5. Difference-in-Differences

LPO 8852: Regression II

Sean P. Corcoran

Matching recap

Matching methods seek to construct a comparison group where the conditional independence assumption is satisfied:

$$Y(0), Y(1) \perp\!\!\!\perp D|X$$

That is, conditional on X (or a one-number summary like the propensity score), potential outcomes are independent of treatment status D. If this holds, we can use mean outcomes of the matched comparison group as a stand-in for the treated group counterfactual.

$$\underbrace{E[Y(0)|D=1,X]}_{\text{unobserved}} = \underbrace{E[Y(0)|D=0,X]}_{\text{matched comparison group}}$$

Matching recap

Challenges:

- The conditional independence assumption (selection on observables) is strong! In most settings we have to be concerned about selection on *unobservables*.
- Constructing matched samples is somewhat of an art, and results may be sensitive to specification of the matching model.
- We are typically comparing outcomes at one point in time (e.g., post treatment).

Difference-in-differences

Difference-in-differences (DD) is a research design that most often contrasts *changes over time* for treated and untreated groups. The approach is fruitfully applied to **natural experiments**, settings in which an external force naturally assigns units into treatment and control groups.

Figure: Scott Cunningham's (of *Mixtape* fame) bumper sticker

DD models are often estimated with *panel* data but can also be used with *repeated cross-sections.*

Natural experiments

Examples of natural experiments:

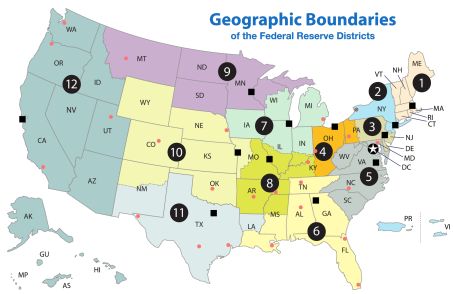
- John Snow's cholera study (1855)
- Natural and other disasters (hurricanes, earthquakes, COVID, 9/11)
- Policy implementation (e.g., graduated drivers license laws, EZ Pass)
- Investments (e.g., school construction)
- Idiosyncratic policy rules (e.g., class size maximum)
- Idiosyncratic differences in location (opposite sides of boundaries)
- Date of birth and eligibility rules

Many natural experiments are analyzed using DD, others are better suited to tools we'll see later.

Federal Reserve policy and bank failures

Mastering 'Metrics considers the effect of FRB lending to troubled banks during the Great Depression. Did this policy stem bank failures?

A potential natural experiment: in 1930 a large Southern bank collapsed, putting others at risk. Did a Fed "easy money" policy help? Note that Mississippi is served by two Fed districts:



The 6th (Atlanta) favored lending to troubled banks while the 8th (St. Louis) did not. "Treatment" is being in the 6th after the 1930 failure.

Federal Reserve policy and bank failures

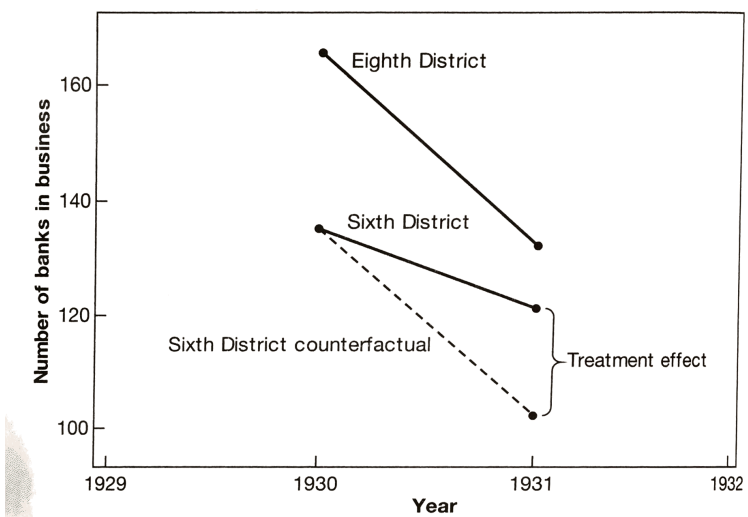
Consider two comparisons within Mississippi:

- Cross-sectional: the number of operating banks in the 6th and 8th districts in 1931
- Interrupted time series (ITS): the number of operating banks in the 6th, before and after 1930

Note a better ITS design would have more data points than this, to establish a trend.

Mastering 'Metrics Fig 5.1

FIGURE 5.1
Bank failures in the Sixth and Eighth Federal Reserve Districts



Notes: This figure shows the number of banks in operation in Mississippi in the Sixth and Eighth Federal Reserve Districts in 1930 and 1931. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district in this period as did in the Eighth.

Federal Reserve policy and bank failures

The cross sectional comparison in 1931 suggests worse outcomes in the 6th district:

$$Y_{6,1931} - Y_{8,1931} = 121 - 132 = -11$$

The **first difference** for the 6th district also shows a decline in the number of banks:

$$Y_{6,1931} - Y_{6,1930} = 121 - 135 = -14$$

Neither looks good for the 6th district "easy money" policy.

Federal Reserve policy and bank failures

Problems:

- The cross sectional comparison fails to recognize that the 6th district had fewer banks *before* the crisis.
- The interrupted time series is unable to differentiate between a treatment effect in the 6th district (if any) and secular changes in the financial sector. (Note the 8th district also experienced declines. This was the Great Depression, after all).

Federal Reserve policy and bank failures

Under the assumption that trends in the 8th district represent what *would have happened* in the 6th district in the absence of treatment, we can contrast *changes* in the two districts, or the **difference-in-differences**:

$$\delta_{DD} = \underbrace{(Y_{6,1931} - Y_{6,1930})}_{\text{Change in 6th District}} - \underbrace{(Y_{8,1931} - Y_{8,1930})}_{\text{Change in 8th District}}$$

$$\delta_{DD} = (121 - 135) - (132 - 165) = +19$$

The second term in the above expression is the **second difference**. The decline was actually much *worse* in the 8th district, suggesting a positive treatment effect for the 6th.

Federal Reserve policy and bank failures

An equivalent way to write δ_{DD}:

$$\delta_{DD} = \underbrace{(Y_{6,1931} - Y_{8,1931})}_{\text{Difference post}} - \underbrace{(Y_{6,1930} - Y_{8,1930})}_{\text{Difference pre}}$$

Writing δ_{DD} this way makes it clear we are "netting out" pre-existing differences between the two groups.

Note in this example δ_{DD} was calculated using only four numbers (counts of banks in the 6th and 8th, pre and post).

Card & Krueger (1994)

A classic DD study of the impact of the minimum wage on fast food employment (an industry likely to be affected by the minimum wage).

- NJ increased its minimum wage in April 1992, PA did not.
- Card & Krueger collected data on employment at fast food restaurants in NJ and Eastern PA before and after the minimum wage hike.

