Lecture 11:
Panel data I

PPHA 34600

Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: finishing IV

Z_i is a valid instrument when the following are satisfied:

1. **First stage**: $Cov(Z_i, D_i) \neq 0$
2. **Exclusion restriction**: $Cov(Z_i, \varepsilon_i) = 0$

When we have these two conditions, we can...:

- Estimate causal effects
- ...but only for compliers!

Further down the SOU rabbit hole

So far, we've focused on data across units. Now we'll add time:

- **Cross-sectional data:**
  - What we've been using
  - Observations across units at a single point in time

Further down the SOU rabbit hole

So far, we've focused on data across units. Now we'll add time:

- **Cross-sectional data:**
  - What we've been using
  - Observations across units at a single point in time
- **Time series data:**
  - Observations on a single unit over time

Further down the SOU rabbit hole

So far, we've focused on data across units. Now we'll add time:

- **Cross-sectional data:**
  - What we've been using
  - Observations across units at a single point in time
- **Time series data:**
  - Observations on a single unit over time
- **Repeated cross-section data:**
  - Repeated sampling of different units over time

Further down the SOU rabbit hole

So far, we've focused on data across units. Now we'll add time:

- **Cross-sectional data:**
  - What we've been using
  - Observations across units at a single point in time
- **Time series data:**
  - Observations on a single unit over time
- **Repeated cross-section data:**
  - Repeated sampling of different units over time
- **Panel data:**
  - Multiple observations of the same unit over time

Why is data over time useful?

We've spilled a lot of ink on the selection problem:

- To isolate the effect of D_i, we need potential outcomes to be the same among treated and untreated units

Why is data over time useful?

We've spilled a lot of ink on the selection problem:

- To isolate the effect of D_i, we need potential outcomes to be the same among treated and untreated units
- With cross-sectional data, this is fundamentally tricky:
    - People, firms, households, etc are different from one another in lots of ways
    - Getting a clean comparison means separating τ from all of these differences

Why is data over time useful?

We've spilled a lot of ink on the selection problem:

- To isolate the effect of D_i, we need potential outcomes to be the same among treated and untreated units
- With cross-sectional data, this is fundamentally tricky:
  - People, firms, households, etc are different from one another in lots of ways
  - Getting a clean comparison means separating τ from all of these differences
- Enter time series data:
  - **Fundamental insight:**
    Rather than comparing i to j, compare i in t to i in $t-1$
  - In this formulation, i serves as a control for itself
  - i am much more similar to myself yesterday than i am to j

Making time-series comparisons

Consider a setting with only one unit:

- We now denote our outcome as $Y_t(D_t)$ (no subscript: only one unit)
- As usual, we want to estimate $\tau^{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$
- But we can't observe both $Y_{t=1}(D_{t=1} = 1)$ and $Y_{t=1}(D_{t=1} = 0)$
  - $\rightarrow$ Remember that fundamental problem of causal inference?

Making time-series comparisons

Consider a setting with only one unit:

- We now denote our outcome as $Y_t(D_t)$ (no subscript: only one unit)
- As usual, we want to estimate $\tau^{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$
- But we can't observe both $Y_{t=1}(D_{t=1} = 1)$ and $Y_{t=1}(D_{t=1} = 0)$
  - $\rightarrow$ Remember that fundamental problem of causal inference?
- Instead, we look for periods before and after treatment begins
- Suppose in $t = 0$, $D_{t=0} = 0$, and in $t = 1$, $D_{t=1} = 1$
- Then we can estimate:

$$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0}$$

Making time-series comparisons

Consider a setting with only one unit:

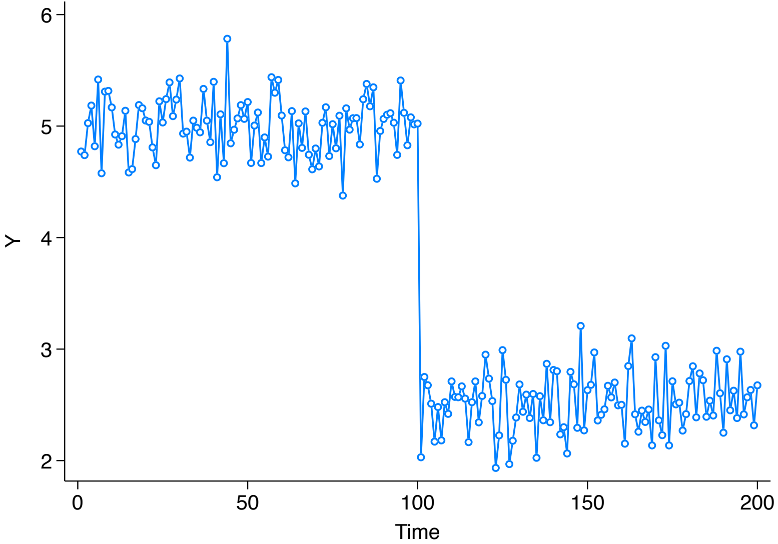
- We now denote our outcome as $Y_t(D_t)$ (no subscript: only one unit)
- As usual, we want to estimate $\tau^{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$
- But we can't observe both $Y_{t=1}(D_{t=1} = 1)$ and $Y_{t=1}(D_{t=1} = 0)$
  - → Remember that fundamental problem of causal inference?
- Instead, we look for periods before and after treatment begins
- Suppose in $t = 0$, $D_{t=0} = 0$, and in $t = 1$, $D_{t=1} = 1$
- Then we can estimate:

