

# Difference-in-differences

EC420 MSU

---

Justin Kirkpatrick

Last updated March 31, 2021

## Solving endogeneity with Instrumental Variables

- Conceptual
  - "Partialling in" exogenous variation in our endogenous variable
  - In practice (2SLS)

## Simultaneity

- A type of endogeneity common in economics

When is IV/2SLS causal?

- *ATE*?
- Counterfactual

**Difference-in-differences** or *DID*

- Why?
- How?

If we go way back, we started with a problem rooted in the Rubin *potential outcomes* framework.

We want to know  $E[Y_1 - Y_0]$ , the *Average Treatment Effect*, and

$$E[Y|D = 1] - E[Y|D = 0] \neq E[Y_1 - Y_0]$$

That is, comparing the average outcome from those who are treated to those who are not treated does not yield the *ATE*

We fail to get the  $ATE$  because of selection bias:

$$E[Y|D = 1] - E[Y|D = 0] = \overbrace{E[Y_1 - Y_0|D = 1]}^{SATE} - \underbrace{\{E[Y_0|D = 1] - E[Y_0|D = 0]\}}_{\text{Selection Bias}}$$

And that selection bias is because the *potential outcomes* for those who get treated are different from those who do not get treatment:

$$(Y_{i0}, Y_{i1}) \not\perp D$$

- The  $\not\perp$  means "(not) independent of"

This statement is important. If the potential outcomes  $Y_{i0}$  and  $Y_{i1}$  determine treatment, then we have selection bias because  $E[Y_0|D = 1] \neq E[Y_0|D = 0]$ .

$D$  is a (potentially endogenous) RHS variable like  $x$

- e.g. think of  $D$  as *Attending KIPP*

Let's tie a concept from IV to this:

The *independence assumption* for the instrument was that it was *as good as randomly assigned* conditional on other exogenous  $x$ 's.

Another way of saying this is:

$$(Y_{i0}, Y_{i1}) \perp Z | X_{\text{exogenous}}$$

"Conditional on exogenous  $X_{\text{exogenous}}$ , the potential outcomes are independent of  $Z$ ."

- The  $X_{\text{exogenous}}$  here is not the "problem", endogenous  $D$  is (e.g. *Attending KIPP*).

Since:

$$(Y_{i0}, Y_{i1}) \perp Z | X_{exogenous}$$

Then  $\hat{D} = \hat{\gamma}_0 + \hat{\gamma}_1 z_1 + \hat{\gamma}_2 x_{exogenous}$  **also** has this property:

$$(Y_{i0}, Y_{i1}) \perp \hat{D} | X_{exogenous}$$

And we have (with IV) solved selection bias.  $\hat{D}$  is *as good as randomly assigned*.

We call this the *conditional independence assumption* or CIA

## Counterfactual

Since  $(Y_{i0}, Y_{i1}) \perp \hat{D} | X_{exogenous}$ , then our selection bias,

$$E[Y_0 | \hat{D} = 1] - E[Y_0 | \hat{D} = 0] = 0$$

- Another way of saying this is that, because  $Z$  is exogenous (and thus so is  $\hat{D}$ ) we have a **valid counterfactual** for  $E[Y_0 | D = 1]$ .
- Remember that a *counterfactual* is defined by:
  - An outcome that has more than one potential value
  - A choice or occurrence that may change the realization of the outcome



## And one more bit of terminology you may run into:

The **intent to treat effect**,  $ITT$  is the numerator of the  $LATE$ . It is the coefficient on:

$$y = \gamma_0 + \gamma_1 z + u$$

The term pops up from time to time - you already know *what* it is, we just want to give it a name.

It is not the first stage. Nor is it the second stage. It is the effect of the the instrument on  $y$  directly.

- "The change in test scores per change in winning the lottery"

# Difference-in- differences

## Difference-in-differences

In the absence of random assignment, treatment and control groups are likely to differ for many reasons

MM, Ch. 5

Our goal is to get a valid counterfactual for  $E[Y_0|D = 1]$

So that we have solved our selection bias problem

But we might not have a good instrument

Or we might worry that there are other unobserved things about the group being treated

DiD is another tool for your econometric toolbox

Diff-in-diff is closely related to our discussion on fixed effects

$$y_{it} = \beta_0 + \beta_1 x_{it} + \phi_i + \delta_t + u_{it}$$

- Where  $\phi_i$  is a vector of effects for each  $i$  (except the base or omitted level)
- And where  $\delta_t$  is a vector of effects for each time period  $t$  (except for a base or omitted level)

$\phi_i$  controls for persistent individual-specific effects

- "persistent" means "constant over time" within  $i$
- $\delta_t$  does the same, but for effects in a time period  $t$  that are constant over all  $i$

Of course, we must have panel data for DID

- We need multiple  $i$  and multiple  $t$

We have to have some "treated"  $i$ , and some "untreated"  $i$

- The "treatment group"
- The "control group"
  - This group helps us build our counterfactual
  - And our counterfactual helps us fix selection bias
  - And fixing selection bias gets us to an *ATE*!

