# Fixed Effects and Difference-in-Differences

Han Zhang

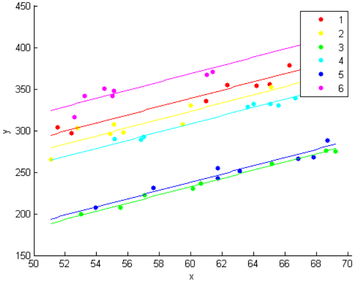
Mar 30, 2021

# Outline

# Readings

- Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricists Companion* . Princeton University Press, 2009. (Chapters 5)

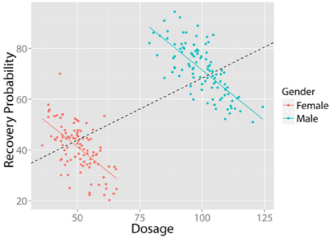
# Fixed Effect Regression




- Each group has its own intercepts
- But slope is the same
- Also called varying intercept (same slope) model
  - contrasting varying intercept and varying slope model

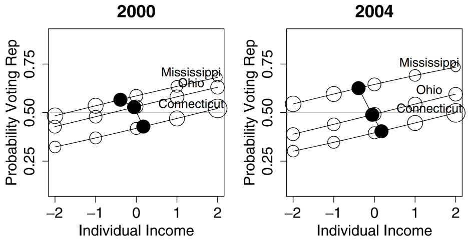
# Fixed effect and Simpson's paradox

- Simpson's paradox: correlation between $X$ and $Y$ is reversed at individual and at macro-level
  - Remember the kidney stone example in the 1st lecture?
- Kiev it et al., 2013, "Simpson's paradox in psychological science"
- The mean of male/female suggests that dosage and recovery prob is *positively* correlated
- Within each group, at individual level, dosage and recovery prob is *negatively* correlated



# Another example: blue state vs red state

- Gelman, 2007, "Red State, Blue State, Rich State, Poor State"
- Within each state, rich people have a high prob to vote for Rep (individual level)
- But richer states (e.g., NY, CA) have a higher prob to vote for Dem (macro level)



# Group Structure

- In these studies, the correct relationship is what you obtained at the micro level
- If we mistakenly model the data at aggregate level, we may obtain a wrong relation
- Two ways of running regression within group structure:
- Fixed effect: same slope, different intercept
  - The simplest way: add group as additional regressor
  - Also often called as least square dummy variable model (LSDV)
- Random effect: different slope, different intercept
  - It's hard to fit; if you are interested, read Gelman and Hill, *Data Analysis using Regression and Multilevel Hierarchical Models*, 2007.

# Counterfactual view of fixed effect

- Selection on observables is often a very strong assumption; how do we make it more realistic?
- Intuition: cluster-randomized experiments, with conditional random assignment
  - Within a certain group (e.g., school), treatment assignment is randomized
- Similarly, we may find natural grouping in real worlds that let you make selection on observable assumptions, within groups

# Twin studies

- Example: we are interested in whether college degree leads to higher wages
  - Unobserved factors we cannot control: many family and genetic backgrounds
- Twin studies: find twins that grow up together, but have different schooling length
  - We can assume that twins are very similar in their unobserved family and genetic backgrounds;
  - In other words, within each group (here, twin pair), units share unobserved but fixed characteristics
  - If we compare within group, we can rule out unobserved but fixed characteristics

# Counterfactual framework

- Use $i$ to denote individuals, and $j$ for twin pairs, $X_{ij}$ for observed individual-level variables, and $Y_{ij}$ is outcome
- $D_{ij} = 1$ if $i$ received college education and 0 if not
- Twin studies assume that:
  - Unobserved factors $\alpha_j$ are at the group level
  - Selection bias is 0 within each group, and conditional on $X$
    - In math, $E(Y^0_{ij}|D_{ij} = 1, \alpha_j, X_{ij}) = E(Y^0_{ij}|D_{ij} = 0, \alpha_j, X_{ij})$
    - That is, if $A$ and $B$ are twins, and $A$ is treated and $B$ is not. We expect that $A$'s counterfactual outcome can be predicted by $B$'s observed outcome
- With the above assumptions, we can estimate ATT using non-parametric estimator
  - Calculate within-group differences in outcome
  - Take the mean of within-group differences