See next figure: the minimum wage increase had a "first stage." That is, it led to higher starting wages in NJ. (This is important—if the minimum wage were not binding, it wouldn't make for a very interesting study).

Card & Krueger (1994)

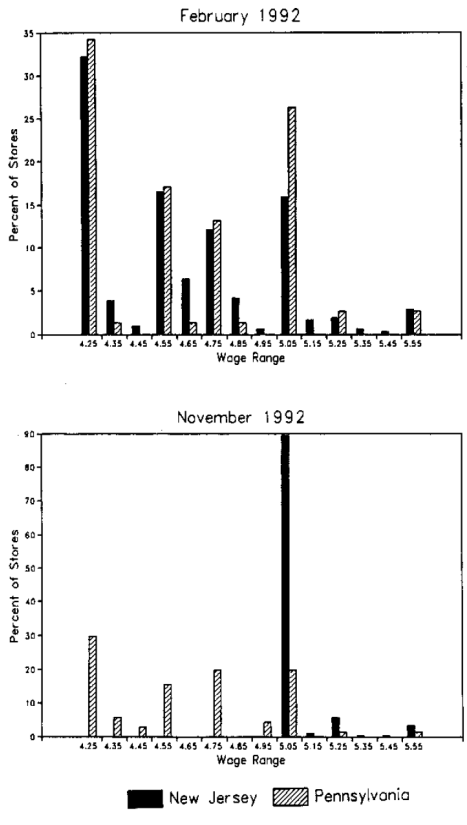


FIGURE 1. DISTRIBUTION OF STARTING WAGE RATES

Card & Krueger (1994)

Main result (portion of Table 3 in C&K):

	Stores by State		
	PA	*NJ*	*NJ* − *PA*
FTE before	23.3	20.44	-2.89
	(1.35)	(-0.51)	(1.44)
FTE after	21.15	21.03	-0.14
	(0.94)	(0.52)	(1.07)
Change in mean FTE	2.16	0.59	**2.76**
	(1.25)	(0.54)	**(1.36)**

Standard errors in parentheses. FTE=full time equivalent employees.

Mean employment fell in PA and *rose* in NJ, for $\delta_{DD} = 2.76$. A surprising result to many economists who expected to see a reduction in employment following an increase in the minimum wage.

2x2 difference-in-differences

The two examples thus far are the simplest type of a difference-in-difference:

- Two groups: treated and untreated
- Two time periods: pre and post, before and after
- Treated units are all treated at the same time

Difference-in-differences estimation

Under what conditions might the difference-in-differences design estimate a *causal parameter*? What causal parameter is it estimating?

Let's return to the potential outcomes framework, applying it to a 2x2 DD example.

Difference-in-differences estimation

Suppose that—in the absence of treatment—the potential outcome for individual i at time t is given by:

$$Y_{it}(0) = \gamma_i + \lambda_t$$

In the *presence* of treatment, the potential outcome for individual i at time t is:

$$Y_{it}(1) = \gamma_i + \delta + \lambda_t$$

Note: portions of this section were drawn from Jakiela & Ozier's excellent ECON 626 lecture notes from the University of Maryland.

Difference-in-differences estimation

$$Y_{it}(0) = \gamma_i + \lambda_t$$
$$Y_{it}(1) = \gamma_i + \delta + \lambda_t$$

A few things to note in this example:

- There are fixed individual differences represented by γ_i
- The time-specific factor λ_t is the same for all individuals
- The impact of the treatment δ is assumed to be the same for all individuals, and does not vary over time

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

Difference-in-differences estimation

In this framework individuals can self-select into treatment, and selection can be related to γ_i.

- Let $D_i = 1$ for those who—*at any point*—are treated
- Let $D_i = 0$ for those who are never treated

Note this indicator is not subscripted with a t. It is important to note that we are grouping i by whether they are *ever* treated, since we observe them in treated/untreated states at different points in time.

Assume for simplicity two time periods, "pre" ($t = 0$) and "post" ($t = 1$), where treatment occurs for the $D_i = 1$ group in $t = 1$.

Difference-in-differences estimation

The causal estimand of interest here is:

$$\begin{aligned} ATT &= \underbrace{E[Y_{it}(1)|D_i = 1, t = 1]}_{\text{observed}} - \underbrace{E[Y_{it}(0)|D_i = 1, t = 1]}_{\textit{unobserved}} \\ &= E[\gamma_i|D_i = 1] + \delta + \lambda_t - E[\gamma_i|D_i = 1] - \lambda_t \\ &= \delta \end{aligned}$$

That is, the mean difference in outcomes in the treated and untreated state—in the "post" period—among those who are treated.

Difference-in-differences estimation

Of course, we can't observe the same i in two different states (0 and 1) in the same period t. Suppose instead we compare the $D_i = 1$ and $D_i = 0$ groups in time period 1:

$$\underbrace{E[Y_{it}(1)|D_i = 1, t = 1]}_{E[\gamma_i|D_i=1]+\delta+\lambda_1} - \underbrace{E[Y_{it}(0)|D_i = 0, t = 1]}_{E[\gamma_i|D_i=0]+\lambda_1}$$

$$= \delta + E[\gamma_i|D_i = 1] - E[\gamma_i|D_i = 0]$$

If treatment were randomly assigned, the $E[\gamma_i]$ would not vary with D_i. However, if there is selection into D related to the fixed characteristics of individuals, then $E[\gamma_i|D_i = 1] \neq E[\gamma_i|D_i = 0]$. The δ is not identified.

Difference-in-differences estimation

Alternatively we might restrict our attention to the $D_i = 1$ group and do a pre-post comparison from time 0 to time 1:

$$\underbrace{E[Y_{it}(1)|D_i = 1, t = 1]}_{E[\gamma_i|D_i=1]+\delta+\lambda_1} - \underbrace{E[Y_{it}(0)|D_i = 1, t = 0]}_{E[\gamma_i|D_i=1]+\lambda_0}$$

$$= \delta + \lambda_1 - \lambda_0$$

This is the first difference or interrupted time series (ITS). Unfortunately, δ is still not identified, since this difference reflects both the impact of the program and the time trend.

Difference-in-differences estimation

Consider now the pre-post comparison for the $D_i = 0$ group:

$$\underbrace{E[Y_{it}(0)|D_i = 0, t = 1]}_{E[\gamma_i|D_i=0]+\lambda_1} - \underbrace{E[Y_{it}(0)|D_i = 0, t = 0]}_{E[\gamma_i|D_i=0]+\lambda_0}$$

$$= \lambda_1 - \lambda_0$$

The comparison group allows us to estimate the time trend.

