Week 12: Repeated Observations and Panel Data

Brandon Stewart[1]

Princeton

December 12 and 14, 2016

[1] These slides are heavily influenced by Matt Blackwell, Adam Glynn, Jens Hainmueller

Where We've Been and Where We're Going...

- Last Week
  - causal inference with unmeasured confounding
- This Week
  - Monday:
    - panel data
    - diff-in-diff
    - fixed effects
  - Wednesday:
    - Q&A
    - fun With
    - wrap-Up
- The Following Week
  - break!
- Long Run
  - probability → inference → regression → causality

Questions?

Gameplan

Gameplan

- Presentations

Gameplan

- Presentations
- Wednesday's Class

Gameplan

- Presentations
- Wednesday's Class
- Final Exam

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but. . .

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but. . .
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier
  - provide benefits more efficiently

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier
  - provide benefits more efficiently
  - possess some cultural trait correlated with better health outcomes

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier
  - provide benefits more efficiently
  - possess some cultural trait correlated with better health outcomes
- If we have data on countries over time, can we make any progress in spite of these problems?

Ross data

```
##       cty_name year democracy infmort_unicef
## 1 Afghanistan 1965         0            230
## 2 Afghanistan 1966         0             NA
## 3 Afghanistan 1967         0             NA
## 4 Afghanistan 1968         0             NA
## 5 Afghanistan 1969         0             NA
## 6 Afghanistan 1970         0            215
```

Notation for panel data

- Units, $i = 1, \ldots, n$

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Slightly different focus than clustered data we covered earlier

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.
  - The main difference is what level of analysis we care about (individual, city, county, state, country, etc).

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.
  - The main difference is what level of analysis we care about (individual, city, county, state, country, etc).
- Time is a typical application, but applies to other groupings:

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.
  - The main difference is what level of analysis we care about (individual, city, county, state, country, etc).
- Time is a typical application, but applies to other groupings:
  - counties within states

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.
  - The main difference is what level of analysis we care about (individual, city, county, state, country, etc).
- Time is a typical application, but applies to other groupings:
  - counties within states
  - states within countries

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Slightly different focus than clustered data we covered earlier
  - Panel: we have repeated measurements of the same units
  - Clustering: units are clustered within some grouping.
  - The main difference is what level of analysis we care about (individual, city, county, state, country, etc).
- Time is a typical application, but applies to other groupings:
  - counties within states
  - states within countries
  - people within countries, etc.

Nomenclature

- Panel data: large n, relatively short T

Nomenclature

- Panel data: large n, relatively short T
- Time series, cross-sectional (TSCS) data: smaller n, large T

Nomenclature

- Panel data: large n, relatively short T
- Time series, cross-sectional (TSCS) data: smaller n, large T
- We are primarily going to focus on similarities today but there are some differences.

Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- $\mathbf{x}_{it}$ is a vector of covariate (possibly time-varying)

Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- $\mathbf{x}_{it}$ is a vector of covariate (possibly time-varying)
- a_i is an unobserved time-constant unit effect ("fixed effect")

Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

- $\mathbf{x}_{it}$ is a vector of covariate (possibly time-varying)
- a_i is an unobserved time-constant unit effect ("fixed effect")
- u_{it} are the unobserved time-varying "idiosyncratic" errors

Model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

- $\mathbf{x}_{it}$ is a vector of covariate (possibly time-varying)
- a_i is an unobserved time-constant unit effect ("fixed effect")
- u_{it} are the unobserved time-varying "idiosyncratic" errors
- $v_{it} = a_i + u_{it}$ is the combined unobserved error:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it}$$

Pooled OLS

- Pooled OLS: pool all observations into one regression

Pooled OLS

- Pooled OLS: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.

Pooled OLS

- Pooled OLS: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:

Pooled OLS

- Pooled OLS: pool all observations into one regression
- Treats all unit-periods (each *it*) as an iid unit.
- Has two problems:
  1. Heteroskedasticity (see clustering from diagnostics week)

Pooled OLS

- Pooled OLS: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:
  1. Heteroskedasticity (see clustering from diagnostics week)
  2. Possible violation of zero conditional mean errors

Pooled OLS

- Pooled OLS: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:
  1. Heteroskedasticity (see clustering from diagnostics week)
  2. Possible violation of zero conditional mean errors
- Both problems arise out of ignoring the unmeasured heterogeneity inherent in a_i

Pooled OLS with Ross data

```
pooled.mod <- lm(log(kidmort_unicef) ~ democracy + log(GDPcur),
                 data = ross)
summary(pooled.mod)

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.76405    0.34491   28.31   <2e-16 ***
## democracy   -0.95525    0.06978  -13.69   <2e-16 ***
## log(GDPcur) -0.22828    0.01548  -14.75   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7948 on 646 degrees of freedom
##   (5773 observations deleted due to missingness)
## Multiple R-squared:  0.5044, Adjusted R-squared:  0.5029
## F-statistic: 328.7 on 2 and 646 DF,  p-value: < 2.2e-16
```

Unmeasured heterogeneity

- Assume that zero conditional mean error holds for the idiosyncratic error:

$$\mathbb{E}[u_{it}|\mathbf{X}] = 0$$

Unmeasured heterogeneity

- Assume that zero conditional mean error holds for the idiosyncratic error:

$$\mathbb{E}[u_{it}|\mathbf{X}] = 0$$

- But time-constant effect, a_i, is correlated with the $\mathbf{X}$:

$$\mathbb{E}[a_i|\mathbf{X}] \neq 0$$

Unmeasured heterogeneity

- Assume that zero conditional mean error holds for the idiosyncratic error:

$$\mathbb{E}[u_{it}|\mathbf{X}] = 0$$

- But time-constant effect, a_i, is correlated with the $\mathbf{X}$:

$$\mathbb{E}[a_i|\mathbf{X}] \neq 0$$

- Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.

Unmeasured heterogeneity

- Assume that zero conditional mean error holds for the idiosyncratic error:

$$\mathbb{E}[u_{it}|\mathbf{X}] = 0$$

- But time-constant effect, a_i, is correlated with the $\mathbf{X}$:

$$\mathbb{E}[a_i|\mathbf{X}] \neq 0$$

- Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.
- Ignore the heterogeneity $\rightsquigarrow$ correlation between the combined error and the independent variables:

$$\mathbb{E}[v_{it}|\mathbf{X}] = \mathbb{E}[a_i + u_{it}|\mathbf{X}] \neq 0$$

Unmeasured heterogeneity

- Assume that zero conditional mean error holds for the idiosyncratic error:

$$\mathbb{E}[u_{it}|\mathbf{X}] = 0$$

- But time-constant effect, a_i, is correlated with the $\mathbf{X}$:

$$\mathbb{E}[a_i|\mathbf{X}] \neq 0$$

- Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.
- Ignore the heterogeneity $\rightsquigarrow$ correlation between the combined error and the independent variables:

$$\mathbb{E}[v_{it}|\mathbf{X}] = \mathbb{E}[a_i + u_{it}|\mathbf{X}] \neq 0$$

- Pooled OLS will be biased and inconsistent because zero conditional mean error fails for the combined error.

First differencing

- First approach: compare changes over time as opposed to levels

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$y_{i1} = \mathbf{x}_{i1}'\boldsymbol{\beta} + a_i + u_{i1}$$
$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta} + a_i + u_{i2}$$

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$\begin{aligned} y_{i1} &= \mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1} \\ y_{i2} &= \mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2} \end{aligned}$$

- Look at the change in y over time:

$$\Delta y_i = y_{i2} - y_{i1}$$

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$y_{i1} = \mathbf{x}_{i1}'\boldsymbol{\beta} + a_i + u_{i1}$$
$$y_{i2} = \mathbf{x}_{i2}'\boldsymbol{\beta} + a_i + u_{i2}$$

- Look at the change in y over time:

$$\Delta y_i = y_{i2} - y_{i1}$$

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$y_{i1} = \mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1}$$
$$y_{i2} = \mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2}$$

- Look at the change in y over time:

$$\begin{aligned}\Delta y_i &= y_{i2} - y_{i1} \\ &= (\mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2}) - (\mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1})\end{aligned}$$

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$\begin{aligned} y_{i1} &= \mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1} \\ y_{i2} &= \mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2} \end{aligned}$$

- Look at the change in y over time:

$$\begin{aligned} \Delta y_i &= y_{i2} - y_{i1} \\ &= (\mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2}) - (\mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1}) \\ &= (\mathbf{x}'_{i2} - \mathbf{x}'_{i1})\boldsymbol{\beta} + (a_i - a_i) + (u_{i2} - u_{i1}) \end{aligned}$$

First differencing

- First approach: compare changes over time as opposed to levels
- Intuitively, the levels include the unobserved heterogeneity, but changes over time should be free of this heterogeneity
- Two time periods:

$$
\begin{aligned}
y_{i1} &= \mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1} \\
y_{i2} &= \mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2}
\end{aligned}
$$

- Look at the change in y over time:

$$
\begin{aligned}
\Delta y_i &= y_{i2} - y_{i1} \\
&= (\mathbf{x}'_{i2}\boldsymbol{\beta} + a_i + u_{i2}) - (\mathbf{x}'_{i1}\boldsymbol{\beta} + a_i + u_{i1}) \\
&= (\mathbf{x}'_{i2} - \mathbf{x}'_{i1})\boldsymbol{\beta} + (a_i - a_i) + (u_{i2} - u_{i1}) \\
&= \Delta \mathbf{x}'_i\boldsymbol{\beta} + \Delta u_i
\end{aligned}
$$

First differences model

$$\Delta y_i = \Delta \mathbf{x}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{x}_{it}$ is the same as the coefficient on the changes $\Delta \mathbf{x}_i$
- fixed effect/unobserved heterogeneity, a_i drops out (depends on time-constancy!)