# Fixed Effects Regression

- For non-parametric estimators, the above assumptions are enough
- If we want to derive a regression estimator, we need additional assumptions.
- Linear and additive counterfactual outcome for untreated units

$$E(Y_{ij}^0 | j, X_{ij}) = \alpha_j + X_{ij}\beta$$

- Constant individual-level treatment effect $\rho$

$$E(Y_{ij}^1 | A_j, X_{ij}, D_{ij}) = E(Y_{ij}^0 | A_j, X_{ij}) + \rho = \alpha_j + X_{ij}\beta + \rho$$

- Constant treatment effect assumptions is shared by regression estimator of experiments
- Linear and additive counterfactual is unique here; we do not need it to derive regression estimator for randomized experiments

# Fixed effects Regressions

- The above assumptions mean that:

$$E(Y_{ij}|X_{ij}, D_{ij}) = \alpha_j + D_{ij}\rho + X_{ij}\beta \tag{1}$$

- Or alternatively, $Y_{ij} = \alpha_j + D_{ij}\rho + X_{ij}\beta + \varepsilon_{ij}$, $E(\varepsilon_{ij}) = 0$, $E(\varepsilon_{ij}|D_{ij}) = 0$ and $E(\varepsilon_{ij}|X_{ij}) = 0$
- This is a regression with group-specific fixed effects ($\alpha_j$);
- From a pure prediction perspective, fixed-effect regressions means there is separate intercept $\alpha_j$ for each group $j$
- From causal inference perspective, $\alpha_j$ were motivated as unobserved factors at the group-level
  - From now on we take the causal inference perspective

# Fixed effect regression

- With all assumptions we have so far, regression coefficient $\rho$ identifies causal effect $ATT$

$$\begin{aligned} \rho &= E(Y_{ij}^1|A_j, X_{ij}) - E(Y_{ij}^0|A_j, X_{ij}) \\ &= E(Y_{ij}^1|A_j, X_{ij}, D_{ij}=1) - E(Y_{ij}^0|A_j, X_{ij}, D_{ij}=1) \\ &= ATT \end{aligned} \tag{2}$$

- A recap of assumptions we needed:
  - Assumptions about treatment assignment process:
    - Unobserved factors are at the group level;
    - Selection on observables, within groups
  - Assumptions about parametric model (non-parametric estimator does not require these; regression estimator need these)
    - Linear and additive counterfactual outcome
    - Constant treatment effect

## Group Structure in panel data

- Panel data often exhibit natural group structure
  - Repeated measure for a single unit
- Dell, Jones, and Olken (2011): Temperature Shocks and Economic Growth: Evidence from the Last Half Century, AEJ.
- Key question: whether hot temperature leads to poor economy

  *There are countries where the excess of heat enervates the body, and renders men so slothful and dispirited that nothing but the fear of chastisement can oblige them to perform any laborious duty. [Montesquieu, 1750]*

- Problem: many unobserved factors can explain poor economy
- Solution: assume that unobserved factors are at the unit level but time-invariant
- Then, within each unit:
  - Whether hotter years result in lower GDP

# Fixed effects as a dummy variable model

$$g_{ct} = \alpha_c + \beta TEMP_{ct} + \varepsilon_{ct} \quad (3)$$

- We have learned that $\beta = ATT$, if all assumptions hold

$$g_{ct} = \alpha_c + \beta TEMP_{ct} + \varepsilon_{ct}$$

- $\alpha_c$ are unobserved unit-level factors.
- How do we estimate the above model in software?
- We can simply treat $\alpha_c$ as parameters to estimate for each unit (i.e., least square dummy variable model)

| | $g_{ct}$ | $TEMP_{ct}$ | $\alpha_{USA}$ | $\alpha_{INDO}$ | $a_{NIGER}$ |
|---|---|---|---|---|---|
| USA $_{2010}$ | 3 | 12 | 1 | 0 | 0 |
| USA $_{2011}$ | 3.2 | 14 | 1 | 0 | 0 |
| Indonesia2010 | 1 | 22 | 0 | 1 | 0 |
| Indonesia$_{2011}$ | 1.3 | 23 | 0 | 1 | 0 |
| Niger $_{2010}$ | 0.1 | 28 | 0 | 0 | 1 |
| Niger $_{2011}$ | 0.1 | 27 | 0 | 0 | 1 |