Difference-in-differences estimation

Now subtract the pre-post comparison for the *untreated* group from the pre-post comparison for the *treated* group:

$$\underbrace{E[Y_{it}(1)|D_i = 1, t = 1]}_{E[\gamma_i|D_i=1]+\delta+\lambda_1} - \underbrace{E[Y_{it}(0)|D_i = 1, t = 0]}_{E[\gamma_i|D_i=1]+\lambda_0} -$$

$$(\underbrace{E[Y_{it}(0)|D_i = 0, t = 1]}_{E[\gamma_i|D_i=0]+\lambda_1} - \underbrace{E[Y_{it}(0)|D_i = 0, t = 0]}_{E[\gamma_i|D_i=0]+\lambda_0})$$

$$= (\delta + \lambda_1 - \lambda_0) - (\lambda_1 - \lambda_0)$$

$$= \delta$$

The difference-in-differences estimator recovers the ATT. The **parallel trends assumption** is critical here.

Difference-in-differences estimation

To see this a different way, the ATT again is:

$$ATT = \underbrace{E[Y(1)|D_i = 1, t = 1]}_{\text{observed}} - \underbrace{E[Y(0)|D_i = 1, t = 1]}_{\textit{unobserved}}$$

The DD estimates:

$$\underbrace{E[Y(1)|D = 1, t = 1] - E[Y(0)|D = 1, t = 0]}_{\text{change over time for treated group}}$$

$$- \underbrace{(E[Y(0)|D = 0, t = 1] - E[Y(0)|D = 0, t = 0])}_{\text{change over time for untreated group}}$$

From this, subtract and add the *unobserved* term from above right:

Difference-in-differences estimation

$$E[Y(1)|D=1,t=1] - E[Y(0)|D=1,t=0] - \underbrace{E[Y(0)|D_i=1,t=1]}_{\text{unobserved}}$$

$$-(E[Y(0)|D=0,t=1] - E[Y(0)|D=0,t=0]) + \underbrace{E[Y(0)|D_i=1,t=1]}_{\text{unobserved}}$$

Gathering terms, this equals:

$$ATT + \underbrace{(E[Y(0)|D=1,t=1] - E[Y(0)|D=1,t=0])}_{\text{pre to post change in Y(0) for D=1 group}}$$

$$- \underbrace{(E[Y(0)|D=0,t=1] - E[Y(0)|D=0,t=0])}_{\text{pre to post change in Y(0) for D=0 group}}$$

The second term is counterfactual (unobserved). However if parallel trends holds, the second and third term cancel each other out.

Difference-in-differences estimation

To summarize:

- Changes over time in the $D = 0$ group provide the counterfactual
- Selection into treatment related to fixed unobserved differences is OK
- The outcome *levels* are not important, only the *changes*

DD is probably the most commonly used quasi-experimental design in the social sciences and education.

- Its use precedes the RCT
- The "comparative interrupted time series" (CITS) design is similar, though not the same. See Section 3 of the MDRC paper by Somers et al. (2013) for a good delineation between the two in the context of an educational intervention.

Regression difference-in-differences (2x2)

With only two groups and time periods (pre-post):

$$Y_{it} = \alpha + \beta D_i + \lambda Post_t + \delta(D_i \times Post_t) + u_{it}$$

where $D_i = 1$ for units i who are ultimately treated, and $POST_t = 1$ for observations in the "post" period.

Very easy to implement in Stata, especially with factor variable notation:
`reg y i.treated##i.post`

Regression difference-in-differences (2x2)

How does this map onto our earlier notation? There are four expectations estimated in this regression:

$$\begin{aligned}
E[Y_{it}|D_i = 0, t = 0] &= \alpha \\
E[Y_{it}|D_i = 1, t = 0] &= \alpha + \beta \\
E[Y_{it}|D_i = 0, t = 1] &= \alpha + \lambda \\
E[Y_{it}|D_i = 1, t = 1] &= \alpha + \beta + \lambda + \delta
\end{aligned}$$

- α is the pre-period mean for the $D_i = 0$ group
- $\alpha + \beta$ is the pre-period mean for the $D_i = 1$ group
- β is the baseline mean difference between the $D_i = 0$ and $D_i = 1$
- $\alpha + \lambda$ is the *post*-period mean for the $D_i = 0$ group
- $\alpha + \beta + \lambda + \delta$ is the *post*-period mean for the $D_i = 1$ group
- λ is the change over time for the $D_i = 0$ group

Regression difference-in-differences (2x2)

The four expectations being estimated in this regression and their differences:

	Pre	Post	Diff
Untreated ($D_i = 0$)	α	$\alpha + \lambda$	λ
Treated ($D_i = 1$)	$\alpha + \beta$	$\alpha + \beta + \lambda + \delta$	$\lambda + \delta$
Diff	β	$\beta + \delta$	δ

DD is effectively a comparison of four cell-level means.

Regression difference-in-differences (2x2)

The 2x2 DD regression:

- Is estimating a CEF since the model is fully saturated.
- The CEF is not necessarily causal (depends on parallel trends).
- OLS will always (mechanically) estimate δ as the differential change in the $D_i = 1$ vs. $D_i = 0$ group.
- Whether that δ can be interpreted as the ATT depends on the parallel trends assumption.

Regression difference-in-differences (2x2)

We could also estimate a regression using first differences for each observation i, subtracting Y_{i0} from Y_{i1} (again assuming two periods):

$$Y_{i1} = \alpha + \beta D_i + \lambda + \delta(D_i) + u_{i1}$$
$$Y_{i0} = \alpha + \beta D_i + u_{i0}$$

$$Y_{i1} - Y_{i0} = \lambda + \delta D_i + \epsilon_{it}$$

$$\Delta Y_i = \lambda + \delta D_i + \epsilon_{it}$$

This regression is equivalent to the standard DD regression shown earlier. The intercept here represents the time trend λ, and δ is the DD. The baseline differences wash out in the first difference (Δ)

Regression difference-in-differences (2x2)

The 2x2 regression model can also include covariates:

$$Y_{it} = \alpha + \beta D_i + \lambda Post_t + \delta(D_i \times POST_t) + \mathbf{X_{it}}\eta + u_{it}$$

Example: Dynarski (2003)

Prior to 1982, 18- to 22-year old children of deceased Social Security beneficiaries were eligible for survivor's benefits that could be applied toward college. This practice ended in 1982. Dynarski (2003) used this policy change to estimate the effect of financial aid on college enrollment.

- Table 8.1 from Murnane & Willett on next page begins with the ITS design, focusing only on survivors (a first difference)
- Data: NLSY high school seniors who would be eligible for benefits just before (N=137) and after (N=54) the policy change.

Example: Dynarski (2003)

Table 8.1 "First difference" estimate of the causal impact of an offer of $6,700 in financial aid (in 2000 dollars) on whether high-school seniors whose fathers were deceased attended college by age 23 in the United States

(a) *Direct Estimate*

H.S. Senior Cohort	Number of Students	Was Student's Father Deceased	Did H.S. Seniors Receive an Offer of SSSB Aid?	Avg Value of *COLL* (standard error)	Between-Group Difference in Avg Value of *COLL*	H_0: $\mu_{OFFER} = \mu_{NO\ OFFER}$	
						t-statistic	*p*-value
1979–81	137	Yes	Yes (*Treatment Group*)	0.560 (0.053)	**0.208***	2.14	0.017†
1982–83	54	Yes	No (*Control Group*)	0.352 (0.081)			

˙p <0.10; * p <0.05; ** p <0.01; *** p <0.001.
†One-tailed test.