## The intuition:

Let's start with a simple difference. We observe some before and after data, where  $\tau$  is the time that the treatment starts:

$$y_t = \beta_0 + \beta_1 D_t + u$$

and

$$D_t = 1(t > \tau)$$

Since  $\beta_1$  is the difference in  $y_{t < \tau}$  and  $y_{t > \tau}$ , then:

$$E[y|D = 1] = E[y|t > \tau]$$

$$E[y|D = 0] = E[y|t < \tau]$$

Which means

$$\beta_1 = E[y|D = 1] - E[y|D = 0]$$

plus:

$$\beta_0 = E[y|D = 0]$$

The simple difference is a naive estimator of the treatment effect of interest.

Even if we take a bunch of before-after observations:  $y_{it} = \beta_0 + \beta_1 D_{it} + u_{it}$ , we might still have a problem.

The problem here is that *treatment is not necessarily as good as randomly assigned*.

The biggest threat is that there is something *unobserved* that *varies between treated units* that might be correlated with the start of treatment,  $\tau$ .



## An example

Let's say  $\tau$  is the time that a minimum wage increases in a state, and  $y$  is total hours worked in the city

- Interesting because we want to see the effect of minimum wage on employment
- **But** when states decide to increase their minimum wage, it's *may be* because they think the national economy is on the rise
- So treatment is correlated with *better economy* and that can drive employment up.
- Treatment time is endogenous!

## Some more examples:

$\tau$  may be the time in which a superfund site is cleaned up, and  $y$  is the average home value around the superfund site

- Economic rationale: If cleaning up a highly polluted site increases the value of homes (by more than the cost of cleanup), we are increasing welfare by cleaning up pollution.
- **But** it may be the house prices are on the way up when there are good economic times, and government has more revenue during good economic times for cleanup. Or people who move into this neighborhood might be better at lobbying for cleanup.
- Treatment time is endogenous!

## Yet another example:

$\tau$  may be the start of an incentive program for solar panel installation in a state, and  $y$  is the number of solar panel installations in a given month.

- If this incentive program is effective, we would expect to see an increase in the number of solar panels installed after  $\tau$
- **But** it might be that the program was enacted because policymakers thought solar would be becoming more popular everywhere
- Treatment time is endogenous!

The common theme is that treatment time may be endogenous, so  $\beta_1$  is biased.

- $E[u|D] \neq 0$
- Those time periods (months, years) that are treated have different  $(Y_{t0}, Y_{t1})$
- Here, potential outcomes are over **time**.

## Counfounder:

### We might say:

There is an unobserved, time-variant trend confounding our estimate

**Confounder** is something that is biasing our estimate by "looking like" our treatment variable  $D$ .

**National** trends **confound** our estimate when they coincide with treatment time  $\tau$ .

When we have panel data, we have multiple  $i$ 's with treatment:

$$y_{1t} = \beta_0 + \beta_1 D_{1t} + u_{1t}$$

$$y_{2t} = \beta_0 + \beta_1 D_{2t} + u_{2t}$$

Unfortunately, we can't just fixed-effect away the problem. We would worry that:

$$y_{it} = \beta_0 + \beta_1 D_t + \phi_i + u$$

is still potentially biased because *some other time-variant effect possibly correlated to the treatment time* confounds  $\beta_1$ .

- "There were going to be lots of solar panels installed anyways", etc.
- National trend in employment going up

## The control group

- Observed  $i$  that are not treated
- They provide the counterfactual
- They are affected by the same unobserved time trend
- We should be able to "subtract off" that time trend
- Let's use a new dummy variable for members of this group:
  - $TREAT_i = 0$  for those in the control group
  - $TREAT_i = 1$  for those in the treatment group

Since treatment is time-variant, let's re-name  $D_t$ :

- $POST_t = 0$  in the pre-treatment period
- $POST_t = 1$  in the post-treatment period.

