

IV and Panel Data

Antonella Bandiera

March 14, 2023

Last class

- ▶ We discussed the basics of IV
- ▶ Let's start with an example

An example: Conflict in Africa

Why do we care what causes conflict?

An example: Conflict in Africa

Why do we care what causes conflict?

- ▶ Conflict is a big problem
- ▶ but what are the causes of conflict?
- ▶ How much support is there for economic channels?

An example: Conflict in Africa

Why do we care what causes conflict?

- ▶ Conflict is a big problem
- ▶ but what are the causes of conflict?
- ▶ How much support is there for economic channels?

Approach

- ▶ We need a shock to economic conditions
 - ▶ What does this mean?

An example: Conflict in Africa

Why do we care what causes conflict?

- ▶ Conflict is a big problem
- ▶ but what are the causes of conflict?
- ▶ How much support is there for economic channels?

Approach

- ▶ We need a shock to economic conditions
 - ▶ What does this mean?
- ▶ Randomization is not an option, so we use IV
- ▶ Instrument of choice: **changes in rainfall**
- ▶ We will discuss if we believe this later

Estimating treatment effects of economic growth on conflict

How does economic growth affect civil conflict?

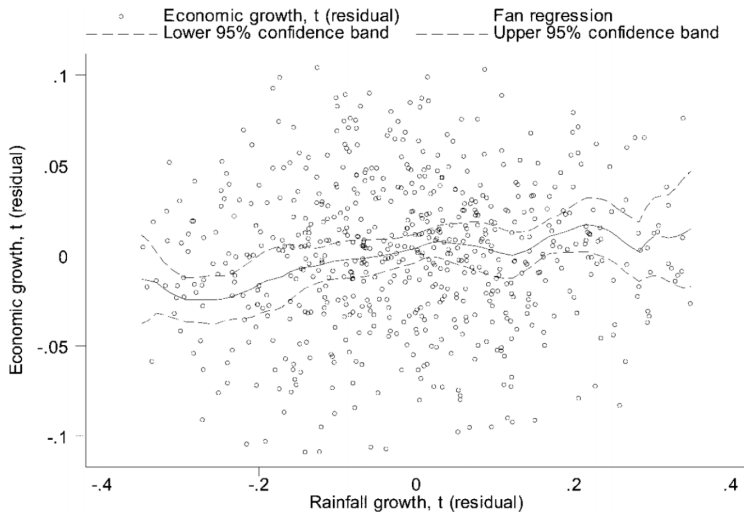
First stage

$$\text{growth}_{it} = \alpha + \gamma \Delta R_{it} + \beta X_{it} + \eta_{it}$$

where

- ▶ growth is economic growth in country i , time t
- ▶ $\Delta R_{it} = R_{i,t} - R_{i,t-1}$ is the change in rainfall
- ▶ η_{it} is an error term

First Stage



First Stage

RAINFALL AND ECONOMIC GROWTH (First-Stage)
Dependent Variable: Economic Growth Rate, t

EXPLANATORY VARIABLE	ORDINARY LEAST SQUARES				
	(1)	(2)	(3)	(4)	(5)
Growth in rainfall, t	.055*** (.016)	.053*** (.017)	.049*** (.017)	.049*** (.018)	.053*** (.018)
Growth in rainfall, $t - 1$.034** (.013)	.032** (.014)	.028** (.014)	.028* (.014)	.037** (.015)
Growth in rainfall, $t + 1$.001 (.019)	
Growth in terms of trade, t					-.002 (.023)
Log(GDP per cap- ita), 1979		-.011 (.007)			
Democracy (Polity IV), $t - 1$.0000 (.0007)			
Ethnolinguistic fractionalization		.006 (.044)			
Religious fractionalization		.045 (.044)			
Oil-exporting country		.007 (.019)			
Log(mountainous)		.001 (.005)			
Log(national popu- lation), $t - 1$		-.009 (.009)			
Country fixed effects	no	no	yes	yes	yes
Country-specific time trends	no	yes	yes	yes	yes
R^2	.02	.08	.13	.13	.16
Root mean square error	.07	.07	.07	.07	.06
Observations	743	743	743	743	661

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. A country-specific year time trend is included in all specifications (coefficient estimates not reported).

* Significantly different from zero at 90 percent confidence.

** Significantly different from zero at 95 percent confidence.

*** Significantly different from zero at 99 percent confidence.

Estimating treatment effects of economic growth on conflict

How does economic growth affect civil conflict?

