Differences in Differences

EC3720: Introduction to Econometrics

Jonathan Colmer

University of Virginia

Natural Experiments

- By randomly assigning a treatment we can ensure that it is uncorrelated with other determinants of the outcome.
- Unfortunately, there are many situations in which it is difficult or impossible to conduct randomized experiments.
- When we cannot conduct randomized experiments, we might be able to exploit "natural experiments".
- The Key idea: exploit naturally occurring exogenous variation to mimic a randomized experiment.
- Today's topic is difference-in-difference estimation, which is very widely used.
- Technically this is closely related to fixed effects panel regressions (but not the same).

The Most Famous Natural Experiment

- In the 19th century, London (and many other cities) experienced severe cholera epidemics, which claimed thousands of lives.
- Until the mid-19th century there were two competing theories about the causes of cholera:
  - The dominant view was that cholera is caused by particles (“miasmas”) suspended in the air.
  - The alternative view was that cholera was caused by a germ that is transmitted through the water supply.
- Both views obviously had very different policy implications.

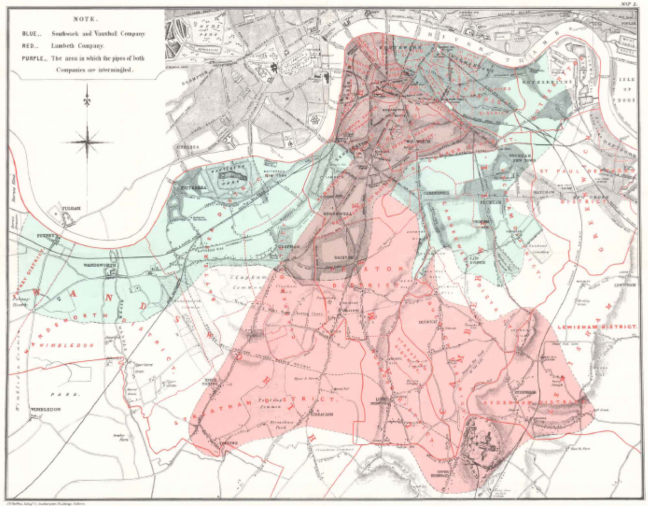
The Most Famous Natural Experiment

- John Snow, a medical doctor and self-trained scientist, exploited two natural experiments to provide evidence for the second theory.
- At the time, he was laughed at; today he is regarded as the founder of modern epidemiology.
- His work is a true masterpiece of statistical detective work and is summarized in Snow (1855).
- In popular accounts of his work he is most famous for his map of the outbreak around the Broad Street pump in Soho.
- We will focus on two even more interesting and less known pieces of his evidence.

The Most Famous Natural Experiment

- Snow's first piece of evidence exploits a very interesting property of the London Water Supply during the 1854 cholera epidemic.
- At the time, some parts of London were simultaneously supplied by two water companies:
  - The Southwark & Vauxhall Water Company pumped water from a part of the Thames that was contaminated with sewage.
  - The Lambeth Water Company pumped its water from a place further upstream the Thames that was not contaminated.

London’s Water Supply



Collecting the Data

- Snow used a very time-intensive approach to gather his data.
- He obtained a list of all cholera deaths during the first seven weeks of the 1854 epidemic.
- He considered all deaths in districts that were either supplied by the Lambeth, Southwark & Vauxhall, or both companies.
- For each death he determined (through a sometimes lengthy process) the water supplier to the house of the deceased.

Snow's Results

	Number of Houses	Deaths from Cholera	Deaths per 10,000 houses
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

What might be wrong with this simple comparison?

The Role of Competition

- In the areas in which the two companies compete this evidence is powerful.
- Adjacent houses often had different water suppliers and in many cases did not even know who their supplier was.
- Therefore, residents of these houses should have breathed in approximately the same amount of “miasmas"
- This “as if” random assignment of the water suppliers makes it unlikely that the people connected to one or the other company differ systematically in any other dimension.
- This is a very special natural experiment; rarely will our settings be so fortuitous

A Second Natural Experiement

- Snow also makes use of a second natural experiment.
- The Lambeth Water Company also used to pump its water from the contaminated part of the Thames.
- It only relocated the water intake upstream in 1852.
- Comparing the 1849 and 1854 cholera epidemics Snow shows:
  - Areas exclusively supplied by the Southwark & Vauxhall company had similar numbers during two epidemics.
  - Areas either partly or exclusively supplied by the Lambeth company experienced a substantially lower number of deaths in 1854 compared to 1849.
- This is effectively a difference-in-differences estimator.