$$\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0}$$

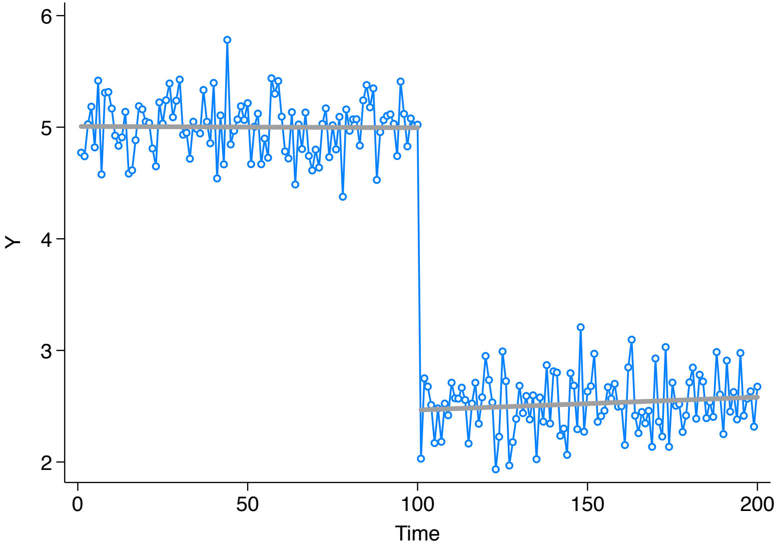
- We can also extend this to many periods:

$$\hat{\tau}^{TS} = \bar{Y}_{t \in \text{post}} - \bar{Y}_{t \in \text{pre}}$$

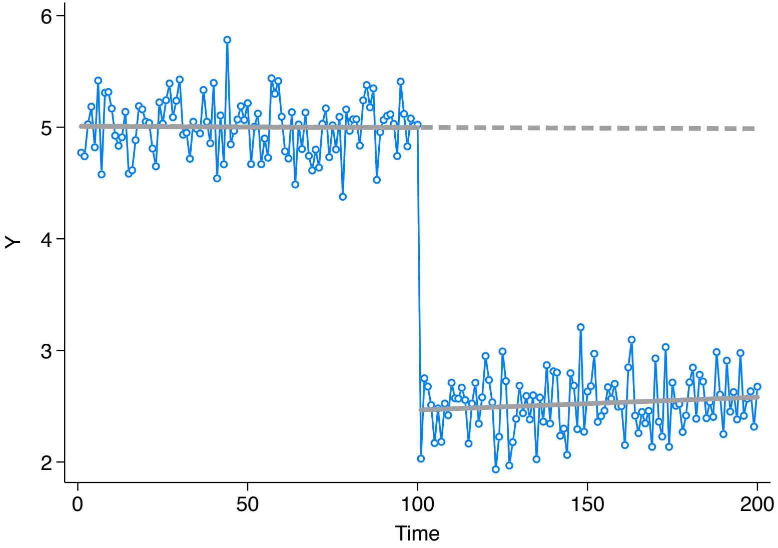
Time series, visually



Time series, visually



Time series, visually



What's good about the time series?

This time series approach compares unit i to itself over time:

- Consider a simple data-generating process:

$$Y_{it} = \beta_i X_i$$

- In this model, we have time-invariant characteristics (X_i)

What's good about the time series?

This time series approach compares unit i to itself over time:

- Consider a simple data-generating process:

$$Y_{it} = \beta_i X_i$$

- In this model, we have time-invariant characteristics (X_i)
- Now add treatment:

$$Y_{it} = \tau D_{it} + \beta X_i$$

  where $D_{it} = 0$ in $t = 0$ and $D_{it} = 1$ in $t = 1$
- We want to separate treatment from the other characteristics

What's good about the time series?

This time series approach compares unit i to itself over time:

- Consider a simple data-generating process:

$$Y_{it} = \beta_i X_i$$

- In this model, we have time-invariant characteristics (X_i)
- Now add treatment:

$$Y_{it} = \tau D_{it} + \beta X_i$$

  where $D_{it} = 0$ in $t = 0$ and $D_{it} = 1$ in $t = 1$
- We want to separate treatment from the other characteristics
- Enter the difference estimator:

$$\begin{aligned} Y_{i,t=1} - Y_{i,t=0} &= \tau(D_{i,t=1} - D_{i,t=0}) + \beta(X_i - X_i) \\ &= \tau(D_{i,t=1} - D_{i,t=0}) \\ &= \tau(1 - 0) \\ &= \tau \end{aligned}$$

We can pull the same trick with unobservables

- Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \gamma U_i$$

- In this model, we have time-invariant observable characteristics (X_i)
- We also have time-invariant unobserved characteristics (U_i)

We can pull the same trick with unobservables

- Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \gamma U_i$$

- In this model, we have time-invariant observable characteristics (X_i)
- We also have time-invariant unobserved characteristics (U_i)
- Now add treatment:

$$Y_{it} = \tau D_{it} + \beta X_i + \gamma U_i$$

  where $D_{it} = 0$ in $t = 0$ and $D_{it} = 1$ in $t = 1$
- We want to separate treatment from the other characteristics

We can pull the same trick with unobservables

- Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \gamma U_i$$

- In this model, we have time-invariant observable characteristics (X_i)
- We also have time-invariant unobserved characteristics (U_i)
- Now add treatment:

$$Y_{it} = \tau D_{it} + \beta X_i + \gamma U_i$$

  where $D_{it} = 0$ in $t = 0$ and $D_{it} = 1$ in $t = 1$
- We want to separate treatment from the other characteristics
- Enter the difference estimator:

$$\begin{aligned} Y_{i,t=1} - Y_{i,t=0} &= \tau(D_{i,t=1} - D_{i,t=0}) + \beta(X_i - X_i) + \gamma(U_i - U_i) \\ &= \tau(D_{i,t=1} - D_{i,t=0}) \\ &= \tau \end{aligned}$$

Time series: identifying assumptions

In order for

$$\hat{\tau}^{TS} = Y_{i,t=1} - Y_{i,t=0}$$

to recover the true τ, we need an important assumption.

Consider the DGP:

$$Y_{it} = \tau D_{it} + \beta X_i + \gamma U_i + \delta V_{it}$$

where V_{it} is a set of observed and unobserved **time-varying** characteristics

Time series: identifying assumptions

In order for

$$\hat{\tau}^{TS} = Y_{i,t=1} - Y_{i,t=0}$$

to recover the true τ, we need an important assumption.