First differences model

$$\Delta y_i = \Delta \mathbf{x}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{x}_{it}$ is the same as the coefficient on the changes $\Delta \mathbf{x}_i$
- fixed effect/unobserved heterogeneity, a_i drops out (depends on time-constancy!)
- Now if $\mathbb{E}[u_{it}|\mathbf{X}] = 0$, then, $\mathbb{E}[\Delta u_i|\Delta X] = 0$ and zero conditional mean error holds.

First differences model

$$\Delta y_i = \Delta \mathbf{x}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{x}_{it}$ is the same as the coefficient on the changes $\Delta \mathbf{x}_i$
- fixed effect/unobserved heterogeneity, a_i drops out (depends on time-constancy!)
- Now if $\mathbb{E}[u_{it}|\mathbf{X}] = 0$, then, $\mathbb{E}[\Delta u_i | \Delta X] = 0$ and zero conditional mean error holds.
- No perfect collinearity: $\mathbf{x}_{it}$ has to change over time for some units

First differences model

$$\Delta y_i = \Delta \mathbf{x}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{x}_{it}$ is the same as the coefficient on the changes $\Delta \mathbf{x}_i$
- fixed effect/unobserved heterogeneity, a_i drops out (depends on time-constancy!)
- Now if $\mathbb{E}[u_{it}|\mathbf{X}] = 0$, then, $\mathbb{E}[\Delta u_i | \Delta X] = 0$ and zero conditional mean error holds.
- No perfect collinearity: $\mathbf{x}_{it}$ has to change over time for some units
- Differencing will reduce the variation in the independent variables and increase standard errors

First differences in R

```
library(plm)

fd.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur), data = ross,
                      index = c("id", "year"), model = "fd")
summary(fd.mod)

## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = log(kidmort_unicef) ~ democracy + log(GDPcur),
##     data = ross, model = "fd", index = c("id", "year"))
##
## Unbalanced Panel: n=166, T=1-7, N=649
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -0.9060 -0.0956  0.0468  0.1410  0.3950
##
## Coefficients :
##              Estimate Std. Error  t-value Pr(>|t|)
## (intercept) -0.149469   0.011275 -13.2567  < 2e-16 ***
## democracy   -0.044887   0.024206  -1.8544  0.06429 .
## log(GDPcur) -0.171796   0.013756 -12.4886  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    23.545
## Residual Sum of Squares: 17.762
## R-Squared      :  0.24561
##       Adj. R-Squared :  0.24408
## F-statistic: 78.1367 on 2 and 480 DF, p-value: < 2.22e-16
```

Differences-in-differences

Differences-in-differences

- Often called “diff-in-diff”, it is a special kind of FD model

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:

Differences-in-differences

- Often called “diff-in-diff”, it is a special kind of FD model
- Let x_{it} be an indicator of a unit being “treated” at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i
  - $x_{i2} = 1$ for the "treated group"

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i
  - $x_{i2} = 1$ for the "treated group"
- Here is the basic model:

$$y_{it} = \beta_0 + \delta_0 d_t + \beta_1 x_{it} + a_i + u_{it}$$

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i
  - $x_{i2} = 1$ for the "treated group"
- Here is the basic model:

$$y_{it} = \beta_0 + \delta_0 d_t + \beta_1 x_{it} + a_i + u_{it}$$

- d_t is a dummy variable for the second time period

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i
  - $x_{i2} = 1$ for the "treated group"
- Here is the basic model:

$$y_{it} = \beta_0 + \delta_0 d_t + \beta_1 x_{it} + a_i + u_{it}$$

- d_t is a dummy variable for the second time period
  - $d_2 = 1$ and $d_1 = 0$

Differences-in-differences

- Often called "diff-in-diff", it is a special kind of FD model
- Let x_{it} be an indicator of a unit being "treated" at time t.
- Focus on two-periods where:
  - $x_{i1} = 0$ for all i
  - $x_{i2} = 1$ for the "treated group"
- Here is the basic model:

$$y_{it} = \beta_0 + \delta_0 d_t + \beta_1 x_{it} + a_i + u_{it}$$

- d_t is a dummy variable for the second time period
  - $d_2 = 1$ and $d_1 = 0$
- β_1 is the quantity of interest: it's the effect of being treated

Diff-in-diff mechanics

- Let's take differences:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

Diff-in-diff mechanics

- Let's take differences:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

- δ_0: the difference in the average outcome from period 1 to period 2 in the untreated group

Diff-in-diff mechanics

- Let's take differences:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

- δ_0: the difference in the average outcome from period 1 to period 2 in the untreated group
- $(x_{i2} - x_{i1}) = 1$ only for the treated group

Diff-in-diff mechanics

- Let's take differences:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

- δ_0: the difference in the average outcome from period 1 to period 2 in the untreated group
- $(x_{i2} - x_{i1}) = 1$ only for the treated group
- $(x_{i2} - x_{i1}) = 0$ only for the control group

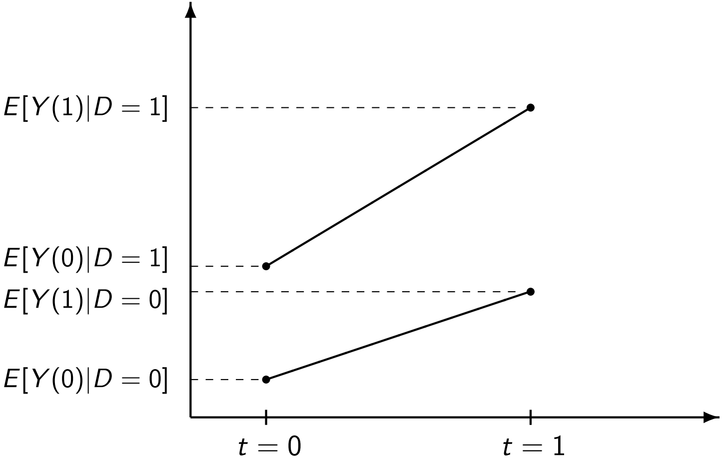
Diff-in-diff mechanics

- Let's take differences:

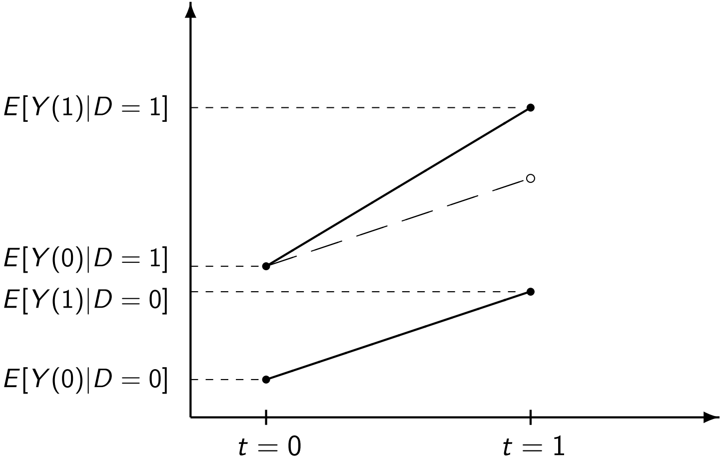
$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

- δ_0: the difference in the average outcome from period 1 to period 2 in the untreated group
- $(x_{i2} - x_{i1}) = 1$ only for the treated group
- $(x_{i2} - x_{i1}) = 0$ only for the control group
- β_1 represents the additional change in y over time (on top of δ_0) associated with being in the treatment group.

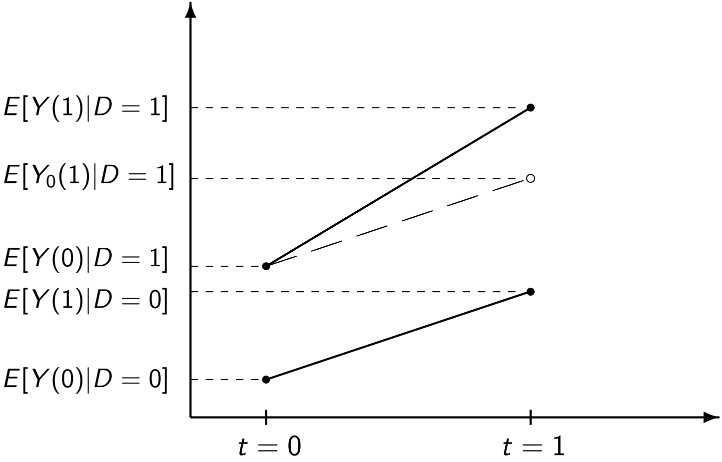
Graphical Representation: Difference-in-Differences



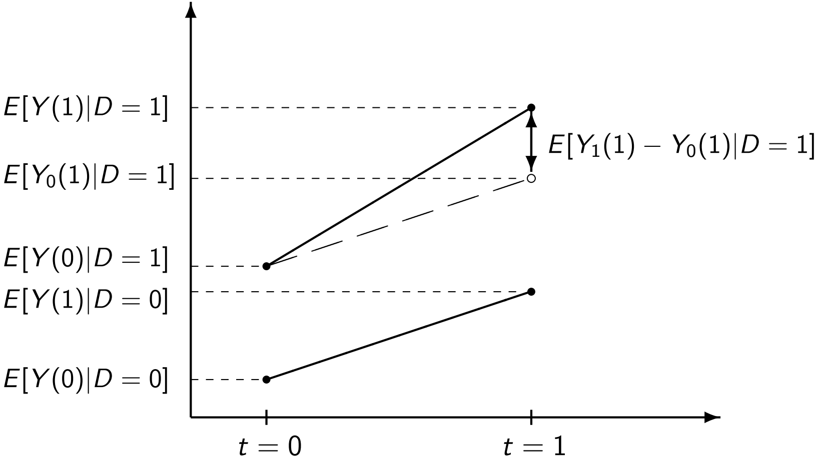
Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



Graphical Representation: Difference-in-Differences



where we define $D = 1$ when $x_{i2} - x_{i1} = 1$ and 0 otherwise

Identification with Difference-in-Differences

Identification Assumption (parallel trends)

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

Identification Result

Given parallel trends the ATT is identified as:

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

Identification with Difference-in-Differences

Identification Assumption (parallel trends)

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

Proof.