The Simplest Difference-in-Differences Setup

- Two groups:
  - D=1 Treated units
  - D=0 Control units
- Two periods:
  - T=0 Pre-Treatment period
  - T=1 Post-Treatment period
- Potential outcome $Y_d(t)$
  - $Y_{1i}(t)$ is the outcome that unit i attains in period t when treated between t and $t-1$
  - $Y_{0i}(t)$ is the outcome that unit i attains in period t when the control between t and $t-1$

Difference-in-Differences Setup

- The causal effect for unit i at time t is,

$$\alpha_i(t) = Y_{1i}(t) - Y_{0i}(t)$$

- Observed outcomes $Y_i(t)$ are realized as,

$$Y_i(t) = Y_{0i}(t) \cdot (1 - D_i(t)) + Y_{1i}(t) \cdot D_i(t)$$

- We again face the fundamental problem of causal inference!

- We will focus on estimating the average treatment effect on the treated:

$$\alpha_{ATT} = E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1]$$

Difference-in-Differences Setup

$$\alpha_{ATT} = E[Y_{1i}(1) - Y_{0i}(1) | D_i = 1]$$

	Post (t=1)	Pre (t=0)
Treated $D = 1$	$E[Y_1(1)\|D = 1]$	$E[Y_0(0)\|D = 1]$
Control $D = 0$	$E[Y_0(1)\|D = 0]$	$E[Y_0(0)\|D = 0]$

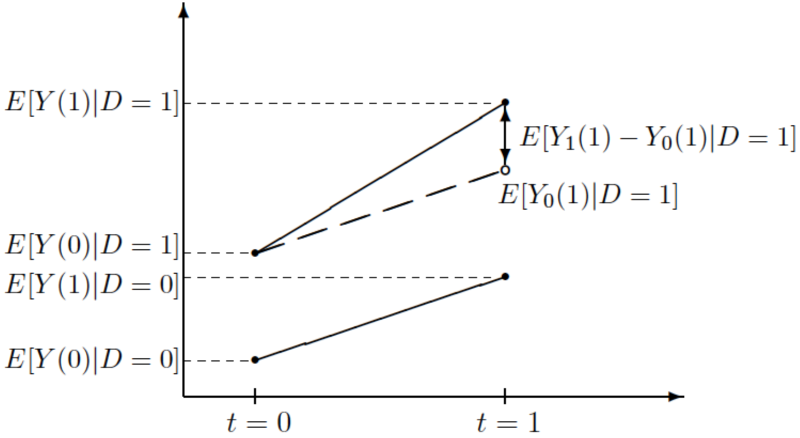
- We're missing potential outcome $E[Y_0(1)|D = 0]$ (the average post-period outcome for the treated in the absence of the treatment)

Difference-in-Differences Setup

- Control Strategy: Before vs. After
  - Estimator: $E[Y(1)|D=1] - E[Y(0)|D=1]$
  - Assumes: $E[Y_0(1)|D=1] = E[Y_0(0)|D=1]$
- Control Strategy: Treated vs. Control in Post-Period
  - Estimator: $E[Y(1)|D=1] - E[Y(1)|D=0]$
  - Assumes: $E[Y_0(1)|D=1] = E[Y_0(1)|D=0]$
- Control Strategy: Difference-in-Differences
  - Estimator:

    $$\{E[Y(1)|D=1] - E[Y(1)|D=0]\} - \{E[Y(1)|D=1] - E[Y(0)|D=1]\}$$
  - Assumes:
    $E[Y_0(1) - Y_0(0)|D=1] = E[Y_0(1) - Y_0(0)|D=0]$
    (parallel trends)

Graphical Representation: Difference-in-Differences



Identification with Difference-in-Differences

- Identification Assumption:

$$E[Y_0(1) - Y_0(0)|D = 1] = E[Y_0(1) - Y_0(0)|D = 0]$$

- This is known as the parallel trends assumption.

