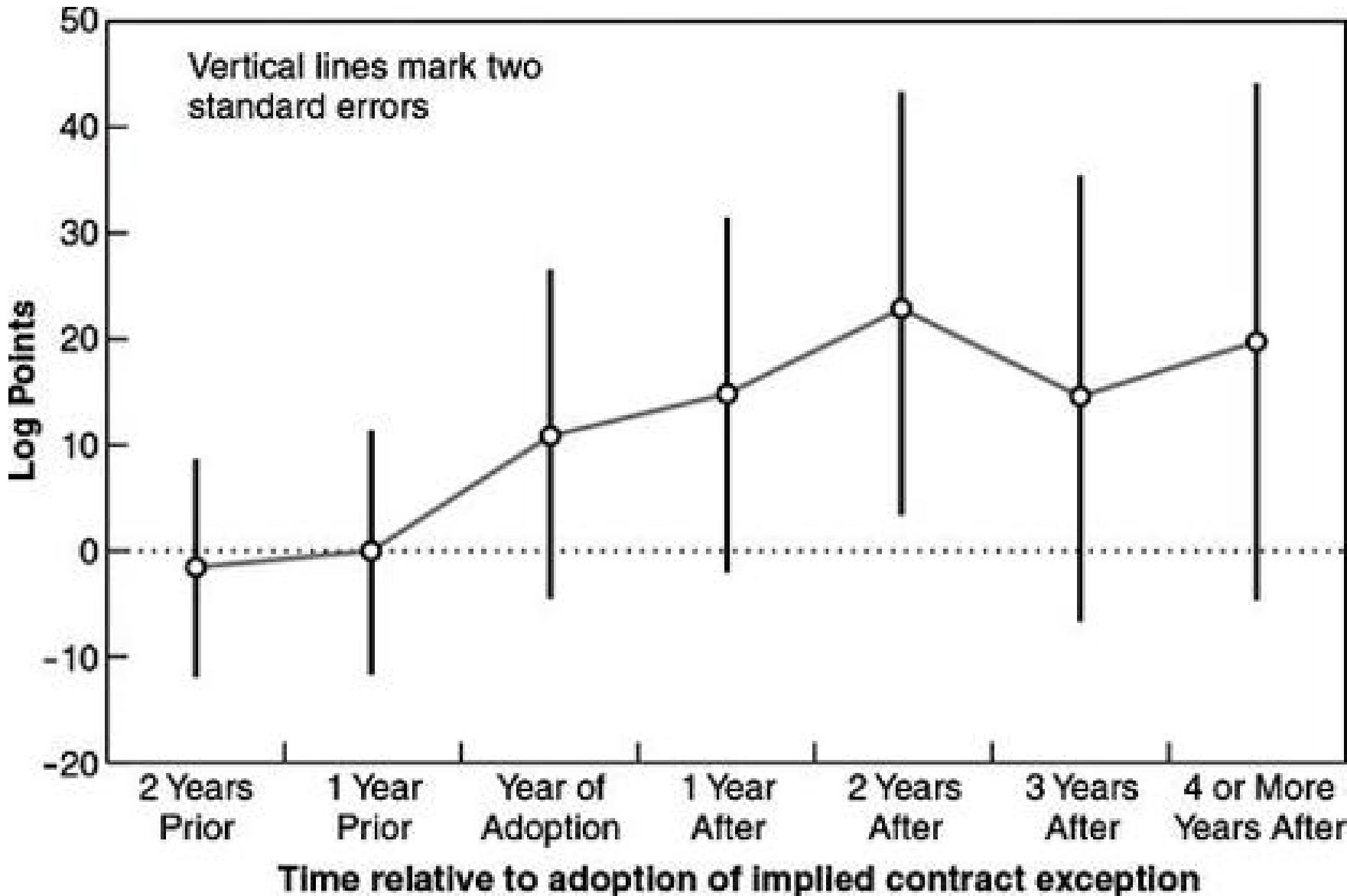
- When the sample includes many years, the regression-DD model lends itself to a test for causality in the spirit of Granger (1969).
- The Granger idea is to see whether causes happen before consequences, and not vice versa
- Suppose the policy variable of interest, D_{st} , changes at different times in different states. In this context, Granger causality testing means a check on whether, conditional on state and year effects, past D_{st} predicts Y_{ist} while future D_{st} does not. If it holds true, then dummies for future policy changes should not matter in an equation like:

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=0}^m \delta_{-\tau} D_{s,t-\tau} + \sum_{\tau=1}^q \delta_{+\tau} D_{s,t+\tau} + X_{ist}' \beta + \varepsilon_{ist}$$

where the sums on the right-hand side allow for m lags ($\delta_{-1}, \delta_{-2}, \dots, \delta_{-m}$) or posttreatment effects and q leads ($\delta_{+1}, \delta_{+2}, \dots, \delta_{+q}$) or anticipatory effects. The pattern of lagged effects is usually of substantive interest as well. We might, for example, believe that causal effects should grow or fade as time passes.

Autor (2003)

- The effect of employment protection on firms' use of temporary help.
- In the US, Employment protection is a type of labor law (Particularly true in China)—promulgated by state legislatures or, more typically, through common law as made by state courts—that makes it harder to fire workers.
- As a rule, US labor law allows employment at will, which means that workers can be fired for just cause or no cause, at the employer's whim. But some state courts have allowed a number of exceptions to the employment-at-will doctrine, leading to lawsuits for unjust dismissal.
- Autor is interested in whether fear of employee lawsuits makes firms more likely to use temporary workers for tasks for which they would otherwise have increased their workforce. Temporary workers are employed by someone else besides the firm for which they are executing tasks. As a result, firms using them cannot be sued for unjust dismissal when they let temporary workers go.



Miller et al., 2019

Miller, S., Altekroose, S., Johnson, N., and Wherry, L. R. (2019). Medicaid and mortality: New evidence from linked survey and administrative data. NBER working paper

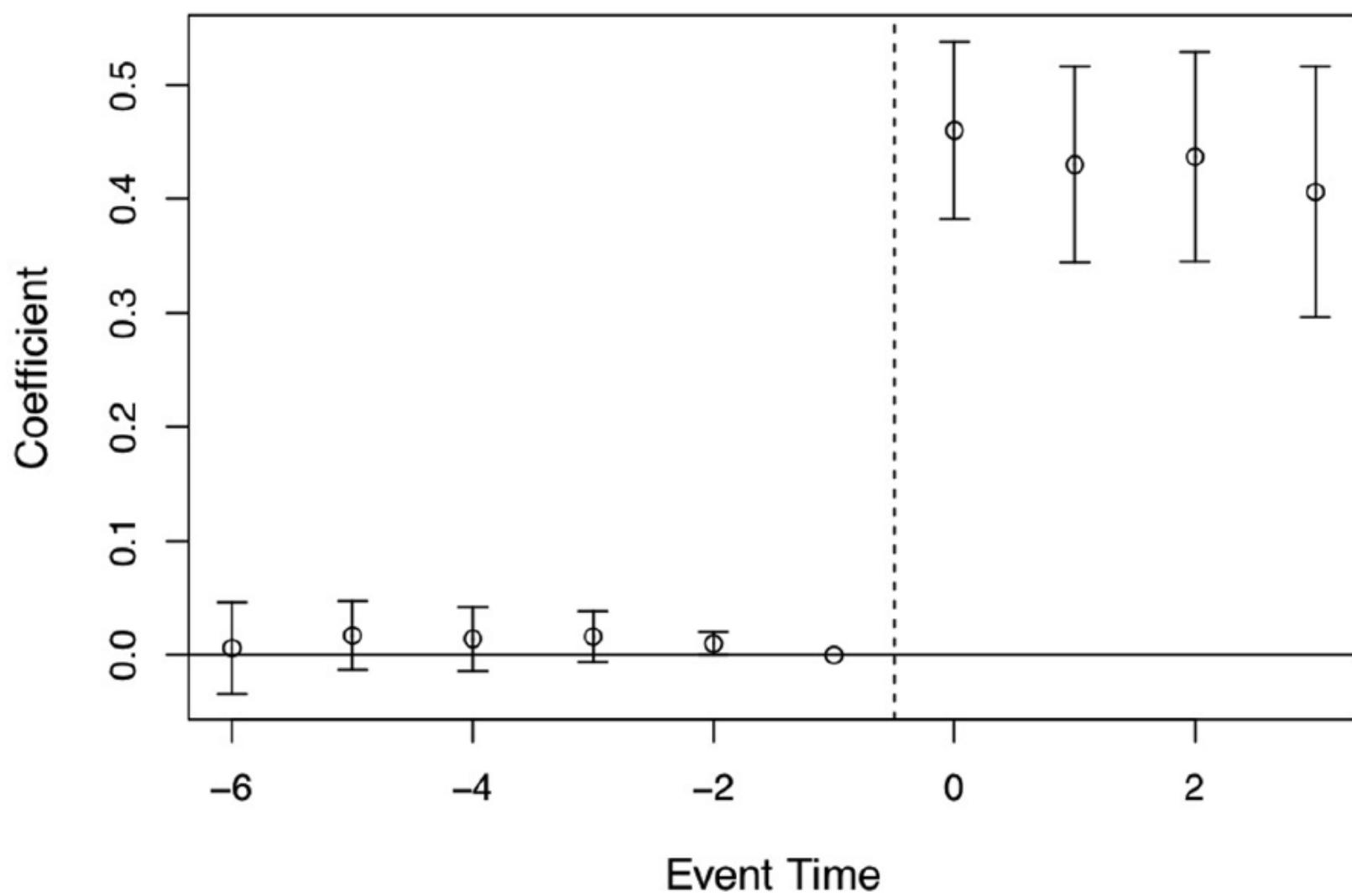


Figure 57. Estimates of Medicaid expansion's effects on eligibility using leads and lags in an event-study model

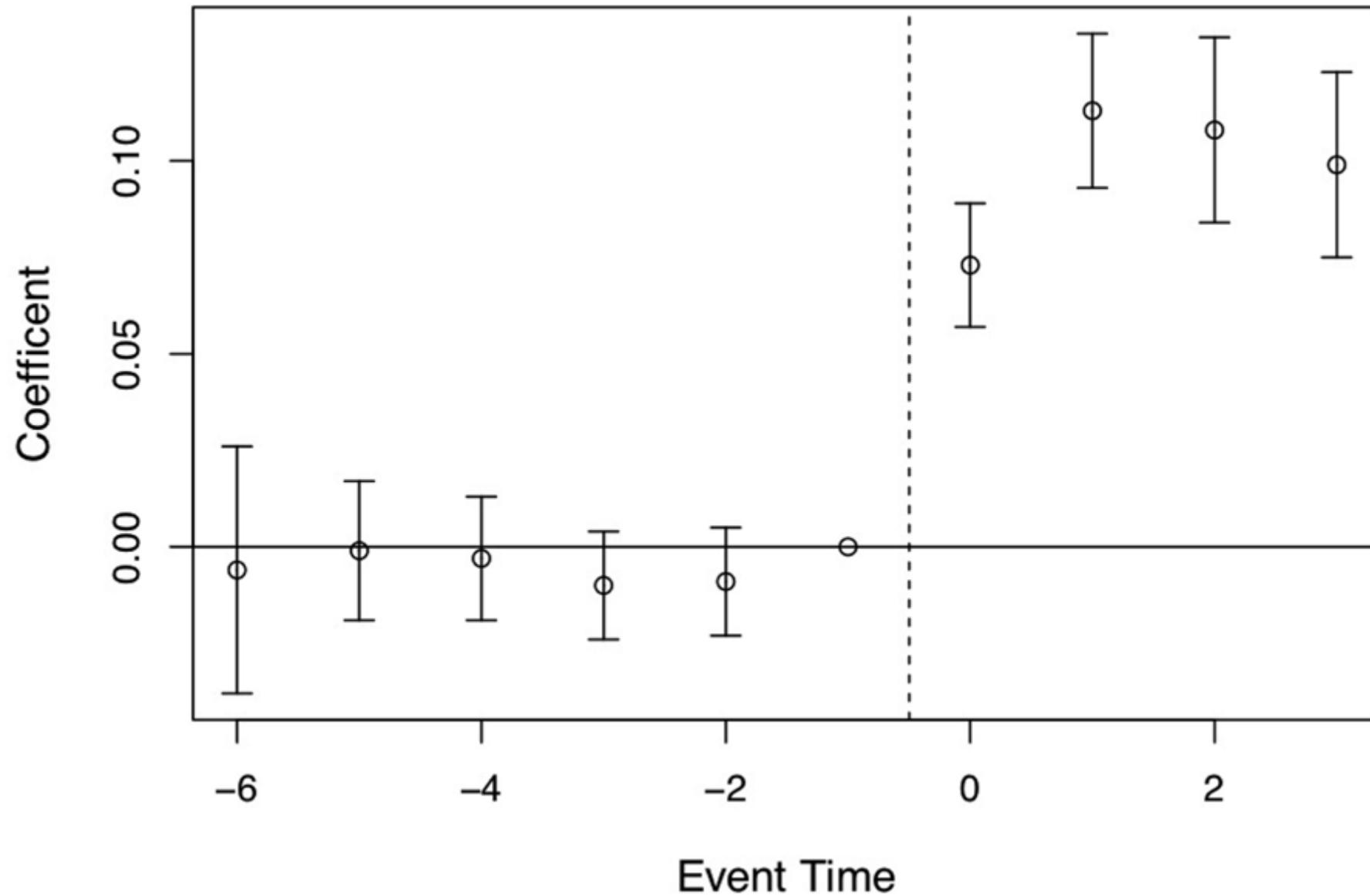


Figure 58. Estimates of Medicaid expansion's effects on coverage using leads and lags in an event-study model.