(b) *Linear-Probability Model (OLS) Estimate*

Predictor	Estimate	Standard Error	H_0: $\beta = 0$;	
			t-statistic	*p*-value
Intercept	0.352***	0.081	4.32	0.000
OFFER	0.208*	0.094	2.23	0.013†
R^2	0.036			

˙p <0.10; * p <0.05; ** p <0.01; *** p <0.001.
†One-tailed test.

Example: Dynarski (2003)

Table 8.2 from Murnane & Willett reports the DD estimate, incorporating data for high school seniors that were not survivors, before (N=2,745) and after (N=1,050) the policy change—a second difference.

Example: Dynarski (2003)

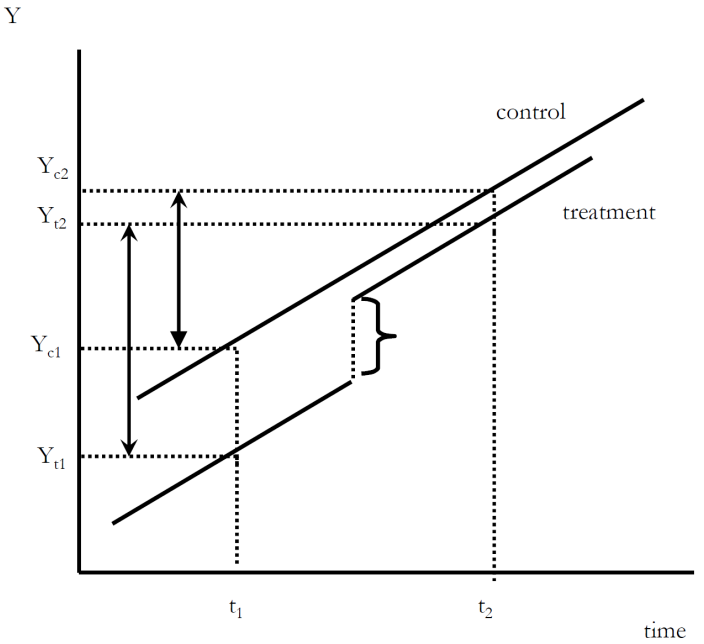
Table 8.2 Direct "difference-in-differences" estimate of the impact of an offer of $6,700 in financial aid (in 2000 dollars) on whether high-school seniors whose fathers were deceased attended college by age 23, in the United States

H.S. Senior Cohort	Number of Students	Was Student's Father Deceased?	Did H.S. Seniors Receive an Offer of SSSB Aid?	Avg Value of *COLL* *(standard error)*	Between-Group Difference in Avg Value of *COLL*	"Difference in Differences"	
						Estimate *(standard error)*	*p*-value
1979–81	137	Yes	Yes *(Treatment Group)*	0.560 (0.053)	0.208 *(First Diff)*		
1982–83	54	Yes	No *(Control Group)*	0.352 (0.081)		0.182* (0.099)	0.033†
1979–81	2,745	No	No	0.502 (0.012)			
1982–83	1,050	No	No	0.476 (0.019)	0.026 *(Second Diff)*		

~ p <0.10; * p <0.05; ** p <0.01; *** p <0.001.
†One-tailed test.

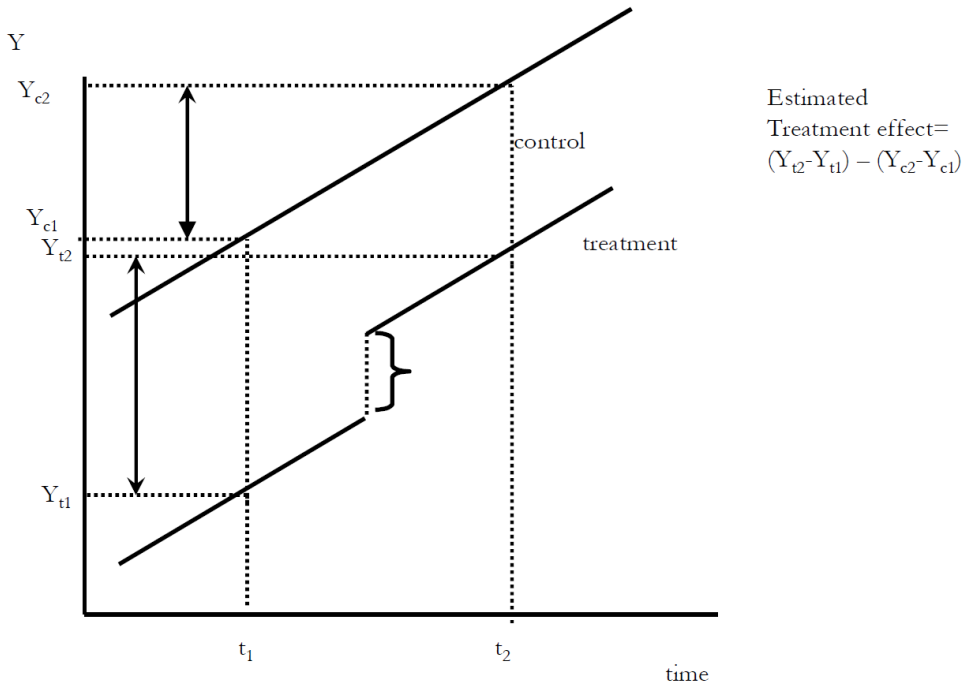
Parallel trends assumption

The key assumption in DD is parallel trends: that the time trend in the absence of treatment would be the same in both groups.