Consider the DGP:

$$Y_{it} = \tau D_{it} + \beta X_i + \gamma U_i + \delta V_{it}$$

where V_{it} is a set of observed and unobserved **time-varying** characteristics

In this case,

$$\hat{\tau}^{TS} = Y_{i,t=1} - Y_{i,t=0} = \tau + \delta(V_{i,t=1} - V_{i,t=0})$$

In order for $\hat{\tau}^{TS}$ to equal τ, we need $\delta = 0$ or $V_{i,t=1} = V_{i,t=0} (= V_i)$

Time series: identifying assumptions

In order for

$$\hat{\tau}^{TS} = Y_{i,t=1} - Y_{i,t=0}$$

to recover the true τ, we need an important assumption.

Consider the DGP:

$$Y_{it} = \tau D_{it} + \beta X_i + \gamma U_i + \delta V_{it}$$

where V_{it} is a set of observed and unobserved **time-varying** characteristics

In this case,

$$\hat{\tau}^{TS} = Y_{i,t=1} - Y_{i,t=0} = \tau + \delta(V_{i,t=1} - V_{i,t=0})$$

In order for $\hat{\tau}^{TS}$ to equal τ, we need $\delta = 0$ or $V_{i,t=1} = V_{i,t=0} (= V_i)$

→ **Any time-varying variables will create bias in $\hat{\tau}^{TS}$**

→ Also true for observables: we can't separate D_{it} from coincident V_{it}

Time series: identifying assumptions

Another way to think about this assumption:

- Y_i would be unchanged in the absence of treatment
- To see this, we can write:

$$\hat{\tau}^{TS} = Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)$$

Time series: identifying assumptions

Another way to think about this assumption:

- Y_i would be unchanged in the absence of treatment
- To see this, we can write:

$$\hat{\tau}^{TS} = Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) + (Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

Time series: identifying assumptions

Another way to think about this assumption:

- Y_i would be unchanged in the absence of treatment
- To see this, we can write:

$$\hat{\tau}^{TS} = Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) + (Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + (Y_{i,t=1}(0) - Y_{i,t=0}(0))$$

Time series: identifying assumptions

Another way to think about this assumption:

- Y_i would be unchanged in the absence of treatment
- To see this, we can write:

$$\hat{\tau}^{TS} = Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) + (Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + (Y_{i,t=1}(0) - Y_{i,t=0}(0))$$

$$\approx E[Y_{i,t=1}(1) - Y_{i,t=1}(0)|D_i = 1] + E[Y_{i,t=1}(0) - Y_{i,t=0}(0)|D_i = 1]$$

Time series: identifying assumptions

Another way to think about this assumption:

- Y_i would be unchanged in the absence of treatment
- To see this, we can write:

$$\hat{\tau}^{TS} = Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) + (Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

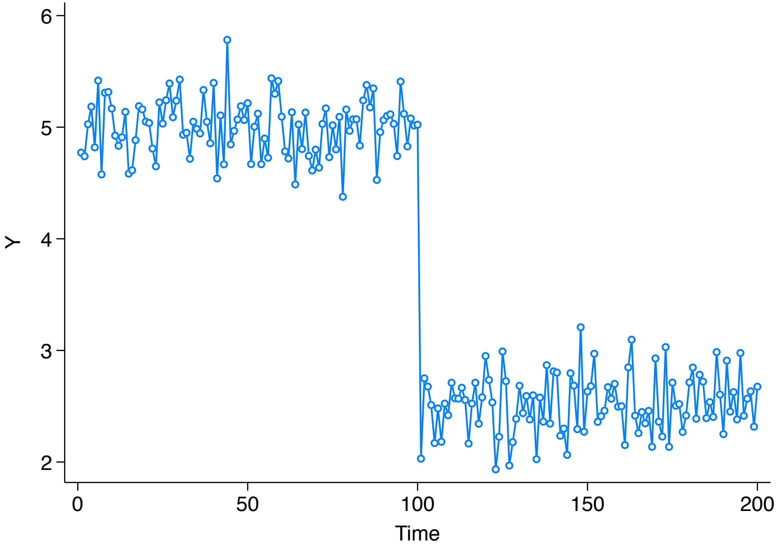
$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + (Y_{i,t=1}(0) - Y_{i,t=0}(0))$$

$$\approx E[Y_{i,t=1}(1) - Y_{i,t=1}(0)|D_i = 1] + E[Y_{i,t=1}(0) - Y_{i,t=0}(0)|D_i = 1]$$

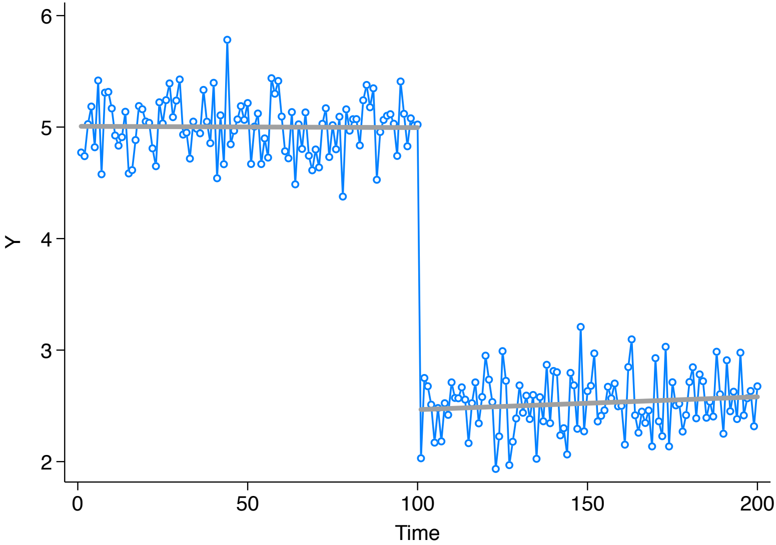
$$= \tau + \text{counterfactual trend}$$