(4)

# Fixed effects as deviation from means

- We have learned that $\beta = ATT$, if all assumptions hold

$$g_{ct} = \alpha_c + \beta TEMP_{ct} + \varepsilon_{ct}$$

- $\alpha_c$ are unobserved unit-level factors.
- In fact, $\alpha_c$ do not matter, if the goal is to estimate $\beta$
- Take the mean within each country:

$$\overline{g_c} = \overline{\alpha_c} + \beta \overline{TEMP_c} + \bar{\varepsilon}_c = \alpha_c + \beta \overline{TEMP_c} + \bar{\varepsilon}_c \tag{5}$$

- Subtracting to get deviation from means:

$$g_{ct} - \overline{g_c} = \beta \left( TEMP_{ct} - \overline{TEMP_c} \right) + \varepsilon'_{ct} \tag{6}$$

- In other words, the estimation of causal effect $\beta$ can be purely done within each group
- this representation is called within estimator

# Fixed effects as deviation from means

$$
\begin{array}{lrr}
 & g_{ct} - \overline{g_c} & TEMP_{ct} - \overline{TEMP_c} \\
\text{USA}_{2010} & -0.1 & -1 \\
\text{USA}_{2011} & 0.1 & 1 \\
\text{Indonesia}_{2010} & -0.15 & -0.5 \\
\text{Indonesia}_{2011} & 0.15 & 0.5 \\
\text{Niger}_{2010} & 0 & 0.5 \\
\text{Niger}_{2011} & 0 & -0.5
\end{array} \tag{7}
$$

- Is the overall relationship between temperature and growth positive or negative?

# Fixed effects as difference in time

$$g_{ct} = \alpha_c + \beta TEMP_{ct} + \varepsilon_{ct}$$

- In panel data, from a causal inference perspective, an alternative way to rule out unit-level unobserved factors is to take difference by time
- Use $\Delta$ to represent difference between the time $t$ and time $t-1$
  - e.g., $\Delta g_{ct} = g_{ct} - g_{c,t-1}$
- Apply $\Delta$ on both sides:

$$\Delta g_{ct} = \beta \Delta TEMP_{ct} + \Delta \varepsilon_{ct} \tag{8}$$

- Again, taking the first difference within groups cancels out the group-level unobserved factors
- In panel data, this is often known as taking the first difference

# Fixed effects as different in time

- Taking the first difference with respect to time:

$$\begin{array}{lrr} & \Delta g_{ct} & \Delta TEMP_{ct} \\ \text{USA} & 0.2 & 2 \\ \text{Indonesia} & 0.3 & 1 \\ \text{Niger} & 0 & -1 \end{array} \tag{9}$$

- This is known as first difference estimator
- Is the overall relationship between temperature and growth positive or negative?

# Violation of one-way fixed effect assumption

- We have seen cases with one grouping (one-way fixed effect). One key assumption is
  - unobserved confoundings is fixed within group
  - E.g., in Dell et al. example, group is country
    - all unobserved factors are fixed for each country
- What if the assumption is not correct? Say, unobserved can also vary by time?
- E.g., each year there may be some common global trends (e.g., financial crisis in 2008) that impact all countries

# Two-way fixed effects

- In such case, we have two natural groups: country and time
- If we further assume that unobserved factors are either fixed by country (but vary by time) or fixed by time (but vary by country)
- Two-way fixed effect regression

$$Y_{it} = \alpha_i + \beta_t + D_{it}\rho + X_{it}\beta + \varepsilon_{it} \tag{10}$$

- $\alpha_i$: country fixed effect
- $\beta_t$: time fixed effect
- $\rho = ATT$ if:
    - Unobserved confounders are either fixed by country or fixed by time
    - all other assumptions of one-way fixed effects (linear and additive counterfactual outcome, constant treatment effect) are also needed.