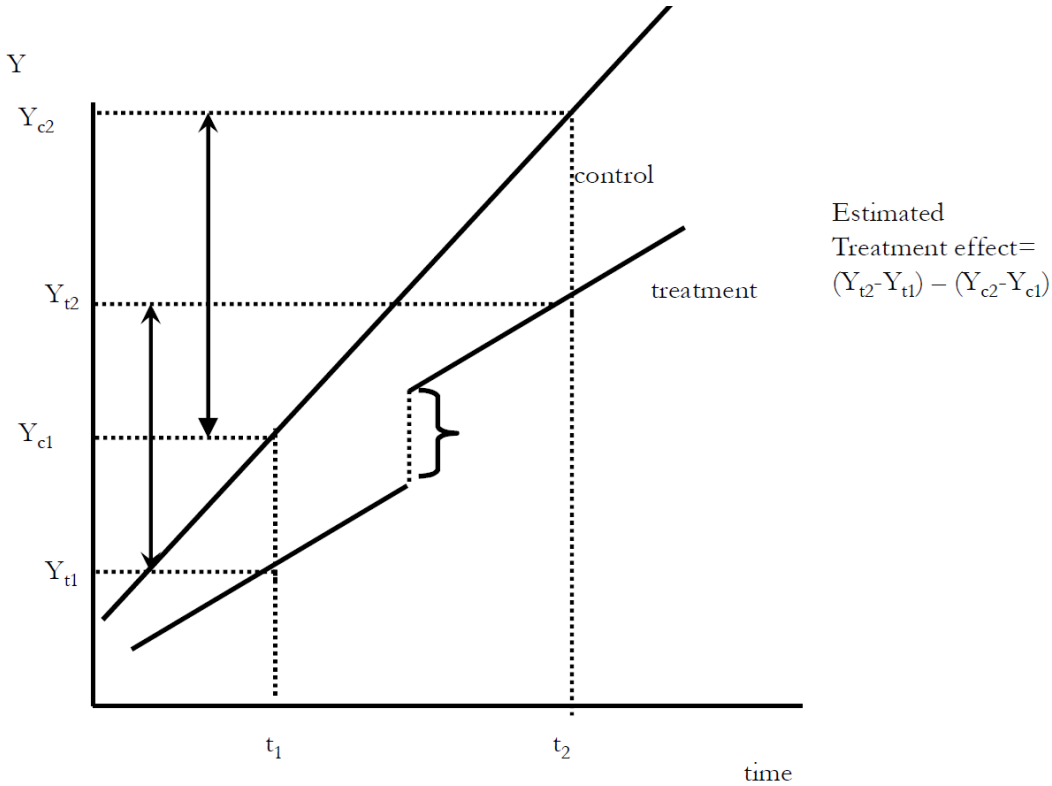


Parallel trends assumption

Size of baseline difference in treated and untreated groups doesn't matter.



Violation of parallel trends assumption



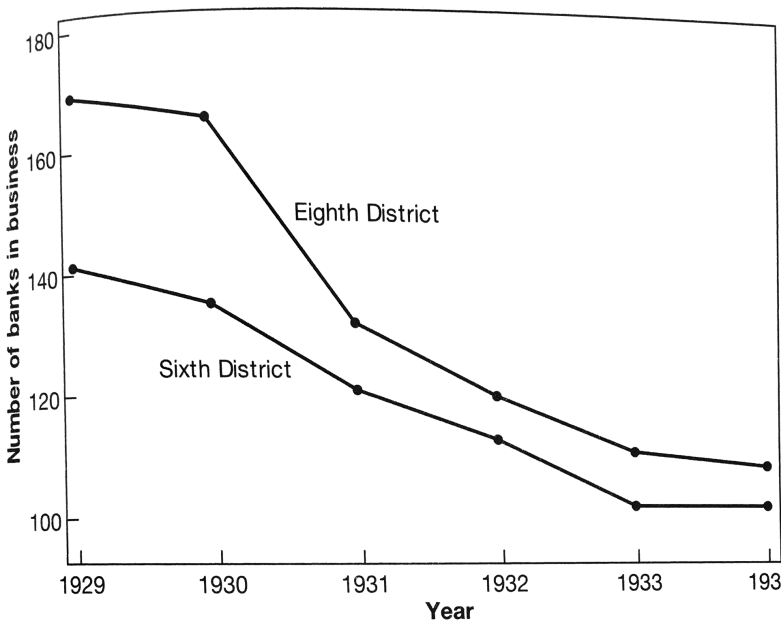
Parallel trends assumption

We can't verify the parallel trends assumption directly, but researchers typically defend it in a variety of ways:

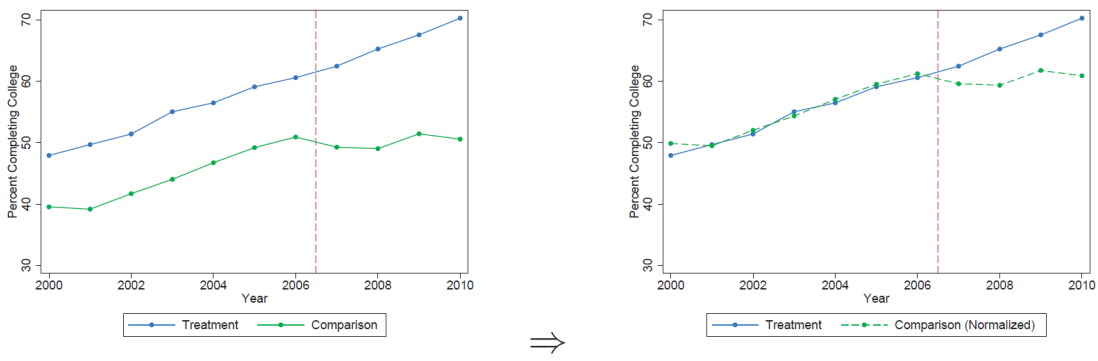
- A compelling graph: pointing to similar trends prior to the treatment. Note: common trends prior to treatment are neither necessary nor sufficient for parallel trends assumption!
- Event study regression and graph
- A placebo / falsification test
- Controlling for time trends directly (leans heavily on functional form)
- Triple-difference model
- Understanding the context of your study! Ruling out alternative explanations

Federal Reserve policy and bank failures

FIGURE 5.2
Trends in bank failures in the Sixth and Eighth Federal Reserve Districts

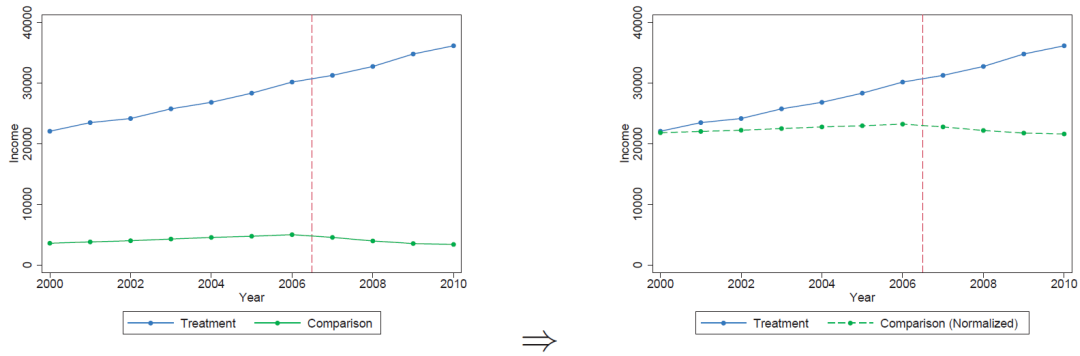


Checking the parallel trends assumption (1)



The graph on the right ("normalized") subtracts baseline difference between Treated and Comparison group, to help see the parallel trend.

