Frontier Research on Multinational Activities and Global Governance

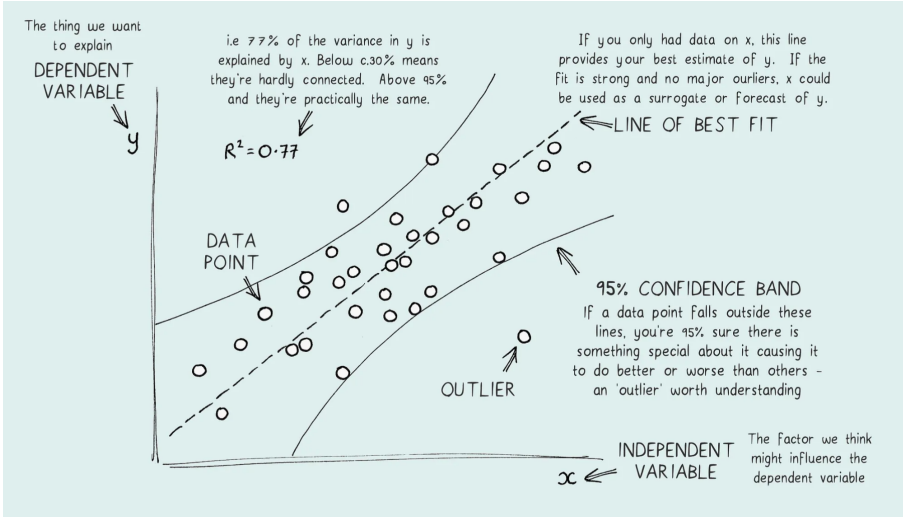
Yuan Zi[1]

[1]Graduate Institute of International Studies (yuan.zi@graduateinstitute.ch)

LX. Popular Empirical Methods: A VERY QUICK Layperson's Guide

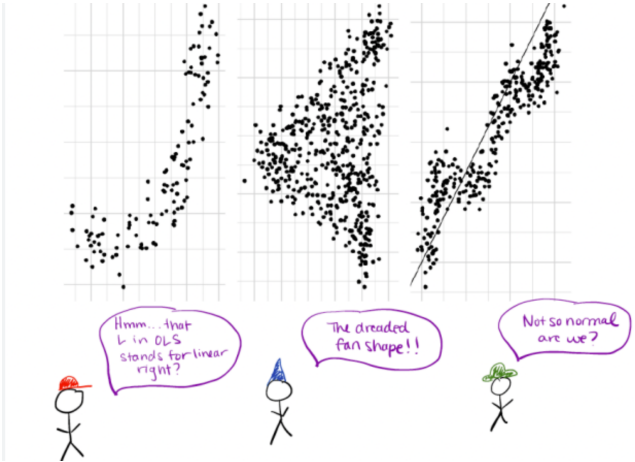
Ordinary Least Square (OLS): Overview

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$



OLS: Assumption

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$



OLS: Essentials

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

- How to interpret β? (keywords: "condition on")
- Log or not log?
- 1 percent vs. 1 unit vs. 1 standard deviation *increase of* x_{i1}...
- Source of bias
  - Omitted variable bias. e.g. Grade $\leftarrow$ IQ (time to study)
  - Simultaneously. e.g. Price $\leftarrow$ Demand