First stage

$$\text{growth}_{it} = \alpha + \gamma \Delta R_{it} + \beta X_{it} + \eta_{it}$$

where

- ▶ growth is economic growth in country i , time t
- ▶ $\Delta R_{it} = R_{i,t} - R_{i,t-1}$ is the change in rainfall
- ▶ η_{it} is an error term

Second stage

$$\text{conflict}_{it} = \alpha + \delta \hat{\text{growth}}_{it} + \beta X_{it} + \eta_{it}$$

where $\hat{\text{growth}}_{it}$ are the fitted values from the first stage

Second stage

ECONOMIC GROWTH AND CIVIL CONFLICT

EXPLANATORY VARIABLE	DEPENDENT VARIABLE: Civil Conflict ≥ 25 Deaths						DEPENDENT VARIABLE: Civil Conflict $\geq 1,000$ Deaths
	Probit (1)	OLS (2)	OLS (3)	OLS (4)	IV-2SLS (5)	IV-2SLS (6)	IV-2SLS (7)
Economic growth rate, t	-.37 (.26)	-.33 (.26)	-.21 (.20)	-.21 (.16)	-.41 (1.48)	-1.13 (1.40)	-1.48* (.82)
Economic growth rate, $t-1$	-.14 (.23)	-.08 (.24)	.01 (.20)	.07 (.16)	-2.25** (1.07)	-2.55** (1.10)	-.77 (.70)
Log(GDP per cap- ita), 1979	-.067 (.061)	-.041 (.050)	.085 (.084)		.053 (.098)		
Democracy (Polity IV), $t-1$.001 (.005)	.001 (.005)	.003 (.006)		.004 (.006)		
Ethnolinguistic fractionalization	.24 (.26)	.23 (.27)	.51 (.40)		.51 (.39)		
Religious fractionalization	-.29 (.26)	-.24 (.24)	.10 (.42)		.22 (.44)		
Oil-exporting country	.02 (.21)	.05 (.21)	-.16 (.20)		-.10 (.22)		
Log(mountainous)	.077** (.041)	.076* (.039)	.057 (.060)		.060 (.058)		
Log(national pop- ulation), $t-1$.080 (.051)	.068 (.051)	.182* (.086)		.159* (.093)		
Country fixed effects	no	no	no	yes	no	yes	yes
Country-specific time trends	no	no	yes	yes	yes	yes	yes
R^213	.53	.71
Root mean square error42	.31	.25	.36	.32	.24
Observations	743	743	743	743	743	743	743

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. Regression 1 presents marginal probit effects, evaluated at explanatory variable mean values. The instrumental variables for economic growth in regressions 5–7 are growth in rainfall, t and growth in rainfall, $t-1$. A country-specific year time trend is included in all specifications (coefficient estimates not reported), except for regressions 1 and 2, where a single linear time trend is included.

*. Significant at the 10 percent confidence level.

The reduced form

Reduced form

$$\text{conflict}_{it} = \alpha + \theta \Delta R_{it} + \pi X_{it} + \eta_{it}$$

Reduced form

RAINFALL AND CIVIL CONFLICT (Reduced-Form)		
EXPLANATORY VARIABLE	DEPENDENT VARIABLE	
	Civil Conflict ≥ 25 Deaths (OLS) (1)	Civil Conflict $\geq 1,000$ Deaths (OLS) (2)
Growth in rainfall, t	-.024 (.043)	-.062** (.030)
Growth in rainfall, $t - 1$	-.122** (.052)	-.069** (.032)
Country fixed effects	yes	yes
Country-specific time trends	yes	yes
R^2	.71	.70
Root mean square error	.25	.22
Observations	743	743

NOTE.—Huber robust standard errors are in parentheses. Regression disturbance terms are clustered at the country level. A country-specific year time trend is included in all specifications (coefficient estimates not reported).

* Significantly different from zero at 90 percent confidence.

** Significantly different from zero at 95 percent confidence.

*** Significantly different from zero at 99 percent confidence.

The exclusion restriction is very important

When looking at an IV analysis you should always ask: what is the exclusion restriction in this analysis saying? Do we believe this?
Why or why not?

Other examples of IV applications

- ▶ Lotteries (school assignment)
- ▶ Random experiments with imperfect compliance
- ▶ Judge assignment
- ▶ Compulsory schooling
- ▶ Bartik share

Social Networks and the Decision to Insure (Cai, De Janvry, and Sadoulet, 2015)

- ▶ The authors are looking into the decision that farmers make about whether to buy insurance against weather events: does information about insurance travels through social networks?
- ▶ Two rounds of different informational sessions about insurance in rural China: how much your friends learn about insurance, affects your decision to take up insurance?
- ▶ Does your friends actually buying insurance make you more likely to buy?
- ▶ This example comes from “The Effect” you will be able to find python code

Social Networks and the Decision to Insure (Cai, De Janvry, and Sadoulet, 2015)