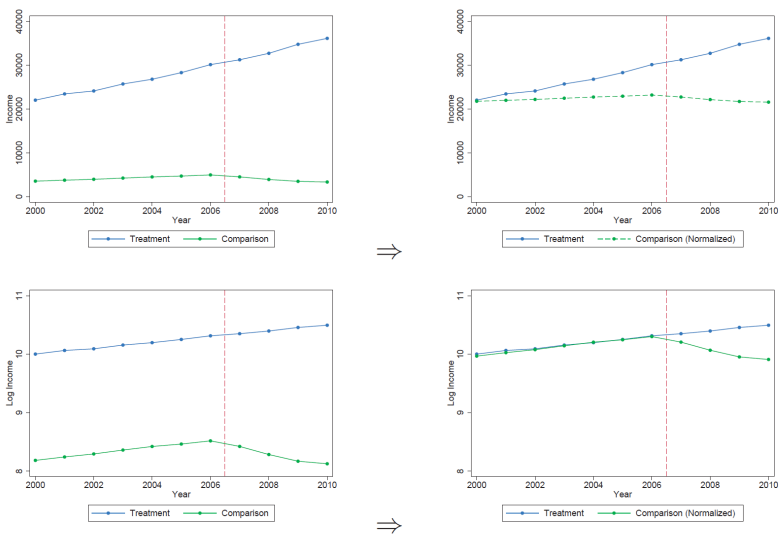
Checking the parallel trends assumption (2)



The graph on the right ("normalized") makes the lack of a parallel trend more apparent than the graph on the left.

Checking the parallel trends assumption (3)

A variable transformation may help satisfy the parallel trends assumption. The bottom panels use the *log*:



Note: if trends are parallel in levels they will *not* be parallel in logs, and vice versa!

Event study

An **event study** regression is like the DD model shown earlier, except it includes separate indicators and treatment interactions for *all* pre and post periods.

Assume you have observations on q periods before treatment, which occurs at $t = 0$, and m periods after treatment:

$$Y_{it} = \beta D_i + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_\tau D_i + \sum_{\tau=0}^{m} \delta_\tau D_i + u_{it}$$

The δ_τ are differences between the $D = 1$ and $D = 0$ groups in the post period. Like the earlier DD coefficient, but separate for each time period.

The γ_τ are differences in the pre period—beyond those already captured by the β. If the groups were on similar time paths, you would not expect to see these differ from zero or show any sort of trend.

Event study

If all treated units are treated in the same period, this is easy to implement in Stata (assume *year* is the time period):
`reg y i.treated##i.year`

A full factorial of treatment group status and time period—includes main effects for *treated* and individual years, and their interaction. See also the user-written package `eventdd`

If timing varies, need to center *year* so that it equals 0 in the first year of treatment. Less obvious what to do with never-treated units. (More on this later).

Event study example

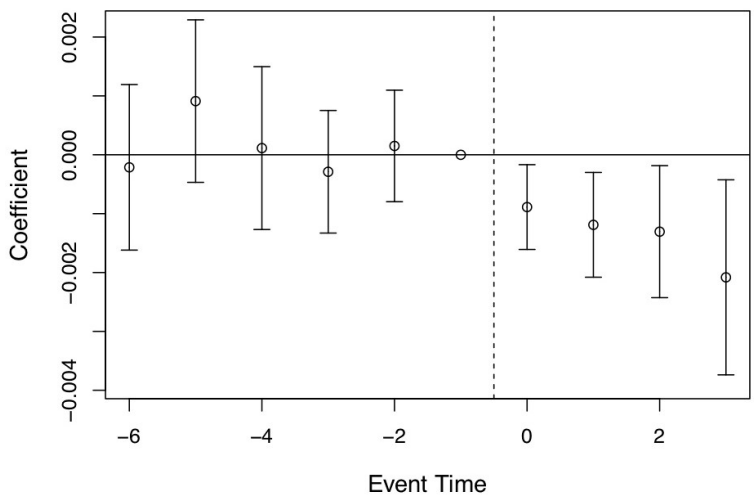
The following figures are from Miller et al. (forthcoming), via the *Mixtape*. The authors estimate the impact of state expansion of Medicaid under ACA on the annual mortality rates of older persons under 65 in the U.S.

Causal inference from DD assumes changes over time in states that did *not* expand Medicaid provide the counterfactual for those that did.

They find a 0.13 percentage-point decline in annual mortality, a 9.3% reduction over the sample mean, as a result of Medicaid expansion.

Event study example

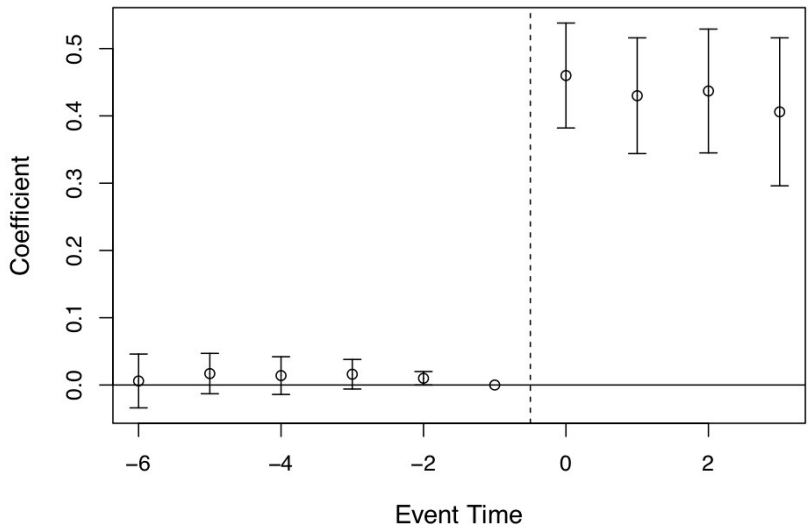
Plotted points are event study coefficients, shown with 95% confidence intervals. (Time zero is the first year of expansion).



There is no evidence these states were on different trajectories prior to Medicaid expansion.

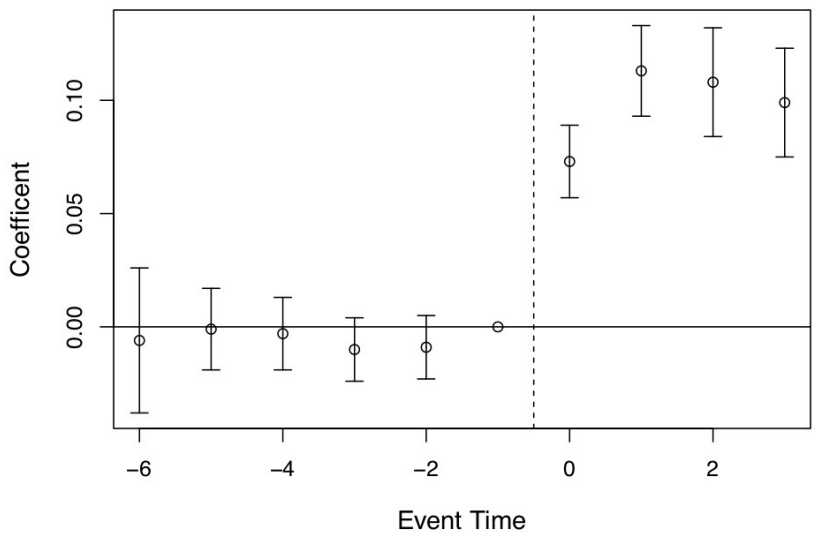
Event study example

The authors first look for a "first stage": did the expansion of Medicaid actually increase rates of eligibility for Medicaid? Did it increase Medicaid coverage? Did it lower the uninsured rate? Here: eligibility