A treated  $t$  (time) for  $i$  (individual/unit) is  $POST_t \times TREAT_i == 1$

First, take the before-after differences for the  $i$  that get treated:

$$\begin{aligned}\delta_{TREAT} = & E[y|POST == 1, TREAT == 1] \\ & - E[y|POST == 0, TREAT == 1]\end{aligned}$$

$\delta_{TREAT}$  is what we had before just looking at treated units before-after.

As usual, we could get this from a regression of  $Y$  on  $TREAT$  and  $POST$

Second, take the before-after differences for the  $i$  that did not get treated:

$$\begin{aligned}\delta_{CONTROL} = & E[y|POST == 1, TREAT == 0] \\ & - E[y|POST == 0, TREAT == 0]\end{aligned}$$

If we assume that *this*  $\delta_{CONTROL}$  is "what would have happened to the treated in the absence of treatment", then it provides a good counterfactual.



Finally, take the *difference* in the *differences*:

$$\delta_{DID} = \delta_{TREAT} - \delta_{CONTROL}$$

That's it!

Plugging in those two equations and using  $\bar{y}$  for  $E[y]$ :

$$\begin{aligned}\hat{\delta}_{DID} = & (\bar{y}_{TREAT=1,POST=1} - \bar{y}_{TREAT=1,POST=0}) \\ & - (\bar{y}_{TREAT=0,POST=1} - \bar{y}_{TREAT=0,POST=0})\end{aligned}$$

We know that we can get an unbiased and consistent estimate of

$$E[y|TREAT = 1, POST = 1] - E[y|TREAT = 1, POST = 0]$$

using a regression **only** on  $TREAT_i == 1$ :

$$y_{it} = \beta_0 + \beta_2 POST_t + u \quad (\text{using only treated } i)$$

Similarly, we could do the same for those observations where  $TREAT_i == 0$

**But** there is an easier way

We could regress:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \underbrace{\beta_3 TREAT_i * POST_t}_{\delta_{DID}} + u$$

We can see this by calculating:

- $E[Y|TREAT = 1, POST = 1] = \beta_0 + \beta_1 + \beta_2 + \beta_3$
- $E[Y|TREAT = 1, POST = 0] = \beta_0 + \beta_1$
- $E[Y|TREAT = 0, POST = 1] = \beta_0 + \beta_2$
- $E[Y|TREAT = 0, POST = 0] = \beta_0$

$$\begin{aligned} E[y|TREAT = 1, POST = 1] - E[y|TREAT = 1, POST = 0] &= \\ (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) &= \beta_2 + \beta_3 \end{aligned}$$

and

$$\begin{aligned} E[y|TREAT = 0, POST = 1] - E[y|TREAT = 0, POST = 0] &= \\ (\beta_0 + \beta_2) - (\beta_0) &= \beta_2 \end{aligned}$$

In the regression:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \underbrace{\beta_3 TREAT_i * POST_t}_{\delta_{DID}} + u$$

$\delta_{DID}$  is the difference between the before-after differences. It is the unexpected "jump" that occurs in the treatment  $i$  at time of treatment  $t = \tau$  relative to those that don't get treatment.

We can see this visually, and see the counterfactual.