- ▶ We want to identify the effect of *FriendsPurchaseBehavior* on *YourPurchaseBehavior* among people in the second-round informational sessions, looking at the average purchasing behavior of their friends who were in the first-round informational sessions. What are possible problems?
- ▶ As an instrument they use the variable *FirstRoundDefault*, which is a binary indicator for whether your friends were assigned to a *default buy* (buy insurance by default, had to specify they did not want it), as opposed to default not buy

```
# we'll use feols from fixest for speed and ease of fe implementation
options(width = 55)
library(tidyverse);library(fixest);
library(causaldata);library(estimatr); library(car)
d <- causaldata::social_insure
first <- lm_robust(pre_takeup_rate ~ default, data=d)
summary(first)
```

```
##
## Call:
## lm_robust(formula = pre_takeup_rate ~ default, data = d)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept)   0.3587   0.007887   45.48 1.329e-278
## default       0.1446   0.012324   11.73 2.108e-30
##              CI Lower CI Upper    DF
## (Intercept)   0.3432   0.3741 1408
## default       0.1204   0.1688 1408
##
## Multiple R-squared:  0.08985 ,    Adjusted R-squared:  0.0892
## F-statistic: 137.7 on 1 and 1408 DF,  p-value: < 2.2e-16
```

```
options(width = 55)
# Include just the outcome and controls first, then endogenous ~ instru
# in the second part, and for this study we cluster on address
iv <- feols(takeup_survey ~ male + age + agpop + ricearea_2010 +
            literacy + intensive + risk_averse + disaster_prob +
            factor(village) | pre_takeup_rate ~ default,
            cluster = ~address, data = d)
summary(iv, stage=1)
```

```
## TSLS estimation, Dep. Var.: pre_takeup_rate, Endo.: pre_takeup_rate,
## First stage: Dep. Var.: pre_takeup_rate
## Observations: 1,378
## Standard-errors: Clustered (address)
##
```

	Estimate	Std. Error
## (Intercept)	0.124837	0.161638
## default	0.118026	0.034433
## male	-0.006192	0.016915
## age	0.000406	0.000409
## agpop	-0.001643	0.003045
## ricearea_2010	-0.000394	0.000214
## literacy	0.012140	0.014036
## intensive	0.014640	0.005157
## risk_averse	-0.021254	0.017649
## disaster_prob	0.000042	0.000242
## factor(village)beixing	0.181503	0.157497

```
summary(iv)
```

```
## TSLS estimation, Dep. Var.: takeup_survey, Endo.: pre_takeup_rate, I
```

```
## Second stage: Dep. Var.: takeup_survey
```

```
## Observations: 1,378
```

```
## Standard-errors: Clustered (address)
```

```
##               Estimate Std. Error
```

```
## (Intercept)      0.020484    0.128308
```

```
## fit_pre_takeup_rate  0.791097    0.273127
```

```
## male              0.037540    0.053741
```

```
## age               0.005245    0.001291
```

```
## agpop             -0.002354    0.006720
```

```
## ricearea_2010      0.002768    0.001137
```

```
## literacy           0.082109    0.035793
```

```
## intensive         -0.009016    0.026729
```

```
## risk_averse        0.236723    0.045202
```

```
## disaster_prob      0.001272    0.000754
```

```
## factor(village)beixing -0.492735    0.076087
```

```
## factor(village)caijia -0.392660    0.071720
```

```
## factor(village)daqiao -0.493662    0.178831
```

```
## factor(village)daxi   -0.270923    0.087071
```

```
## factor(village)dayu   -0.328883    0.108998
```

```
## factor(village)dazhou -0.505534    0.099899
```

```
## factor(village)dongan -0.427896    0.103527
```

```
## factor(village)dukou  -0.498722    0.111219
```

Recap

Z_i is a valid instrument when the following are satisfied:

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

When we have both, we can:

- ▶ handle OVB
- ▶ handle measurement error

IV unravelling

With $\tau_i = \tau$ for all i , life is good:

- ▶ We need a first stage,
- ▶ and an exclusion restriction

IV unravelling

With $\tau_i = \tau$ for all i , life is good:

- ▶ We need a first stage,
- ▶ and an exclusion restriction

what happens with heterogeneous treatment effects?

- ▶ What are we actually recovering with τ_{IV} ?

IV unravelling

With $\tau_i = \tau$ for all i , life is good:

- ▶ We need a first stage,
- ▶ and an exclusion restriction

what happens with heterogeneous treatment effects?

- ▶ What are we actually recovering with τ_{IV} ?

Recall that we set up IV with the following decomposition

$$D_i = B_i \varepsilon_i + C_i$$

IV unravelling

With $\tau_i = \tau$ for all i , life is good:

- ▶ We need a first stage,
- ▶ and an exclusion restriction

what happens with heterogeneous treatment effects?

- ▶ What are we actually recovering with τ_{IV} ?

Recall that we set up IV with the following decomposition

$$D_i = B_i \varepsilon_i + C_i$$

- ▶ For the IV to work, it must be correlated with C_i but not $B_i \varepsilon_i$
- ▶ But Z_i is generating variation in *part* of C_i

IV unravelling

With $\tau_i = \tau$ for all i , life is good:

- ▶ We need a first stage,
- ▶ and an exclusion restriction

what happens with heterogeneous treatment effects?

- ▶ What are we actually recovering with τ_{IV} ?