Note that the identification assumption implies
$E[Y_0(1)|D = 0] = E[Y_0(1)|D = 1] - E[Y_0(0)|D = 1] + E[Y_0(0)|D = 0]$
plugging in we get

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

□

Diff-in-diff interpretation

- Key idea: comparing the changes over time in the control group to the changes over time in the treated group.

Diff-in-diff interpretation

- Key idea: comparing the changes over time in the control group to the changes over time in the treated group.
- The differences between these differences is our estimate of the causal effect:

$$\beta_1 = \overline{\Delta y}_{\text{treated}} - \overline{\Delta y}_{\text{control}}$$

Diff-in-diff interpretation

- Key idea: comparing the changes over time in the control group to the changes over time in the treated group.
- The differences between these differences is our estimate of the causal effect:

$$\beta_1 = \overline{\Delta y}_{\text{treated}} - \overline{\Delta y}_{\text{control}}$$

- Why more credible than simply looking at the treatment/control differences in period 2?

Diff-in-diff interpretation

- Key idea: comparing the changes over time in the control group to the changes over time in the treated group.
- The differences between these differences is our estimate of the causal effect:

$$\beta_1 = \overline{\Delta y}_{\text{treated}} - \overline{\Delta y}_{\text{control}}$$

- Why more credible than simply looking at the treatment/control differences in period 2?
- Unmeasured reasons why the treated group has higher or lower outcomes than the control group

Diff-in-diff interpretation

- Key idea: comparing the changes over time in the control group to the changes over time in the treated group.
- The differences between these differences is our estimate of the causal effect:

$$\beta_1 = \overline{\Delta y}_{\text{treated}} - \overline{\Delta y}_{\text{control}}$$

- Why more credible than simply looking at the treatment/control differences in period 2?
- Unmeasured reasons why the treated group has higher or lower outcomes than the control group
- $\rightsquigarrow$ bias due to violation of zero conditional mean error

Example: Lyall (2009)

Journal of Conflict Resolution
Volume 53 Number 3
June 2009 331-362

10.1177/0022002708330881
http://jcr.sagepub.com
hosted at
http://online.sagepub.com

Does Indiscriminate Violence Incite Insurgent Attacks?

Evidence from Chechnya

Jason Lyall
Department of Politics and the Woodrow Wilson School
Princeton University, New Jersey

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest
- That is, part of the village fixed effect, a_i might be correlated with whether or not shelling occurs, x_{it}

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest
- That is, part of the village fixed effect, a_i might be correlated with whether or not shelling occurs, x_{it}
- This would cause our pooled estimates to be biased

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest
- That is, part of the village fixed effect, a_i might be correlated with whether or not shelling occurs, x_{it}
- This would cause our pooled estimates to be biased
- Instead Lyall takes a diff-in-diff approach: compare attacks over time for shelled and non-shelled villages:

$$\Delta \text{attacks}_i = \beta_0 + \beta_1 \Delta \text{shelling}_i + \Delta u_i$$

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest
- That is, part of the village fixed effect, a_i might be correlated with whether or not shelling occurs, x_{it}
- This would cause our pooled estimates to be biased
- Instead Lyall takes a diff-in-diff approach: compare attacks over time for shelled and non-shelled villages:

$$\Delta\text{attacks}_i = \beta_0 + \beta_1 \Delta\text{shelling}_i + \Delta u_i$$

Example: Lyall (2009)

- Does Russian shelling of villages cause insurgent attacks?

$$\text{attacks}_{it} = \beta_0 + \beta_1 \text{shelling}_{it} + a_i + u_{it}$$

- We might think that artillery shelling by Russians is targeted to places where the insurgency is the strongest
- That is, part of the village fixed effect, a_i might be correlated with whether or not shelling occurs, x_{it}
- This would cause our pooled estimates to be biased
- Instead Lyall takes a diff-in-diff approach: compare attacks over time for shelled and non-shelled villages:

$$\Delta \text{attacks}_i = \beta_0 + \beta_1 \Delta \text{shelling}_i + \Delta u_i$$

- Counterintuitive findings: shelled villages experience a 24% reduction in insurgent attacks relative to controls.

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage
- Diff-in-diff approach: regress changes in employment on store being in NJ

$$\Delta \text{employment}_i = \beta_0 + \beta_1 NJ_i + \Delta u_i$$

Example: Card and Krueger (2000)

- Do increases to the minimum wage depress employment at fast-food restaurants?

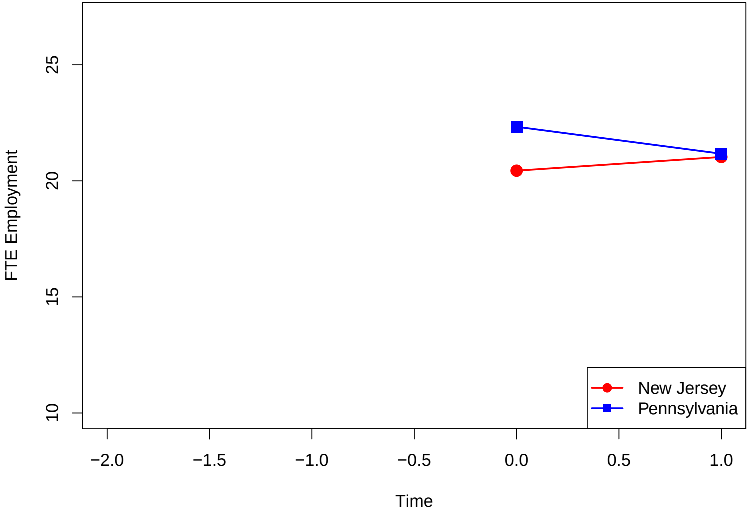
$$\text{employment}_{it} = \beta_0 + \beta_1 \text{minimum wage}_{it} + a_i + u_{it}$$

- Each i here is a different fast food restaurant in either New Jersey or Pennsylvania
- Between $t = 1$ and $t = 2$ NJ raised its minimum wage
- Employment in fast food might be driven by other state-level policies correlated with minimum wage
- Diff-in-diff approach: regress changes in employment on store being in NJ

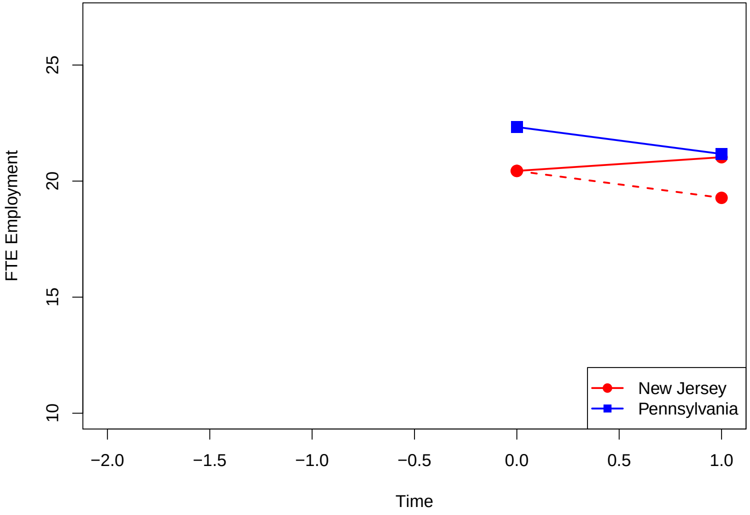
$$\Delta \text{employment}_i = \beta_0 + \beta_1 NJ_i + \Delta u_i$$

- NJ_i indicates which stores received the treatment of a higher minimum wage at time period $t = 2$

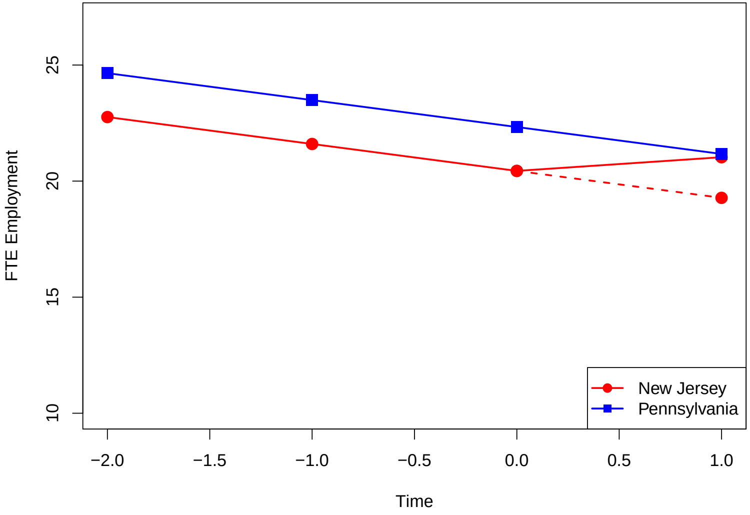
Parallel Trends?



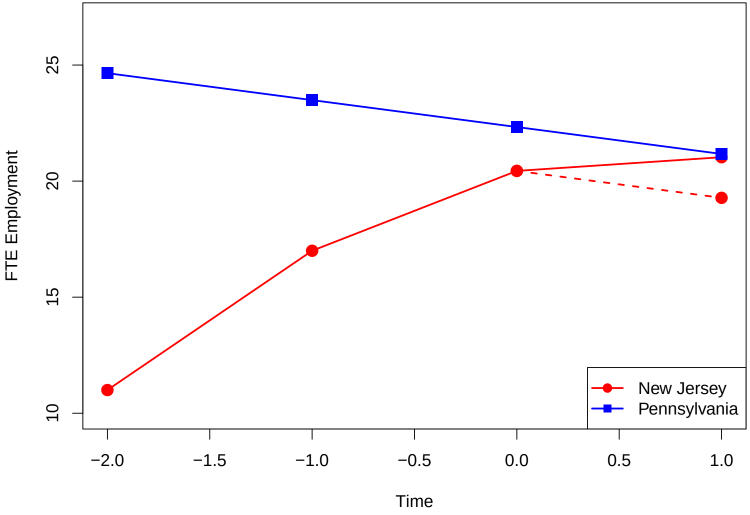
Parallel Trends?