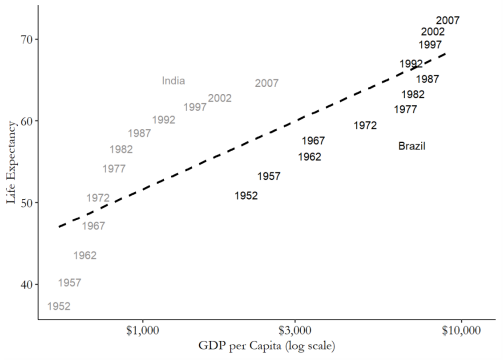
Fix Effects: Essentials

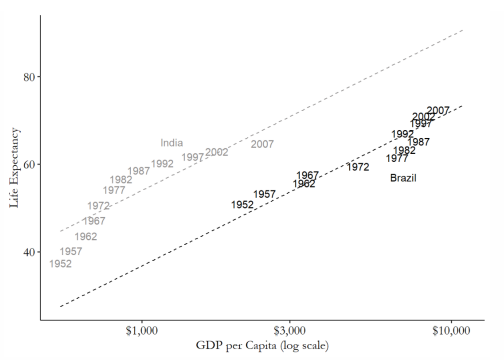
$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_p x_{itp} + D_i + \epsilon_i$$

- An easy way to fix (some) omitted variable bias
- E.g. How "study time" affect grade? Many omitted variables ("confounders"): IQ, family background, past education.. all those time-invariant confounders can be controlled by fixed effects
- Intrepretation: within-comparison
- Extension: cross-section w. group FE; multi-dim with multiple FEs
- Limitation: lose variations; limited fix for omitted variable bias; problems with heterogeneity etc

Fix Effects: Graphical Explanation

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_p x_{itp} + D_i + \epsilon_i$$





Difference in Difference

- Card and Krueger (1992) investigate the important question of how minimum wage affecting employment.
- On April 1, 1992, New Jersey's minimum wage rose from \$4.25 to \$5.05 per hour. Meanwhile the minimum wage in Pennsylvania remained unchanged.
- So NJ is the treatment group, while PA is the control group.
- Two waves of survey were conducted before and after the change in minimum wage.
- Card and Krueger' idea: NJ and PA have different employment levels; but we can compare the difference in their employment changes after the reform date.

Difference in Difference: Basic Idea + Basic Specification

	Before Reform	After Reform	
NJ	$e\bar{m}p_{NJ}$	$e\bar{m}p_{NJ} + \beta + \epsilon_1$	
PA	$e\bar{m}p_{PA}$	$e\bar{m}p_{PA} + \epsilon_2$	

Table: Difference in Difference

Difference in Difference: Basic Idea + Basic Specification

	Before Reform	After Reform	After-Before
NJ	$e\bar{m}p_{NJ}$	$e\bar{m}p_{NJ} + \beta + \epsilon_1$	$\beta + \epsilon_1$
PA	$e\bar{m}p_{PA}$	$e\bar{m}p_{PA} + \epsilon_2$	ϵ_2

Table: Difference in Difference

Difference in Difference: Basic Idea + Basic Specification

	Before Reform	After Reform	After-Before
NJ	$e\bar{m}p_{NJ}$	$e\bar{m}p_{NJ} + \beta + \epsilon_1$	$\beta + \epsilon_1$
PA	$e\bar{m}p_{PA}$	$e\bar{m}p_{PA} + \epsilon_2$	ϵ_2
$NJ_{After-Before} - PA_{After-Before}$			$\beta + \epsilon$

Table: Difference in Difference

Difference in Difference: Basic Idea + Basic Specification

	Before Reform	After Reform	After-Before
NJ	$e\bar{m}p_{NJ}$	$e\bar{m}p_{NJ} + \beta + \epsilon_1$	$\beta + \epsilon_1$
PA	$e\bar{m}p_{PA}$	$e\bar{m}p_{PA} + \epsilon_2$	ϵ_2
$NJ_{After-Before} - PA_{After-Before}$			$\beta + \epsilon$