FIGURE 5.1

Bank failures in the Sixth and Eighth Federal Reserve Districts

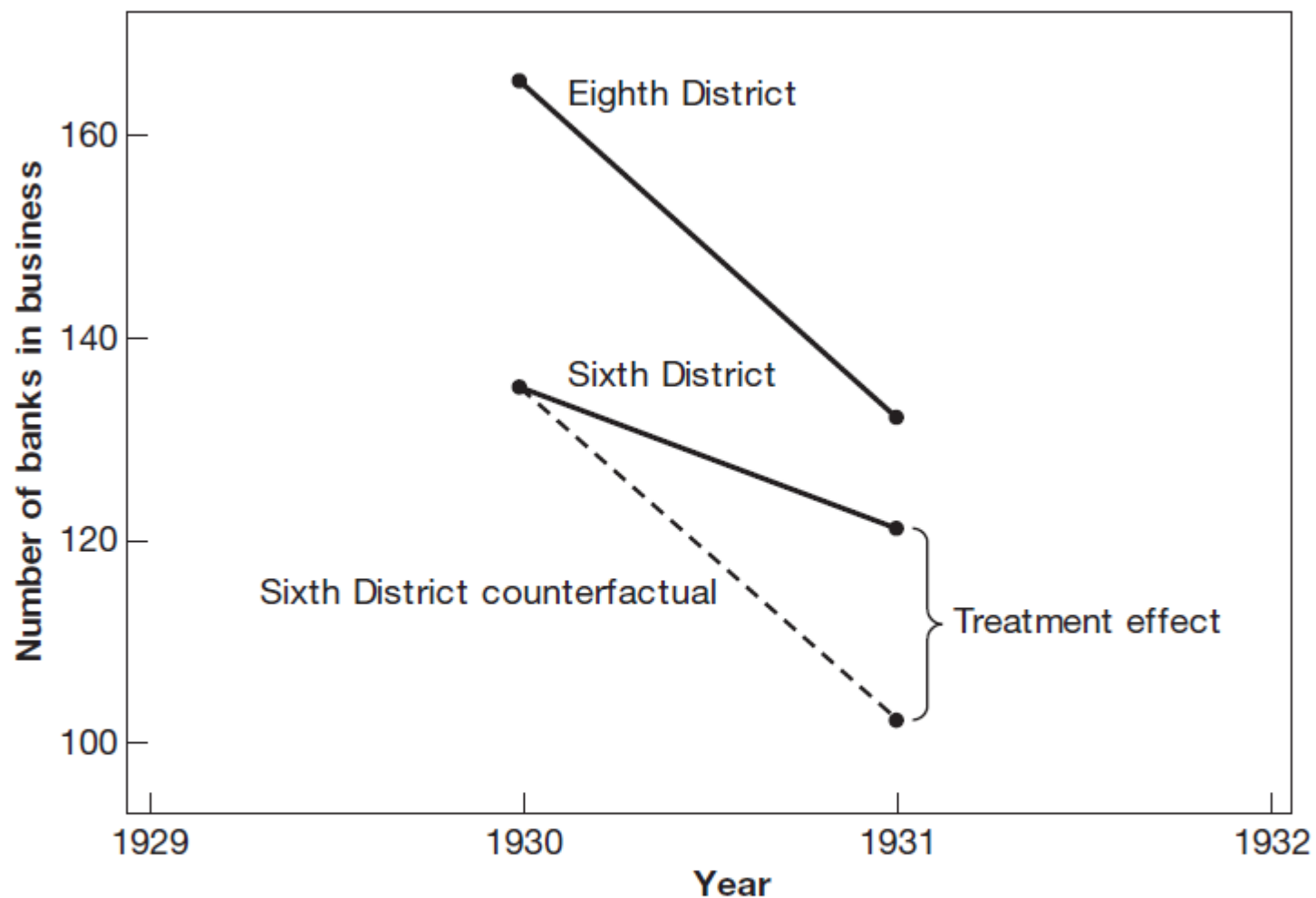


FIGURE 5.3

Trends in bank failures in the Sixth and Eighth Federal Reserve Districts, and the Sixth District's DD counterfactual