- Given parallel trends the ATT is identified as:

$$\begin{aligned} ATT &= E[Y_1(1) - Y_0(1)|D = 1] \\ &= \{E[Y(1)|D = 1] - E(Y(1)|D = 0]\} \\ &\quad -\{E[Y(0)|D = 1] - E[Y(0)|D = 0]\} \end{aligned}$$

Proof

$$
\begin{aligned}
ATT &= \{E[Y(1)|D=1] - E(Y(1)|D=0]\} - \{E[Y(0)|D=1] - E[Y(0)|D=0]\} \\
&= \{E[Y_1(1)|D=1] - E[Y_0(1)|D=0]\} - \{E[Y_0(0)|D=1] - E[Y_0(0)|D=0]\} \\
&= E[Y_1(1)|D=1] - E[Y_0(1)|D=1] + E[Y_0(1)|D=1] - E[Y_0(1)|D=0] \\
&\quad - E[Y_0(0)|D=1] - E[Y_0(0)|D=0] \\
&= E[Y_1(1) - Y_0(1)|D=1] + \underbrace{\{E[Y_0(1) - Y_0(0)|D=1] - E[Y_0(1) - Y_0(0)|D=0]\}}_{=0} \\
&= E[Y_1(1) - Y_0(1)|D=1] \qquad \blacksquare
\end{aligned}
$$

- Parallel Trends Assumption:
  $E[Y_0(1) - Y_0(0)|D=1] = E[Y_0(1) - Y_0(0)|D=0]$

What is the Effect of Tenancy Reform on Agricultural Productivity?

- In 1978, the newly-elected, left-wing government of West Bengal launched Operation Barga.
- Program of tenancy reform to regulate rents and improve security of tenure for sharecroppers
- How did this reform affect agricultural productivity?
- We will look at rice yields measured in kg/hectare
- Perhaps one could have run an experiment, but this reform has already occurred

We Could Look at Simple Differences Across Time

	Pre-Reform	Post-Reform	Difference
West Bengal	1,308	1,650	342

- In a regression framework with two periods, $t \in \{PRE, POST\}$

$$Y_{it} = \alpha + \beta \cdot \mathbb{1}(t = POST) + \epsilon_{it}$$

- The OLS estimate of β is the difference in means, $\bar{Y}_{POST} - \bar{Y}_{PRE}$, before and after the change.

- Would this give us a good measure of the program's effect?

Can We Make the Simple Difference More Convincing?

- With only two periods, the answer is no
- With more data $t \in \{0, 1, 2, \ldots, T\}$, you may be able to demonstrate a trend break
- Run a regression with all the year dummies

$$Y_{it} = \alpha + \sum_{\tau=1}^{T} \beta_\tau \cdot \mathbb{1}(t = \tau) + \epsilon_{it}$$

- This is just a fancy way to estimate averages for each period
- You look to see if there is a break in the pattern of $\hat{\beta}_\tau$ around the reform date.
- But with gradual reform or uncertain lags, this strategy will not work

We Could Instead Consider a Comparison Group

- Bangladesh
  - Neighbouring country
  - Similar agricultural characteristics
  - No reform

	Post-Reform
West Bengal	1,650
Bangladesh	1,562
Difference	88

- Would this give us a good measure of the program's effect?

The Difference-in-Differences Estimator

- The difference-in-differences estimator utilizes both comparison groups.

	Pre-Reform	Post-Reform	Difference
West Bengal	1,308	1,650	342
Bangladesh	1,297	1,562	265
Difference	11	88	77

The Difference-in-Differences Estimator

- The DID estimator is easy to calculate
- Call West Bengal the **treatment group** and Bangladesh the **control group**
- The DID estimator is:

$$\hat{\beta}_{DID} = [\bar{y}_{tPOST} - \bar{y}_{tPRE}] - [\bar{y}_{cPOST} - \bar{y}_{cPRE}]$$

  where

  - $\bar{y}_{tPOST}$ is average productivity in the treatment group post reform
  - $\bar{y}_{tPRE}$ is average productivity in the treatment group pre reform
  - $\bar{y}_{cPOST}$ is average productivity in the control group post reform
  - $\bar{y}_{cPRE}$ is average productivity in the control group pre reform
- This estimator compares the **change** in the outcome variables in the treatment group with the **change** in the control group.