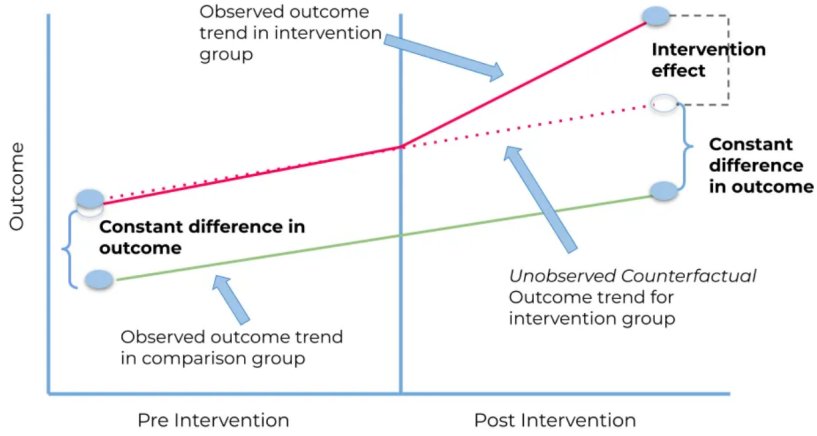
Table: Difference in Difference

Specification

$$y_{it} = \beta_1 + \beta_2 Post_t + \beta_3 Treat_i + \beta Post_t \cdot Treat_i + \Gamma X_i t + \epsilon_i$$

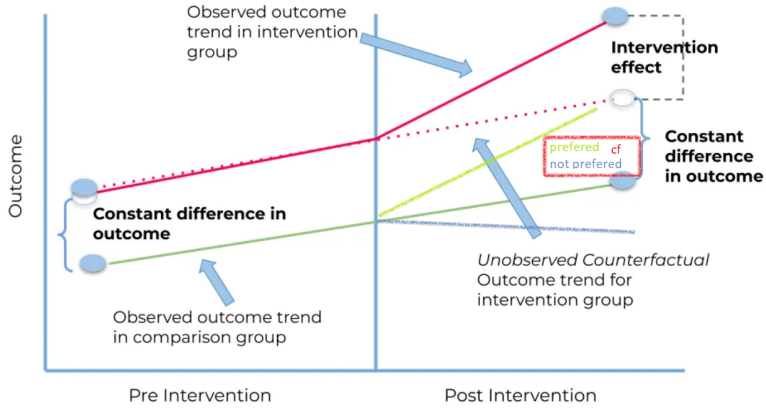
- $Treat_i$: dummy=1 if i is in the treatment group
- $Post_t$: dummy=1 if t is in the post-shock period

Difference in Difference: Parallel Trend Assumption



Parallel-Trend Assumption

Difference in Difference: Balancing Test



Why Balancing Test

Difference in Difference: Variations

$$y_{it} = \beta_1 + \beta_2 Post_t + \beta_3 Treat_i + \beta Post_t \cdot Treat_i + \Gamma X_i t + \epsilon_i$$

- $Treat_i$ is a continuous treatment
- $Post_t$ is different to different groups; e.g., different region imposed minimum wage in different times (Staggered DID)
- Triple DID
- DID are not restricted to one with time
- Combine DID with other methods, such as PSM, or IV
- PS. With controls ΓX_{it}, it would be hard to eyeball the parallel trend assumption; use event study (next slides) plots.

(Panel) Event Study: Specification

Consider a panel covering states s and time periods t. We are interested in estimating the impact of the passage of an event which may occur at different times in different states.

Denoting the outcome of interest as y_{st}, the specification can be written as:

$$\sum_{y_{st}} = \alpha + \sum_{j=2}^{J} \beta_j (\text{Lag } j)_{st} + \sum_{k=1}^{K} \gamma_k (\text{Lead } k)_{st} + \mu_s + \lambda_t + X'_{st}\Gamma + \epsilon_{st} \qquad (1)$$

- Here μ_s and λ_t are state and time fixed effects, X_{st} are (optionally) time-varying controls, and ϵ_{st} is an unobserved error term.

