Lecture 13: Panel data III

PPHA 34600

Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: The event study design

In the general version of this model, we include pre-treatment "effects":

$$Y_{it} = \sum_{s=-R}^{S} \tau_s D_i \times \mathbf{1}[\text{periods to treatment} = s]_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

- (Note the number of pre-treatment periods, R, and the number of post-treatment periods, S, need not be the same)

From last time: The event study design

In the general version of this model, we include pre-treatment "effects":

$$Y_{it} = \sum_{s=-R}^{S} \tau_s D_i \times \mathbf{1}[\text{periods to treatment} = s]_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

- (Note the number of pre-treatment periods, R, and the number of post-treatment periods, S, need not be the same)

What's so great about this design?

- This "lines up" treatment at the same time for everyone
- We can still use fixed effects to soak up confounders
- We get a (partial) test of the identifying assumption

From last time: The event study design

In the general version of this model, we include pre-treatment "effects":

$$Y_{it} = \sum_{s=-R}^{S} \tau_s D_i \times \mathbf{1}[\text{periods to treatment} = s]_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

- (Note the number of pre-treatment periods, R, and the number of post-treatment periods, S, need not be the same)

What's so great about this design?

- This "lines up" treatment at the same time for everyone
- We can still use fixed effects to soak up confounders
- We get a (partial) test of the identifying assumption
  - $\rightarrow$ We want pre-treatment τ_s's to be centered on 0 and not trending

Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

What would be a problem for this assumption?

Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

What would be a problem for this assumption?

- Selection into treatment:
  - Treated and untreated individuals on different trends

Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

What would be a problem for this assumption?

- Selection into treatment:
  - Treated and untreated individuals on different trends
- Coincident treatments
  - If a treated unit does two things at once, we can't cleanly estimate treatment

Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

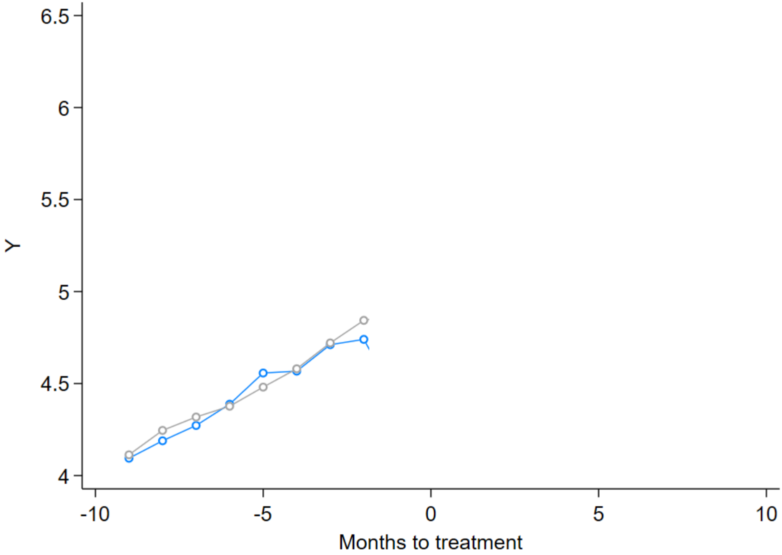
For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

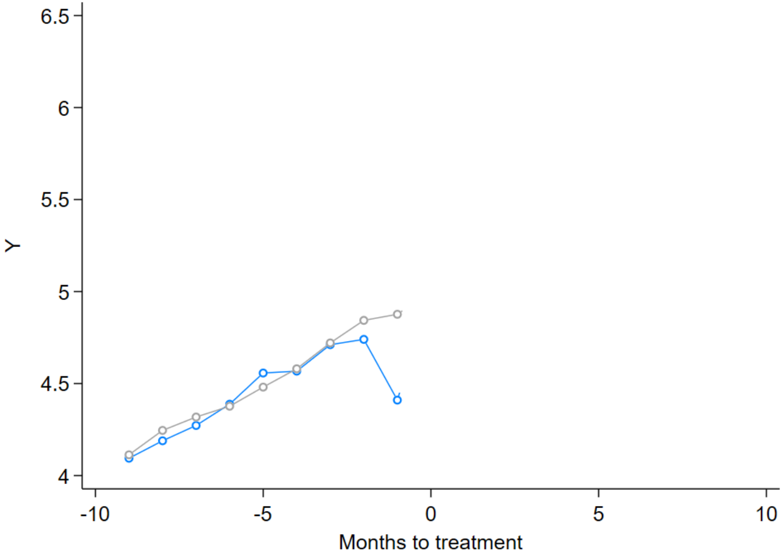
What would be a problem for this assumption?