Difference-in-Differences

- The difference-in-differences estimator is typically estimated with the following regression:

  $$Y = \alpha + \beta_1 TREAT + \beta_2 POST + \beta_3 (TREAT \times POST) + \epsilon$$

  where

  - Y is the outcome of interest (here, agricultural productivity)
  - TREAT is a dummy, which is equal to one if the observation is from the treatment group (here, West Bengal)
  - POST is a dummy, which is equal to one if the observation is from the period after the treatment (here, post Operation Barga reforms)

- It is not difficult to show that β_3 is our DID estimator β_{DID}.

The Benefits of Difference-in-Differences

- In contrast to the simple difference estimator ($\bar{y}_{tPOST} - \bar{y}_{cPOST}$) it controls for pre-existing differences between the treatment and control group

- In contrast to a first difference estimator ($\bar{y}_{tPOST} - \bar{y}_{tPRE}$) is controls for any time trends that affected both groups similarly.

- You may hear the phrase: "Difference-in-differences is robust to time-invariant heterogeneity".

The Key Identifying Assumption

- The critical assumption behind the difference-in-differences estimator is very simple:
  - In the absence of treatment the treatment group would have developed in exactly the same way as the control group
  - This is the "parallel trends" assumption that we discussed above.
- We are using the change in the control group as our counterfactual for what that change in the treatment group **would have been** had it not been treated.
- In any application of the difference-in-differences approach your first thought should be: Is this a "valid" control group?
- i.e., do we have a valid counterfactual?

Nice Work If You Can Get It?

- In his 2014 SOTU address, President Obama proposed an increase in the federal minimum wage, to \$10.10/hour from \$7.25.
- In February 2014, the president used his executive authority to raise the minimum wage that must be paid by federal contractors to this level, an action that does not require congressional approval.
- Congressional democrats have recently rallied around a \$12/hour minimum wage.
- Nice work if you can get it, but whether you can indeed get it, is the \$20,000 question.
- Economists and policy makers have long worried that while a higher minimum wage benefits those who find work, some workers (especially the low-skilled) may lose their jobs.
- What's the counterfactual? How can we know?

Learning from History

- On April 1, 1992, New Jersey imposed a state minimum wage of $5.05. The federal minimum wage was then $4.25.
- Card and Krueger (1994) surveyed fast food restaurants in NJ and Eastern PA before this change (February, 1992) and after (November, 1992).
- They use the data from these surveys to study the impact of the higher minimum wage on employment at fast food establishments like Wendy's and Burger King.
- The NJ increase in the minimum wage had a dramatic effect on starting wages at NJ fast food stores.

Means of Key Variables

	Stores in:		
	NJ	PA	t-stat
Distribution of Store Types (%):			
Burger King	41.1	44.3	-0.5
KFC	20.5	15.2	1.2
Roy Rogers	24.8	21.5	0.6
Wendy's	13.6	19.0	-1.1
Company-owned	34.1	35.4	-0.2