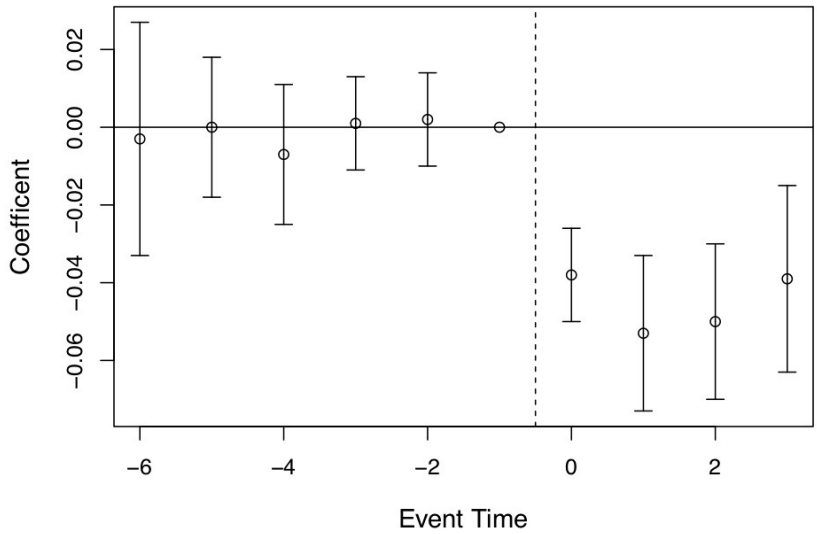
Event study example

Here: Medicaid coverage rates



Event study example

Here: Medicaid uninsured rates



Taken together, these graphs are compelling: Medicaid expansion increased eligibility and coverage, and reduced the uninsured. One would hope to see these first stage effects before expecting an effect on health outcomes.

Parallel trends assumption

When covariates are included in the model, the parallel trends assumption is *conditional* on the covariates. It is possible that the unconditional outcomes do not follow a parallel trend, but the conditional outcomes do.

Put another way, controlling for covariates allows you to account for factors that might produce different time trends.

Common violations of parallel trends assumption

Two common scenarios that would violate the parallel trends assumption:

- Targeted treatments: often programs are targeted at subjects who are most likely to benefit from it. In many cases, the fact that a subject was on a different trajectory is what made them a good candidate from the program (e.g., a struggling student).
- Ashenfelter's dip: treated cases may experience a "dip" just prior to treatment that results in a reversion to the mean after treatment (e.g, job training).

Placebo/falsification tests

The DD design assumes that any change over time beyond that predicted by the untreated group is the ATT, and not some other time-varying factor specific to the treated group.

If there is indeed an unobserved time-varying factor specific to the treated group, one might see its effects show up on *other* outcomes that shouldn't have been affected by the treatment.

- Card & Krueger: employment in higher-wage firms
- Miller et al.: mortality of populations not eligible for Medicaid
- Cheng & Hoekstra (2013): effects of Stand Your Ground laws on other non-homicide crimes (see Mixtape)

Estimate the same DD model for these outcomes. If there is an "effect", this may indicate an unobserved, time-varying confounder specific to the treated group.

Placebo/falsification tests

Another approach is to apply the same treatment assignment to an earlier period, well before the treatment actually occurred, and re-estimate the DD model on this earlier data. If there is an apparent treatment "effect" in these untreated years, there may well be unobserved, group-specific trends driving the result.

Triple difference

The **triple difference** uses an additional untreated group to difference out time trends unique to the treatment group that are also experienced by the added untreated group. For example:

- In C&K, suppose we were concerned that the (treated) state of NJ was on a different time trend from the (untreated) state of PA.
- The lack of parallel trends could make the DD invalid.
- The minimum wage treatment should only affect *low-wage* workers.
- We can contrast *higher-wage* workers in NJ and PA to identify any differential time trend in NJ.
- The treatment effect of the minimum wage on low wage workers would be any *additional* change over time experienced by low-wage workers in NJ.

Triple difference

Let $G = 1$ be the focal group and $G = 0$ be the additional untreated group (e.g., low-wage workers are $G = 1$ and higher-wage workers are $G = 0$).

The triple difference regression model is:

$$Y_{it} = \alpha + \beta_1 POST_t + \beta_2 G_i + \beta_3 D_i + \beta_4 (G_i \times POST_t)$$
$$+\beta_5 (D_i \times POST_t) + \beta_6 (G_i \times D_i) + \beta_7 (G_i \times D_i \times POST_t) + u_{it}$$

Triple difference

First, consider the focal group $G = 1$ (e.g., low-wage workers)

$$Y_{it} = \alpha + \beta_1 POST_t + \beta_2 G_i + \beta_3 D_i + \beta_4 (G_i \times POST_t)$$
$$+\beta_5 (D_i \times POST_t) + \beta_6 (G_i \times D_i) + \beta_7 (G_i \times D_i \times POST_t) + u_{it}$$

$D = 0$,	$E[Y\|D = 0, t = 0, G = 1]$	$= \alpha + \beta_2$
focal group	$E[Y\|D = 0, t = 1, G = 1]$	$= \alpha + \beta_1 + \beta_2 + \beta_4$
$D = 1$,	$E[Y\|D = 1, t = 0, G = 1]$	$= \alpha + \beta_2 + \beta_3 + \beta_6$
focal group	$E[Y\|D = 1, t = 1, G = 1]$	$= \alpha + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7$

The traditional DD here would be $\beta_5 + \beta_7$, reflecting the time trend specific to $D = 1$ *and* the ATT (β_7). The latter is not identified.

Triple difference

Now, consider the non-focal group $G = 0$ (e.g., higher-wage workers)

$$Y_{it} = \alpha + \beta_1 POST_t + \beta_2 G_i + \beta_3 D_i + \beta_4 (G_i \times POST_t)$$

$$+\beta_5 (D_i \times POST_t) + \beta_6 (G_i \times D_i) + \beta_7 (G_i \times D_i \times POST_t) + u_{it}$$

$D = 0$,	$E[Y\|D = 0, t = 0, G = 0]$	$= \alpha$
non-focal group	$E[Y\|D = 0, t = 1, G = 0]$	$= \alpha + \beta_1$
$D = 1$,	$E[Y\|D = 1, t = 0, G = 0]$	$= \alpha + \beta_3$
non-focal group	$E[Y\|D = 1, t = 1, G = 0]$	$= \alpha + \beta_1 + \beta_3 + \beta_5$

The DD for the non-focal group is β_5. The difference between this and the focal group DD ($\beta_5 + \beta_7$) is β_7, the ATT we are looking for.

Put more simply, the β_7 coefficient gives you the triple difference.