Event Study: Lags and leads

$$\sum_{y_{st}} = \alpha + \sum_{j=2}^{J} \beta_j (\text{Lag } j)_{st} + \sum_{k=1}^{K} \gamma_k (\text{Lead } k)_{st} + \mu_s + \lambda_t + X'_{st}\Gamma + \epsilon_{st} \qquad (1)$$

- Here lags and leads to the event of interest are defined as follows:

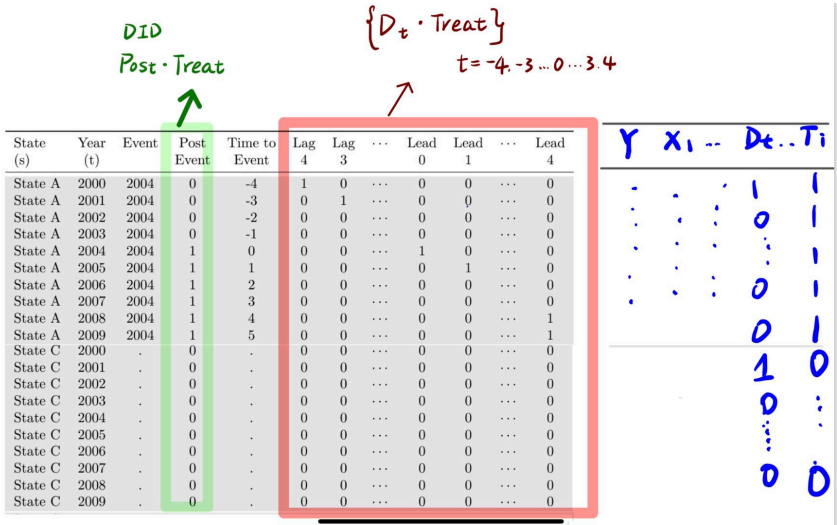
$$(\text{Lag } j)_{st} = 1[t = \text{Events} - j] \text{ for } j \in \{1, ..., J-1\}$$
$$(\text{Lead } k)_{st} = 1[t = \text{Events} + k] \text{ for } k \in \{1, ..., K-1\}$$

- A single lag or lead variable is omitted to capture the baseline difference between areas where the event does and does not occur. In specification 1, as standard, this baseline omitted case is the first lag, where $j = 1$.

Event Study: Shock (lag and leads) structure

State (s)	Year (t)	Event	Post Event	Time to Event	Lag 4	Lag 3	⋯	Lead 0	Lead 1	⋯	Lead 4
State A	2000	2004	0	-4	1	0	⋯	0	0	⋯	0
State A	2001	2004	0	-3	0	1	⋯	0	0	⋯	0
State A	2002	2004	0	-2	0	0	⋯	0	0	⋯	0
State A	2003	2004	0	-1	0	0	⋯	0	0	⋯	0
State A	2004	2004	1	0	0	0	⋯	1	0	⋯	0
State A	2005	2004	1	1	0	0	⋯	0	1	⋯	0
State A	2006	2004	1	2	0	0	⋯	0	0	⋯	0
State A	2007	2004	1	3	0	0	⋯	0	0	⋯	0
State A	2008	2004	1	4	0	0	⋯	0	0	⋯	1
State A	2009	2004	1	5	0	0	⋯	0	0	⋯	1
State B	2000	2005	0	-5	1	0	⋯	0	0	⋯	0
State B	2001	2005	0	-4	1	0	⋯	0	0	⋯	0
State B	2002	2005	0	-3	0	1	⋯	0	0	⋯	0
State B	2003	2005	0	-2	0	0	⋯	0	0	⋯	0
State B	2004	2005	0	-1	0	0	⋯	0	0	⋯	0
State B	2005	2005	1	0	0	0	⋯	1	0	⋯	0
State B	2006	2005	1	1	0	0	⋯	0	1	⋯	0
State B	2007	2005	1	2	0	0	⋯	0	0	⋯	0
State B	2008	2005	1	3	0	0	⋯	0	0	⋯	0
State B	2009	2005	1	4	0	0	⋯	0	0	⋯	1
State C	2000	.	0	.	0	0	⋯	0	0	⋯	0
State C	2001	.	0	.	0	0	⋯	0	0	⋯	0
State C	2002	.	0	.	0	0	⋯	0	0	⋯	0
State C	2003	.	0	.	0	0	⋯	0	0	⋯	0
State C	2004	.	0	.	0	0	⋯	0	0	⋯	0
State C	2005	.	0	.	0	0	⋯	0	0	⋯	0
State C	2006	.	0	.	0	0	⋯	0	0	⋯	0
State C	2007	.	0	.	0	0	⋯	0	0	⋯	0
State C	2008	.	0	.	0	0	⋯	0	0	⋯	0
State C	2009	.	0	.	0	0	⋯	0	0	⋯	0