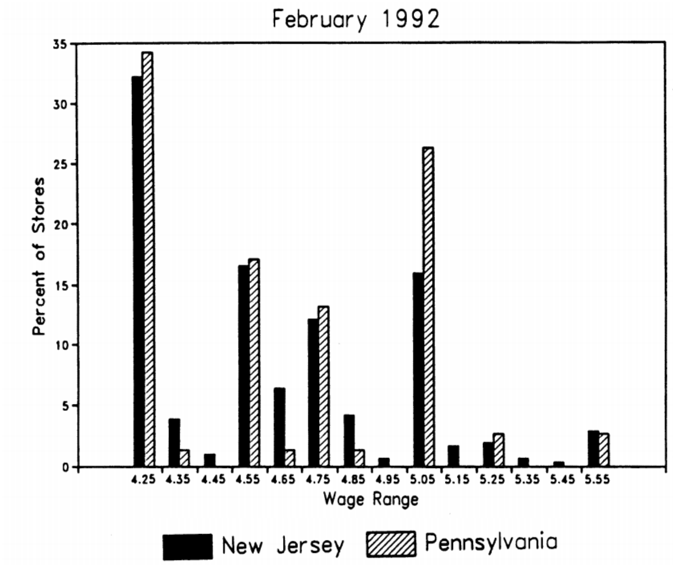
Means of Key Variables

	Stores in: NJ	PA	t-stat
Means in Pre-Period:			
FTE Employment	20.4	23.3	-2.0
	(0.51)	(1.35)	
% full-time employees	32.8	35.0	-0.7
	(1.3)	(2.7)	
Starting wage	4.61	4.63	-0.4
	(0.02)	(0.04)	
Wage = \$4.25 (%)	30.5	32.9	-0.4
	(2.5)	(5.3)	
Price of full meal	3.35	3.04	4.0
	(0.04)	(0.07)	
Hours open (weekday)	14.4	14.5	-0.3
	(0.2)	(0.3)	
Recruiting bonus	23.6	29.1	-1.0
	(2.3)	(5.1)	

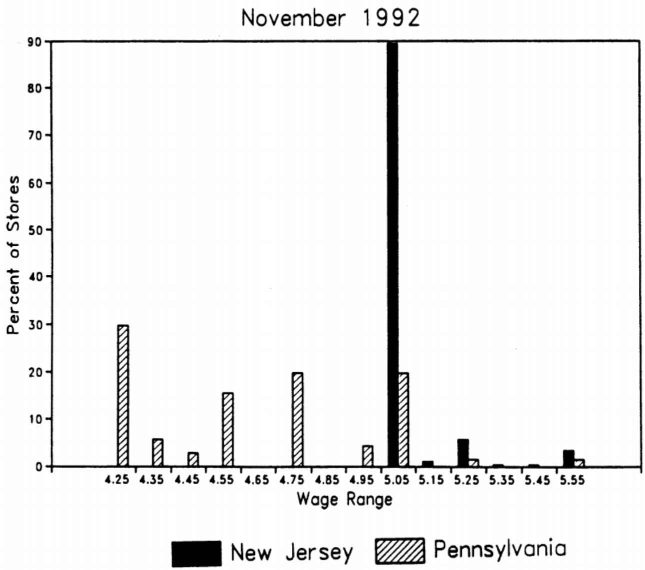
Means of Key Variables

	Stores in:		
	NJ	PA	t-stat
Means in Pre-Period:			
FTE Employment	21.0	21.2	-0.2
	(0.52)	(0.94)	
% full-time employees	35.9	30.4	1.8
	(1.4)	(2.8)	
Starting wage	5.08	4.62	10.8
	(0.01)	(0.04)	
Wage = $4.25 (%)	0.0	25.3	–
		(4.9)	
Price of full meal	3.41	3.03	5.0
	(0.04)	(0.07)	
Hours open (weekday)	14.4	14.7	-0.8
	(0.2)	(0.3)	
Recruiting bonus	20.3	23.4	-0.6
	(2.3)	(4.9)	

The Distribution of Wages - Pre



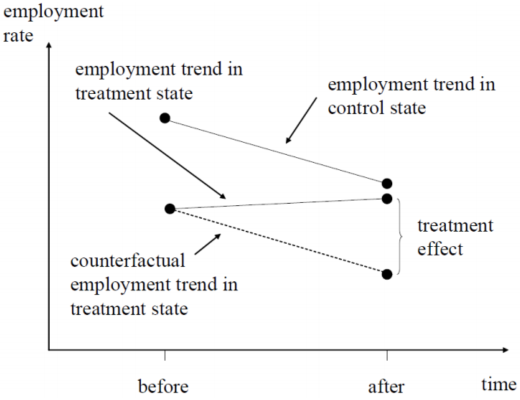
The Distribution of Wages - Post



Key Result

	PA	NJ	Difference (NJ - PA)
FTE Employment Before	23.33	20.44	-2.89
	(1.35)	(0.51)	(1.44)
FTE Employment After	21.17	21.03	-0.14
	(0.94)	(0.52)	(1.07)
Change in mean FTE Employment	-2.16	0.59	2.75
	(1.25)	(0.54)	(1.36)

Difference-in-Differences



We assume the change in the control state (PA) is what the change in the treatment state (NJ) would have been if there had been no treatment. The **Parallel Trends** assumption.