We have to assume that the counterfactual trend is zero

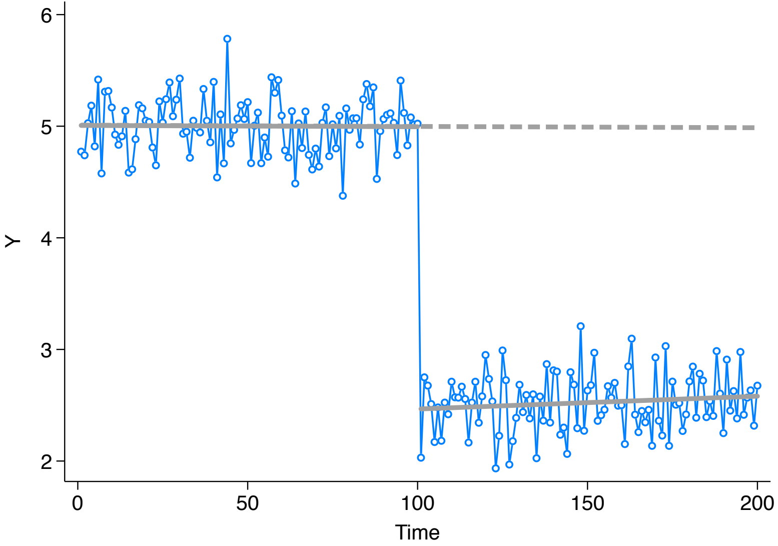
Identifying assumptions, visually



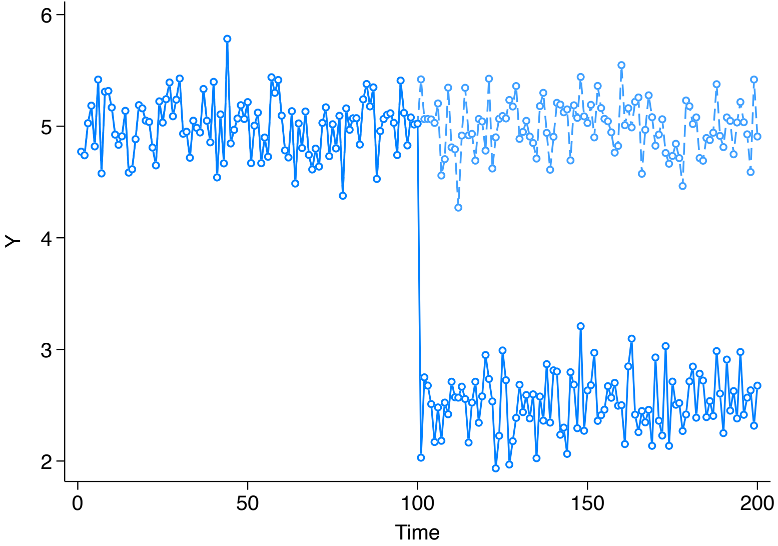
Identifying assumptions, visually



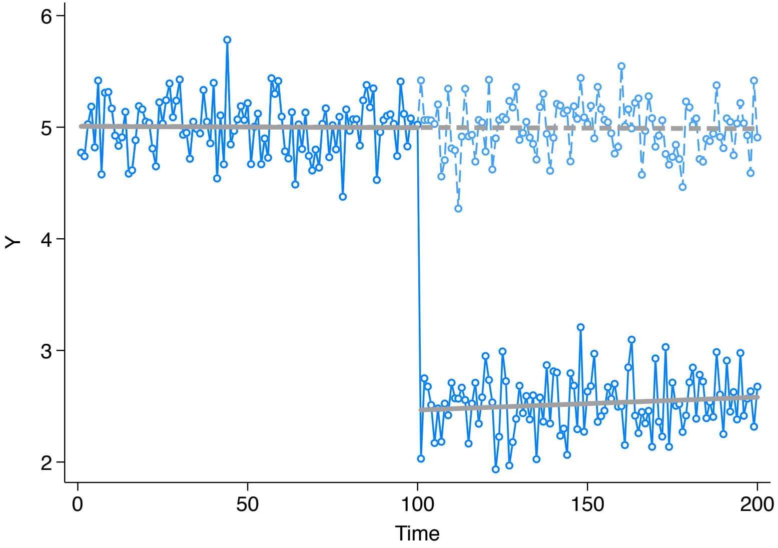
Identifying assumptions, visually



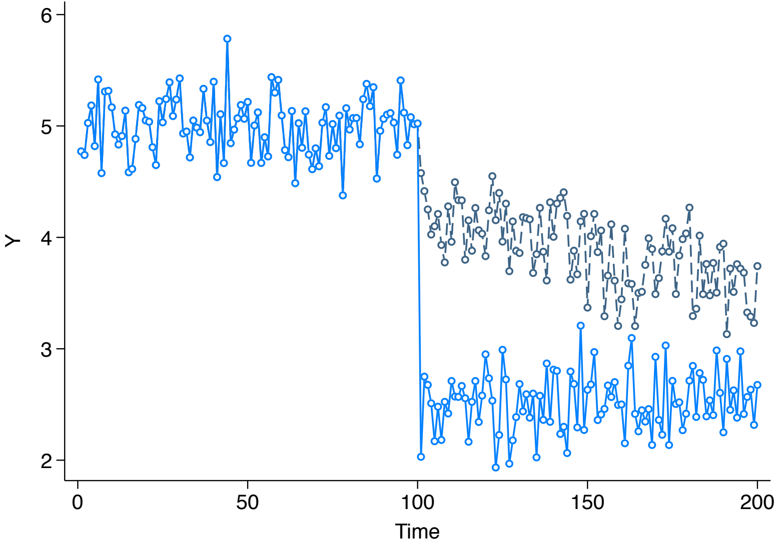
Identifying assumptions, visually



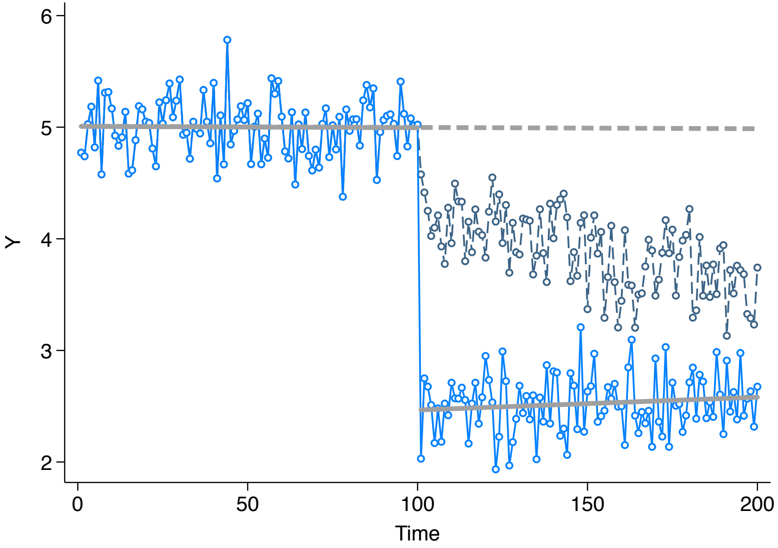
Identifying assumptions, visually



Identifying assumptions, visually



Identifying assumptions, visually



Validating the identifying assumption

We can never prove the identifying assumption:

- To do this, we would need to **observe the counterfactual**
- We can't do this → no test (☠)

Validating the identifying assumption

We can never prove the identifying assumption:

- To do this, we would need to **observe the counterfactual**
- We can't do this → no test (☠)
- You could try to look at pre-treatment trends...
- ...but this is likely unsatsifying

Validating the identifying assumption

We can never prove the identifying assumption:

- To do this, we would need to **observe the counterfactual**
- We can't do this $\rightarrow$ no test (☠)
- You could try to look at pre-treatment trends...
- ...but this is likely unsatsifying

Why does this matter?

- If the assumption is not satisfied, we will confound the trend with τ
- We cannot eliminate the trend with one time series

Two wrongs make a right?

Using panel data, we can combine two bad estimators into a good one:

- Our naive (cross-sectional) estimator:
  - Compare i to j (static)
  - Suffers from selection bias (i and j are systematically different)

Two wrongs make a right?

Using panel data, we can combine two bad estimators into a good one:

- Our naive (cross-sectional) estimator:
  - Compare i to j (static)
  - Suffers from selection bias (i and j are systematically different)
- The time-series estimator:
  - Compare i to itself over time
  - Suffers from time-varying unobservables
  - AKA non-zero trends

Two wrongs make a right?

Using panel data, we can combine two bad estimators into a good one:

- Our naive (cross-sectional) estimator:
  - Compare i to j (static)
  - Suffers from selection bias (i and j are systematically different)
- The time-series estimator:
  - Compare i to itself over time
  - Suffers from time-varying unobservables
  - AKA non-zero trends
- We can combine these into the **difference-in-differences** estimator:
  - Uses across-unit, within time, comparisons
  - And within-unit, across time, comparisons

Differences-in-differences (DD)

The problem with time series is the counterfactual trend:

- How would treated i have behaved in $t = 1$ without treatment?
- This is the missing counterfactual
- Using other individuals j, we can make a guess