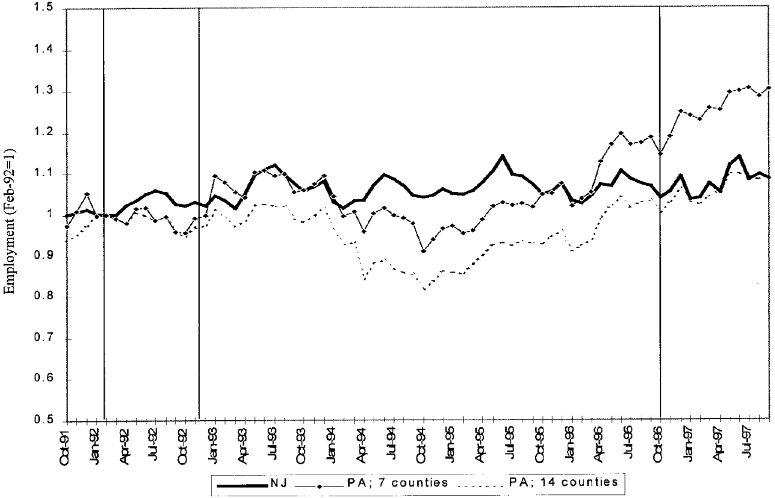
Parallel Trends?



Parallel Trends?



Longer Trends in Employment (Card and Krueger 2000)



First two vertical lines indicate the dates of the Card-Krueger survey. October 1996 line is the federal minimum wage hike which was binding in PA but not NJ

Threats to identification

- Treatment needs to be independent of the idiosyncratic shocks:

$$\mathbb{E}[(u_{i2} - u_{i1})|x_{i2}] = 0$$

Threats to identification

- Treatment needs to be independent of the idiosyncratic shocks:

$$\mathbb{E}[(u_{i2} - u_{i1})|x_{i2}] = 0$$

- Variation in the outcome over time is the same for the treated and control groups

Threats to identification

- Treatment needs to be independent of the idiosyncratic shocks:

$$\mathbb{E}[(u_{i2} - u_{i1})|x_{i2}] = 0$$

- Variation in the outcome over time is the same for the treated and control groups
- Non-parallel dynamics such as Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)

Threats to identification

- Treatment needs to be independent of the idiosyncratic shocks:

$$\mathbb{E}[(u_{i2} - u_{i1})|x_{i2}] = 0$$

- Variation in the outcome over time is the same for the treated and control groups
- Non-parallel dynamics such as Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)
- In the Lyall paper, it might be the case that insurgent attacks might be falling in places where there is shelling because rebels attacked in those areas and have moved on.

Threats to identification

- Treatment needs to be independent of the idiosyncratic shocks:

$$\mathbb{E}[(u_{i2} - u_{i1})|x_{i2}] = 0$$

- Variation in the outcome over time is the same for the treated and control groups
- Non-parallel dynamics such as Ashenfelter's dip: people who enroll in job training programs see their earnings decline prior to that training (presumably why they are entering)
- In the Lyall paper, it might be the case that insurgent attacks might be falling in places where there is shelling because rebels attacked in those areas and have moved on.
- Could add covariates, sometimes called "regression diff-in-diff"

$$y_{i2} - y_{i1} = \delta_0 + \mathbf{z}_i'\tau + \beta(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

Concluding Thoughts on Panel Differencing Models

Concluding Thoughts on Panel Differencing Models

- Useful toolkit for leveraging panel data

Concluding Thoughts on Panel Differencing Models

- Useful toolkit for leveraging panel data
- Be cautious of assumptions required

Concluding Thoughts on Panel Differencing Models

- Useful toolkit for leveraging panel data
- Be cautious of assumptions required
- Always think through "what is the counterfactual" or "what variation lets me identify this effect"
- Parallel trends assumptions are most likely to hold over a shorter time-window. Methods primarily helpful for short, one-shot style effects
- On Wednesday we will discuss a diff-in-diff approach where we don't have a good counterfactual unit.

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right]$$

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right]$$

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$
\begin{aligned}
\overline{y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}'_{it}\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T} a_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}
\end{aligned}
$$

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$
\begin{aligned}
\overline{y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}'_{it}\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T} a_i + \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
&= \overline{\mathbf{x}}'_i\boldsymbol{\beta} + a_i + \overline{u}_i
\end{aligned}
$$

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}
\overline{y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}'_{it}\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}a_i + \frac{1}{T}\sum_{t=1}^{T}u_{it} \\
&= \overline{\mathbf{x}}'_i\boldsymbol{\beta} + a_i + \overline{u}_i
\end{aligned}$$

- Key fact: mean of the time-constant a_i is just a_i

Fixed effects models

- Fixed effects model: alternative way to remove unmeasured heterogeneity
- Focuses on within-unit comparisons: changes in y_{it} and x_{it} relative to their within-group means
- First note that taking the average of the y's over time for a given unit leaves us with a very similar model:

$$
\begin{aligned}
\overline{y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}'_{it}\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}a_i + \frac{1}{T}\sum_{t=1}^{T}u_{it} \\
&= \overline{\mathbf{x}}'_i\boldsymbol{\beta} + a_i + \overline{u}_i
\end{aligned}
$$

- Key fact: mean of the time-constant a_i is just a_i
- This regression is sometimes called the "between regression"

Within transformation

- The "fixed effects," "within," or "time-demeaning" transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \overline{y}_i) = (\mathbf{x}'_{it} - \overline{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

Within transformation

- The "fixed effects," "within," or "time-demeaning" transformation is when we subtract off the over-time means from the original data:

$$(y_{it} - \overline{y}_i) = (\mathbf{x}'_{it} - \overline{\mathbf{x}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

- If we write $\ddot{y}_{it} = y_{it} - \overline{y}_i$, then we can write this more compactly as:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

Fixed effects with Ross data

```
fe.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur), data = ross, index = c("id", "year"),
 model = "within")
summary(fe.mod)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log(kidmort_unicef) ~ democracy + log(GDPcur),
##     data = ross, model = "within", index = c("id", "year"))
##
## Unbalanced Panel: n=166, T=1-7, N=649
##
## Residuals :
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -0.70500 -0.11700  0.00628  0.12200  0.75700
##
## Coefficients :
##              Estimate Std. Error  t-value  Pr(>|t|)
## democracy   -0.143233   0.033500  -4.2756 2.299e-05 ***
## log(GDPcur) -0.375203   0.011328 -33.1226 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    81.711
## Residual Sum of Squares: 23.012
## R-Squared      :  0.71838
##       Adj. R-Squared :  0.53242
## F-statistic: 613.481 on 2 and 481 DF, p-value: < 2.22e-16
```

Strict exogeneity

- FE models are valid if $\mathbb{E}[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$\mathbb{E}[\ddot{u}_{it}|\ddot{\mathbf{X}}] = \mathbb{E}[u_{it}|\ddot{\mathbf{X}}] - \mathbb{E}[\overline{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

Strict exogeneity

- FE models are valid if $\mathbb{E}[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$\mathbb{E}[\ddot{u}_{it}|\ddot{\mathbf{X}}] = \mathbb{E}[u_{it}|\ddot{\mathbf{X}}] - \mathbb{E}[\overline{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

- This is because the composite errors, $\ddot{u}_{it}$ are function of the errors in every time period through the average, $\overline{u}_i$

Strict exogeneity

- FE models are valid if $\mathbb{E}[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$\mathbb{E}[\ddot{u}_{it}|\ddot{\mathbf{X}}] = \mathbb{E}[u_{it}|\ddot{\mathbf{X}}] - \mathbb{E}[\overline{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

- This is because the composite errors, $\ddot{u}_{it}$ are function of the errors in every time period through the average, $\overline{u}_i$
- This rules out, for instance, lagged dependent variables, since $y_{i,t-1}$ has to be correlated with $u_{i,t-1}$. Thus it can't be a covariate for y_{it}.

Strict exogeneity

- FE models are valid if $\mathbb{E}[\mathbf{u}|\mathbf{X}] = 0$: all errors are uncorrelated with covariates in every period:

$$\mathbb{E}[\ddot{u}_{it}|\ddot{\mathbf{X}}] = \mathbb{E}[u_{it}|\ddot{\mathbf{X}}] - \mathbb{E}[\overline{u}_i|\ddot{\mathbf{X}}] = 0 - 0 = 0$$

- This is because the composite errors, $\ddot{u}_{it}$ are function of the errors in every time period through the average, $\overline{u}_i$
- This rules out, for instance, lagged dependent variables, since $y_{i,t-1}$ has to be correlated with $u_{i,t-1}$. Thus it can't be a covariate for y_{it}.
- Degrees of freedom: $nT - n - k - 1$, which accounts for within transformation

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \overline{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \overline{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t.
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept violate full rank. R/Stata and the like will drop it from the regression.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \overline{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t.
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept violate full rank. R/Stata and the like will drop it from the regression.
- Basic message: any time-constant variable gets "absorbed" by the fixed effect

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
- Then $x_{it} = \overline{x}_i$ and $\ddot{x}_{it} = 0$ for all periods t.
- If the time-demeaned covariate is always 0, then it will be perfectly collinear with the intercept violate full rank. R/Stata and the like will drop it from the regression.
- Basic message: any time-constant variable gets "absorbed" by the fixed effect
- Can include interactions between time-constant and time-varying variables, but lower order term of the time-constant variables get absorbed by fixed effects too

Time-constant variables

- Pooled model with a time-constant variable, proportion Islamic:

```
library(lmtest)
p.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam,
             data = ross, index = c("id", "year"), model = "pooling")
coeftest(p.mod)
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 10.30607817  0.35951939  28.6663 < 2.2e-16 ***
## democracy   -0.80233845  0.07766814 -10.3303 < 2.2e-16 ***
## log(GDPcur) -0.25497406  0.01607061 -15.8659 < 2.2e-16 ***
## islam        0.00343325  0.00091045   3.7709 0.0001794 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time-constant variables

- FE model, where the islam variable drops out, along with the intercept:

```
fe.mod2 <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam,
               data = ross, index = c("id", "year"), model = "within")
coeftest(fe.mod2)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error  t value  Pr(>|t|)
## democracy   -0.129693   0.035865  -3.6162 0.0003332 ***
## log(GDPcur) -0.379997   0.011849 -32.0707 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix: Relating to PO Model Setup

- Units $i = 1, \dots, N$

Appendix: Relating to PO Model Setup

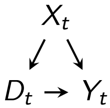
- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,

Appendix: Relating to PO Model Setup

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, D_{it} are the outcome and treatment for unit i in period t We have a set of covariates in each period, as well,

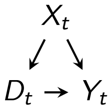
Appendix: Relating to PO Model Setup

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, D_{it} are the outcome and treatment for unit i in period t We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to D_{it}.