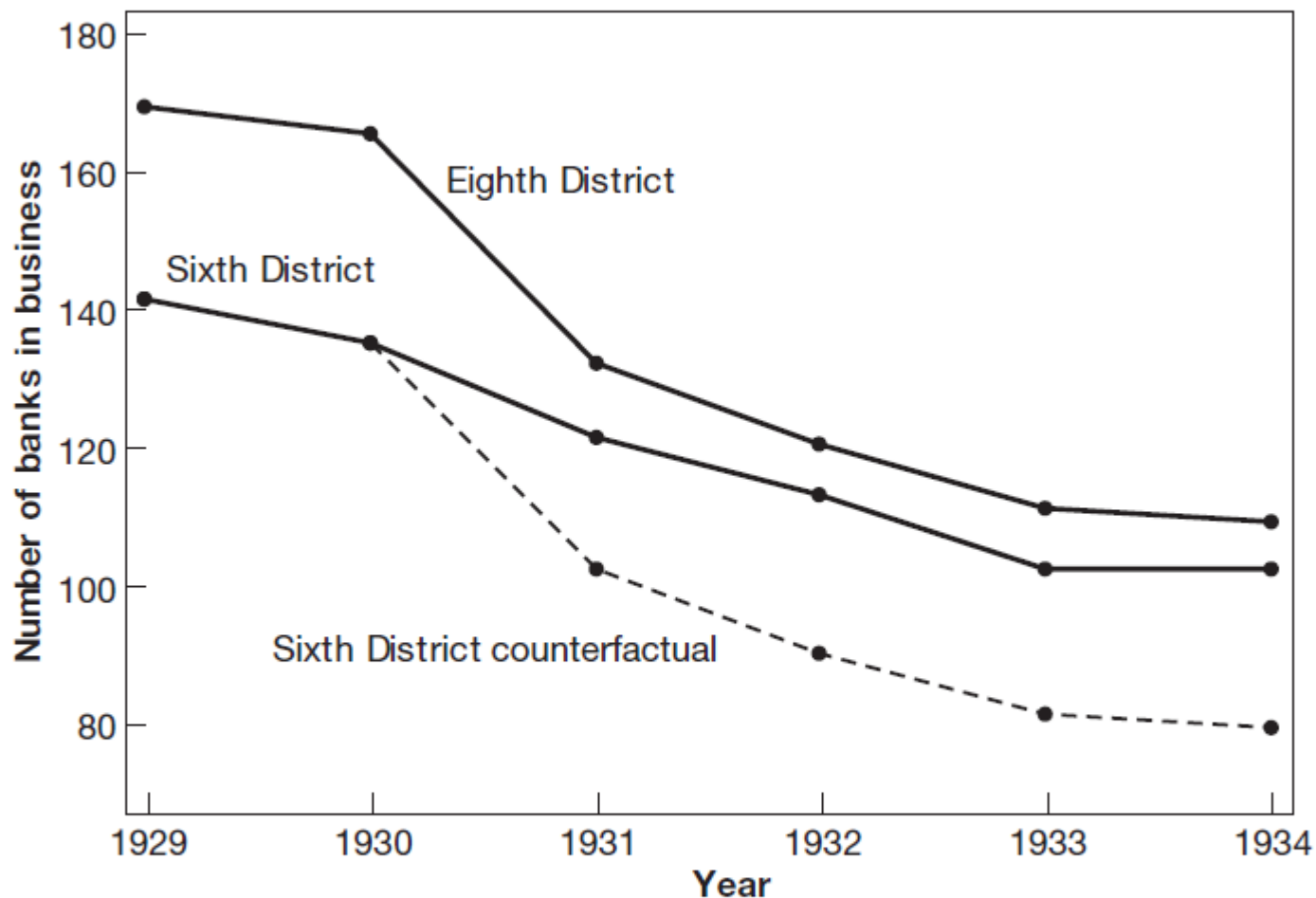
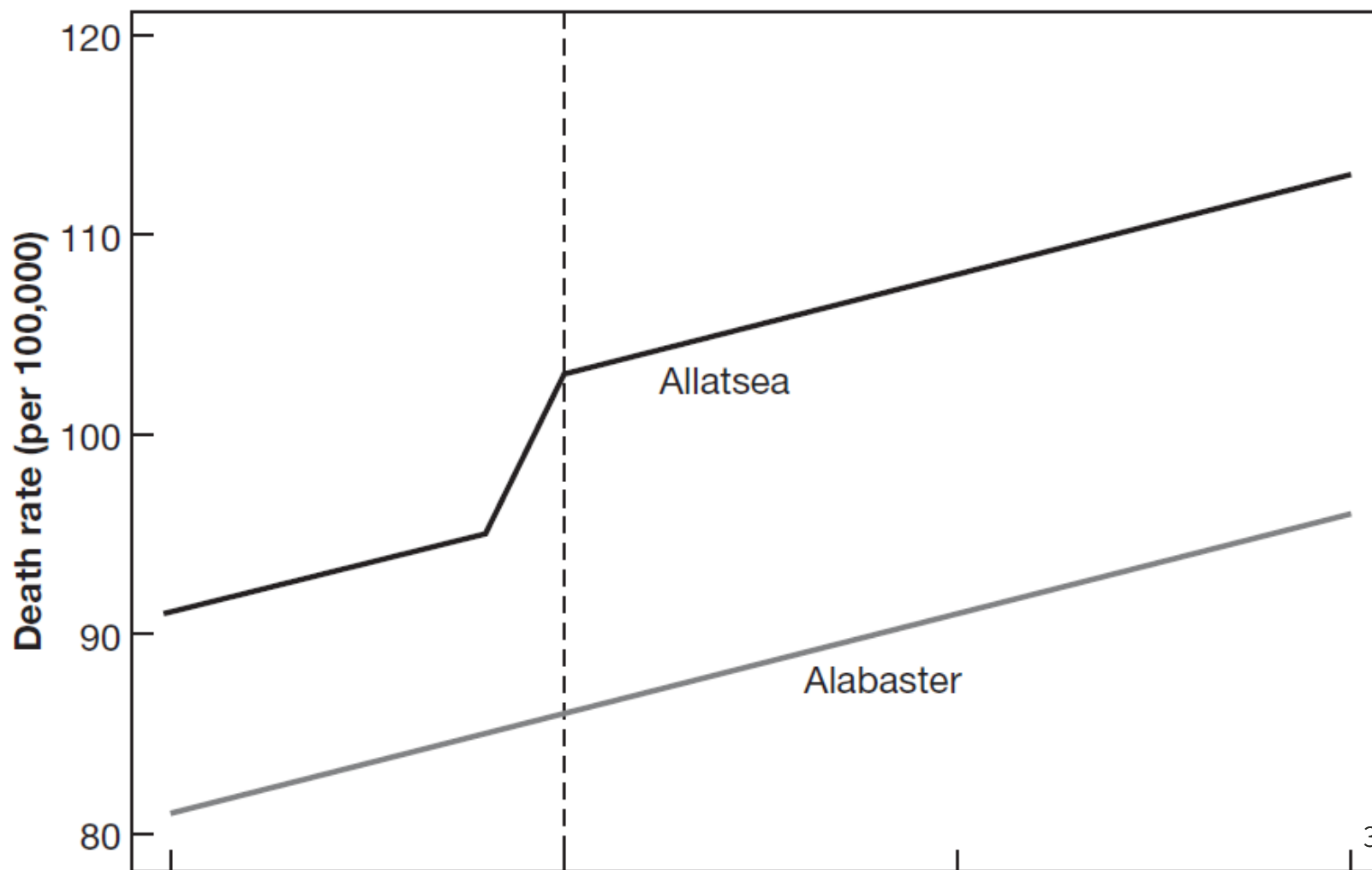