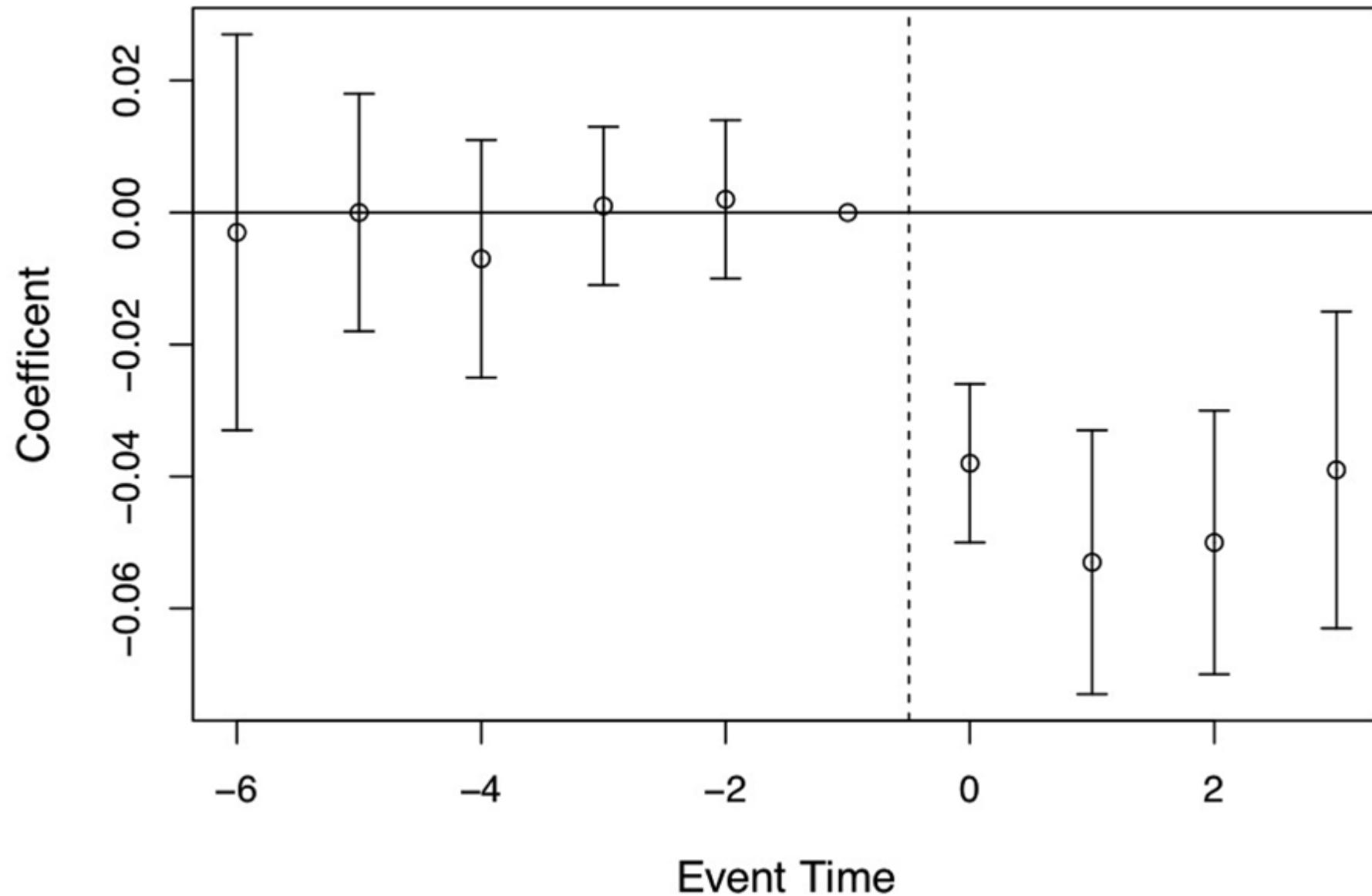


Figure 59. Estimates of Medicaid expansion's effects on the uninsured state using leads and lags in an event-study model.

Cunningham, S. and Cornwell, C. (2013)

- Cunningham, S. and Cornwell, C. (2013). “The Long-Run Effect of Abortion on Sexually Transmitted Infections.” *American Law and Economics Review*, 15(1):381–407.
- The design exploited the early repeal of abortion in five states in 1970 and compared those states to the states that were legalized under *Roe v. Wade* in 1973.
- To do this, I needed cohort-specific data on gonorrhea incidence by state and year, but as those data are not collected by the CDC, I had to settle for second best. That second best was the CDC’s gonorrhea data broken into five-year age categories (e.g., age 15–19, age 20–24). But this might still be useful because even with aggregate data, it might be possible to test the model I had in mind.
- Literature context:
- Gruber et al.(1999): the characteristics of the marginal child aborted had that child reached their teen years. ---any far-reaching effects?
- Donohue and Levitt (2001), Levitt (2004); abortion legalization –decrease in crime rates

Age in calendar year	CDC Surveillance Data in Calendar Year															
	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
15	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
16	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
17	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
18	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
19	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
20	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
21	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
22	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
23	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77
24	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
25	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
26	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
27	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
28	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
29	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
Repeal (1)	0	1	2	3	4	5	5	5	5	5	5	5	5	5	5	5
No Repeal (2)	0	0	0	0	1	2	3	4	5	5	5	5	5	5	5	5
Difference (3)	0	1	2	3	3	2	1	0	0	0	0	0	0	0	0	0

Number of cohorts (age 15-19) exposed, reforms in 71, 74

Figure 62. Theoretical predictions of abortion legalization on age profiles of gonorrhea incidence.

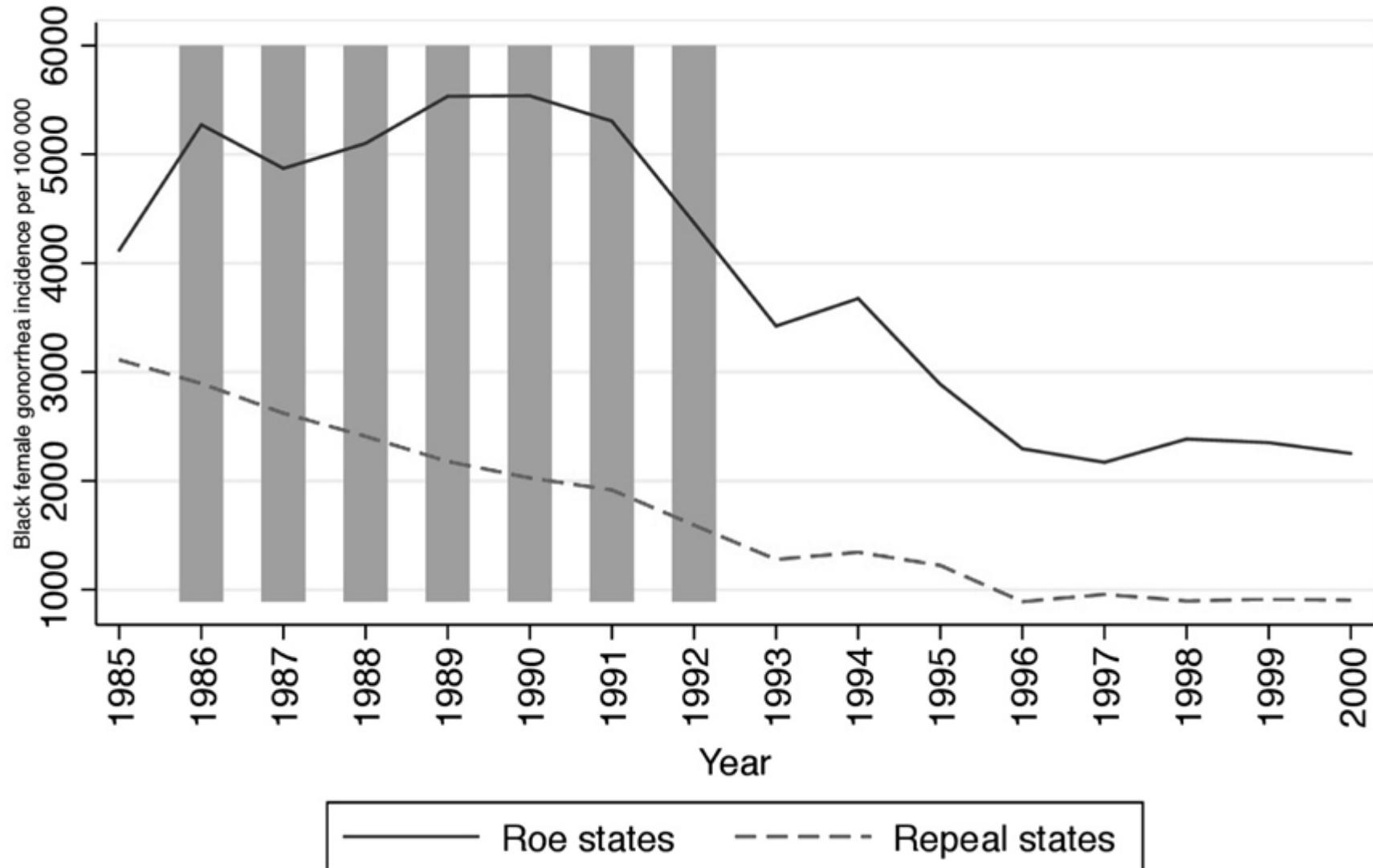


Figure 63. Differences in gonorrhea incidence among black females between repeal and *Roe* cohorts expressed as coefficient plots.

Now let's look at regression coefficients. Our estimating equation is as follows:

$$Y_{st} = \beta_1 Repeals + \beta_2 DT_t + \beta_{3t} Repeal_s \times DT_t + X_{st}\psi + \alpha_s DS_s + \varepsilon_{st}$$

where Y is the log number of new gonorrhea cases for 15- to 19-year-olds (per 100,000 of the population); $Repeal_s$ equals 1 if the state legalized abortion prior to *Roe*; DT_t is a year dummy; DS_s is a state dummy; t is a time trend; X is a matrix of covariates. In the paper, I sometimes included state-specific linear trends, but for this analysis, I present the simpler model. Finally, ε_{st} is a structural error term assumed to be conditionally independent of the regressors. All standard errors, furthermore, were clustered at the state level allowing for arbitrary serial correlation.