Differences-in-differences (DD)

The problem with time series is the counterfactual trend:

- How would treated i have behaved in $t = 1$ without treatment?
- This is the missing counterfactual
- Using other individuals j, we can make a guess

The differences-in-differences estimator:

$$\hat{\tau}^{DD} = \hat{\tau}^{TS}_{D_i=1} - \hat{\tau}^{TS}_{D_i=0}$$

Differences-in-differences (DD)

The problem with time series is the counterfactual trend:

- How would treated i have behaved in $t = 1$ without treatment?
- This is the missing counterfactual
- Using other individuals j, we can make a guess

The differences-in-differences estimator:

$$\hat{\tau}^{DD} = \hat{\tau}^{TS}_{D_i=1} - \hat{\tau}^{TS}_{D_i=0}$$

$$= (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{i,t=0}(D_{jt} = 0))$$

Differences-in-differences (DD)

The problem with time series is the counterfactual trend:

- How would treated i have behaved in $t = 1$ without treatment?
- This is the missing counterfactual
- Using other individuals j, we can make a guess

The differences-in-differences estimator:

$$\hat{\tau}^{DD} = \hat{\tau}^{TS}_{D_i=1} - \hat{\tau}^{TS}_{D_i=0}$$

$$= (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{i,t=0}(D_{jt} = 0))$$

$$= (Y(D_i = 1, post) - Y(D_i = 1, pre)) - (Y(D_i = 0, post) - Y(D_i = 0, pre))$$

Differences-in-differences (DD)

The problem with time series is the counterfactual trend:

- How would treated i have behaved in $t = 1$ without treatment?
- This is the missing counterfactual
- Using other individuals j, we can make a guess

The differences-in-differences estimator:

$$\hat{\tau}^{DD} = \hat{\tau}^{TS}_{D_i=1} - \hat{\tau}^{TS}_{D_i=0}$$

$$= (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{i,t=0}(D_{jt} = 0))$$

$$= (Y(D_i = 1, post) - Y(D_i = 1, pre)) - (Y(D_i = 0, post) - Y(D_i = 0, pre))$$

This compares **treated** to **untreated** units **over time**

DD compares *i* to itself over time to *j* to itself over time

Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \delta S_t$$

- We have time-invariant characteristics (X_i) and time-varying S_t

DD compares i to itself over time to j to itself over time

Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \delta S_t$$

- We have time-invariant characteristics (X_i) and time-varying S_t
- Now add treatment (only for unit i):

$$Y_{it} = \tau D_{it} + \beta X_i + \delta S_t$$

$$Y_{jt} = \beta X_j + \delta S_t$$

DD compares *i* to itself over time to *j* to itself over time

Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \delta S_t$$

- We have time-invariant characteristics (X_i) and time-varying S_t
- Now add treatment (only for unit i):

$$Y_{it} = \tau D_{it} + \beta X_i + \delta S_t$$
$$Y_{jt} = \beta X_j + \delta S_t$$