Differences-in-Differences Regression

- We can also implement the DID design using a regression.
- Let NJ_s be a dummy variable for restaurants in New Jersey, and d_t be a dummy variable for observations from November.
- Then we can write the model as,

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s \times d_t) + \epsilon_{ist}$$

where $(NJ_s \times d_t)$ is our treatment variable D_{st}.

Potential Outcomes

- In terms of the potential outcomes, and the regression coefficients, the 4 groups are:

$$\begin{aligned}
E[y_{ist}|s = PA, t = Feb] &= \gamma_{PA} + \lambda_{Feb} = \alpha \\
E[y_{ist}|s = PA, t = Nov] &= \gamma_{PA} + \lambda_{Nov} = \alpha + \lambda \\
E[y_{ist}|s = NJ, t = Feb] &= \gamma_{NJ} + \lambda_{Feb} = \alpha + \gamma \\
E[y_{ist}|s = NJ, t = Nov] &= \gamma_{NJ} + \lambda_{Nov} + \delta = \alpha + \gamma + \lambda + \delta
\end{aligned}$$

- and so we can see that,

$$\begin{aligned}
\delta &= \{E[y_{ist}|s = NJ, t = Nov] - E[y_{ist}|s = NJ, t = Feb]\} \\
&\quad -\{E[y_{ist}|s = PA, t = Nov] - E[y_{ist}|s = PA, t = Feb]\}
\end{aligned}$$

Might there be Potential Problems with this Result?

1) The change in the minimum wage could be correlated with other policy changes

    - This can only be ruled out with detailed institutional knowledge (and seems not to be a problem in this context).

2) The parallel trends assumption could be violated: For example, NJ and PA could have developed differently, because of different business cycles.

    - To try and rule out the second problem, DID papers tend to use a number of approaches.
    - We will consider three of the most popular and powerful

Alternative Control Groups

- Can we find a similar effect with a different control group?
- To provide further support for the parallel trends assumption, Card and Krueger consider different control groups.
- In particular, they compare stores that were paying below \$5/hour before the change to those that were paying more.
- They find that high-wage stores in New Jersey had an almost identical change in employment to Pennsylvania stores.

More Time Periods?

- The most powerful way to provide evidence supporting the parallel trends assumption is to collect data for more than two periods.
- Imagine that Card and Krueger had collected their data for three years prior to the change in minimum wages.
- If employment in fast food restaurants in NJ and PA moved in parallel during the years prior to the minimum wage change, this would support the parallel trends assumption

Placebo Tests

- “Placebo” tests provide a final and very popular strategy to support the parallel trends assumption
- Suppose for example that Card and Krueger had also collected data for Delware, which also border on Pennsylvania and New Jersey.
- Furthermore, suppose that Delaware did not experience any change in its minimum wage like Pennsylvania.
- We could then estimate a DID model between Delaware and Pennsylvania.
- What would be expect to find?

Your Turn

If we estimated a DID model between Delaware and Pennsylvannia, what would we expect to find?

A) An increase (decrease) in employment in Delaware (Pennsylvania)?

B) A decrease (increase) in employment in Delaware (Pennsylvania)?