# Difference-in-differences

- With two groups and two periods, two-way fixed effects become difference-in-differences (DiD) estimator
- Card and Kruger, 1994, “Minimum wages and employment”, AER
- Whether raising mimimum wage lead to decrease in employment?
- Treatment $D$: New Jersey raised the state minimum wage from \$4.25 to \$5.05 on April 1, 1992. Nearby Pennsylvania did not.
- Outcome $Y$: average change in Full-time Equivalent (FTE) jobs in fast food restaurants
- DiD approximates experimental ideal explicitly:
  - One treatment and one control group
  - Outcome is measured as change of $Y$ after/before the treatment

# DiD as differences-in-means

- MHE, Table 5.2.1

| time/state | PA (control) | NJ (treated) | NJ - PA |
|---|---|---|---|
| Before ($t = 0$) | 23.33 | 20.44 | -2.89 |
| After ($t = 1$) | 21.17 | 21.03 | -0.14 |
| After - Before | -2.16 | 0.59 | 2.76 |

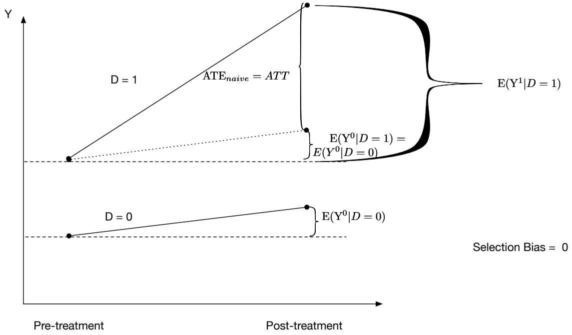
- First calculate difference within state by time, which cancels out state fixed effect
- Next calculate difference of the differences, which cancels out time fixed effect

# DiD as two-way fixed effect

$$Y_{st} = \alpha_s + \beta_t + \rho D_{st} + \epsilon_{st}, D_{st} = 1 \text{ when } s = \text{treated } \& t = 1 \tag{11}$$

- First calculate difference within state by time, which cancels out time-invariant confounders $\alpha_s$
  - $Y_{NJ,1} - Y_{NJ,0} = \beta_1 - \beta_0 + \rho(1-0) = -2.16$
  - $Y_{PA,1} - Y_{PA,0} = \beta_1 - \beta_0 + \rho(0-0) = 0.59$
- Next calculate difference of the differences, which cancels out unit-invariant confounders $\beta_t$
  - $(Y_{NJ,1} - Y_{NJ,0}) - (Y_{PA,1} - Y_{PA,0}) = \rho$

# DiD Assumptions: Parallel Trends




- In DiD setting, selection bias $= 0$ $(E(Y^0|D=1) = E(Y^0|D=0))$ assumption can be visualized
  - $Y^0 = Y^0_{i,t=1} - Y^0_{i,t=0}$ is the change of $i$ after/before treatment
- It's often called parallel trends assumption: if there were no treatment, treated units $(D=1)$'s counterfactual outcome $Y^0$ would be the same as that of the control units

# Difference-in-differences as a linear regression with interaction

- Difference-in-differences (DiD later) is a special case of two-way fixed effects, with two units and two time periods
- We can further simplify DiD as an interaction model, which is easier to work with
- Define $d_s$ as treatment unit dummy (here, $d_s = 1$ if the state is NJ)
- Define $d_t$ as time dummy($d_t = 1$ if after treatment)
- $D_{st} = d_s \cdot d_t$ ($D_{st} = 1$ when both $s$ is treated and $t$ is 1)

# DiD as a linear regression with interaction

The two models are equivalent:

$$
\begin{aligned}
Y_{st} &= \alpha_s + \beta_t + \rho D_{st} + \epsilon_{st}, \text{ fixed effect} \\
Y_{st} &= \gamma + \lambda d_s + \tau d_t + \delta(d_s \cdot d_t) + \epsilon_{st}, \text{ interaction}
\end{aligned}
\tag{12}
$$

- Because the previous DiD table can be exactly expressed as a linear regression with interactive model

| time/state | PA (control) | NJ (treated) | NJ - PA |
|---|---|---|---|
| Before (t = 0) | $\gamma$ | $\gamma + \lambda$ | $\lambda$ |
| After (t = 1) | $\gamma + \tau$ | $\gamma + \lambda + \tau + \delta$ | $\lambda + \delta$ |
| After - Before | $\tau$ | $\tau + \lambda$ | $\delta$ |