Appendix: Relating to PO Model Setup

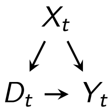
- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, D_{it} are the outcome and treatment for unit i in period t We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to D_{it}.



- U_i = unobserved, time-invariant unit effects (causally prior to everything)

Appendix: Relating to PO Model Setup

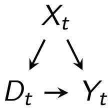
- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, D_{it} are the outcome and treatment for unit i in period t We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to D_{it}.



- U_i = unobserved, time-invariant unit effects (causally prior to everything)
- History of some variable: $\underline{D}_{it} = (D_1, \ldots, D_t)$.

Appendix: Relating to PO Model Setup

- Units $i = 1, \ldots, N$
- Time periods $t = 1, \ldots, T$ with $T \geq 2$,
- Y_{it}, D_{it} are the outcome and treatment for unit i in period t We have a set of covariates in each period, as well,
- Covariates X_{it}, causally "prior" to D_{it}.



- U_i = unobserved, time-invariant unit effects (causally prior to everything)
- History of some variable: $\underline{D}_{it} = (D_1, \ldots, D_t)$.
- Entire history: $\underline{D}_i = \underline{D}_{iT}$

Appendix: Relating to PO Model Assumptions

- **Potential outcomes**: $Y_{it}(1) = Y_{it}(d_t = 1)$ is the potential outcome for unit i at time t if they were treated at time t.

Appendix: Relating to PO Model Assumptions

- **Potential outcomes**: $Y_{it}(1) = Y_{it}(d_t = 1)$ is the potential outcome for unit i at time t if they were treated at time t.
  - Here we focus on contemporaneous effects, $Y_{it}(d_t = 1) - Y_{it}(d_t = 0)$

Appendix: Relating to PO Model Assumptions

- **Potential outcomes**: $Y_{it}(1) = Y_{it}(d_t = 1)$ is the potential outcome for unit i at time t if they were treated at time t.
  - Here we focus on contemporaneous effects, $Y_{it}(d_t = 1) - Y_{it}(d_t = 0)$
  - Harder when including lags of treatment, $Y_{it}(d_t = 1, d_{t-1} = 1)$

Appendix: Relating to PO Model Assumptions

- **Potential outcomes**: $Y_{it}(1) = Y_{it}(d_t = 1)$ is the potential outcome for unit i at time t if they were treated at time t.
  - Here we focus on contemporaneous effects, $Y_{it}(d_t = 1) - Y_{it}(d_t = 0)$
  - Harder when including lags of treatment, $Y_{it}(d_t = 1, d_{t-1} = 1)$
- **Consistency** for each time period:

$$Y_{it} = Y_{it}(1)D_{it} + Y_{it}(0)(1 - D_{it})$$

Appendix: Relating to PO Model Assumptions

- **Potential outcomes**: $Y_{it}(1) = Y_{it}(d_t = 1)$ is the potential outcome for unit i at time t if they were treated at time t.
    - Here we focus on contemporaneous effects, $Y_{it}(d_t = 1) - Y_{it}(d_t = 0)$
    - Harder when including lags of treatment, $Y_{it}(d_t = 1, d_{t-1} = 1)$
- **Consistency** for each time period:

$$Y_{it} = Y_{it}(1)D_{it} + Y_{it}(0)(1 - D_{it})$$

- **Strict ignorability**: potential outcomes are independent of the entire history of treatment conditional on the history of covariates and the time-constant heterogeneity:

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

Appendix: Relating to PO Model

- Assume that the CEF for the mean potential outcome under control is:

$$\mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i] = X_{it}'\beta + U_i$$

Appendix: Relating to PO Model

- Assume that the CEF for the mean potential outcome under control is:

$$\mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i] = X_{it}'\beta + U_i$$

- And then assume a constant treatment effects:

$$\mathbb{E}[Y_{it}(1)|\underline{X}_i, U_i] = \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i] + \tau$$

Appendix: Relating to PO Model

- Assume that the CEF for the mean potential outcome under control is:

$$\mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i] = X_{it}'\beta + U_i$$

- And then assume a constant treatment effects:

$$\mathbb{E}[Y_{it}(1)|\underline{X}_i, U_i] = \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i] + \tau$$

- With consistency and strict ignorability, we can write this as a CEF of the observed outcome:

$$\mathbb{E}[Y_{it}|\underline{X}_i, \underline{D}_i, U_i] = X_{it}'\beta + \tau D_{it} + U_i$$

Appendix: Relating to PO Model

- We can now write the observed outcomes in a traditional regression format:

$$Y_{it} = X_{it}'\beta + \tau D_{it} + U_i + \varepsilon_{it}$$

Appendix: Relating to PO Model

- We can now write the observed outcomes in a traditional regression format:

$$Y_{it} = X_{it}'\beta + \tau D_{it} + U_i + \varepsilon_{it}$$

- Here, the error is similar to what we had for regression:

$$\varepsilon_{it} \equiv Y_{it}(0) - \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i]$$

Appendix: Relating to PO Model

- We can now write the observed outcomes in a traditional regression format:

$$Y_{it} = X_{it}'\beta + \tau D_{it} + U_i + \varepsilon_{it}$$

- Here, the error is similar to what we had for regression:

$$\varepsilon_{it} \equiv Y_{it}(0) - \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i]$$

- In traditional FE models, we skip potential outcomes and rely on a **strict exogeneity** assumption:

$$\mathbb{E}[\varepsilon_{it}|\underline{X}_i, \underline{D}_i, U_i] = 0$$

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

$$\mathbb{E}[\varepsilon_{it} | \underline{X}_i, \underline{D}_i, U_i] = \mathbb{E}\left[\left(Y_{it}(0) - \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i]\right) | \underline{X}_i, \underline{D}_i, U_i\right]$$

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

$$\begin{aligned}
\mathbb{E}[\varepsilon_{it}|\underline{X}_i, \underline{D}_i, U_i] &= \mathbb{E}\left[\left(Y_{it}(0) - \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i]\right)|\underline{X}_i, \underline{D}_i, U_i\right] \\
&= \mathbb{E}[Y_{it}(0)|\underline{X}_i, \underline{D}_i, U_i] - \mathbb{E}[Y_{it}(0)|\underline{X}_i, U_i]
\end{aligned}$$

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

$$\begin{aligned}
\mathbb{E}[\varepsilon_{it} | \underline{X}_i, \underline{D}_i, U_i] &= \mathbb{E}\left[\left(Y_{it}(0) - \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] \right) | \underline{X}_i, \underline{D}_i, U_i \right] \\
&= \mathbb{E}[Y_{it}(0) | \underline{X}_i, \underline{D}_i, U_i] - \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] \\
&= \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] - \mathbb{E}[Y_{it}(0) | \underline{X}_{iT}, U_i]
\end{aligned}$$

Appendix: Relating to PO Model: Strict ignorability vs strict exogeneity

$$Y_{it}(d) \perp\!\!\!\perp \underline{D}_i | \underline{X}_i, U_i$$

- Easy to show to that strict ignorability implies strict exogeneity:

$$\begin{aligned}
\mathbb{E}[\varepsilon_{it} | \underline{X}_i, \underline{D}_i, U_i] &= \mathbb{E}\left[\left(Y_{it}(0) - \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] \right) | \underline{X}_i, \underline{D}_i, U_i \right] \\
&= \mathbb{E}[Y_{it}(0) | \underline{X}_i, \underline{D}_i, U_i] - \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] \\
&= \mathbb{E}[Y_{it}(0) | \underline{X}_i, U_i] - \mathbb{E}[Y_{it}(0) | \underline{X}_{iT}, U_i] \\
&= 0
\end{aligned}$$

Least squares dummy variable

- As an alternative to the within transformation, we can also include a series of $n-1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + d1_i\alpha_1 + d2_i\alpha_2 + \cdots + dn_i\alpha_n + u_{it}$$

Least squares dummy variable

- As an alternative to the within transformation, we can also include a series of $n-1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + d1_i\alpha_1 + d2_i\alpha_2 + \cdots + dn_i\alpha_n + u_{it}$$

- Here, $d1_i$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.

Least squares dummy variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + d1_i\alpha_1 + d2_i\alpha_2 + \cdots + dn_i\alpha_n + u_{it}$$

- Here, $d1_i$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the exact same estimates/standard errors as with time-demeaning

Least squares dummy variable

- As an alternative to the within transformation, we can also include a series of $n - 1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + d1_i\alpha_1 + d2_i\alpha_2 + \cdots + dn_i\alpha_n + u_{it}$$

- Here, $d1_i$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the exact same estimates/standard errors as with time-demeaning
- Advantage: easy to implement in R

Least squares dummy variable

- As an alternative to the within transformation, we can also include a series of $n-1$ dummy variables for each unit:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + d1_i\alpha_1 + d2_i\alpha_2 + \cdots + dn_i\alpha_n + u_{it}$$

- Here, $d1_i$ is a binary variable which is 1 if $i = 1$ and 0 otherwise—just a unit dummy.
- Gives the exact same estimates/standard errors as with time-demeaning
- Advantage: easy to implement in R
- Disadvantage: computationally difficult with large N, since we have to run a regression with $n+k$ variables.