- Enter the DD estimator:

$$Y_{i,t=1} - Y_{i,t=0} = \tau(D_{i,t=1} - D_{i,t=0}) + \beta(X_i - X_i) + \delta(S_{t=1} - S_{t=0})$$
$$Y_{j,t=1} - Y_{j,t=0} = \beta(X_j - X_j) + \delta(S_{t=1} - S_{t=0})$$

DD compares i to itself over time to j to itself over time

Consider a simple data-generating process:

$$Y_{it} = \beta X_i + \delta S_t$$

- We have time-invariant characteristics (X_i) and time-varying S_t
- Now add treatment (only for unit i):

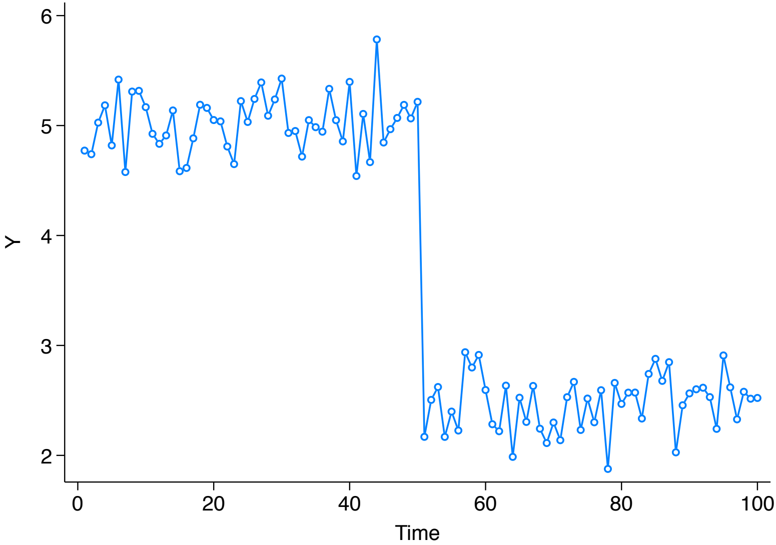
$$Y_{it} = \tau D_{it} + \beta X_i + \delta S_t$$
$$Y_{jt} = \beta X_j + \delta S_t$$

- Enter the DD estimator:

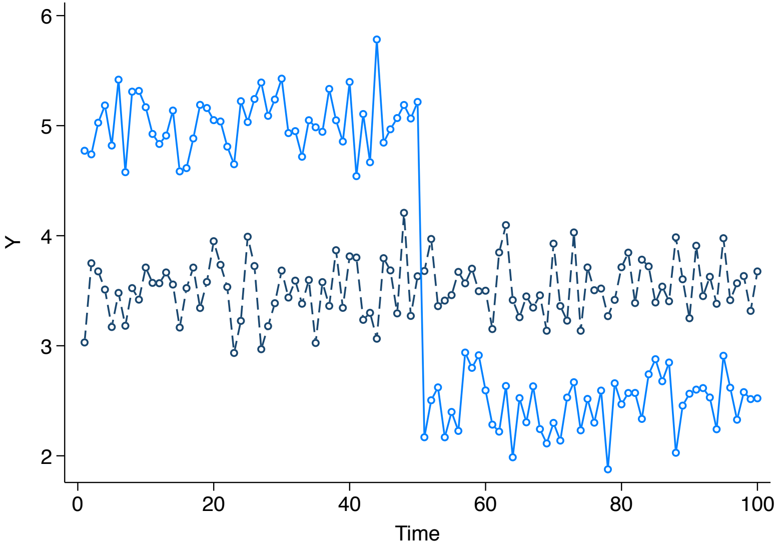
$$Y_{i,t=1} - Y_{i,t=0} = \tau(D_{i,t=1} - D_{i,t=0}) + \beta(X_i - X_i) + \delta(S_{t=1} - S_{t=0})$$
$$Y_{j,t=1} - Y_{j,t=0} = \beta(X_j - X_j) + \delta(S_{t=1} - S_{t=0})$$

$$\hat{\tau}^{DD} = (Y_{i,t=1} - Y_{i,t=0}) - (Y_{j,t=1} - Y_{j,t=0})$$
$$= \tau(D_{i,t=1} - D_{i,t=0}) = \tau$$

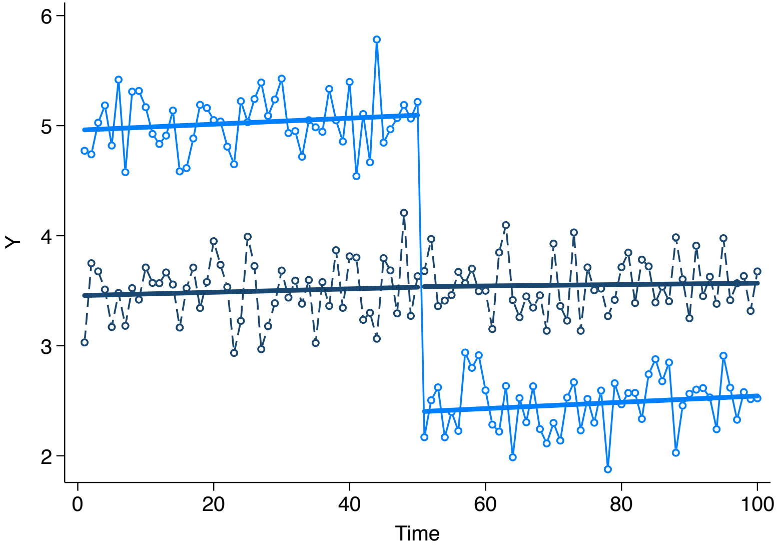
DD, visually



DD, visually



DD, visually



In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))$$

In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))$$

$$+(Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))$$

$$+(Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + [(Y_{i,t=1}(0) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))]$$

In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))$$

$$+(Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + [(Y_{i,t=1}(0) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))]$$

$$\approx E[Y_{i,t=1}(1) - Y_{i,t=1}(0)|D_i = 1] + E[Y_{i,t=1}(0) - Y_{i,t-1}(0)|D_i = 1]$$
$$-E[Y_{j,t=1}(0) - Y_{j,t=0}(0)|D_j = 0]$$

In other notation...