C) No differences in employment between Delaware and Pennsylvania

Go to **www.menti.com** and use the code **36 76 5**

Supporting the Parallel Trends Assumption

- We have three ways to support (not *prove*) the parallel trends assumption:

    1) Alternative control groups

    2) More data from periods before policy treatment: establish that the treatment and control groups really are moving in **parallel** before the treatment

    3) Placebo tests: should not find an effect

Panel IV: Food Aid and Civil Conflict

- All of the mechanics of IV carry over to the panel data setting.
- Nunn & Qian (2014) study the effects of US food aid on civil conflict:

$$C_{irt} = \beta F_{irt} + X_{irt}\Gamma + \delta r Y_t + \phi_{ir} + \epsilon_{irt}$$

- $C_{irt} = \{\text{conflict in country } i \text{ in region } r \text{ in year } t\}$, F_{irt} is quantity of food aid, X_{irt} are covariates, $\delta r Y_t$ are region-specific time trends, and ϕ_{ir} are country fixed effects.
- Basic idea: use level of wheat production in US one year before, P_{t-1} as an instrument for F_{irt},

$$F_{irt} = \alpha P_{t-1} + X_{irt}\Gamma + \delta r Y_t + \phi_{ir} + \nu_{irt}$$

Panel IV: Food Aid and Civil Conflict

- The exclusion restriction requires that the only channel through which P_{t-1} affects C_{irt} is F_{irt}.
  - Other time trends in C_{irt} that happen to be correlated with P_{t-1} would violate this.

- Develop a DID strategy:

$$\begin{aligned} C_{irt} &= \beta F_{irt} + X_{irt}\Gamma + \varphi_{rt} + \phi_{ir} + \epsilon_{irt} \\ F_{irt} &= \alpha(P_{t-1} \times \bar{D}_{ir}) + X_{irt}\Gamma + \varphi_{rt} + \phi_{ir} + \nu_{irt} \end{aligned}$$

  where $\bar{D}_{ir}$ is country i's propensity to receive food aid, and φ_{rt} are region-year fixed effects.

- Countries that tend to receive more food aid are more affected by swings in US wheat production.

Panel IV: Food Aid and Civil Conflict

TABLE 2—THE EFFECT OF FOOD AID ON CONFLICT: BASELINE SPECIFICATION WITH $P_{t-1} \times D_{ir}$ AS THE INSTRUMENT

	Parsimonious specifications				Baseline specification		
Dependent variable (panels A, B, and C):	Any conflict (1)	Any conflict (2)	Any conflict (3)	Any conflict (4)	Any conflict (5)	Intrastate (6)	Interstate (7)
Panel A. OLS estimates							
US wheat aid (1,000 MT)	−0.00006	−0.00007	−0.00005	−0.00007	−0.00011	−0.00005	−0.00011
	(0.00018)	(0.00018)	(0.00017)	(0.00017)	(0.00017)	(0.00017)	(0.00004)
R^2	0.508	0.508	0.518	0.534	0.549	0.523	0.385
Panel B. Reduced form estimates (× 1,000)**							
Lag US wheat production (1,000 MT) × avg. prob. of any US food aid	0.00829	0.01039	0.01070	0.01133	0.01071	0.00909	−0.00158
	(0.00257)	(0.00263)	(0.00262)	(0.00318)	(0.00320)	(0.00322)	(0.00121)
R^2	0.511	0.512	0.521	0.536	0.551	0.525	0.382
Panel C. 2SLS estimates							
US wheat aid (1,000 MT)	0.00364	0.00303	0.00312	0.00343	0.00299	0.00254	−0.00044
	(0.00174)	(0.00125)	(0.00117)	(0.00106)	(0.00096)	(0.00088)	(0.00033)
Dependent variable (panel D):	US wheat aid (1,000 MT)						
Panel D. First-stage estimates							
Lag US wheat production (1,000 MT) × avg. prob. of any US food aid	0.00227	0.00343	0.00343	0.00330	0.00358	0.00358	0.00358
	(0.00094)	(0.00126)	(0.00120)	(0.00092)	(0.00103)	(0.00103)	(0.00103)
Kleibergen-Paap F-statistic	5.84	7.37	8.24	12.76	12.10	12.10	12.10

Summary

- DID estimation is a popular way of estimating causal effects in the absence of experimental data
- The basic idea is simple: we compare changes in our treatment group to changes in a plausible control group
- The change in the control group provides our counterfactual for what would have happened to the treatment group had it not been treated.
- While the key assumption of parallel trends between the treatment and control group cannot be tested directly, good papers will provide additional evidence to strengthen this assumption
- IV can be combined with panel data.