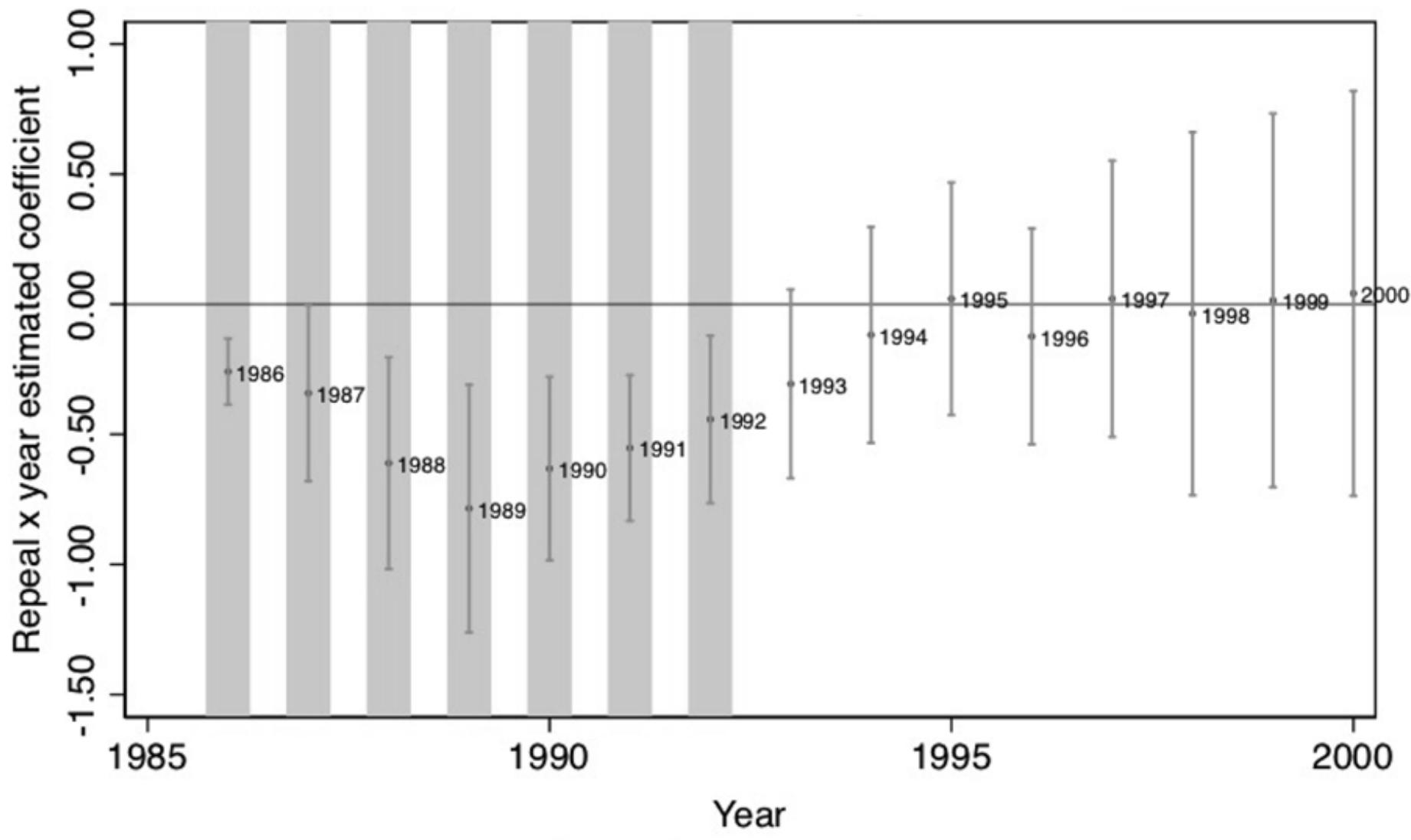


Figure 64. Coefficients and standard errors from DD regression equation.

Extension 5: Difference-in-difference-in-differences(DDD) strategy (2->N)

- In some cases, a more convincing analysis of a policy change is available by further refining the definition of treatment and control groups. For example, suppose a state implements a change in health care policy aimed at the elderly, say people 65 and older, and the response variable, y , is a health outcome.
- DD strategy 1: using data only on people in the state with the policy change
 - treatment group: 65 and older in the state
 - control group: people under 65 in the state
- DD strategy 2: Use another state as the control and use the elderly from the non-policy state as the control group
- Difference-in-difference-in-differences(DDD) strategy

DDD

- We again label the two time periods as 1 and 2, let B represent the state implementing the policy (vs. A, the states not implementing the policy), and let E denote the group of elderly (vs. N, nonelderly), then an expanded version

$$y = \beta_0 + \beta_1 dB + \beta_2 dE + \beta_3 dB \cdot dE \\ + \delta_0 d2 + \delta_1 d2 \cdot dB + \delta_2 d2 \cdot dE + \delta_3 d2 \cdot dB \cdot dE + u$$

The coefficient of interest is now δ_3 . The OLS estimate $\hat{\delta}_3$ can be expressed as follows:

$$\hat{\delta}_3 = (\bar{y}_{B,E,2} - \bar{y}_{B,E,1}) - (\bar{y}_{A,E,2} - \bar{y}_{A,E,1}) - (\bar{y}_{B,N,2} - \bar{y}_{B,N,1})$$

- The hope is that this controls for two kinds of potentially confounding trends: (1) changes in health status of elderly across states which would have nothing to do with the policy; (2) changes in health status of all people living in the policy-change state (possibly due to other state policies that affect everyone's health).

Table 74. Triple differences design.

States	Group	Period	Outcomes	D_1	D_2	D_3
NJ	Low-wage workers	After	$NJ_l + T + NJ_t + l_t + D$	$T + NJ_t +$	$(l_t - h_t) + D$	D
		Before	NJ_l	$l_t + D$		
	High-wage workers	After	$NJ_h + T + NJ_t + h_t$	$T + NJ_t + h_t$		
		Before	NJ_h			
PA	Low-wage workers	After	$PA_l + T + PA_t + l_t$	$T + PA_t + l_t$	$l_t - h_t$	
		Before	PA_l			
	High-wage workers	After	$PA_h + T + PA_t + h_t$	$T + PA_t + h_t$		
		Before	PA_h			

From DD to DDD

JS (Treated =1)			ZJ (Treated =0)				
	t0	t1		t0	t1	DD	DDD
Elder	r_JS,E	$r_{JS,E} + T + T_{JS+E} + ATE$		Elder	$r_{ZJ,E}$	$r_{ZJ,E} + T + T_{ZJ+E}$	
	D	$T + T_{JS+E} + ATE$			D	$T + T_{ZJ+E}$	$(T_{JS} - T_{ZJ}) + ATE$
Youth	r_JS,Y	$r_{JS,Y} + T + T_{JS+T_Y}$		Youth	$r_{ZJ,Y}$	$r_{ZJ,Y} + T + T_{ZJ+T_Y}$	
	D	$T + T_{JS+T_Y}$			D	$T + T_{ZJ+T_Y}$	$(T_{JS} - T_{ZJ})$
	DD	$(T_E - T_Y) + ATE$			DD	$(T_E - T_Y)$	
	DDD		ATE(Average treatme nt on the treated)				

Gruber [1994]

- Study state-level policies providing maternity benefits.
- He uses as his treatment group married women of childbearing age in treatment and control states, but he also uses a set of placebo units (older women and single men 20–40) as within-state controls.
- He then goes through the differences in means to get the difference-in-differences for each set of groups, after which he calculates the DDD as the difference between these two difference-in-differences.

Table 75. DDD Estimates of the Impact of State Mandates on Hourly Wages.

Location/year	Pre-law	Post-law	Difference
<i>A. Treatment: Married women, 20–40yo</i>			
Experimental states	1.547 (0.012)	1.513 (0.012)	-0.034 (0.017)
Control states	1.369 (0.010)	1.397 (0.010)	0.028 (0.014)
Difference	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference		-0.062 (0.022)	
<i>B. Control: Over 40 and Single Males 20–40</i>			
Experimental states	1.759 (0.007)	1.748 (0.007)	-0.011 (0.010)
Control states	1.630 (0.007)	1.627 (0.007)	-0.003 (0.010)
Difference	1.09 (0.010)	1.21 (0.010)	
Difference-in-difference		-0.008 (0.014)	
DDD		-0.054 (0.026)	

3/5/ Note: Standard errors in parentheses.

Aaron S. Yelowitz (1995)

$$y_{iast} = \gamma_{st} + \lambda_{at} + \theta_{as} + \delta D_{ast} + X'_{iast} \beta + \varepsilon_{iast}$$

Where s index states, t indexes time, and a is the age of the youngest child in a family. This model provides full nonparametric control for state-specific time effects that are common across age groups (γ_{st}), time-varying age effects (λ_{at}), and state-specific age effects (θ_{as}). The regressor of interest, D_{ast} , indicates families with children in affected age groups in states and periods where Medicaid coverage is provided. This triple-differences model may generate a more convincing set of results than a traditional DD analysis that exploits differences by state and time alone.

Extension 6: Placebo test and DDD

- Observable pre-treatment dynamics
- Event study
- First stage
- Placebo test:
 - In most cases, the start point of analysis of DD design is the ITT (Intention-to-treat) estimate.
 - That is, the treated group contains treated individuals (the target subject of the policy) and untreated individuals (nontarget subjects). ITT is the average treatment effect on the treated group. Further disentangling the effects on these two subgroups is interesting and important
 - Suppose ITT exists, a natural expectation is:
 - the treatment effect on the treated individuals should be much stronger—as an alternative, first-stage effect should be provided
 - the treatment effect on the untreated individuals should be zero
- Placebo test is increasingly emphasized in modern empirical research in economics; it is even viewed as being necessary by many researchers; neat placebo test is extremely crucial for the credibility of one research.

How to implement the placebo test in the framework of DD

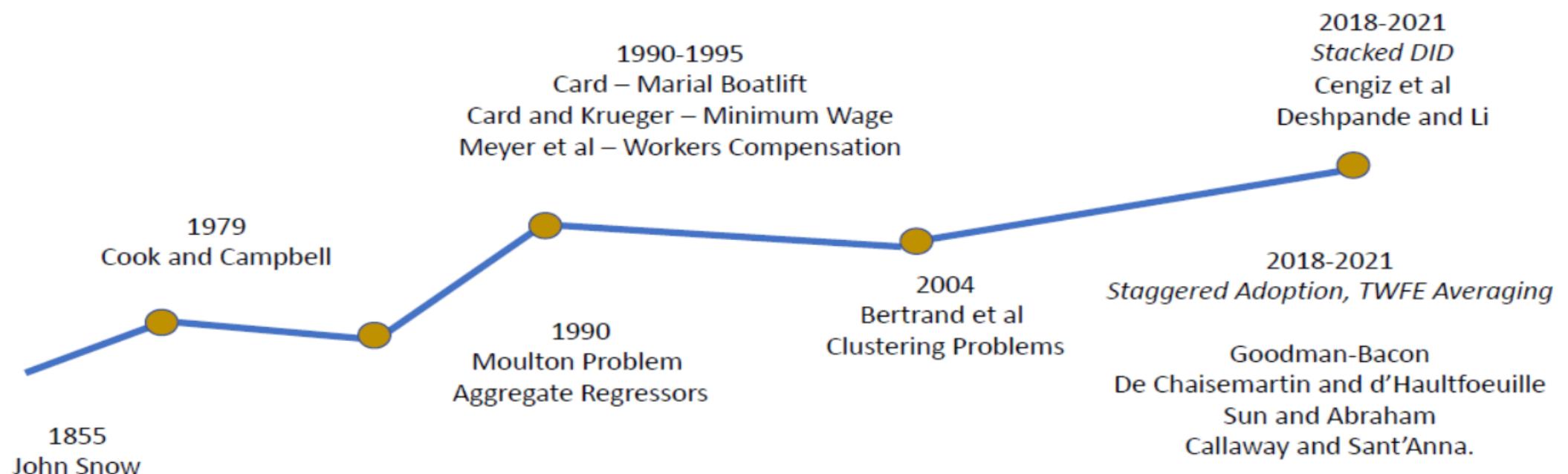
- Subgroup analysis
- DDD analysis: many times not a literally DDD, some form of DDD
- Picking controls and picking different life meanings
- Economic research is partly scientific and partly artistic, art of persuasion.
 - Pre-trends
 - Event-study
 - Placebo test

Inference in DD

- Anytime, keep in mind to report robust standard error clustered at group level
- Sometimes, it gets tricky about this points, such as Card and Kruger (1994).
- Sometimes clustered level is set to be group-year is acceptable.
- Bootstrap standard errors often serve as the last resort.

Extension 7: Most recent development in DID: Staggered adoption and Stacked DID

One history of difference in differences...