The identifying assumption of the DD is "parallel trends":

$$\hat{\tau}^{DD} = (Y_{i,t=1}(D_{it} = 1) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt} = 0) - Y_{j,t=0}(D_{jt} = 0))$$

$$= (Y_{i,t=1}(1) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))$$

$$+(Y_{i,t=1}(0) - Y_{i,t=1}(0))$$

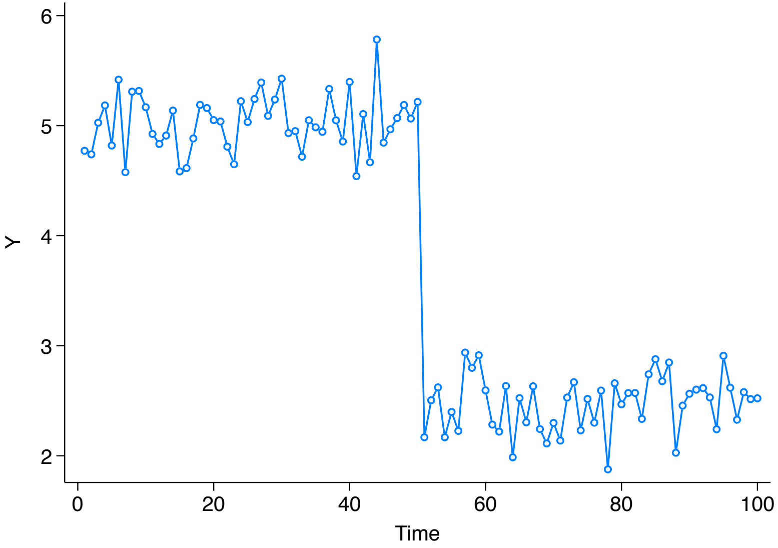
$$= (Y_{i,t=1}(1) - Y_{i,t=1}(0)) + [(Y_{i,t=1}(0) - Y_{i,t=0}(0)) - (Y_{j,t=1}(0) - Y_{j,t=0}(0))]$$

$$\approx E[Y_{i,t=1}(1) - Y_{i,t=1}(0)|D_i = 1] + E[Y_{i,t=1}(0) - Y_{i,t-1}(0)|D_i = 1]$$
$$-E[Y_{j,t=1}(0) - Y_{j,t=0}(0)|D_j = 0]$$

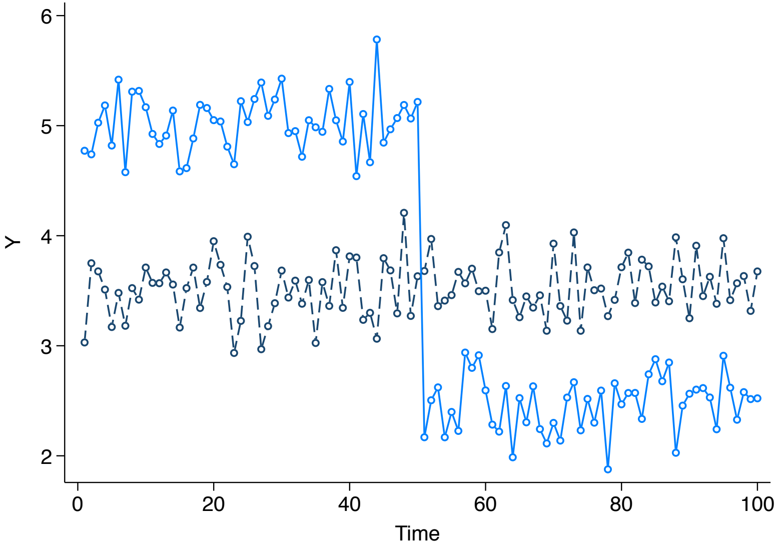
$$= \tau + \text{counterfactual trend} - \text{untreated trend}$$