FIGURE 5.4  
An MLDA effect in states with parallel trends



## Assumptions:

We made an assumption when we were talking about controls and counterfactuals. Specifically, we made the assumption that the *unobserved time trend* would affect both the **treatment** and the **control** the same.

Another way of saying this is (from MM):

Absent any policy differences (treatment), the **trend** in the control group is what we would have seen in the treatment group.

Or, from Wooldridge:

The average trends would be the same for the control and treatment in the absence of treatment



This is the **parallel trends assumption**

In the absence of treatment, the trends in both groups would have been the same.

## We do not get to observe this

We never see the treatment group's untreated trend after treatment.

- That's why this is an assumption.

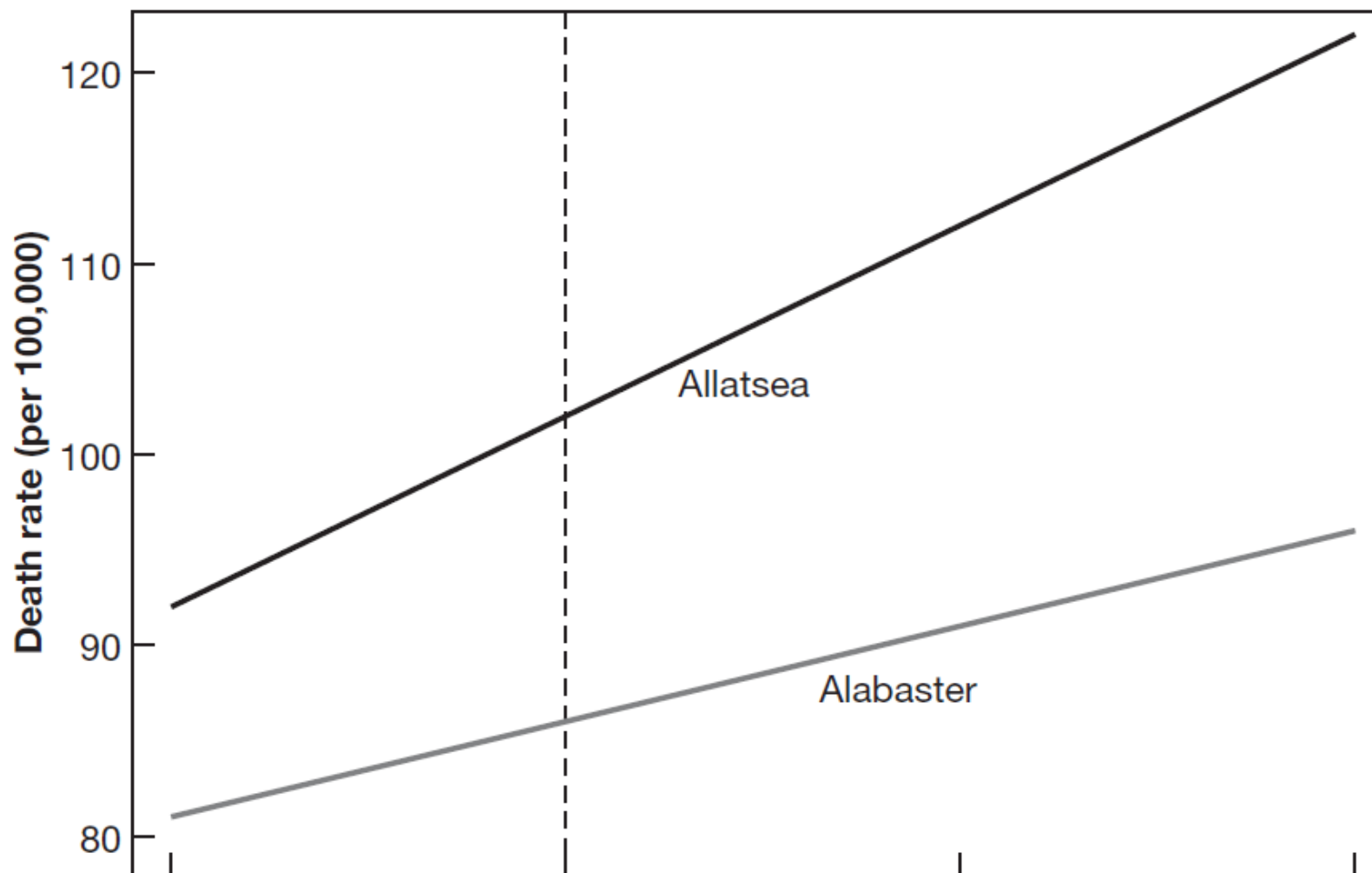
## But we do see the pre-treatment trend