Staggered adoption or differential timing

- Staggered assignment of treatment across geographic units over time
- That is, geographic units or groups receive treatments at different points in time
- The DD strategy is thus divided into two types
 - DD with a universal timing / adoption
 - DD with a differential timing / staggered adoptions
- We have a good understanding of the first type of DD; but we did not as good an understanding of type II DD until Goodman-Bacon (2021)

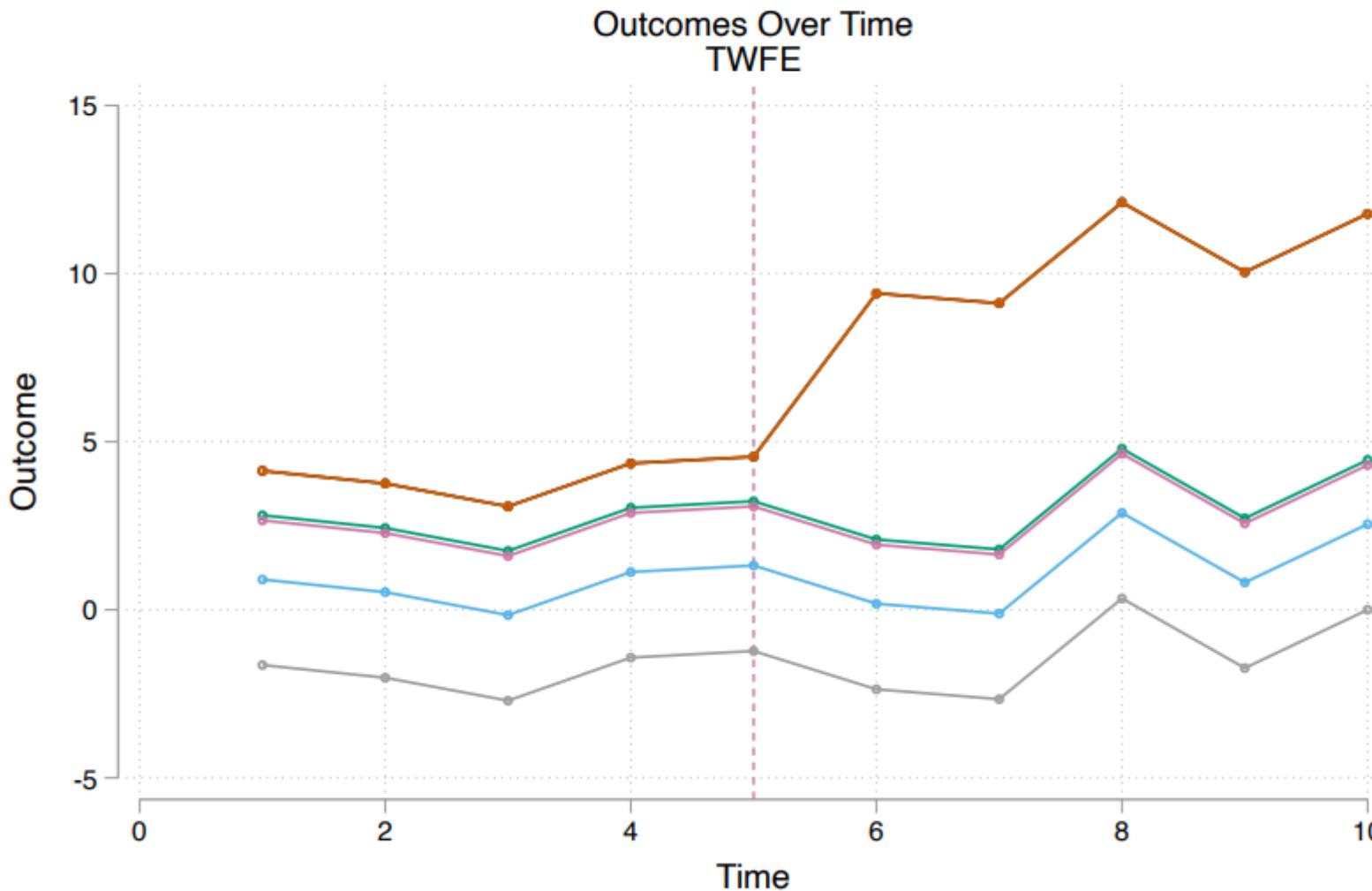
Two-way Fixed Effect DD estimation

- Suppose you have $s = 1, 2, \dots$ and $t = 1, \dots, T$ periods
- Treatment “turns on” at different times in different states.

$$Y_{st} = \beta^{FE} D_{st} + \alpha_s + \beta_t + \varepsilon_{st}$$

- D_{st} , dummy variable set to 1 if the policy is enforced in state s during period t .
 - α_s , state fixed effect (time invariant factor)
 - β_t , time fixed effect (time varying common factor)
- “Two-way Fixed Effects: The current workhorse”

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.

Now, staggered adoption or differential timing

- Things become messy, actually very messy.
- Multiple states adopt the treatment at different times, then
 - What is pre or post? No clear “pre” and “post”.
 - What is control group or treatment group? No clear “control” or “post”.
 - What if treatment effects are heterogeneous?
 - Across Units?
 - Over Time (phase-in effects, interactions with calendar time)
 - How will you do the event study without the universal adoption time? E.g., How to define pre or post, if state s never adopt the policy?

Consider a panel covering a group, indexed as g and time periods t . We are interested in estimating the impact of the passage of an event which may occur at different times in different groups. We will denote as Event_g , a variable recording the time period t in which the event is adopted in group g . Denoting the outcome of interest as y_{gt} , the panel event study specification can be written as

$$y_{gt} = \alpha + \sum_{j=2}^J \beta_j (\text{Lead } j)_{gt} + \sum_{k=1}^K \gamma_k (\text{Lag } k)_{gt} + \mu_g + \lambda_t + X'_{gt} \Gamma + \varepsilon_{gt}. \quad (1)$$

Here μ_g and λ_t are group and time fixed effects, X_{gt} are (optionally) time-varying controls, and ε_{gt} is an unobserved error term. In equation 1, leads and lags to the event of interest are defined as follows:

$$(\text{Lead } J)_{gt} = \mathbb{1}[t \leq \text{Event}_g - J], \quad (2)$$

$$(\text{Lead } j)_{gt} = \mathbb{1}[t = \text{Event}_g - j] \text{ for } j \in \{1, \dots, J-1\}, \quad (3)$$

$$(\text{Lag } k)_{gt} = \mathbb{1}[t = \text{Event}_g + k] \text{ for } k \in \{1, \dots, K-1\}, \quad (4)$$

$$(\text{Lag } K)_{gt} = \mathbb{1}[t \geq \text{Event}_g + K]. \quad (5)$$

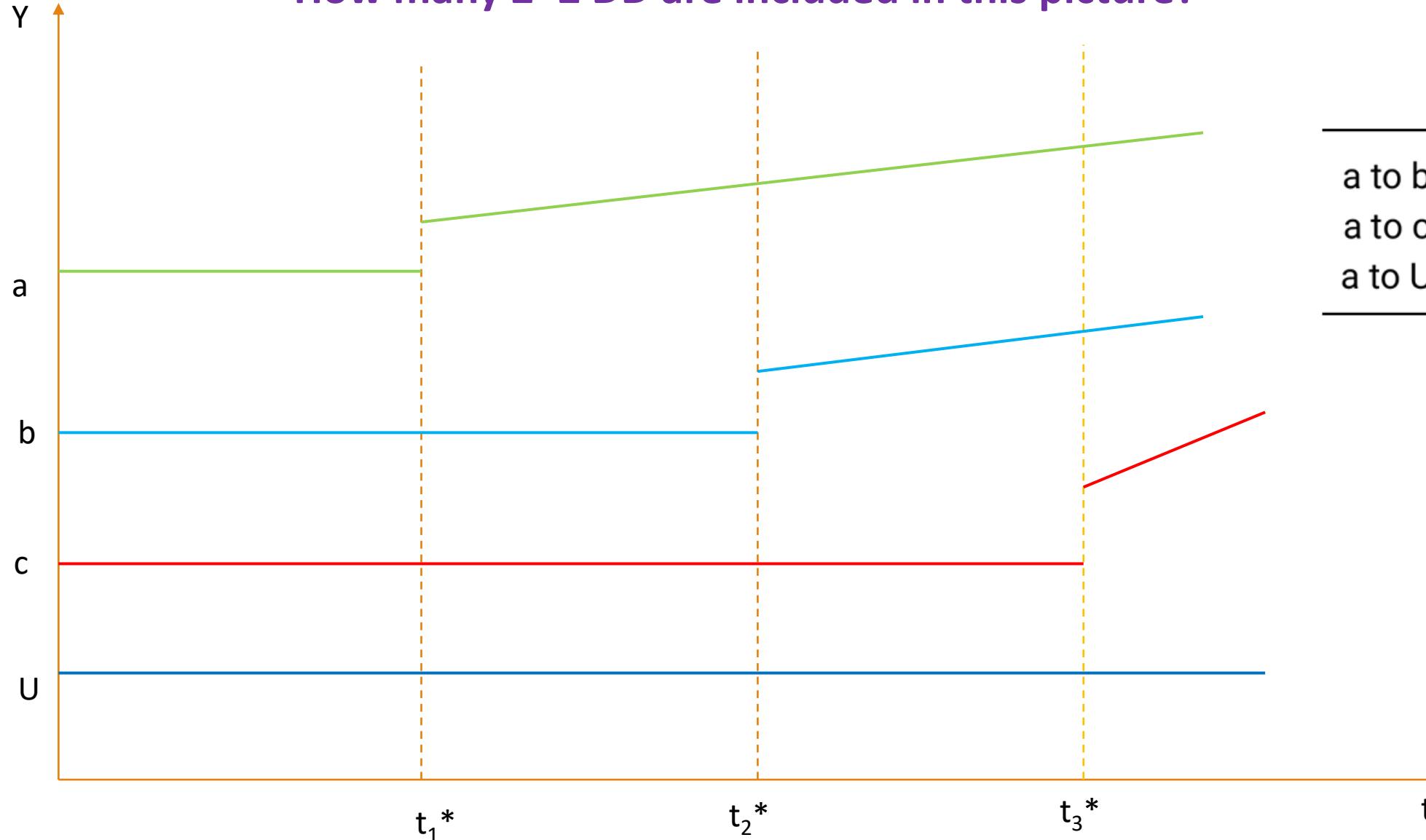
Group (g)	Year (t)	Event	Post Event	Time to Event	Lead 4	Lead 3	...	Lag 0	Lag 1	...	Lag 4
Group A	2000	2004	0	-4	1	0	...	0	0	...	0
Group A	2001	2004	0	-3	0	1	...	0	0	...	0
Group A	2002	2004	0	-2	0	0	...	0	0	...	0
Group A	2003	2004	0	-1	0	0	...	0	0	...	0
Group A	2004	2004	1	0	0	0	...	1	0	...	0
Group A	2005	2004	1	1	0	0	...	0	1	...	0
Group A	2006	2004	1	2	0	0	...	0	0	...	0
Group A	2007	2004	1	3	0	0	...	0	0	...	0
Group A	2008	2004	1	4	0	0	...	0	0	...	1
Group A	2009	2004	1	5	0	0	...	0	0	...	1
Group B	2000	2005	0	-5	1	0	...	0	0	...	0
Group B	2001	2005	0	-4	1	0	...	0	0	...	0
Group B	2002	2005	0	-3	0	1	...	0	0	...	0
Group B	2003	2005	0	-2	0	0	...	0	0	...	0
Group B	2004	2005	0	-1	0	0	...	0	0	...	0
Group B	2005	2005	1	0	0	0	...	1	0	...	0
Group B	2006	2005	1	1	0	0	...	0	1	...	0
Group B	2007	2005	1	2	0	0	...	0	0	...	0
Group B	2008	2005	1	3	0	0	...	0	0	...	0
Group B	2009	2005	1	4	0	0	...	0	0	...	1
Group C	2000	.	0	.	0	0	...	0	0	...	0
Group C	2001	.	0	.	0	0	...	0	0	...	0
Group C	2002	.	0	.	0	0	...	0	0	...	0
Group C	2003	.	0	.	0	0	...	0	0	...	0
Group C	2004	.	0	.	0	0	...	0	0	...	0
Group C	2005	.	0	.	0	0	...	0	0	...	0
Group C	2006	.	0	.	0	0	...	0	0	...	0
Group C	2007	.	0	.	0	0	...	0	0	...	0
Group C	2008	.	0	.	0	0	...	0	0	...	0
Group C	2009	.	0	.	0	0	...	0	0	...	0
Group D	2000	2007	0	-7	1	0	...	0	0	...	0
Group D	2001	2007	0	-6	1	0	...	0	0	...	0
Group D	2002	2007	0	-5	1	0	...	0	0	...	0
Group D	2003	2007	0	-4	1	0	...	0	0	...	0
Group D	2004	2007	0	-3	0	1	...	0	0	...	0
Group D	2005	2007	0	-2	0	0	...	0	0	...	0
Group D	2006	2007	0	-1	0	0	...	0	0	...	0
Group D	2007	2007	1	0	0	0	...	1	0	...	0
Group D	2008	2007	1	1	0	0	...	0	1	...	0
Group D	2009	2007	1	2	0	0	...	0	0	...	0

Table 1: A Stylized Example

Goodman-Bacon decomposition: Goodman-Bacon(2019)

- What does the usually adopted TWFE DD estimation mean?
- The punchline of the Bacon decomposition theorem is that the two-way fixed effects estimator is a weighted average of all potential 2×2 DD estimates where weights are both based on group sizes and variance in treatment.
- A simple example.
 - Assume that there are three groups: an early treatment group (k), a group treated later (l), and a group that is never treated (U). Groups k and l are similar in that they are both treated but they differ in that k is treated earlier than l .
 - Say there are 5 periods, and k is treated in period 2. Then it spends 80% of its time under treatment, or 0.8. But let's say l is treated in period 4. Then it spends 40% of its time treated, or 0.4. I represent this time spent in treatment for a group as $D_k = 0.4$ and $D_l = 0.8$.

How many 2×2 DD are included in this picture?



a to b	b to a	c to a
a to c	b to c	c to b
a to U	b to U	c to U

9!!