- Selection into treatment:
  - Treated and untreated individuals on different trends
- Coincident treatments
  - If a treated unit does two things at once, we can't cleanly estimate treatment
- Anticipatory effects and the Ashenfelter dip

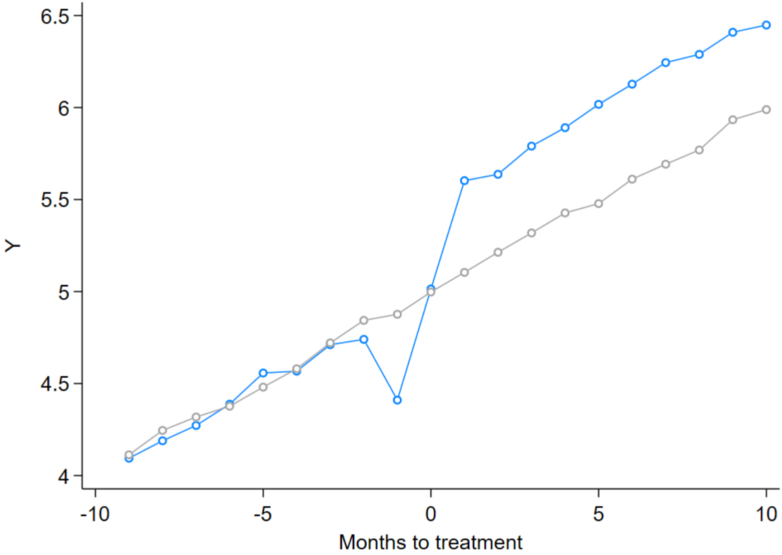
Threats to identification: Ashenfelter dip



Threats to identification: Ashenfelter dip



Threats to identification: Ashenfelter dip



Threats to identification

Recall our general fixed effects model:

$$Y_{it} = \tau D_{it} + \alpha_i + \delta_t + \beta X_{it} + \varepsilon_{it}$$

For the FE model to work, we need:

$$E[\varepsilon_{it} | \alpha_i, \delta_t, X_{it}] = 0$$

What would be a problem for this assumption?

- Selection into treatment:
  - Treated and untreated individuals on different trends
- Coincident treatments
  - If a treated unit does two things at once, we can't cleanly estimate treatment
- Anticipatory effects and the Ashenfelter dip

A note on standard errors

OLS standard errors will be wrong with panel data:

- The OLS standard error assumes each observation is an IID draw

A note on standard errors

OLS standard errors will be wrong with panel data:

- The OLS standard error assumes each observation is an IID draw
- But of course, unit i is correlated with itself over time
- And some of this correlation sits in the error term
- Intuition: you have less data than you think you've got!

A note on standard errors

OLS standard errors will be wrong with panel data:

- The OLS standard error assumes each observation is an IID draw
- But of course, unit i is correlated with itself over time
- And some of this correlation sits in the error term
- Intuition: you have less data than you think you've got!

$\rightarrow$ Clustered standard errors

  - These allow for correlations within cluster (eg, within individual)
  - Clustered SEs are usually (not always) bigger than OLS SEs
  - You need at least $\approx$ 45 clusters for this to work

A note on standard errors

OLS standard errors will be wrong with panel data:

- The OLS standard error assumes each observation is an IID draw
- But of course, unit *i* is correlated with itself over time
- And some of this correlation sits in the error term
- Intuition: you have less data than you think you've got!