- And estimate effect $\rho = \delta = ATT$

# Additional Controls

- Non-parametric model needs fewer assumptions
- But with the regression setup (either fixed-effect or interaction model), it is easier to add additional controls
- E.g., we may have some individual-level controls $X_{ist}$ (characteristics of employers);
- We can easily include these $X_{ist}$, and use micro-level data (instead of aggregate state-year level)

$$
\begin{aligned}
Y_{ist} &= \alpha_s + \beta_t + \rho D_{st} + \mu X_{ist} + \epsilon_{ist}, \text{ fixed effect} \\
&= \text{or, equivalently} \\
Y_{ist} &= \gamma + \lambda d_s + \tau d_t + \delta(d_s \cdot d_t) + \mu X_{ist} + \epsilon_{ist}, \text{ interaction}
\end{aligned} \tag{13}
$$

# Standard errors in fixed effects and DiD

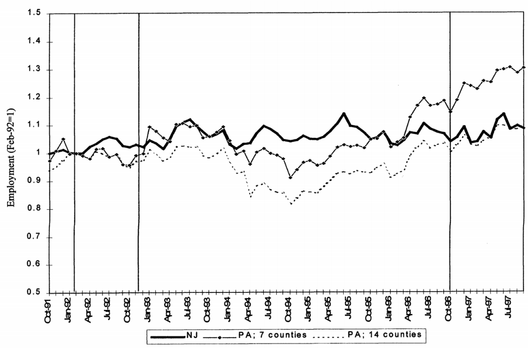
- Last, standard errors should be clustered at group levels
  - Here, we used two-way fixed effects so that standard errors should be clustered at both state and time level
  - It is very easy to see this if we formulate DiD as an extension of two-way fixed effect regression: group structure are clearly specified
  - But it is not so transparent if you formuate DiD as a linear regression with interactions
- Bertrand, Duflo and Mullainathan, 2004, "How much should we trust differences-in-differences estimates", QJE
  - DiD estimates from a linear regression with interactions will seriously underestimate standard error of coefficients, if standard errors are not clustered at group level

# Testing Parallel Trends Assumption

- Strictly speaking, parallel trends assumption (or more general, unobserved factors are either fixed at unit or time-level) are not testable
  - Because it assumes something about $E(Y^0|D=1)$ which can never be observed
- In simple setting such as DiD, and when we have more than one period before the treatment
- We can perform check on the pre-trends
  - If the two groups truly would have exhibited parallel trends in the absence of treatment
  - Then their trends in $Y$ would be parallel prior to treatment
- Note that parallel pre-trends does not mean parallel trends
  - To have the parallel trends assumption we need help from theory
- But violation of parallel pre-trends suggests that parallel trends assumption do not hold

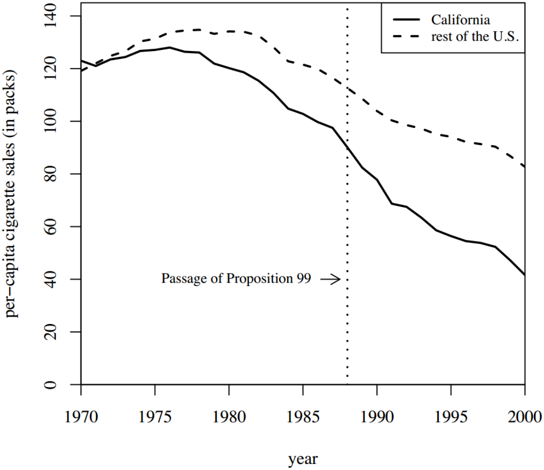
# Graphical Test of parallel pre-trends

- A bad one: MHE, Card and Krueger, 1994, AER (Figure 5.2.2 in MHE)



# Graphical test of parallel pre-trends

- A good one: Abadie, Diamond and Hainmuller, 2010, JANA.
- Effect of Proposition 99 (a tobacco control program in Cali.)



# Fixed Effects and DiD

- Two-way Fixed Effects and DiD share the same set of assumptions (if you use regression estimator)
- DiD is
  - simpler (two periods, two units)
  - approximates experimental ideal explicitly; force you to think through the language of randomized controlled experiments
  - More easily to visualize and thus to recognize wrong assumptions
- Two-way fixed effects can incorporate more years and units and thus produce more precise and robust results
  - But you may just easily assume that the unobserved factors are fixed at unit or time level, without thinking it clearly as an experiment
  - And it's harder to visualize its assumptions