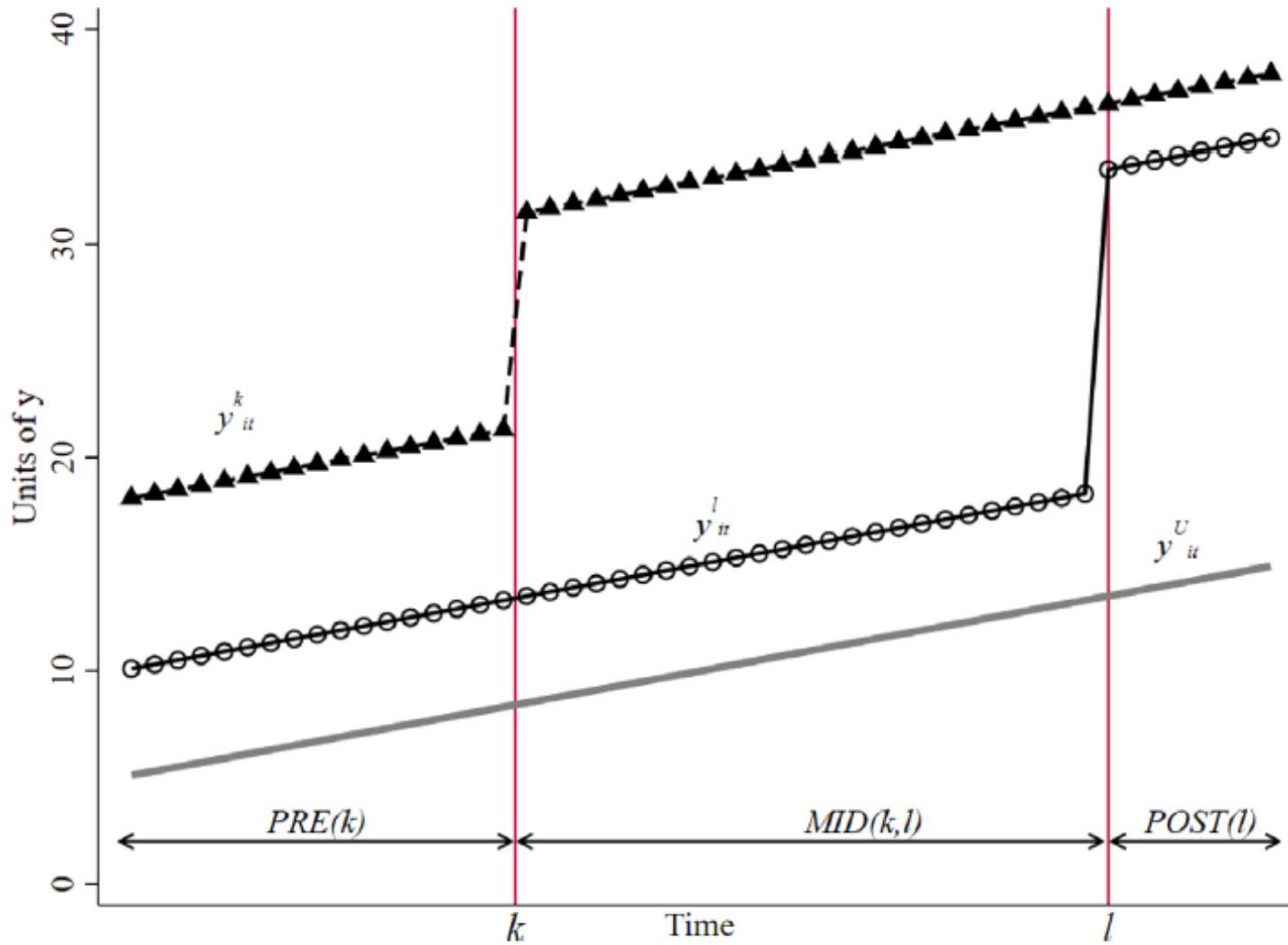


Fig. 1. Difference-in-Differences with variation in treatment Timing: Three groups. Notes: The figure plots outcomes in three timing groups: an untreated group, U ; an early treatment group, k , which receives a binary treatment at $k = \frac{34}{100}T$; and a late treatment group, ℓ , which receives the binary treatment at $\ell = \frac{85}{100}T$. The x -axis notes the three sub-periods: the pre-period for timing group k , $[1, k - 1]$, denoted by $PRE(k)$; the middle period when timing group k is treated and timing group ℓ is not, $[k, \ell - 1]$, denoted by $MID(k, \ell)$; and the post-period for timing group ℓ , $[\ell, T]$, denoted by $POST(\ell)$. The treatment effect is 10 in timing group k and 15 in timing group ℓ .

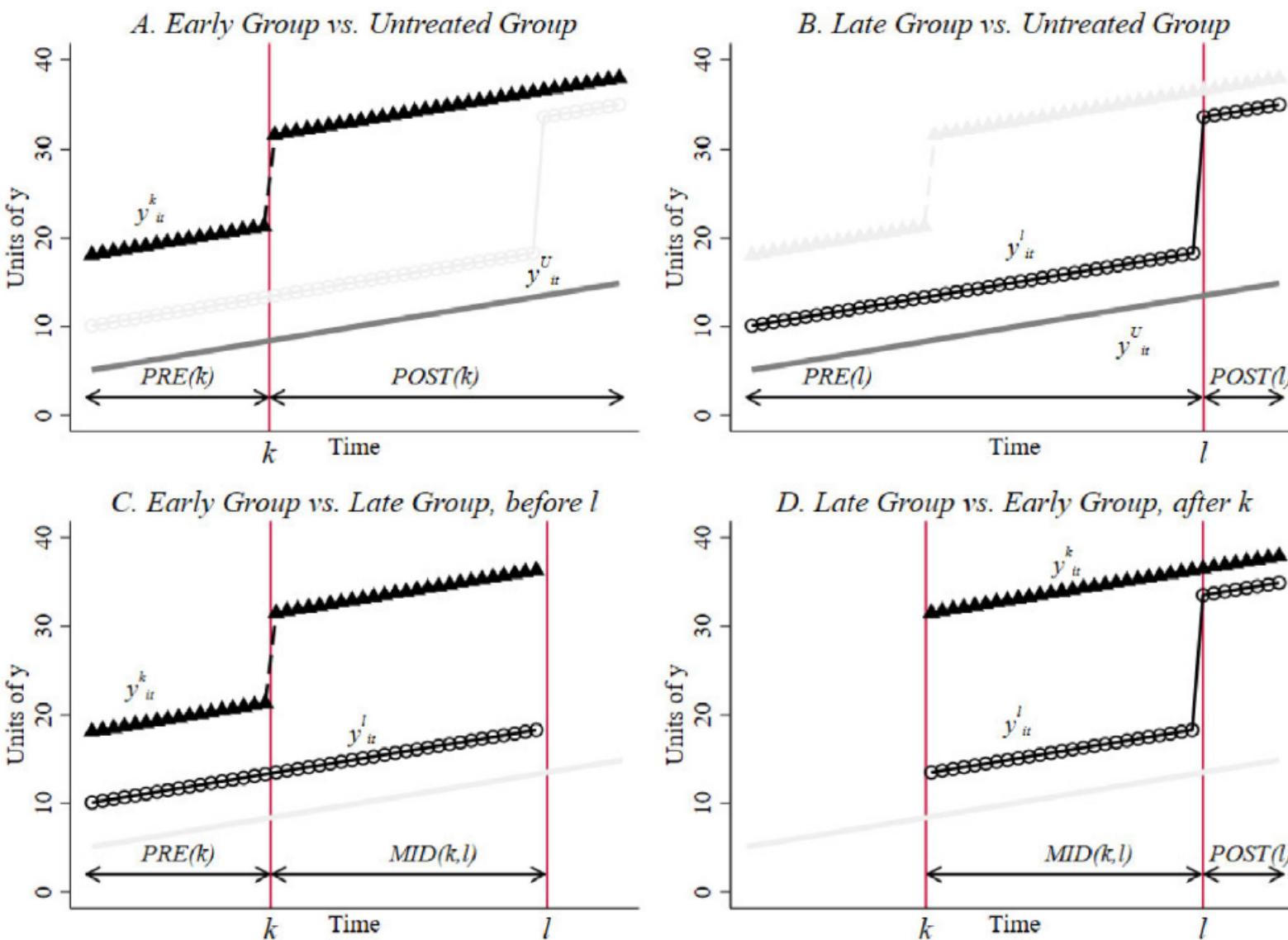


Fig. 2. The four simple (2x2) difference-in-differences estimates in the three group case. Notes: The figure plots outcomes for the subsamples that generate the four simple 2x2 difference-in-difference estimates in the three timing group case from Fig. 1. Each panel plots the data structure for one 2x2 DD. Panel A compares early treated units to untreated units ($\hat{\beta}_{kU}^{DD}$); panel B compares late treated units to untreated units ($\hat{\beta}_{eU}^{DD}$); panel C compares early treated units to late treated units during the late timing group's pre-period ($\hat{\beta}_{k\ell}^{DD,k}$); panel D compares late treated units to early treated units during the early timing group's post-period ($\hat{\beta}_{k\ell}^{DD,\ell}$). The treatment times mean that $\bar{D}_k = 0.67$ and $\bar{D}_\ell = 0.16$, 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.

It turns out that

The OLS estimate, $\hat{\beta}^{DD}$, in a two-way fixed-effects regression is a weighted average of all possible two-by-two DD estimators

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU} \hat{\beta}_{kU}^{2x2} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^k \hat{\beta}_{k\ell}^{2x2,k} + s_{k\ell}^\ell \hat{\beta}_{k\ell}^{2x2,\ell} \right].$$

where the 2x2 DD estimators are:

$$\hat{\beta}_{kU}^{2x2} \equiv \left(\bar{y}_k^{POST(k)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_U^{POST(k)} - \bar{y}_U^{PRE(k)} \right), \quad \text{or } |U$$

$$\hat{\beta}_{k\ell}^{2x2,k} \equiv \left(\bar{y}_k^{MID(k,\ell)} - \bar{y}_k^{PRE(k)} \right) - \left(\bar{y}_\ell^{MID(k,\ell)} - \bar{y}_\ell^{PRE(k)} \right),$$

$$\hat{\beta}_{k\ell}^{2x2,\ell} \equiv \left(\bar{y}_\ell^{POST(\ell)} - \bar{y}_\ell^{MID(k,\ell)} \right) - \left(\bar{y}_k^{POST(\ell)} - \bar{y}_k^{MID(k,\ell)} \right).$$

The weights are:

$$s_{kU} = \frac{(n_k + n_U)^2 \overbrace{n_{kU}(1 - n_{kU}) \bar{D}_k(1 - \bar{D}_k)}^{\hat{V}_{kU}^D}}{\hat{V}^D},$$

$$s_{k\ell}^k = \frac{(n_k + n_\ell)(1 - \bar{D}_\ell)^2 \overbrace{n_{k\ell}(1 - n_{k\ell}) \frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}^{\hat{V}_{k\ell}^{D,k}}}{\hat{V}^D},$$

$$s_{k\ell}^\ell = \frac{\left((n_k + n_\ell) \bar{D}_k \right)^2 n_{k\ell} (1 - n_{k\ell}) \overbrace{\frac{\bar{D}_\ell}{\bar{D}_k} \frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}}^{\hat{V}_{k\ell}^{D,\ell}}}{\hat{V}^D}.$$

$$\text{and } \sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} [s_{k\ell}^k + s_{k\ell}^\ell] = 1.$$

$$\bar{D}_k \equiv \sum_t 1\{t \geq k\} / T$$

$$n_k \equiv \sum_i 1\{t_i \geq k\} / N$$

The weights

$$s_{kU} = \frac{(n_k + n_U)^2 \overbrace{n_{kU}(1 - n_{kU}) \bar{D}_k(1 - \bar{D}_k)}^{\hat{V}_{kU}^D}}{\hat{V}^D},$$

$$s_{k\ell}^k = \frac{\overbrace{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell} (1 - n_{k\ell}) \frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}^{\hat{V}_{k\ell}^{D,k}}}{\hat{V}^D},$$

$$s_{k\ell}^\ell = \frac{\overbrace{\left((n_k + n_\ell) \bar{D}_k\right)^2 n_{k\ell} (1 - n_{k\ell}) \frac{\bar{D}_\ell}{\bar{D}_k} \frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}}^{\hat{V}_{k\ell}^{D,\ell}}}{\hat{V}^D}.$$

and $\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} [s_{k\ell}^k + s_{k\ell}^\ell] = 1$.

$$\bar{D}_k \equiv \sum_t 1\{t \geq k\} / T$$

$$n_k \equiv \sum_i 1\{t_i \geq k\} / N$$

Takeaways from the Bacon decomposition

- Two things immediately pop out of these weights
- First, notice how “group” variation matters, as opposed to unit-level variation. The Bacon decomposition shows that it’s group variation that twoway fixed effects is using to calculate that parameter you’re seeking. *The more states that adopted a law at the same time, the bigger they influence that final aggregate estimate itself.*
- The other thing that matters in these weights is *within-group* treatment variance.
 - Treatment variance is maximized at $\bar{D}=0.5$
 - being treated in the *middle* of the panel actually directly influences the numerical value you get when twoway fixed effects are used to estimate the ATT.
 - That therefore means lengthening or shortening the panel can actually change the point estimate purely by changing group treatment variance and nothing more.
 - Isn’t that kind of strange though? What criteria would we even use to determine the best length?