Example with Ross data

```
library(lmtest)
lsdv.mod <- lm(log(kidmort_unicef) ~ democracy + log(GDPcur) +
               as.factor(id), data = ross)
coeftest(lsdv.mod)[1:6,]
coeftest(fe.mod)[1:2,]
```

```
##                    Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)      13.7644887 0.26597312  51.751427 1.008329e-198
## democracy        -0.1432331 0.03349977  -4.275644  2.299393e-05
## log(GDPcur)      -0.3752030 0.01132772 -33.122568 3.494887e-126
## as.factor(id)AGO  0.2997206 0.16767730   1.787485  7.448861e-02
## as.factor(id)ALB -1.9309618 0.19013955 -10.155498  4.392512e-22
## as.factor(id)ARE -1.8762909 0.17020738 -11.023558  2.386557e-25
```

```
##               Estimate Std. Error    t value      Pr(>|t|)
## democracy   -0.1432331 0.03349977  -4.275644  2.299393e-05
## log(GDPcur) -0.3752030 0.01132772 -33.122568 3.494887e-126
```

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation
  - Matched pairs
    - Twin fixed effects to control for unobserved effects of family background

Applying Fixed Effects

- We can use fixed effects for other data structures to restrict comparisons to within unit variation
  - Matched pairs
    - Twin fixed effects to control for unobserved effects of family background
  - Cluster fixed effects in hierarchical data
    - School fixed effects to control for unobserved effects of school

Problems that (even) fixed effects do not solve

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Where y_{it} is murder rate and x_{it} is police spending per capita
- What happens when we regress y on x and city fixed effects?

Problems that (even) fixed effects do not solve

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Where y_{it} is murder rate and x_{it} is police spending per capita
- What happens when we regress y on x and city fixed effects?
  - $\hat{\boldsymbol{\beta}}_{FE}$ inconsistent unless strict exogeneity conditional on c_i holds
    - $E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i] = 0, \; t = 1, 2, ..., T$
    - implies ε_{it} uncorrelated with past, current, and future regressors

Problems that (even) fixed effects do not solve

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Where y_{it} is murder rate and x_{it} is police spending per capita
- What happens when we regress y on x and city fixed effects?
  - $\hat{\boldsymbol{\beta}}_{FE}$ inconsistent unless strict exogeneity conditional on c_i holds
    - $E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i] = 0, \;\; t = 1, 2, ..., T$
    - implies ε_{it} uncorrelated with past, current, and future regressors
- Most common violations:
  1. Time-varying omitted variables
     - economic boom leads to more police spending and less murders
     - can include time-varying controls, but avoid post-treatment bias

Problems that (even) fixed effects do not solve

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Where y_{it} is murder rate and x_{it} is police spending per capita
- What happens when we regress y on x and city fixed effects?
  - $\hat{\boldsymbol{\beta}}_{FE}$ inconsistent unless strict exogeneity conditional on c_i holds
    - $E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i] = 0, \ t = 1, 2, ..., T$
    - implies ε_{it} uncorrelated with past, current, and future regressors
- Most common violations:
  1. Time-varying omitted variables
     - economic boom leads to more police spending and less murders
     - can include time-varying controls, but avoid post-treatment bias
  2. Simultaneity
     - if city adjusts police based on past murder rate, then spending$_{t+1}$ is correlated with ε_t (since higher ε_t leads to higher murder rate at t)
     - strictly exogenous x cannot react to what happens to y in the past or the future!

Problems that (even) fixed effects do not solve

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Where y_{it} is murder rate and x_{it} is police spending per capita
- What happens when we regress y on x and city fixed effects?
  - $\hat{\boldsymbol{\beta}}_{FE}$ inconsistent unless strict exogeneity conditional on c_i holds
    - $E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i] = 0, \ t = 1, 2, ..., T$
    - implies ε_{it} uncorrelated with past, current, and future regressors
- Most common violations:
  1. Time-varying omitted variables
     - economic boom leads to more police spending and less murders
     - can include time-varying controls, but avoid post-treatment bias
  2. Simultaneity
     - if city adjusts police based on past murder rate, then spending$_{t+1}$ is correlated with ε_t (since higher ε_t leads to higher murder rate at t)
     - strictly exogenous x cannot react to what happens to y in the past or the future!
- Fixed effects do not obviate need for good research design!

Fixed effects versus first differences

- Key assumptions:

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates
- So which one is better when $T > 2$? Which one is more efficient?

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates
- So which one is better when $T > 2$? Which one is more efficient?
- u_{it} uncorrelated $\rightsquigarrow$ FE is more efficient

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates
- So which one is better when $T > 2$? Which one is more efficient?
- u_{it} uncorrelated $\rightsquigarrow$ FE is more efficient
- $u_{it} = u_{i,t-1} + e_{it}$ with e_{it} iid (random walk) $\rightsquigarrow$ FD is more efficient.

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates
- So which one is better when $T > 2$? Which one is more efficient?
- u_{it} uncorrelated $\rightsquigarrow$ FE is more efficient
- $u_{it} = u_{i,t-1} + e_{it}$ with e_{it} iid (random walk) $\rightsquigarrow$ FD is more efficient.
- In between, not clear which is better

Fixed effects versus first differences

- Key assumptions:
  - Strict exogeneity: $E[u_{it}|\mathbf{X}, a_i] = 0$
  - Time-constant unmeasured heterogeneity, a_i
- Together $\implies$ fixed effects and first differences are unbiased and consistent
- With $T = 2$ the estimators produce identical estimates
- So which one is better when $T > 2$? Which one is more efficient?
- u_{it} uncorrelated $\rightsquigarrow$ FE is more efficient
- $u_{it} = u_{i,t-1} + e_{it}$ with e_{it} iid (random walk) $\rightsquigarrow$ FD is more efficient.
- In between, not clear which is better
- Large differences between FE and FD should make us worry about assumptions

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

Random effects model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

- Key difference: $E[a_i|\mathbf{X}] = E[a_i] = 0$

Random effects model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

- Key difference: $E[a_i|\mathbf{X}] = E[a_i] = 0$
- We also assume that a_i are iid and independent of the u_{it}

Random effects model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

- Key difference: $E[a_i|\mathbf{X}] = E[a_i] = 0$
- We also assume that a_i are iid and independent of the u_{it}
- Like with clustering, we can treat $\mathbf{v}_{it} = a_i + u_{it}$ as a combined error that satisfies zero conditional mean error:

$$E[a_i + u_{it}|\mathbf{X}] = E[a_i|\mathbf{X}] + E[u_{it}|\mathbf{X}] = 0 + 0 = 0$$

Quasi-demeaned data

- Random effects models usually transform the data via what is called quasi-demeaning or partial pooling:

$$y_{it} - \theta\overline{y}_i = (\mathbf{x}'_{it} - \theta\overline{\mathbf{x}}'_i) + (v_{it} - \theta\overline{v}_i)$$

Quasi-demeaned data

- Random effects models usually transform the data via what is called quasi-demeaning or partial pooling:

$$y_{it} - \theta \overline{y}_i = (\mathbf{x}'_{it} - \theta \overline{\mathbf{x}}'_i) + (v_{it} - \theta \overline{v}_i)$$

- Here θ is between zero and one, where $\theta = 0$ implies pooled OLS and $\theta = 1$ implies fixed effects. Doing some math shows that

$$\theta = 1 - \left[\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)\right]^{1/2}$$

Quasi-demeaned data

- Random effects models usually transform the data via what is called quasi-demeaning or partial pooling:

$$y_{it} - \theta\overline{y}_i = (\mathbf{x}'_{it} - \theta\overline{\mathbf{x}}'_i) + (v_{it} - \theta\overline{v}_i)$$

- Here θ is between zero and one, where $\theta = 0$ implies pooled OLS and $\theta = 1$ implies fixed effects. Doing some math shows that

$$\theta = 1 - \left[\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)\right]^{1/2}$$

- the random effect estimator runs pooled OLS on this model replacing θ with an estimate $\widehat{\theta}$.

Example with Ross data

```
re.mod <- plm(log(kidmort_unicef) ~ democracy + log(GDPcur),
        data = ross, index = c("id", "year"), model = "random")
coeftest(re.mod)[1:3,]
coeftest(fe.mod)[1:2,]
coeftest(pooled.mod)[1:3,]
```

```
##               Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 12.3128677 0.25500821  48.284202 1.610504e-216
## democracy   -0.1917958 0.03395696  -5.648203  2.431253e-08
## log(GDPcur) -0.3609269 0.01100928 -32.783891 1.458769e-139
```

```
##               Estimate Std. Error    t value      Pr(>|t|)
## democracy   -0.1432331 0.03349977  -4.275644  2.299393e-05
## log(GDPcur) -0.3752030 0.01132772 -33.122568 3.494887e-126
```

```
##               Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)  9.7640482 0.34490999  28.30898 2.881836e-115
## democracy   -0.9552482 0.06977944 -13.68954  1.222538e-37
## log(GDPcur) -0.2282798 0.01548068 -14.74611  1.244513e-42
```

- More general random effects models using `lmer()` from the `lme4` package

Fixed effects versus random effects

- Random effects:

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables
  - Corrects for clustering

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables
  - Corrects for clustering
  - Doesn't correct for heteroskedasticity (can use cluster-robust SEs)

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables
  - Corrects for clustering
  - Doesn't correct for heteroskedasticity (can use cluster-robust SEs)
  - x_{it} can be arbitrarily related to a_i

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables
  - Corrects for clustering
  - Doesn't correct for heteroskedasticity (can use cluster-robust SEs)
  - x_{it} can be arbitrarily related to a_i
- Wooldridge: "FE is almost always much more convincing than RE for policy analysis using aggregated data."