Recall that we set up IV with the following decomposition

$$D_i = B_i \varepsilon_i + C_i$$

- ▶ For the IV to work, it must be correlated with C_i but not $B_i \varepsilon_i$
- ▶ But Z_i is generating variation in *part* of C_i
- ▶ If this part affects Y_i differently, $\hat{\tau} \neq \tau_{ATE}$

Recap: Assumptions

1. **First stage:** $E[D_i|Z_i = 1] \neq E[D_i|Z_i = 0]$ for some i
2. **Independence:** $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp\!\!\!\perp Z_i$
3. **Exclusion restriction:** $Y_i(Z_i = 1, D) = Y_i(Z_i = 0, D)$ for $D \in \{0, 1\}$
4. **Monotonicity:** $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i

In words

- ▶ Since we are using an instrument to isolate the variation of the endogenous variable (to keep the exogenous variation), we are looking at a local average treatment effect.

In words

- ▶ Since we are using an instrument to isolate the variation of the endogenous variable (to keep the exogenous variation), we are looking at a local average treatment effect.
- ▶ The individual treatment effects are weighted by how responsive an individual observation is to the instrument

In words

- ▶ Since we are using an instrument to isolate the variation of the endogenous variable (to keep the exogenous variation), we are looking at a local average treatment effect.
- ▶ The individual treatment effects are weighted by how responsive an individual observation is to the instrument
- ▶ When we use a standard estimator like 2sls the weights are that the individual effect of the instrument would be in your first stage
- ▶ This means is that 2sls will give us different results if we use different instruments: it depends on how individual observations respond to each instrument
 - ▶ Always takers - Never takers
 - ▶ Compliers
 - ▶ Defiers (we need to assume these don't exist)

What do we get?

- ▶ Constant τ
 - ▶ $\hat{\tau}_{IV} = \tau_{ATE}$
- ▶ Perfect compliance:
 - ▶ $\hat{\tau}_{IV} = \tau_{ATE}$
- ▶ Heterogeneous treatment effects
 - ▶ $\hat{\tau}_{IV} = \tau_{LATE}$

Taking stock

We've come a long way from RCTs

- ▶ Took a detour through SOO
- ▶ Started the discussion of SOU
- ▶ IV is our first SOU design
- ▶ IV help us do causal inference with non-random treatment
- ▶ We just need *some form* of randomness to leverage over treatment

More SOU designs

Until now, we have focused on data across units. Now we will add time:

- ▶ **Cross-sectional data:**
 - ▶ What we have 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 have discussed the selection problem a lot:

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

Why is data over time useful?

We have discussed the selection problem a lot:

- ▶ To isolate the effect of D_i we need potential outcome to be the same among treated and untreated units
- ▶ With cross-sectional data this is 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 have discussed the selection problem a lot:

- ▶ To isolate the effect of D_i we need potential outcome to be the same among treated and untreated units
- ▶ With cross-sectional data this is 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 is a control for itself
 - ▶ I am much more similar to myself yesterday, than to j

Controlling for unobserved confounders

- ▶ **Fixed effects** is a method to control for *all* variables, whether observed or not, as long as they stay constant within some larger category

Controlling for unobserved confounders

- ▶ **Fixed effects** is a method to control for *all* variables, whether observed or not, as long as they stay constant within some larger category
- ▶ How? We control for the larger category

What does it mean constant within category?

- ▶ Imagine we are looking at the effect of providing electricity on productivity
- ▶ An obvious confounder is geography. Why?

What does it mean constant within category?

- ▶ Imagine we are looking at the effect of providing electricity on productivity
- ▶ An obvious confounder is geography. Why?
- ▶ What what if we observe towns *multiple times*
- ▶ Each town will have the same geography every time
- ▶ So what if we control for town? What happens?

What does it mean constant within category?

- ▶ Imagine we are looking at the effect of providing electricity on productivity
- ▶ An obvious confounder is geography. Why?
- ▶ What what if we observe towns *multiple times*
- ▶ Each town will have the same geography every time
- ▶ So what if we control for town? What happens?
- ▶ We will remove all the variation explained by town (its geography, but also all the other things that don't change over time)

Making time-series comparisons

Consider a setting with only one unit:

- ▶ We now denote our outcome as $Y_t(D_t)$
- ▶ As usual we want to estimate
$$\tau_{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$$
- ▶ But we can't observe $Y_{t=1}(D_{t=1} = 1)$ and $Y_{t=1}(D_{t=1} = 0)$