Event Study for DID



State (s)	Year (t)	Event	Post Event	Time to Event	Lag 4	Lag 3	⋯	Lead 0	Lead 1	⋯	Lead 4
State A	2000	2004	0	-4	1	0	⋯	0	0	⋯	0
State A	2001	2004	0	-3	0	1	⋯	0	0	⋯	0
State A	2002	2004	0	-2	0	0	⋯	0	0	⋯	0
State A	2003	2004	0	-1	0	0	⋯	0	0	⋯	0
State A	2004	2004	1	0	0	0	⋯	1	0	⋯	0
State A	2005	2004	1	1	0	0	⋯	0	1	⋯	0
State A	2006	2004	1	2	0	0	⋯	0	0	⋯	0
State A	2007	2004	1	3	0	0	⋯	0	0	⋯	0
State A	2008	2004	1	4	0	0	⋯	0	0	⋯	1
State A	2009	2004	1	5	0	0	⋯	0	0	⋯	1
State C	2000	.	0	.	0	0	⋯	0	0	⋯	0
State C	2001	.	0	.	0	0	⋯	0	0	⋯	0
State C	2002	.	0	.	0	0	⋯	0	0	⋯	0
State C	2003	.	0	.	0	0	⋯	0	0	⋯	0
State C	2004	.	0	.	0	0	⋯	0	0	⋯	0
State C	2005	.	0	.	0	0	⋯	0	0	⋯	0
State C	2006	.	0	.	0	0	⋯	0	0	⋯	0
State C	2007	.	0	.	0	0	⋯	0	0	⋯	0
State C	2008	.	0	.	0	0	⋯	0	0	⋯	0
State C	2009	.	0	.	0	0	⋯	0	0	⋯	0

Event Study Plot for DID

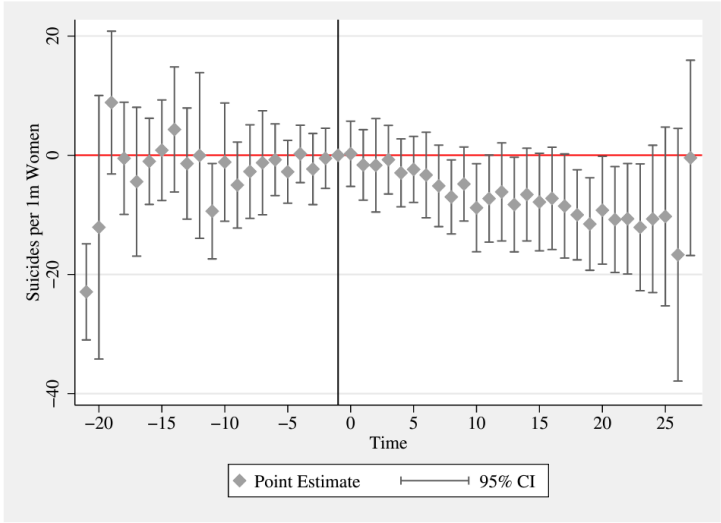


Figure 1: Event Study Example Based on No-fault Divorce Reforms

Propensity Score Matching

- To be added

Instrument Variable Strategy

- To be added