Fixed effects versus random effects

- Random effects:
  - Can include time-constant variables
  - Corrects for clustering/heteroskedasticity
  - Requires x_{it} uncorrelated with a_i
- Fixed effects:
  - Can't include time-constant variables
  - Corrects for clustering
  - Doesn't correct for heteroskedasticity (can use cluster-robust SEs)
  - x_{it} can be arbitrarily related to a_i
- Wooldridge: "FE is almost always much more convincing than RE for policy analysis using aggregated data."
- Correlated random effects: allows for some structured dependence between x_{it} and a_i

Fixed and Random Effects

- We are just scratching the surface here.

Fixed and Random Effects

- We are just scratching the surface here.
- Next semester we will cover more complicated hierarchical models

Fixed and Random Effects

- We are just scratching the surface here.
- Next semester we will cover more complicated hierarchical models
- Although often presented as a method for causal inference, fixed effects can make for some counter-intuitive interpretations: see Kim and Imai (2016) on fixed effects for causal inference.

Fixed and Random Effects

- We are just scratching the surface here.
- Next semester we will cover more complicated hierarchical models
- Although often presented as a method for causal inference, fixed effects can make for some counter-intuitive interpretations: see Kim and Imai (2016) on fixed effects for causal inference.
- Particularly when "two-way" fixed effects are used (e.g. time and country fixed effects) it becomes difficult to tell what the counterfactual is.

Fixed and Random Effects

- We are just scratching the surface here.
- Next semester we will cover more complicated hierarchical models
- Although often presented as a method for causal inference, fixed effects can make for some counter-intuitive interpretations: see Kim and Imai (2016) on fixed effects for causal inference.
- Particularly when "two-way" fixed effects are used (e.g. time and country fixed effects) it becomes difficult to tell what the counterfactual is.
- We have essentially not talked at all about temporal dynamics which is another important area for research with non-short time intervals.

Next Class

Send me questions or write them on cards!

Where We've Been and Where We're Going...

- Last Week
  - causal inference with unmeasured confounding
- This Week
  - Monday:
    - panel data
    - diff-in-diff
    - fixed effects
  - Wednesday:
    - Q&A
    - fun With
    - wrap-Up
- The Following Week
  - break!
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causality

Questions?

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

Q: What conditions do we need to infer causality?

Q: What conditions do we need to infer causality?

A: An identification strategy and an estimation strategy.

Identification Strategies in This Class

Identification Strategies in This Class

- Experiments (randomization)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)
- Instrumental Variables (instrument + exclusion restriction)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences (parallel trends)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences (parallel trends)
- Fixed Effects (time-invariant unobserved heterogeneity, strict ignorability)

Identification Strategies in This Class

- Experiments (randomization)
- Selection on Observables (conditional ignorability)
- Natural Experiments (quasi-randomization)
- Instrumental Variables (instrument + exclusion restriction)
- Regression Discontinuity (continuity assumption)
- Difference-in-Differences (parallel trends)
- Fixed Effects (time-invariant unobserved heterogeneity, strict ignorability)

Essentially everything assumes: consistency/SUTVA (essentially: no interference between units, variation in the treatment is irrelevant).

Some Estimation Strategies

Some Estimation Strategies

- Regression (and relatives)

Some Estimation Strategies

- Regression (and relatives)
- Stratification

Some Estimation Strategies

- Regression (and relatives)
- Stratification
- Matching (next semester)
- Weighting (next semester)

Q: Why is heteroskedasticity a problem?

Q: Why is heteroskedasticity a problem?

A: It keeps us from getting easy standard errors. Sometimes it can cause poor finite sample estimator performance.

Derivation of Variance under Homoskedasticity

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

$$\begin{aligned}V[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= V[\boldsymbol{\beta}|\mathbf{X}] + V[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] \\ &= V[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V[\mathbf{u}|\mathbf{X}]((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \text{ (note: } \mathbf{X} \text{ nonrandom } |\mathbf{X}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V[\mathbf{u}|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \text{ (by homoskedasticity)} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Replacing σ^2 with our estimator $\widehat{\sigma}^2$ gives us our estimator for the $(k+1)\times(k+1)$ variance-covariance matrix for the vector of regression coefficients:

$$\widehat{V[\hat{\boldsymbol{\beta}}|\mathbf{X}]} = \widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Q: Power Analysis?

Q: Power Analysis?

A: Useful for planning experiments and for assessing plausibility of seeing an effect after the fact (retrospective power analysis). Relies on knowledge of some things we don't know.

Q: "If we use fixed effects, aren't we explaining away the thing we care about?"

Q: "If we use fixed effects, aren't we explaining away the thing we care about?"

A: We might be worried about this a little bit. In the causal inference setting we get one thing of interest: the treatment effect estimate. All the coefficients on our confounding variables are uninterpretable (at least as causal estimates). From this perspective fixed effects are just capturing all that background. That said- strong assumptions need to hold to not wash away something of interest.

Q: "t-value, test statistics, compare with standard error"

Q: "*t*-value, test statistics, compare with standard error"

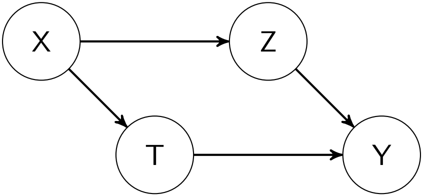
A: The first two relate to hypothesis testing. A t-value is a type of test statistic ($\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ or $\frac{\hat{\beta}-c}{\hat{SE}[\hat{\beta}]}$ depending on context). A test statistic is a function of the sample and the null hypothesis value of the parameter. The standard error is a more general quantity that is the standard deviation of the sampling distribution of the estimator.

Q: What is M-bias? Also could you review mechanics of DAGs, how to follow paths, how to block paths.

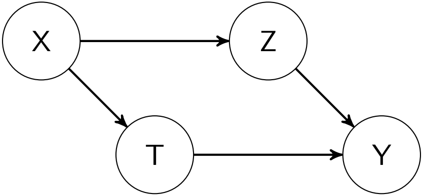
Q: What is M-bias? Also could you review mechanics of DAGs, how to follow paths, how to block paths.

A: Sure

From Confounders to Back-Door Paths

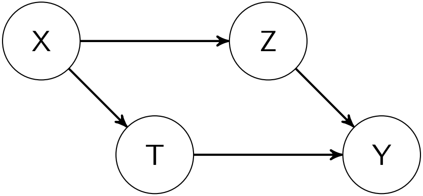


From Confounders to Back-Door Paths



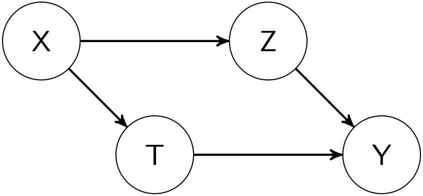
- Identify causal effect of T on Y by conditioning on X, Z or X and Z

From Confounders to Back-Door Paths



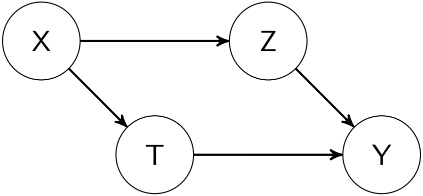
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path

From Confounders to Back-Door Paths



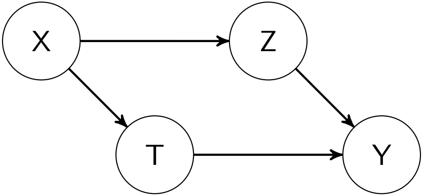
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)

From Confounders to Back-Door Paths



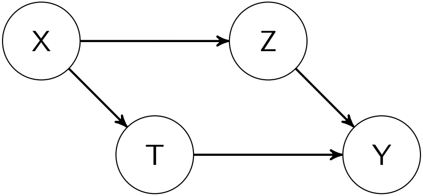
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
  1. $T \rightarrow Y$ (directed or causal path)
  2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)

From Confounders to Back-Door Paths



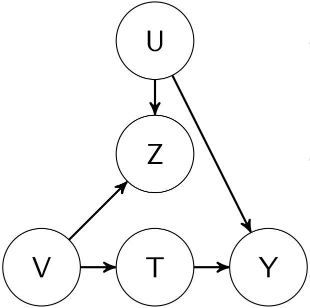
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
  1. $T \rightarrow Y$ (directed or causal path)
  2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)
- Observed marginal association between T and Y is a composite of these two paths and thus does not identify the causal effect of T on Y

From Confounders to Back-Door Paths



- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
  1. $T \rightarrow Y$ (directed or causal path)
  2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)
- Observed marginal association between T and Y is a composite of these two paths and thus does not identify the causal effect of T on Y
- We want to block the back-door path to leave only the causal effect

Colliders and Back-Door Paths



- Z is a collider and it lies along a back-door path from T to Y
- Conditioning on a collider on a back-door path does not help and in fact causes new associations
- Here we are fine unless we condition on Z which opens a path $T \leftarrow V \leftrightarrow U \rightarrow Y$ (this particular case is called M-bias)
- So how do we know which back-door paths to block?

D-Separation

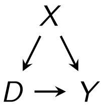
- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is D-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B
- A path p is said to be blocked by a set of vertices C if and only if at least one of the following conditions holds:
  1. p contains a chain structure $a \to c \to b$ or a fork structure $a \leftarrow c \to b$ where the node c is in the set C
  2. p contains a collider structure $a \to y \leftarrow b$ where neither y nor its descendents are in C
- If A is not D-separated from B by C we say that A is D-connected to B by C

Backdoor Criterion

- Backdoor Criterion for X
  1. No node in X is a descendent of T
     (i.e. don't condition on post-treatment variables!)
  2. X D-separates every path between T and Y that has an incoming arrow into T (backdoor path)
- In essence, we are trying to block all non-causal paths, so we can estimate the causal path.

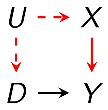
Backdoor paths and blocking paths

- Backdoor path: is a non-causal path from D to Y.
  - Would remain if we removed any arrows pointing out of D.
- Backdoor paths between D and Y $\rightsquigarrow$ common causes of D and Y:



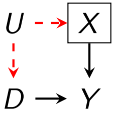
- Here there is a backdoor path $D \leftarrow X \rightarrow Y$, where X is a common cause for the treatment and the outcome.