Which may help us make a convincing argument

- But can never guarantee that we our parallel trend assumption holds

FIGURE 5.5

A spurious MLDA effect in states where trends are not parallel



## Restating the parallel trends assumption a different way:

We have to assume that there are no time-variant,  $i$ -specific effects. We can have constant  $i$ -specific fixed effects, we can have  $t$ -specific fixed effects (trends), but we can't have something that changes over time (e.g. changes right at the treatment time  $\tau$ ) and *only* happens to the treated units.

## The parallel trends assumption is our *identifying assumption*

If it does not hold, we are not identifying the ATE as we cannot write it in terms of population parameters we can estimate from the data.

If it does hold, then we do identify the *ATE* and it is  $\delta_{DiD}$ .

## More flexible DID

We don't have to limit our fixed effects to  $TREAT_i$ . We can be more flexible:

$$y_{it} = \beta_0 + \beta_1 POST_t + \beta_2 TREAT_i + \beta_3 POST_t * TREAT_i + \phi_i + u_{it}$$

Here,  $\beta_2$  wouldn't be identified because it just a combination of  $\phi_i$ 's (remember our *full rank* assumption). Dropping it out results in a **DID** estimator that lets every treated (and control) group have its own intercept instead of one common one for all treated  $i$ .  $\beta_3$  is the **DID** estimator, has the same interpretation, and is identified.

## Time-varying treatment

What if we have a treatment that isn't the same time period for all  $i$ ?

$$y_{it} = \beta_0 + \beta_1 D_t + \beta_2 TREAT_i + \beta_3 TREAT_i * D_t + u_{it}$$

This wouldn't work because there isn't a common  $D_t$ , and if we used  $D_{it}$ , we'd be estimating  $\beta_3$ , which is already  $it$ -specific!

But:

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 TREAT_i * D_t + \delta_t + u_{it}$$

Where  $\delta_t$  flexibly controls for a common time trend. This would result in  $\beta_2$  being the *DID* estimate.

Finally, a very flexible:

$$y_{it} = \beta_0 + \beta_1 D_{it} + \delta_t + \phi_i + u_{it}$$

Where  $D_{it} = 1$  during a post-treatment period in a treated  $i$ .