Making time-series comparisons

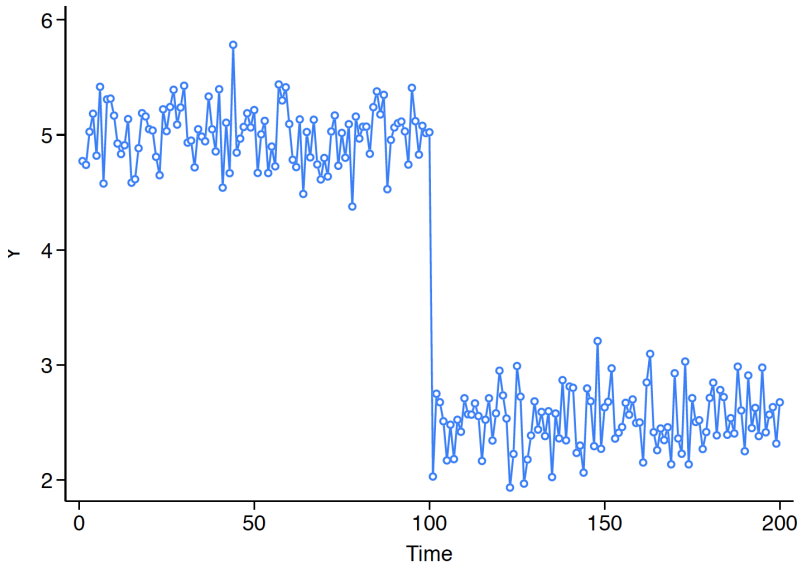
Consider a setting with only one unit:

- ▶ We now denote our outcome as $Y_t(D_t)$
- ▶ As usual we want to estimate
$$\tau_{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)]$$
- ▶ But we can't observe $Y_{t=1}(D_{t=1} = 1)$ and $Y_{t=1}(D_{t=1} = 0)$
- ▶ Instead we look for periods before and after the treatment begins
- ▶ Suppose in $t = 0$, $D_{t=0} = 0$ and in $t = 1$, $D_{t=1} = 1$
- ▶ Then we can estimate:

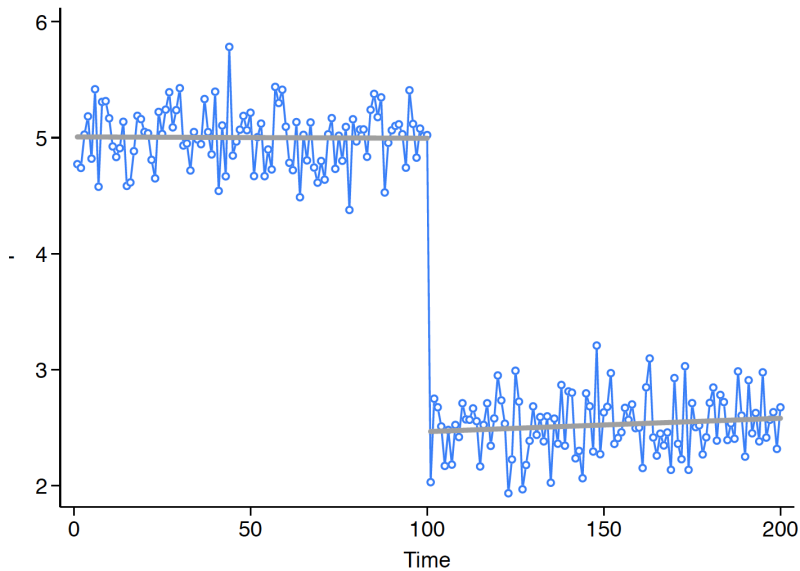
$$\hat{\tau}_{TS} = Y_{t=1} - Y_{t=0}$$

we can extend this to many periods

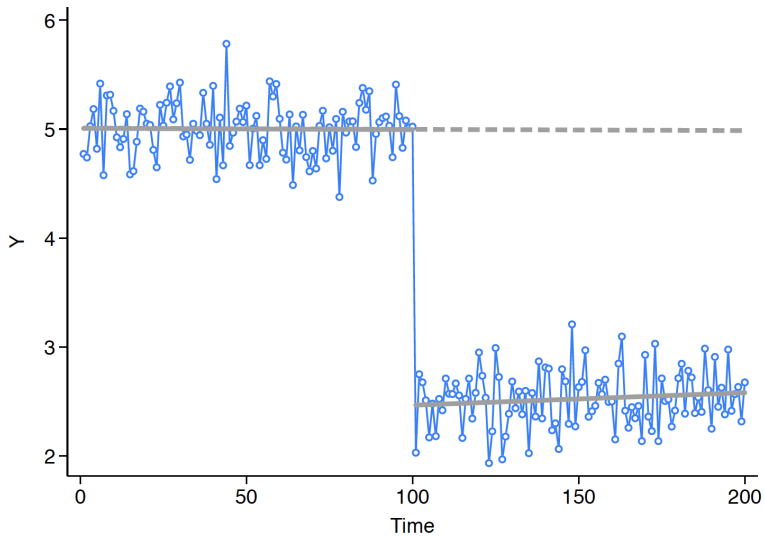
Visually



Visually



Visually



What is good about this?