Identifying assumption: untreated trend = counterfactual trend

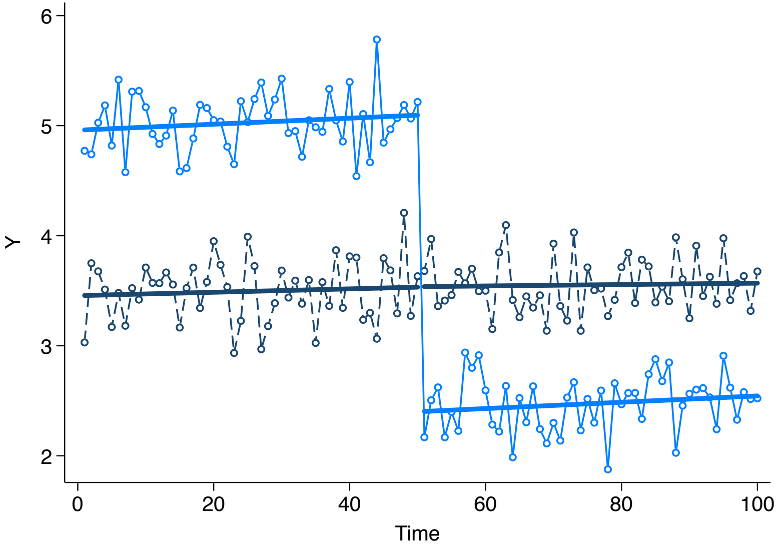
Identifying assumptions, visually



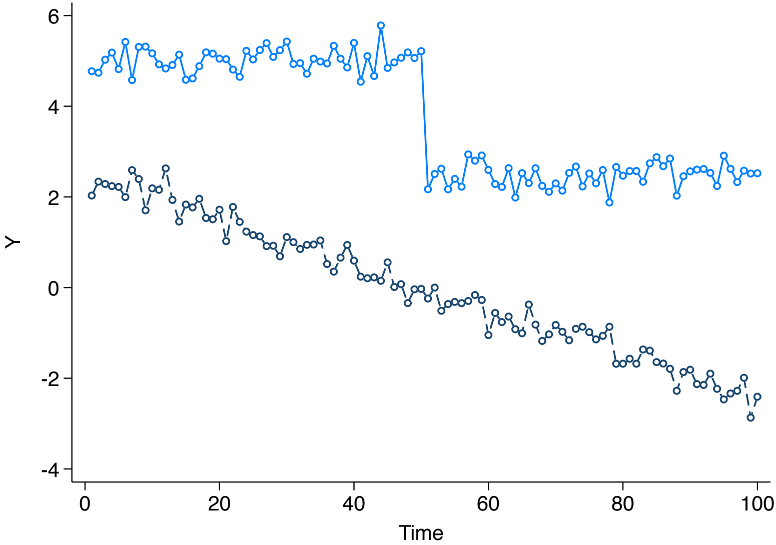
Identifying assumptions, visually



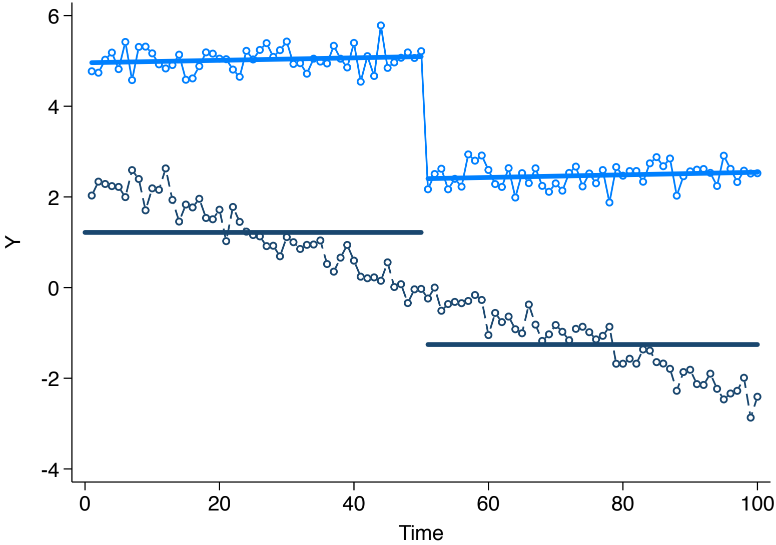
Identifying assumptions, visually



Identifying assumptions, visually



Identifying assumptions, visually



Implementing DD via regression

The simplest implementation of DD is just:

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

Implementing DD via regression

The simplest implementation of DD is just:

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

We can implement this via the following regression:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \varepsilon_{it}$$

Implementing DD via regression

The simplest implementation of DD is just:

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

We can implement this via the following regression:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \varepsilon_{it}$$

To link these together, see:
$\bar{Y}(treat, post) = \hat{\alpha} + \hat{\tau} + \hat{\beta} + \hat{\delta}$
$\bar{Y}(treat, pre) = \hat{\alpha} + \hat{\beta}$
$\rightarrow \bar{Y}(treat, post) - \bar{Y}(treat, pre) = \hat{\delta} + \hat{\tau}$

Implementing DD via regression

The simplest implementation of DD is just:

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

We can implement this via the following regression:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \varepsilon_{it}$$

To link these together, see:

$\bar{Y}(treat, post) = \hat{\alpha} + \hat{\tau} + \hat{\beta} + \hat{\delta}$

$\bar{Y}(treat, pre) = \hat{\alpha} + \hat{\beta}$

$\rightarrow \bar{Y}(treat, post) - \bar{Y}(treat, pre) = \hat{\delta} + \hat{\tau}$

and

$\bar{Y}(untreat, post) = \hat{\alpha} + \hat{\delta}$

$\bar{Y}(untreat, pre) = \hat{\alpha}$

$\rightarrow \bar{Y}(untreat, post) - \bar{Y}(untreat, pre) = \hat{\delta}$

Implementing DD via regression

The simplest implementation of DD is just:

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

We can implement this via the following regression:

$$Y_i = \alpha + \tau \, Treat \times Post_{it} + \beta \, Treat_i + \delta \, Post_t + \varepsilon_{it}$$

To link these together, see:
$\bar{Y}(treat, post) = \hat{\alpha} + \hat{\tau} + \hat{\beta} + \hat{\delta}$
$\bar{Y}(treat, pre) = \hat{\alpha} + \hat{\beta}$
$\rightarrow \bar{Y}(treat, post) - \bar{Y}(treat, pre) = \hat{\delta} + \hat{\tau}$
and
$\bar{Y}(untreat, post) = \hat{\alpha} + \hat{\delta}$
$\bar{Y}(untreat, pre) = \hat{\alpha}$
$\rightarrow \bar{Y}(untreat, post) - \bar{Y}(untreat, pre) = \hat{\delta}$

Running this regression yields $\hat{\tau} = \hat{\tau}^{DD}$

A handy table

We can implement DD via the following regression:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \varepsilon_{it}$$

A handy table

We can implement DD via the following regression:

$$Y_i = \alpha + \tau \, Treat \times Post_{it} + \beta \, Treat_i + \delta \, Post_t + \varepsilon_{it}$$

This gives us:

	Pre	Post	Difference
Treated	$\alpha + \beta + \delta + \tau$	$\alpha + \beta$	$\delta + \tau$
Untreated	$\alpha + \delta$	α	δ
Difference	$\beta + \tau$	β	τ

Adding covariates

We can add covariates:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \gamma X_{it} + \varepsilon_{it}$$

Adding covariates

We can add covariates:

$$Y_i = \alpha + \tau Treat \times Post_{it} + \beta Treat_i + \delta Post_t + \gamma X_{it} + \varepsilon_{it}$$

We do this to:

1. Add precision: like the RCT, we can soak up variation
2. Control for important observables
   - This mixes DD with SOO

Recap

TL;DR:

1. We can leverage time series data for identification
2. This is more powerful when combined with cross-section
3. The resulting diff-in-diff is one of the better quasi-experiments