## *i*-specific trends

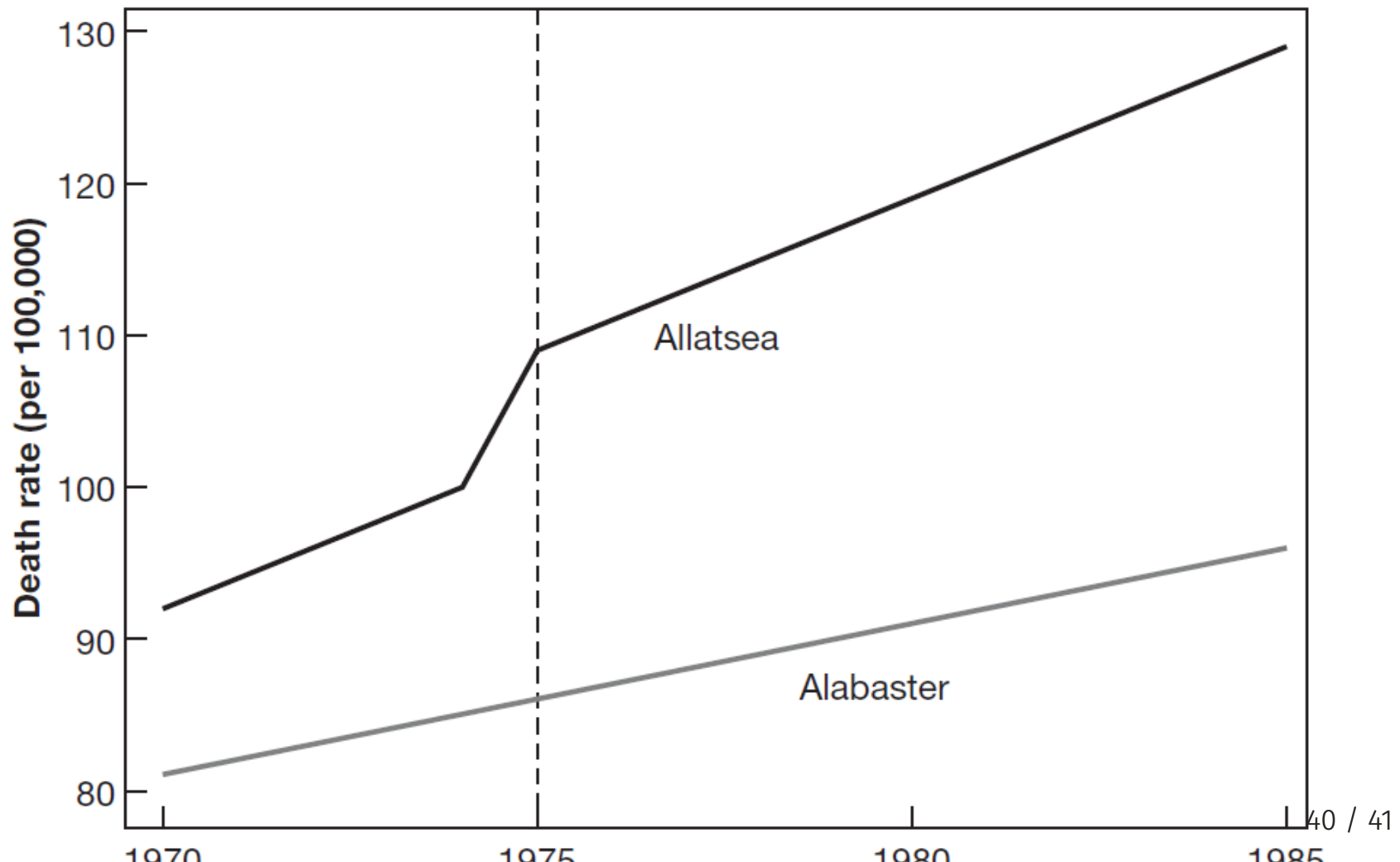
We can relax the parallel trends assumption a little bit by assuming that we can estimate different time trends (using an interaction) and that these trends would *continue to hold even in the absence of treatment*. This is a "parallel trend in trends" assumption.

$$y_{it} = \beta_0 + \beta_1 D_{it} + \phi_i + \underbrace{\sum_{t=1}^T \delta_i * t}_{i\text{-specific time trend}} + u_{it}$$

So, we can have non-parallel trends, but we **do** have to assume that those trends **continue** in the absence of treatment.

FIGURE 5.6

A real MLDA effect, visible even though trends are not parallel





For more examples, see Video 5.1 at

- <https://mattmasten.github.io/bootcamp/>

General panel methods

- Serial correlation
- Clustered std. errors

Synthetic Controls

- Reading is a pair of papers that I will post that will be annotated
- We can ignore some of the technical details.