- ▶ Consider a simple data generating process:

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

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

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

- ▶ The difference estimator separates the treatment effect of D (which = 0 in $t = 0$ and 1 in $t = 1$)

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

The same applies with unobservables

- Consider a simple data generating process:

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

- In this model we have time invariant characteristics (X_i) and time-invariant unobserved (U_i), add treatment D :

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

- The difference estimator separates the treatment effect of D (which = 0 in $t = 0$ and 1 in $t = 1$)

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

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:

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

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

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:

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

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

In this case

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

Thus we need $\delta = 0$ or $V_{i,t=1} = V_{i,t=0}$

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:

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

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

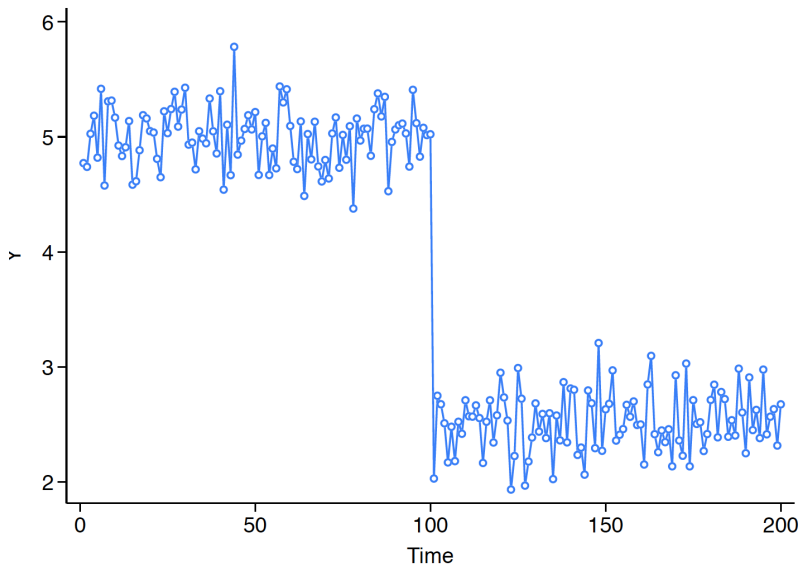
In this case

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

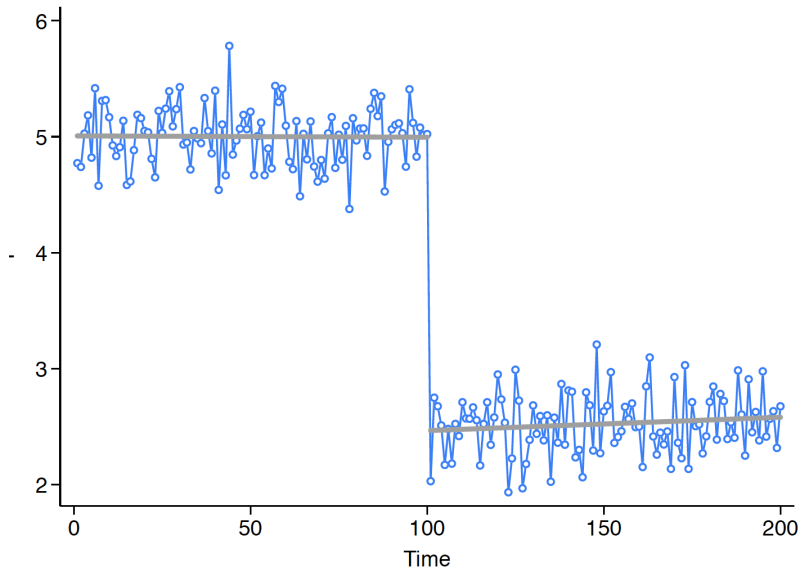
Thus we need $\delta = 0$ or $V_{i,t=1} = V_{i,t=0}$