- Define any year-specific ATT as

$$ATT_k(\tau) = E[Y_{it}^1 - Y_{it}^0 \mid k, t = \tau]$$

- Define it over a time window W (e.g., A post-treatment window)

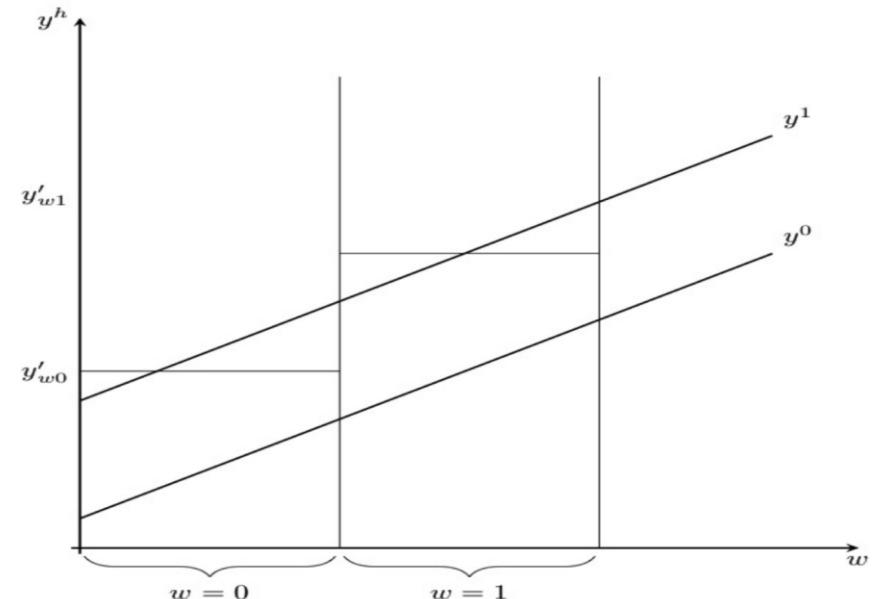
$$ATT_k(\tau) = E[Y_{it}^1 - Y_{it}^0 \mid k, \tau \in W]$$

- Define differences in average potential outcomes over time as

$$\Delta Y_k^h(W_1, W_0) = E[Y_{it}^h \mid k, W_1] - E[Y_{it}^h \mid k, W_0]$$

for $h = 0$ (i.e., Y^0) or $h = 1$ (i.e., Y^1)

- With trends, differences in mean potential outcomes is non-zero.

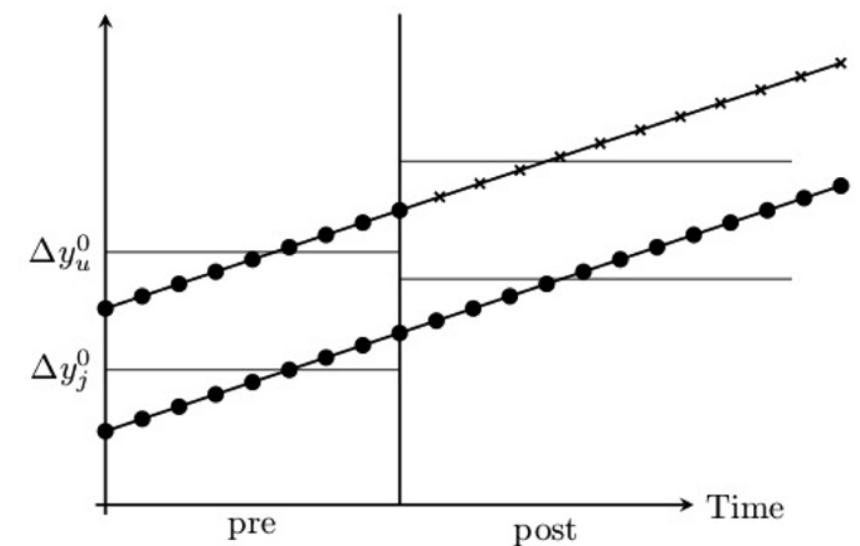


$$\begin{aligned}
\widehat{\delta}_{kU}^{2 \times 2} &= \left(E[Y_j | \text{Post}] - E[Y_j | \text{Pre}] \right) - \left(E[Y_u | \text{Post}] - E[Y_u | \text{Pre}] \right) \\
&= \underbrace{\left(E[Y_j^1 | \text{Post}] - E[Y_j^0 | \text{Pre}] \right) - \left(E[Y_u^0 | \text{Post}] - E[Y_u^0 | \text{Pre}] \right)}_{\text{Switching equation}} \\
&\quad + \underbrace{E[Y_j^0 | \text{Post}] - E[Y_j^0 | \text{Post}]}_{\text{Adding zero}} \\
&= \underbrace{E[Y_j^1 | \text{Post}] - E[Y_j^0 | \text{Post}]}_{\text{ATT}} \\
&\quad + \underbrace{\left[E[Y_j^0 | \text{Post}] - E[Y_j^0 | \text{Pre}] \right] - \left[E[Y_u^0 | \text{Post}] - E[Y_u^0 | \text{Pre}] \right]}_{\text{Non-parallel trends bias in } 2 \times 2 \text{ case}}
\end{aligned}$$

↓

$$\widehat{\delta}_{kU}^{2 \times 2} = ATT_{\text{Post},j} + \underbrace{\Delta Y_{\text{Post},\text{Pre},j}^0 - \Delta Y_{\text{Post},\text{Pre},U}^0}_{\text{Selection bias!}}$$

- The 2×2 DD can be expressed as the sum of the ATT itself plus a parallel trends assumption, and without parallel trends, the estimator is biased.



✖: counterfactual
●: factual

$$\widehat{\delta}_{kU}^{2 \times 2} = ATT_k \text{Post} + \Delta Y_k^0(\text{Post}(k), \text{Pre}(k)) - \Delta Y_U^0(\text{Post}(k), \text{Pre})$$

$$\widehat{\delta}_{kl}^{2 \times 2} = ATT_k(MID) + \Delta Y_k^0(MID, \text{Pre}) - \Delta Y_l^0(MID, \text{Pre})$$

$$\widehat{\delta}_{lk}^{2 \times 2} = ATT_{l, \text{Post}(l)}$$

$$+ \underbrace{\Delta Y_l^0(\text{Post}(l), MID) - \Delta Y_k^0(\text{Post}(l), MID)}_{\text{Parallel-trends bias}}$$

$$- \underbrace{(ATT_k(\text{Post}) - ATT_k(MID))}_{\text{Heterogeneity in time bias!}}$$

- The first line is the ATT that we desperately hope to identify.
- The selection bias zeroes out insofar as Y_0 for k and l has the same parallel trends from mid to post period.
- And the treatment effects bias in the third line zeroes out so long as there are constant treatment effects for a group over time. But if there is heterogeneity in time for a group, then the two ATT terms will not be the same, and therefore will not zero out.

In sum, the decomposition formula for DD is:

$$\widehat{\delta}^{DD} = \sum_{k \neq U} s_{kU} \widehat{\delta}_{kU}^{2 \times 2} + \sum_{k \neq U} \sum_{l > k} s_{kl} \left[\mu_{kl} \widehat{\delta}_{kl}^{2 \times 2,k} + (1 - \mu_{kl}) \widehat{\delta}_{kl}^{2 \times 2,l} \right]$$

$$\widehat{\delta}_{kU}^{2 \times 2} = ATT_k(\text{Post}) + \Delta Y_l^0(\text{Post}, \text{Pre}) - \Delta Y_U^0(\text{Post}, \text{Pre})$$

$$\widehat{\delta}_{kl}^{2 \times 2,k} = ATT_k(\text{Mid}) + \Delta Y_l^0(\text{Mid}, \text{Pre}) - \Delta Y_l^0(\text{Mid}, \text{Post})$$

$$\begin{aligned} \widehat{\delta}_{lk}^{2 \times 2,l} = & ATT_l(\text{Post}(l)) + \Delta Y_l^0(\text{Post}(l), \text{MID}) - \Delta Y_k^0(\text{Post}(l), \text{MID}) \\ & - (ATT_k(\text{Post}) - ATT_k(\text{Mid})) \end{aligned}$$



$$p \lim_{n \rightarrow \infty} \widehat{\delta}_{n \rightarrow \infty}^{DD} = VWATT + VWCT - \Delta ATT$$

- **Variance weighted ATT**

$$\begin{aligned}
 VWATT &= \sum_{k \neq U} \sigma_{kU} ATT_k(\text{Post}(k)) \\
 &\quad + \sum_{k \neq U} \sum_{l > k} \sigma_{kl} \left[\mu_{kl} ATT_k(\text{MID}) + (1 - \mu_{kl}) ATT_l(\text{POST}(l)) \right]
 \end{aligned}$$

- Notice that the VWATT simply contains the three ATTs identified above, each of which was weighted by the weights contained in the decomposition formula. While these weights sum to one, that weighting is irrelevant if the ATT are identical.

- Variance weighted common trends. VWCT stands for variance weighted common trends. This is just the collection of non-parallel-trends biases

$$\begin{aligned}
 VWCT = & \sum_{k \neq U} \sigma_{kU} \left[\Delta Y_k^0(\text{Post}(k), \text{Pre}) - \Delta Y_U^0(\text{Post}(k), \text{Pre}) \right] \\
 & + \sum_{k \neq U} \sum_{l > k} \sigma_{kl} \left[\mu_{kl} \{ \Delta Y_k^0(\text{Mid}, \text{Pre}(k)) - \Delta Y_l^0(\text{Mid}, \text{Pre}(k)) \} \right. \\
 & \quad \left. + (1 - \mu_{kl}) \{ \Delta Y_l^0(\text{Post}(l), \text{Mid}) - \Delta Y_k^0(\text{Post}(l), \text{Mid}) \} \right]
 \end{aligned}$$

- Goodman-Bacon [2019] shows us is that technically you don't need identical trends because the weights can make it hold even if we don't have exact parallel trends.

ATT heterogeneity within time bias.

$$\Delta ATT = \sum_{k \neq U} \sum_{l > k} (1 - \mu_{kl}) [ATT_k(\text{Post}(l)) - ATT_k(\text{Mid})]$$

- Heterogeneity in the ATT has two interpretations: you can have heterogeneous treatment effects across groups, and you can have heterogeneous treatment effects within groups over time. The ATT is concerned with the latter only.
- Time-varying treatment effects, even if they are identical across units, generate cross-group heterogeneity because of the differing post-treatment windows, and the fact that earlier-treated groups are serving as controls for later-treated groups.