Difference-in-differences with variable timing

The examples thus far had 2 groups and 2 time periods. In practice, "treatment" can occur for different groups at different times.

This brings us to the "generalized difference-in-differences" model, or difference-in-difference with variable timing. Usually estimated as a "two-way fixed effects" model with fixed effects for cross-sectional units and time periods. Sometimes written:

$$Y_{it} = \beta_i + \gamma_t + \delta (D_i \times POST_t) + u_{it}$$

Difference-in-differences with variable timing

Mastering 'Metrics: effect of a lower Minimum Legal Drinking Age (MLDA), based on Carpenter & Dobkin (2011).

- Following the 26th Amendment (1971), some states lowered the drinking age to 18
- In 1984, federal legislation pressured states to increase MLDA to 21
- Was a lower MLDA associated with more traffic fatalities among 18-20 year olds?

The authors used panel data (state $\times$ year) to address this question.

Difference-in-differences with variable timing

$$Y_{st} = \alpha + \delta(TREAT_s \times POST_t) + \sum_{k=2}^{50} \beta_k STATE_{ks} + \sum_{j=2}^{T} \gamma_t YEAR_{jt} + u_{st}$$

- $STATE_{ks} = 1$ if observation is from state k. States indexed from $k = 2...50$ as a reminder that one state dummy must be omitted.
- $YEAR_{jt} = 1$ if observation is from year j. Years indexed from $j = 2...T$ as a reminder that one year dummy must be omitted.
- β_k is a *state effect*.
- γ_t is a *year effect*.
- Covariates X_{st} may be included to control for other time-varying factors associated with Y and treatment.

Generalized difference-in-differences

- Analogous to the 2x2 model, each group (state) has its own intercept ($\alpha + \beta_k$ for $k = 2, ..., 50$).
- There need not be a single common post-treatment period. The year effects (γ_t) capture trends in the outcome common to all states.
- The coefficient on the interaction (δ) represents how much, on average, outcomes *differ* in treatment states in the post period from that predicted by the state and year effects.
- In other words, we are contrasting within-state changes over time in the outcome, for treated and untreated states.
- This is a simple example of a panel model, a topic covered more in Lecture 7.

Generalized difference-in-differences

Implementing in Stata: can be done in multiple ways, including `xtreg`:

```
xtreg y x i.year i.treat##i.post, i(state) fe
```

`xtreg` is a panel data command where *state* is the cross-sectional unit and `fe` implements the fixed-effects (within) estimator—covered in Lecture 7. Essentially equivalent to separate intercepts for every state.

The fixed effect variable (*state* here) should be numeric. If it is not, can use `encode`.

Generalized difference-in-differences

Integrating with our earlier notation for *potential outcomes* for a state k:

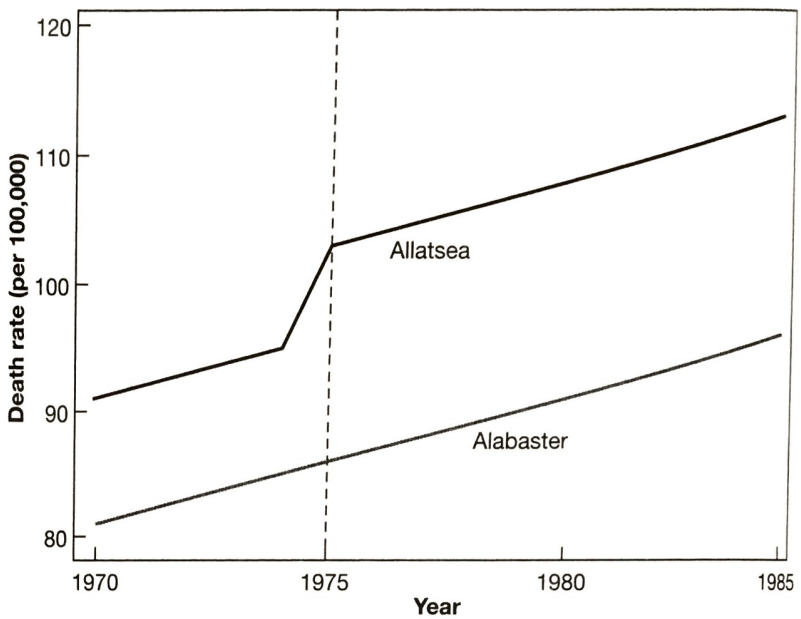
$$Y_{kt}(0) = \alpha + \beta_k + \gamma_t$$
$$Y_{kt}(1) = \alpha + \beta_k + \gamma_t + \delta$$

Potential outcomes are described by a unique intercept for each state $(\alpha + \beta_k)$ and a yearly deviation from this intercept that is common to every state (γ_t). The treatment effect is δ.

Intuitively, under the common trends assumption that changes within states over time would be the same in the absence of treatment, we can estimate δ as the *differential* change over time associated with treatment.

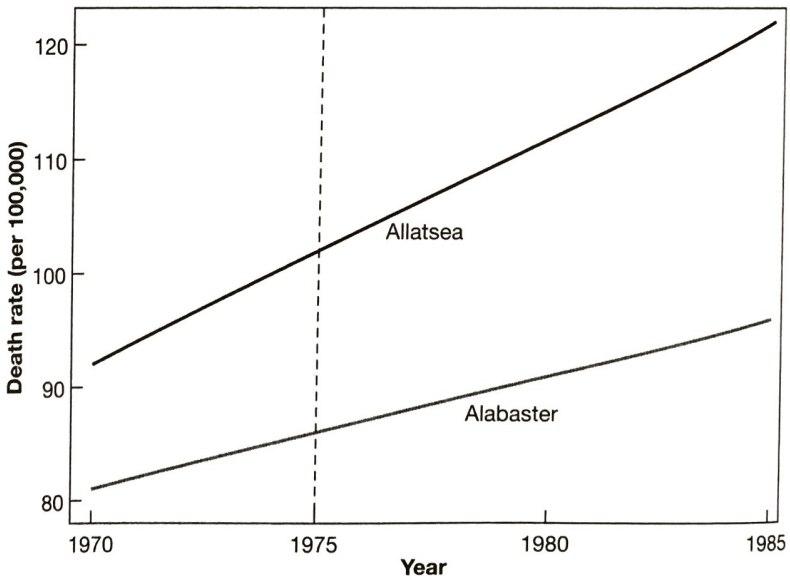
Mastering 'Metrics Fig 5.4

FIGURE 5.4
An MLDA effect in states with parallel trends



Mastering 'Metrics Fig 5.5

FIGURE 5.5
A spurious MLDA effect in states where trends are not parallel



In-class exercise

Replicate the findings in the MLDA study reported in *Mastering 'Metrics*.

- Generalized DD using two-way fixed effects
- Placebo test using other outcomes, age groups

To be added

- Group-specific time trends
- Recent research on the two-way fixed effects model for difference-in-differences (e.g., Goodman-Bacon, forthcoming)
- Inference in difference-in-differences model (standard errors)