- ▶ Any time varying variable that has an effect on Y will create bias

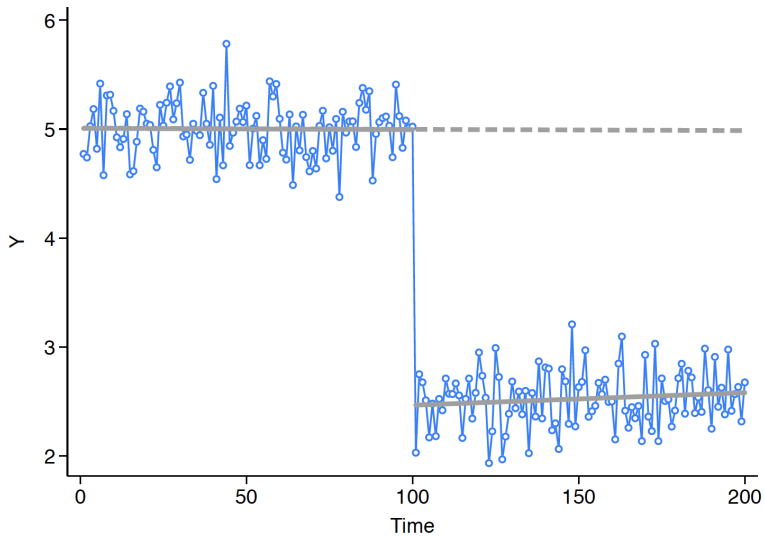
Visually



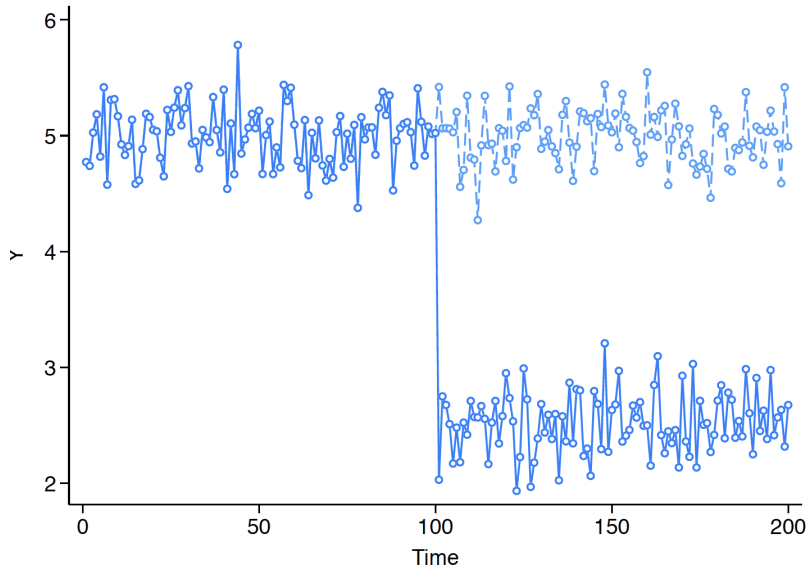
Visually



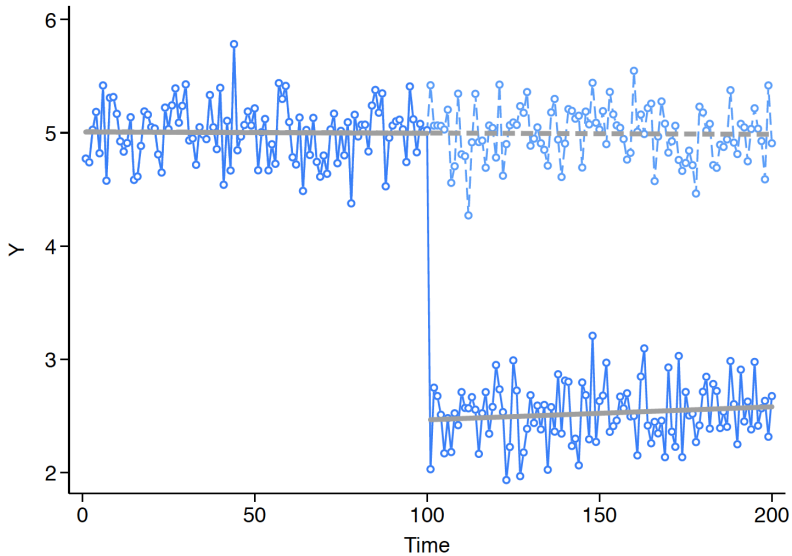
Visually



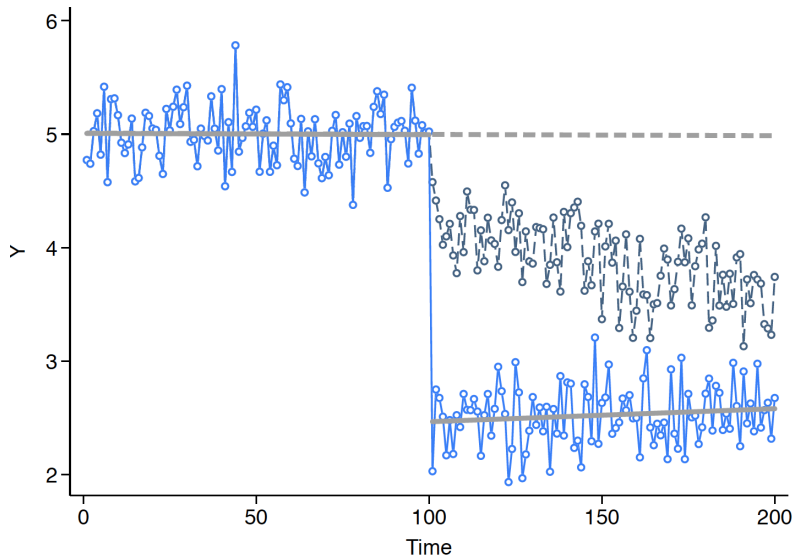
Visually



Visually



Visually



Validating the identifying assumption

We can never validate it

Validating the identifying assumption

We can never validate it

- ▶ We would need to observe the counterfactual
- ▶ We can't do this
- ▶ We could look at pre-treatment trends but it is not enough
- ▶ We cannot eliminate the trend with one time series

Two wrongs make a right?

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

- ▶ Our naive (cross-sectional) estimator:
 - ▶ Compare i to j
 - ▶ 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
 - ▶ Or, non-zero trends

Two wrongs make a right?