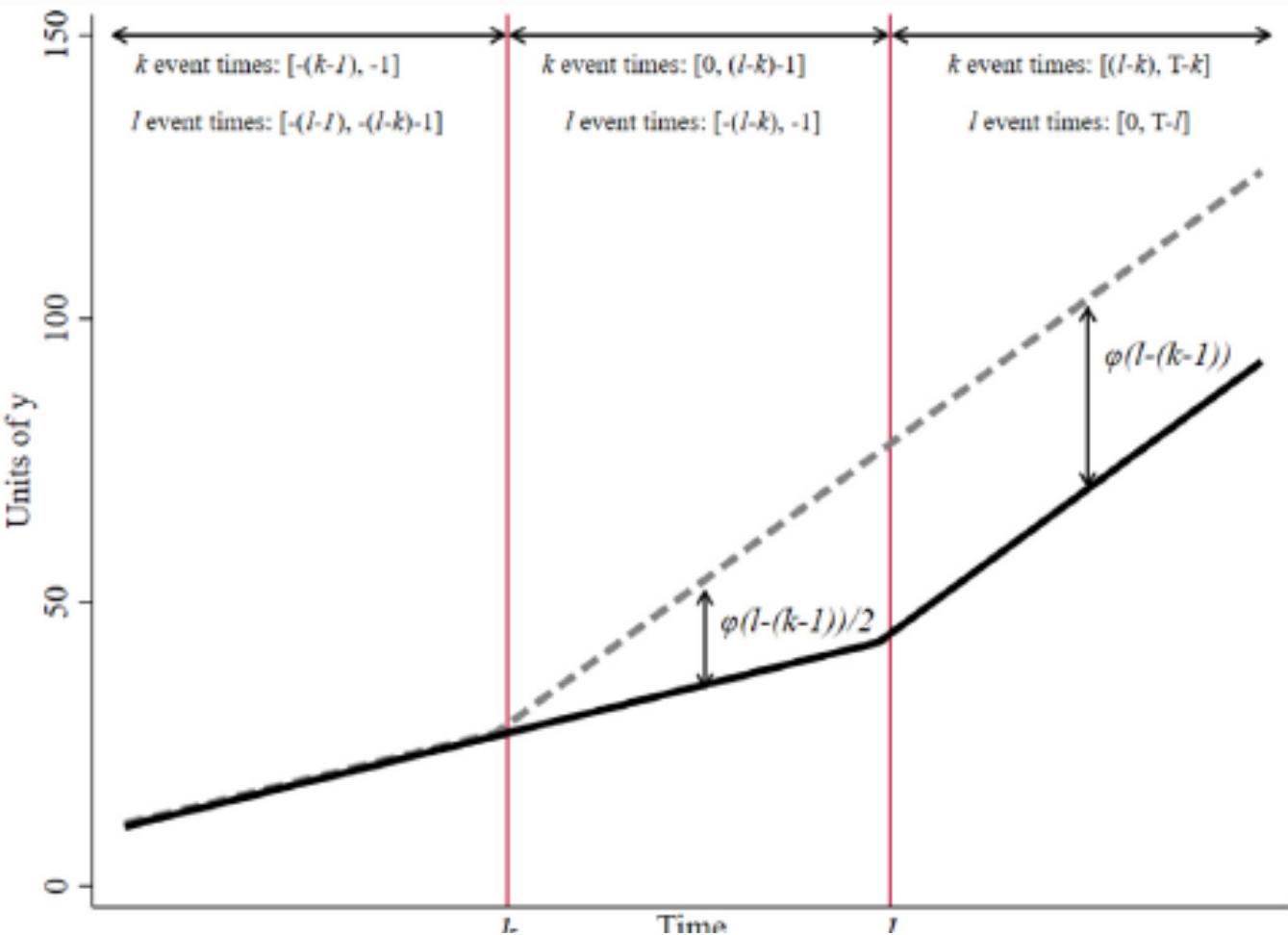


Fig. 3. Difference-in-Differences estimates with variation in timing are biased when treatment effects vary over time. Notes: The figure plots a stylized example of a timing-only DD set up with a treatment effect that is a trend-break rather than a level shift (see Meer and West, 2016). The trend-break effect equals $\phi \cdot (t - t_i + 1)$. The top of the figure notes which event-times lie in the PRE(k), MID(k, ℓ), and POST(ℓ) periods for each unit. The figure also notes the average difference between timing groups in each of these periods. In the MID(k, ℓ) period, outcomes differ by $\frac{\phi}{2}(\ell - (k - 1))$ on average. In the POST(ℓ) period, however, outcomes had already been growing in the early group for $\ell - k$ periods, and so they differ by $\phi(\ell - (k - 1))$ on average. The 2x2 DD that compares the later-treated group to the earlier-treated group is biased and, in the linear trend-break case, weakly negative despite a positive and growing treatment effect.

- When is this a problem?
- It's a problem if there are a lot of those 2×2 s or if their weights are large. If they are negligible portions of the estimate, then even if it exists, then given their weights are small (as group shares are also an important piece of the weighting not just the variance in treatment) the bias may be small.
- Time-varying treatment effects cause bias when you use a TWFE regression to analyze data from a staggered adoption design

Stevenson and Wolfers' (2006)

- Stevenson and Wolfers' (2006) examines the no-fault divorce reforms and female suicide.
- Unilateral (or no-fault) divorce allowed either spouse to end a marriage, redistributing property rights and bargaining power relative to fault-based divorce regimes.
- Stevenson and Wolfers exploit “the natural variation resulting from the different timing of the adoption of unilateral divorce laws” in 37 states from 1969–1985 (see [Table 1](#)) using the “remaining fourteen states as controls” to evaluate the effect of these reforms on female suicide rates.

Table 1

The no-fault divorce rollout: Treatment times, timing group sizes, and treatment shares.

No-fault divorce year (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 to 1996.

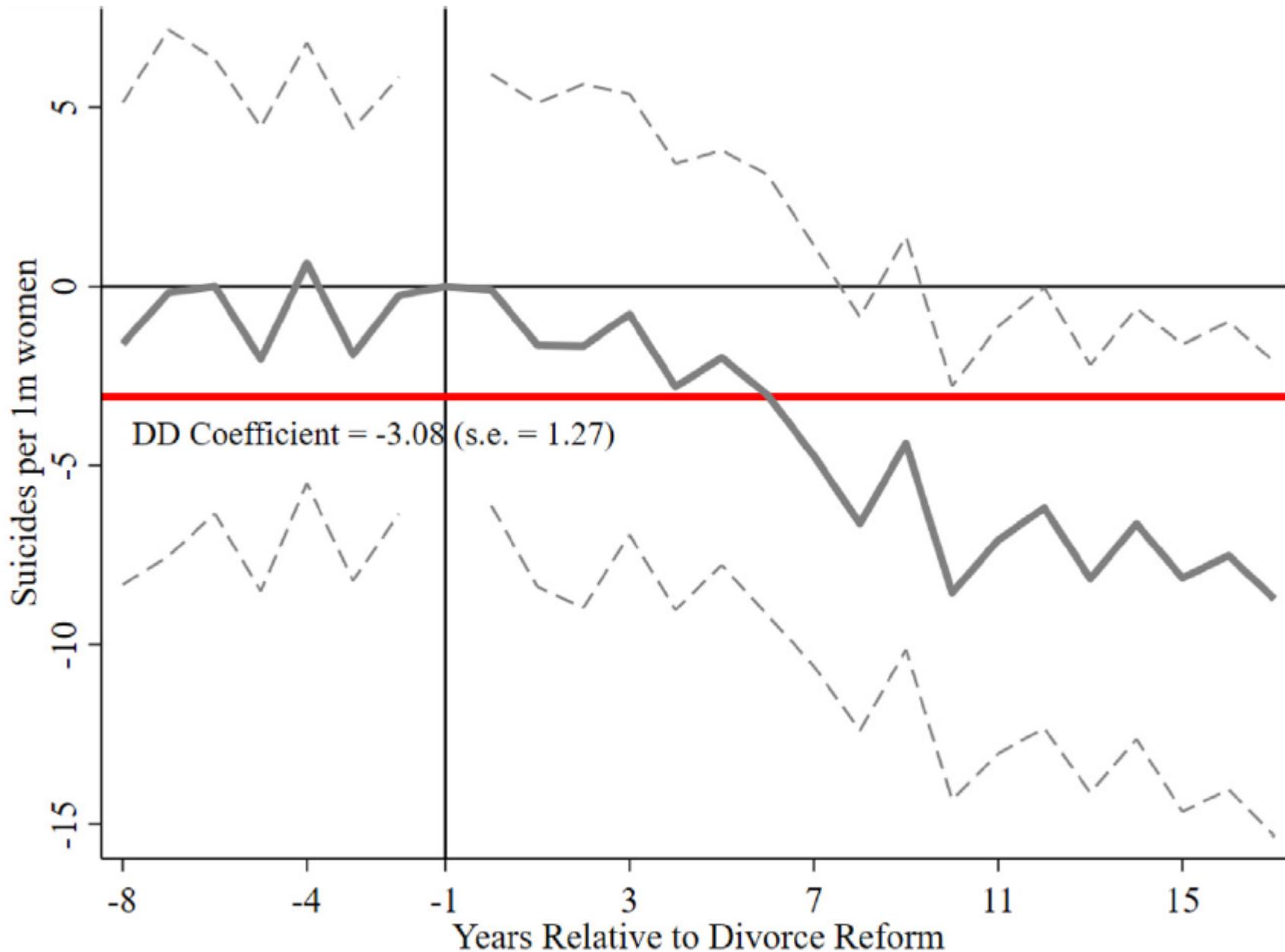


Fig. 5. Event-study and difference-in-differences estimates of the effect of no-fault divorce on female suicide: Replication of Stevenson and Wolfers (2006). Notes: The figure plots event-study estimates from the two-way fixed effects regression equation, along with the DD coefficient. The specification does not include other controls and does not weigh by population.

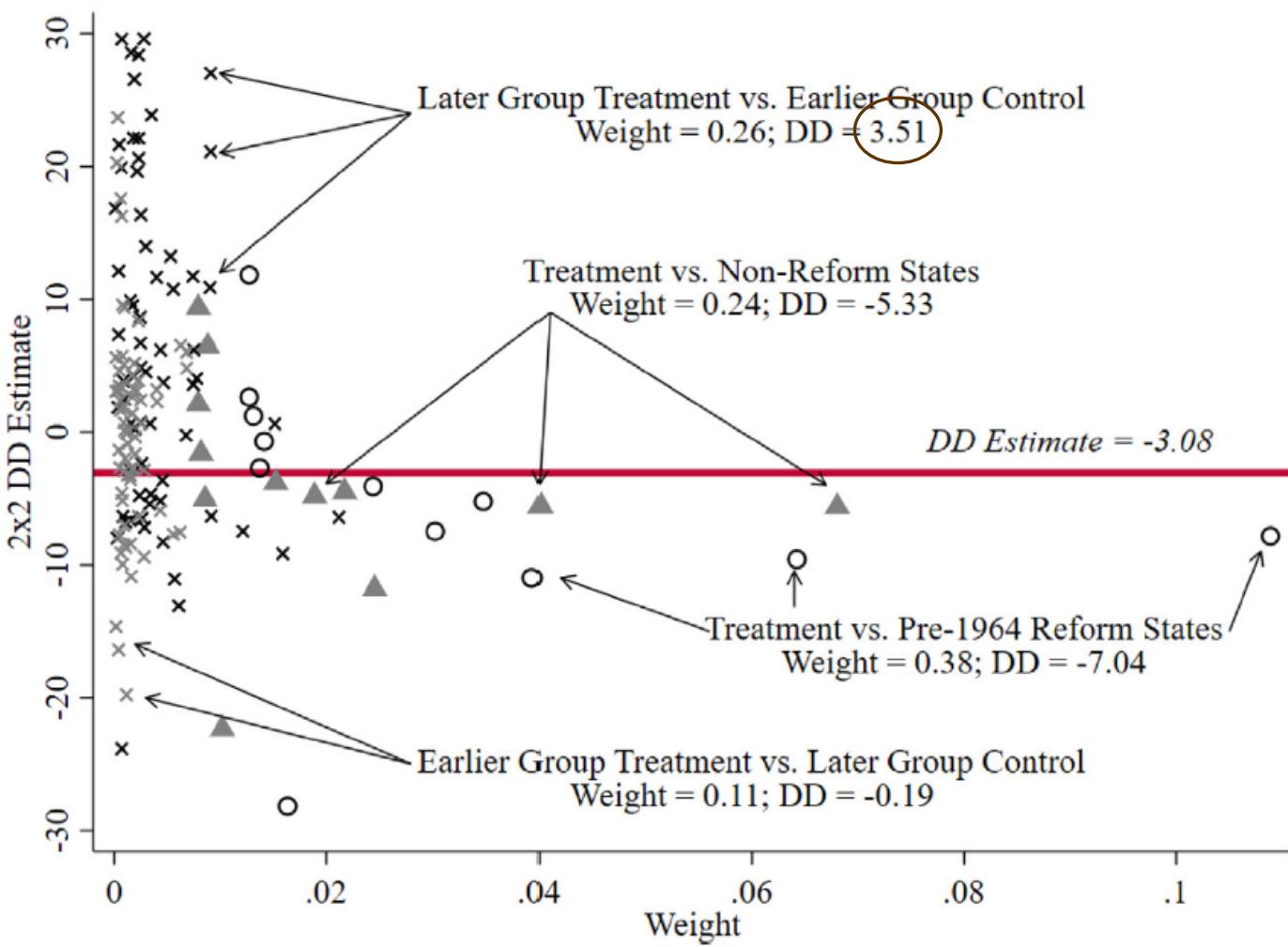


Fig. 6. Difference-in-differences decomposition for unilateral divorce and female suicide.

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.