→ Clustered standard errors

  - These allow for correlations within cluster (eg, within individual)
  - Clustered SEs are usually (not always) bigger than OLS SEs
  - You need at least $\approx$ 45 clusters for this to work

→ Typical panel data: cluster at the unit level

Even more differences

Differences-in-differences-in-differences (DDD)

Our standard DD estimator was:

$$\hat{\tau}^{DD} = (\bar{Y}(treat, post) - \bar{Y}(treat, pre)) - (\bar{Y}(untreat, post) - \bar{Y}(untreat, pre))$$

Differences-in-differences-in-differences (DDD)

Our standard DD estimator was:

$$\hat{\tau}^{DD} = (\bar{Y}(treat, post) - \bar{Y}(treat, pre)) - (\bar{Y}(untreat, post) - \bar{Y}(untreat, pre))$$

We sometimes worry about:

- Non-parallel counterfactual trends
- Coincident treatments

Adding a third comparison can help with this:

- We use a comparison between affected and unaffected units or times

Differences-in-differences-in-differences (DDD)

Using a third difference can sometimes help with identification:

As an example, think about:

- Outcome: SO_x emissions
- Treatment: State-level regulations
- Timing: Pre/post
- *Additional grouping:* Industry type (some industries don't emit SO_x)

Differences-in-differences-in-differences (DDD)

Using a third difference can sometimes help with identification:

As an example, think about:

- Outcome: SO_x emissions
- Treatment: State-level regulations
- Timing: Pre/post
- *Additional grouping:* Industry type (some industries don't emit SO_x)

Rather than just comparing regulated vs. unregulated states pre vs. post, we can layer on emitting vs. non-emitting industries, too

→ This alleviates concerns about non-parallel trends in regulated vs. non-regulated states

Differences-in-differences-in-differences (DDD)

Using a third difference can sometimes help with identification:

As an example, think about:

- Outcome: SO_x emissions
- Treatment: State-level regulations
- Timing: Pre/post
- *Additional grouping:* Industry type (some industries don't emit SO_x)

Rather than just comparing regulated vs. unregulated states pre vs. post, we can layer on emitting vs. non-emitting industries, too

→ This alleviates concerns about non-parallel trends in regulated vs. non-regulated states

→ What is the identifying assumption now?

Differences-in-differences-in-differences (DDD)

Using a third difference can sometimes help with identification:

As an example, think about:

- Outcome: SO_x emissions
- Treatment: State-level regulations
- Timing: Pre/post
- *Additional grouping:* Industry type (some industries don't emit SO_x)

Rather than just comparing regulated vs. unregulated states pre vs. post, we can layer on emitting vs. non-emitting industries, too

→ This alleviates concerns about non-parallel trends in regulated vs. non-regulated states

→ What is the identifying assumption now?

→ The difference in trends over time between regulated and unregulated states is similar for emitters and non-emitters

Differences-in-differences-in-differences (DDD)

We can write this estimator formally as:

$$\hat{\tau}^{DDD} = [(\bar{Y}(treat, post, affected) - \bar{Y}(treat, pre, affected))$$
$$-\underbrace{(\bar{Y}(untreat, post, affected) - \bar{Y}(untreat, pre, affected))]}_{\text{first two lines are our normal DD}}$$

Differences-in-differences-in-differences (DDD)

We can write this estimator formally as:

$$
\begin{aligned}
\hat{\tau}^{DDD} = & [(\bar{Y}(treat, post, affected) - \bar{Y}(treat, pre, affected)) \\
& - \underbrace{(\bar{Y}(untreat, post, affected) - \bar{Y}(untreat, pre, affected))]}_{\text{first two lines are our normal DD}} \\
& -[(\bar{Y}(treat, post, unaffected) - \bar{Y}(treat, pre, unaffected)) \\
& \underbrace{-(\bar{Y}(untreat, post, unaffected) - \bar{Y}(untreat, pre, unaffected))]}_{\text{next two lines are the new difference}}
\end{aligned}
$$

Differences-in-differences-in-differences (DDD)

We can write this estimator formally as:

$$\hat{\tau}^{DDD} = [(\bar{Y}(treat, post, affected) - \bar{Y}(treat, pre, affected))$$
$$\underbrace{-(\bar{Y}(untreat, post, affected) - \bar{Y}(untreat, pre, affected))]}_{\text{first two lines are our normal DD}}$$
$$-[(\bar{Y}(treat, post, unaffected) - \bar{Y}(treat, pre, unaffected))$$
$$\underbrace{-(\bar{Y}(untreat, post, unaffected) - \bar{Y}(untreat, pre, unaffected))]}_{\text{next two lines are the new difference}}$$

This relaxes our identifying assumption:

- **DD:** Treated and untreated groups are on similar (counterfactual) trends

Differences-in-differences-in-differences (DDD)

We can write this estimator formally as:

$$\hat{\tau}^{DDD} = [(\bar{Y}(treat, post, affected) - \bar{Y}(treat, pre, affected))$$

$$\underbrace{-(\bar{Y}(untreat, post, affected) - \bar{Y}(untreat, pre, affected))]}_{\text{first two lines are our normal DD}}$$

$$-[(\bar{Y}(treat, post, unaffected) - \bar{Y}(treat, pre, unaffected))$$

$$\underbrace{-(\bar{Y}(untreat, post, unaffected) - \bar{Y}(untreat, pre, unaffected))]}_{\text{next two lines are the new difference}}$$

This relaxes our identifying assumption:

- **DD:** Treated and untreated groups are on similar (counterfactual) trends
- **DDD:** Differences in trends between treated and untreated impact affected and unaffected groups similarly

Differences-in-differences-in-differences (DDD)

Affected group ($Affected_j = 1$)

	Post	Pre	Difference
Treated	a	b	$a - b$
Untreated	c	d	$c - d$
Difference	$a - c$	$b - d$	$(a - b) - (c - d)$

Differences-in-differences-in-differences (DDD)

Affected group ($Affected_j = 1$)

	Post	Pre	Difference
Treated	a	b	$a - b$
Untreated	c	d	$c - d$
Difference	$a - c$	$b - d$	$(a - b) - (c - d)$

Unaffected group ($Affected_j = 0$)

	Post	Pre	Difference
Treated	e	f	$e - f$
Untreated	g	h	$g - h$
Difference	$e - g$	$f - h$	$(e - f) - (g - h)$

Differences-in-differences-in-differences (DDD)

Affected group ($Affected_j = 1$)

	Post	Pre	Difference
Treated	a	b	$a - b$
Untreated	c	d	$c - d$
Difference	$a - c$	$b - d$	$(a - b) - (c - d)$

Unaffected group ($Affected_j = 0$)

	Post	Pre	Difference
Treated	e	f	$e - f$
Untreated	g	h	$g - h$
Difference	$e - g$	$f - h$	$(e - f) - (g - h)$

DDD estimate is the differences between the two DID estimates:

$$[(a - b) - (c - d)] - [(e - f) - (g - h)].$$

Differences-in-differences-in-differences (DDD)

Just like the DD estimator, we can do the DDD estimator via regression:

$$
\begin{aligned}
Y_{ijt} = \beta_0 &+ \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Affected_j + \beta_4 (Treat_i \times Post_t) \\
&+ \beta_5 (Post_t \times Affected_j) + \beta_6 (Treat_i \times Affected_j) \\
&+ \tau (Treat_i \times Post_t \times Affected_j) + \varepsilon_{ijt}
\end{aligned}
$$

Differences-in-differences-in-differences (DDD)

Just like the DD estimator, we can do the DDD estimator via regression:

$$
\begin{aligned}
Y_{ijt} = \beta_0 &+ \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Affected_j + \beta_4 (Treat_i \times Post_t) \\
&+ \beta_5 (Post_t \times Affected_j) + \beta_6 (Treat_i \times Affected_j) \\
&+ \tau (Treat_i \times Post_t \times Affected_j) + \varepsilon_{ijt}
\end{aligned}
$$

- You can do this with fixed effects too
- You just can't have a $Treat_i \times Post_t \times Affected_j$ FE

Differences-in-differences-in-differences (DDD)

Just like the DD estimator, we can do the DDD estimator via regression:

$$\begin{aligned} Y_{ijt} = \beta_0 + \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Affected_j + \beta_4 (Treat_i \times Post_t) \\ + \beta_5 (Post_t \times Affected_j) + \beta_6 (Treat_i \times Affected_j) \\ + \tau (Treat_i \times Post_t \times Affected_j) + \varepsilon_{ijt} \end{aligned}$$

- You can do this with fixed effects too
- You just can't have a $Treat_i \times Post_t \times Affected_j$ FE

For fun on your own: Show that τ actually gets $\hat{\tau}^{DDD}$

Recap

TL;DR:

1. We like the difference-in-differences approach a lot
2. We discussed estimation with fixed effects
3. And covered the event study and distributed lag versions