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

- ▶ Our naive (cross-sectional) estimator:
 - ▶ Compare i to j
 - ▶ 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
 - ▶ Or, non-zero trends
- ▶ We can combine these two 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?
- ▶ Using other individuals j that did not receive the treatment 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?
- ▶ Using other individuals j that did not receive the treatment we can make a guess

The differences-in-differences estimator:

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

Differences-in-differences (DD)

The problem with time series is the counterfactual trend

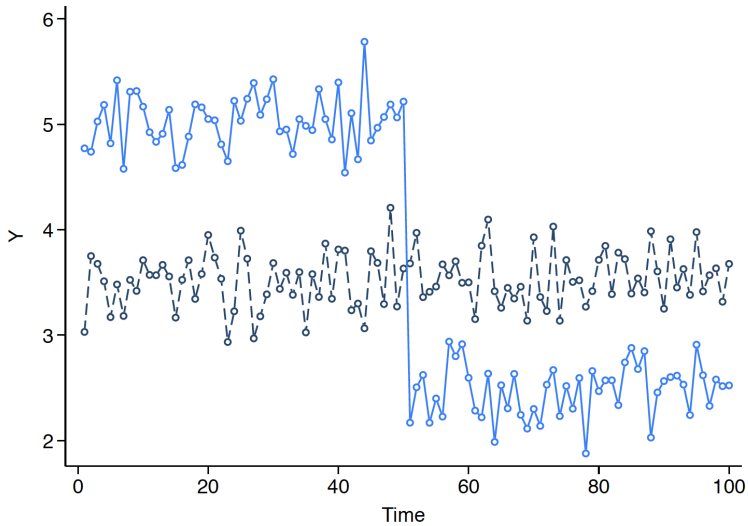
- ▶ How would treated i have behaved in $t = 1$ without treatment?
- ▶ Using other individuals j that did not receive the treatment we can make a guess

The differences-in-differences estimator:

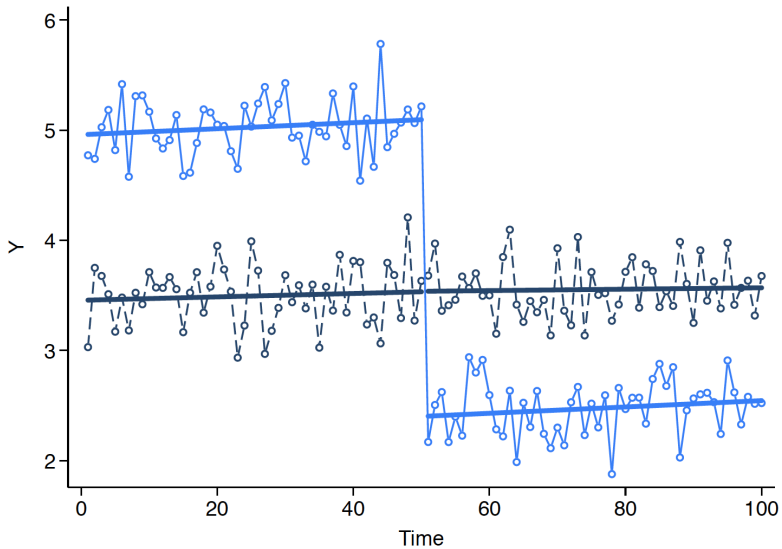
$$\hat{\tau}_{DD} = (Y_{i,t=1}(D_{it=1}) - Y_{i,t=0}(D_{it} = 0)) - (Y_{j,t=1}(D_{jt=0}) - Y_{j,t=0}(D_{jt} = 0))$$

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

Visually

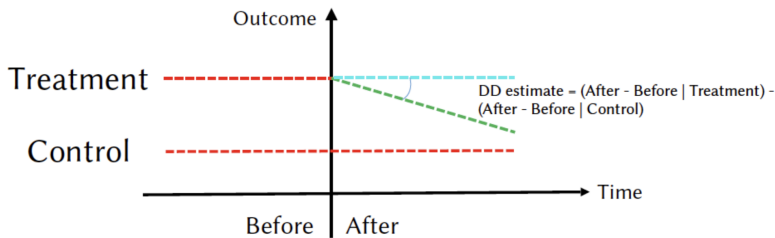


Visually



The identifying assumption

- ▶ Parallel trends: the outcome would have evolved similarly in the treatment group as in the control group, had the treatment not happened



Implementing DD via regression

The simplest implementation of DD is just:

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

We can implement this via the following regression:

$$Y_i = \alpha + \tau \text{Treat}_i \times \text{Post}_t + \beta \text{Treat}_i + \delta \text{Post}_t + \varepsilon_i$$

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

Implementing DD via regression

We can implement this via the following regression:

$$Y_i = \alpha + \tau \text{Treat}_i \times \text{Post}_t + \beta \text{Treat}_i + \delta \text{Post}_t + \varepsilon_i$$

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