Castle-doctrine statutes and homicides: Cheng and Hoekstra (2013)

- Cheng and Hoekstra (2013) evaluated the impact that a gun reform had on violence
- In 2005, Florida made a reformation: before the reform, lethal self-defense was only legal inside the home, a new law, “Stand Your Ground,” had extended that right to other public places.
- Between 2000 and 2010, twenty-one states explicitly expanded the castle-doctrine statute by extending the places outside the home where lethal force could be legally used.
 - After these reforms, though, victims no longer had a duty to retreat in public places if they felt threatened; they could retaliate in lethal self-defense.
 - Assumed to be reasonably afraid
 - Civil liability for those acting under these expansions was also removed.
- From an economic perspective, these reforms lowered the cost of killing someone.
 - The reforms may have, in other words, caused homicides to rise.
 - deterrence of violence -- concealed-carry laws

- A difference-in-differences design for their project where the castle doctrine law was the treatment and timing was differential across states.

$$Y_{it} = \alpha + \delta D_{it} + \gamma X_{it} + \sigma_i + \tau_t + \varepsilon_{it}$$

- where D_{it} is the treatment parameter. it's a variable ranging from 0 to 1, because some states get the law change mid-year. So if they got the law in July, then D_{it} equals 0 before the year of adoption, 0.5 in the year of adoption and 1 thereafter.
- They used the FBI Uniform Crime Reports Summary Part I files from 2000 to 2010. The FBI Uniform Crime Reports is a harmonized data set on eight “index” crimes collected from voluntarily participating police agencies across the country. Crimes were converted into rates, or “offenses per 100,000 population.”

Falsification
Tests: The effect
of castle
doctrine laws on
larceny and
motor vehicle
theft.

	OLS – Weighted by State Population					
	1	2	3	4	5	6
Panel A. Larceny	Log(Larceny Rate)					
Castle Doctrine Law	0.00300 (0.0161)	-0.00600 (0.0147)	-0.00910 (0.0139)	-0.0858 (0.0139)	-0.00401 (0.0128)	-0.00284 (0.0180)
0 to 2 years before adoption of castle doctrine law				0.00112 (0.0105)		
Observation	550	550	550	550	550	550
Panel B. Motor Vehicle Theft	Log(Motor Vehicle Theft Rate)					
Castle Doctrine Law	0.0517 (0.0563)	-0.0389 (0.448)	-0.0252 (0.0396)	-0.0294 (0.0469)	-0.0165 (0.0354)	-0.00708 (0.0372)
0 to 2 years before adoption of castle doctrine law					-0.00896 (0.0216)	
Observation	550	550	550	550	550	550
State and year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-year fixed effects		Yes	Yes	Yes	Yes	Yes
Time-varying controls			Yes	Yes	Yes	Yes
Controls for larceny or motor theft					Yes	
State-specific linear time trends						Yes

The deterrence effects of castle-doctrine laws:
Burglary, robbery, and aggravated assault.

OLS – Weighted by State Population						
	1	2	3	4	5	6
Panel A. Burglary	Log(Burglary Rate)					
Castle-doctrine law	0.0780***	0.0290	0.0223	0.0181	0.0327*	0.0237
0 to 2 years before adoption of castle-doctrine law	(0.0255)	(0.0236)	(0.0223)	(0.0265)	(0.0165)	(0.0207)
				–0.009606		
				(0.0133)		
Panel B. Robbery	Log(Robbery Rate)					
Castle-doctrine law	0.0408	0.0344	0.0262	0.0197	0.0376**	0.0515*
0 to 2 years before adoption of castle-doctrine law	(0.0254)	(0.0224)	(0.0229)	(0.0257)	(0.0181)	(0.0274)
				–0.0138		
				(0.0153)		
Panel C. Aggravated Assault	Log(Aggravated Assault Rate)					
Castle-doctrine law	0.0434	0.0397	0.0372	0.0330	0.0424	0.0414
0 to 2 years before adoption of castle-doctrine law	(0.0387)	(0.0407)	(0.0319)	(0.0367)	(0.0291)	(0.0285)
				–0.00897		
				(0.0147)		
Observation	550	550	550	550	550	550
State and year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-year fixed effects		Yes	Yes	Yes	Yes	Yes

Raw data of log
homicides per 100,000
for Florida versus never-
treated control states.

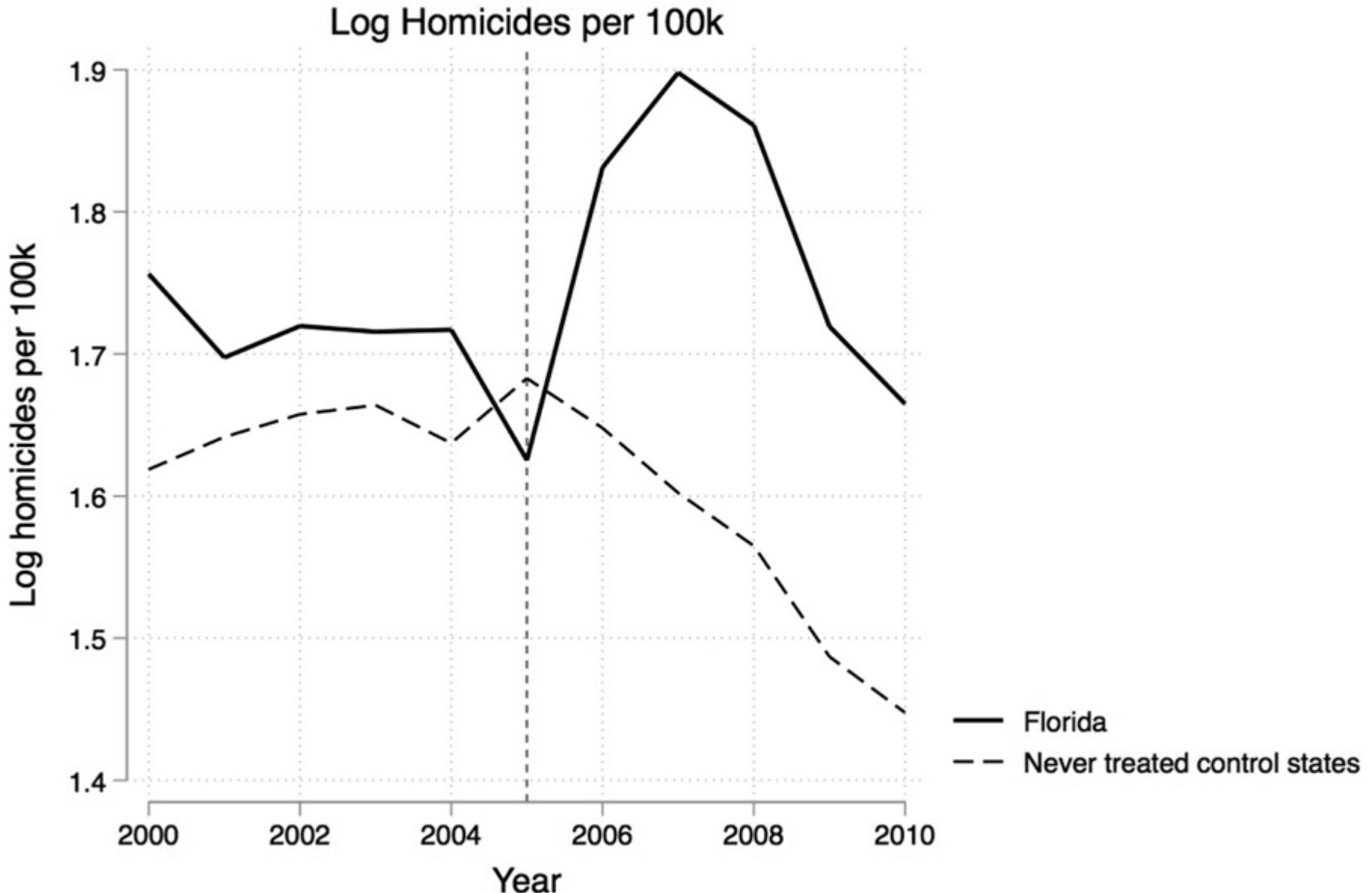


Table 79. The effect of castle-doctrine laws on homicide.

Panel A. Homicide OLS-Weights	Log(Homicide rate)					
	1	2	3	4	5	6
Castle-doctrine law	0.0801** (0.0342)	0.0946*** (0.0279)	0.0937*** (0.0290)	0.0955** (0.0367)	0.0985*** (0.0299)	0.100** (0.0388)
0 to 2 years before adoption of castle-doctrine law					0.00398 (0.0222)	
Observation	550	550	550	550	550	550
State and year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-year fixed effects		Yes	Yes	Yes	Yes	Yes
Time-varying controls			Yes	Yes	Yes	Yes
Controls for larceny or motor theft					Yes	
State-specific linear time trends						Yes

Notes: Each column in each panel represents a separate regression. The unit of observation is state-year. Robust standard errors are clustered at the state level. Time-varying controls include policing and incarceration rates, welfare and public assistance spending, median income, poverty rate, unemployment rate, and demographics. *Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

Log Murder Rate

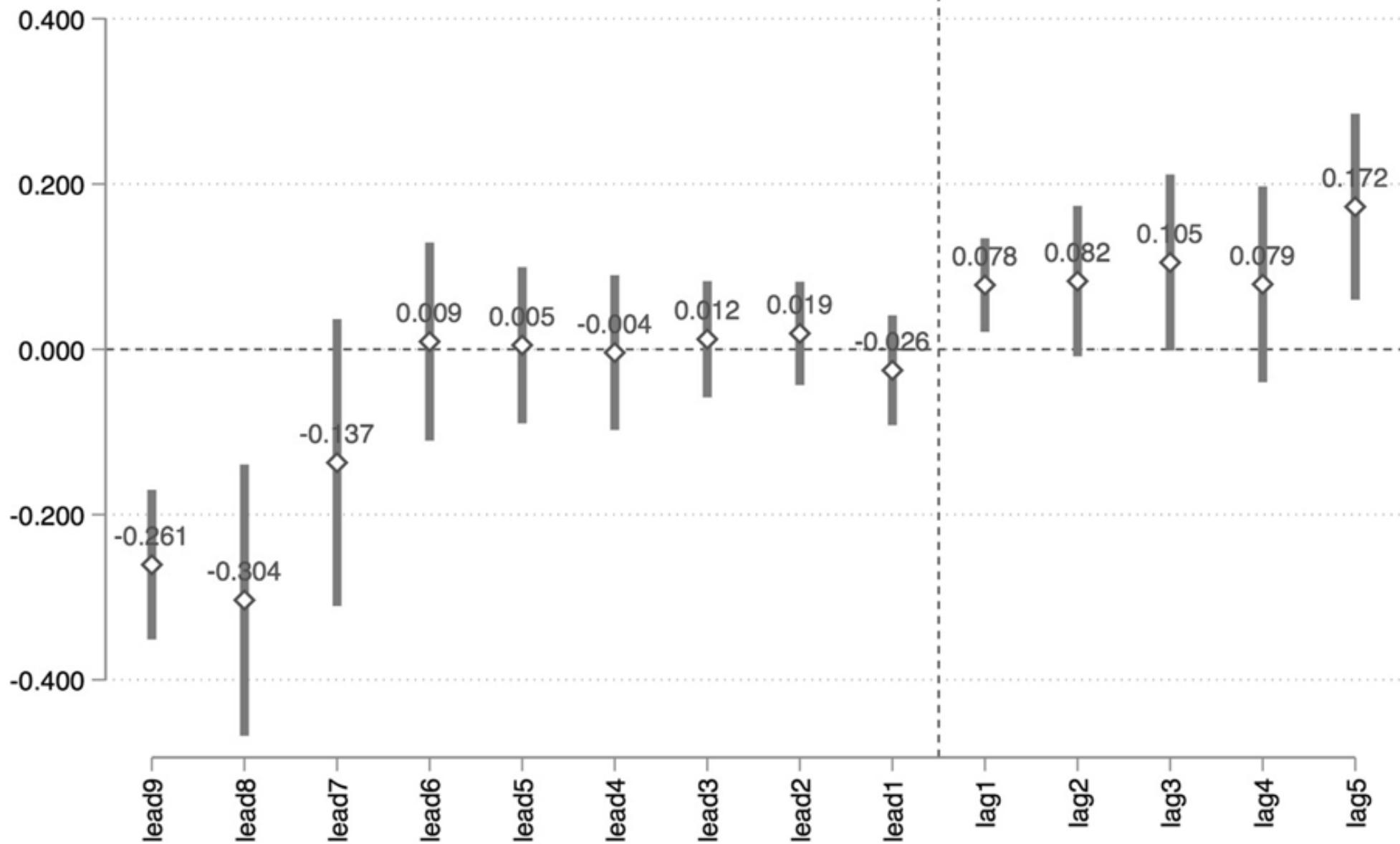


Table 81. Bacon decomposition example.

DD Comparison	Weight	Avg DD Est
Earlier T vs. Later C	0.077	-0.029
Later T vs. Earlier C	0.024	0.046
T vs. Never treated	0.899	0.078

Dep var	Log(homicide rate)
Castle-doctrine law	0.069 (0.034)

Taking these weights, let's just double check that they do indeed add up to the regression estimate we just obtained using our two-way fixed-effects estimator.¹⁹

$$di(0.077 * -0.029) + (0.024 * 0.046) + (0.899 * 0.078) = 0.069$$

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.

Stata package

Coefplot

ddtiming