Other types of confounding



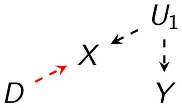
- D is enrolling in a job training program.
- Y is getting a job.
- U is being motivated
- X is number of job applications sent out.
- Big assumption here: no arrow from U to Y.

What's the problem with backdoor paths?



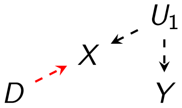
- A path is blocked if:
  1. we control for or stratify a non-collider on that path OR
  2. we do not control for a collider.
- Unblocked backdoor paths $\rightsquigarrow$ confounding.
- In the DAG here, if we condition on X, then the backdoor path is blocked.

Not all backdoor paths



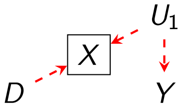
- Conditioning on the posttreatment covariates opens the non-causal path.

Not all backdoor paths



- Conditioning on the posttreatment covariates opens the non-causal path.
  - $\rightsquigarrow$ selection bias.

Not all backdoor paths



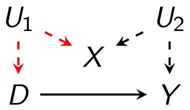
- Conditioning on the posttreatment covariates opens the non-causal path.
  - ⇝ selection bias.

Don't condition on post-treatment variables

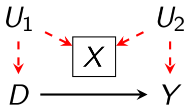
Don't condition on post-treatment variables

Every time you do, a puppy cries.

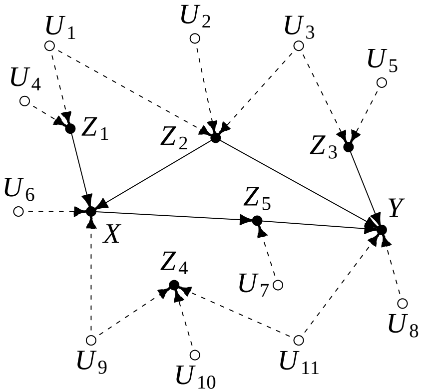
M-bias



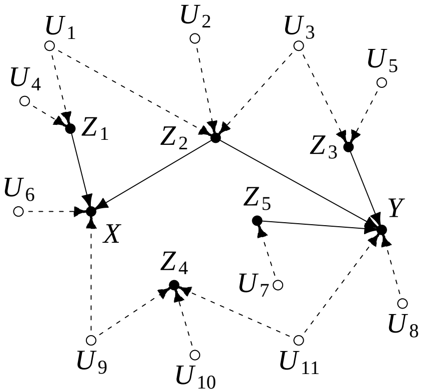
M-bias



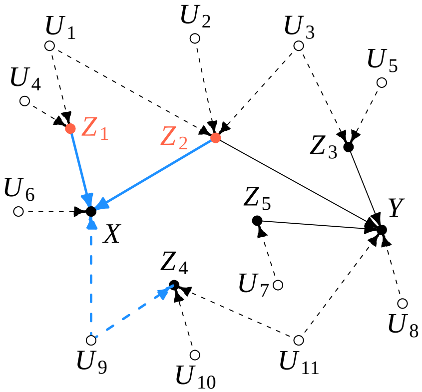
Examples



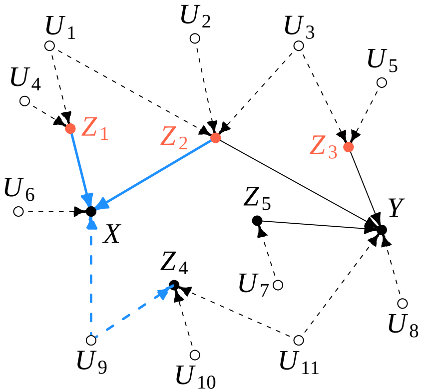
Examples



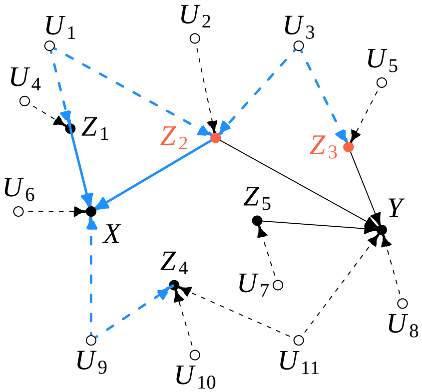
Examples



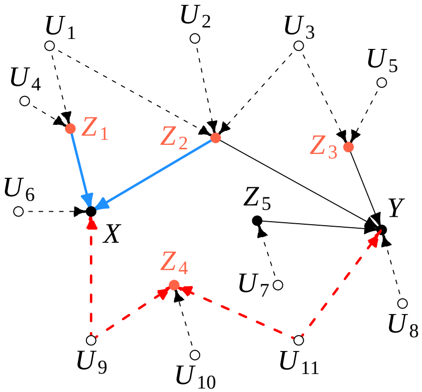
Examples



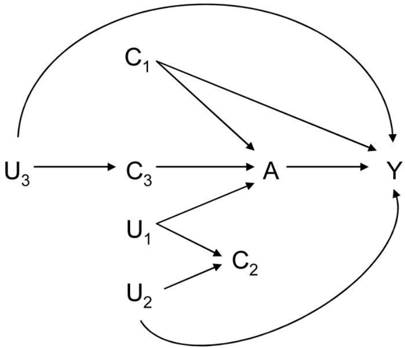
Examples



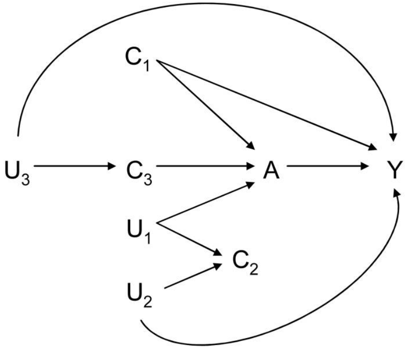
Examples



Implications (via Vanderweele and Shpitser 2011)



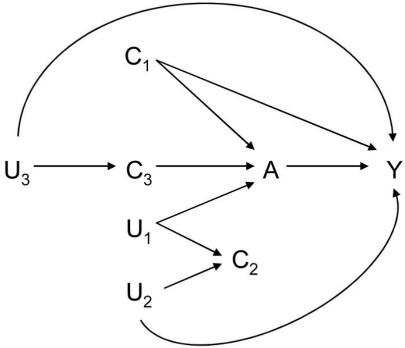
Implications (via Vanderweele and Shpitser 2011)



Two common criteria fail here:

1. Choose all pre-treatment covariates
2. Choose all covariates which directly cause the treatment and the outcome

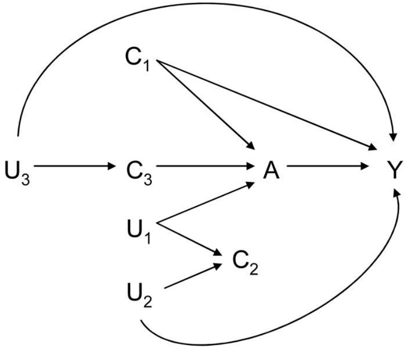
Implications (via Vanderweele and Shpitser 2011)



Two common criteria fail here:

1. Choose all pre-treatment covariates (would condition on C_2 inducing M-bias)
2. Choose all covariates which directly cause the treatment and the outcome

Implications (via Vanderweele and Shpitser 2011)



Two common criteria fail here:

1. Choose all pre-treatment covariates
   (would condition on C_2 inducing M-bias)
2. Choose all covariates which directly cause the treatment and the outcome
   (would leave open a backdoor path $A \leftarrow C_3 \leftarrow U_3 \rightarrow Y$.)

How often are observational studies used for causal inference?

How often are observational studies used for causal inference?

All the time (except maybe psychology)

Can we hear more about your research?

Can we hear more about your research?

Sure.

What are your favorite resources for learning tricky concepts?

What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

1. Identify approx. the best textbook (often can do this via syllabi hunting)

What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

1. Identify approx. the best textbook (often can do this via syllabi hunting)
2. Read the relevant textbook material

What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

1. Identify approx. the best textbook (often can do this via syllabi hunting)
2. Read the relevant textbook material
3. Derive the equations/math

What are your favorite resources for learning tricky concepts?

I've used the following procedure many times:

1. Identify approx. the best textbook (often can do this via syllabi hunting)
2. Read the relevant textbook material
3. Derive the equations/math
4. Try to explain it to someone else

Why would you ever use a linear model instead of something like GAM that can exactly fit the data?

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*
  (but only if a linear approximation is helpful)

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*
  (but only if a linear approximation is helpful)
- Better Sample Complexity*

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*
  (but only if a linear approximation is helpful)
- Better Sample Complexity*
  (but only by assuming away part of the problem)

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*
  (but only if a linear approximation is helpful)
- Better Sample Complexity*
  (but only by assuming away part of the problem)
- Convention*

Why would you ever use a linear model instead of something like GAM that can ~~exactly~~ flexibly fit the data?

The linear model has on its side:

- Unbiasedness*
  (but perhaps high sampling variability)
- Simple Interpretation*
  (but only if a linear approximation is helpful)
- Better Sample Complexity*
  (but only by assuming away part of the problem)
- Convention*
  (not a good reason per se, but a practical one)

Why don't we use maximum likelihood estimation?

Why don't we use maximum likelihood estimation?

We will. Stay tuned for next semester.

For those of us who are considering taking the course next semester, will you tell us what the graded components will be? problem sets? exams? presentations? Thanks!

For those of us who are considering taking the course next semester, will you tell us what the graded components will be? problem sets? exams? presentations? Thanks!

`http://scholar.princeton.edu/bstewart/teaching`

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

Synthetic Control Methods

The Economic Costs of Conflict: A Case Study of the Basque Country

By Alberto Abadie and Javier Gardeazabal*

This article investigates the economic effects of conflict, using the terrorist conflict in the Basque Country as a case study. We find that, after the outbreak of terrorism in the late 1960's, per capita GDP in the Basque Country declined about 10 percentage points relative to a synthetic control region without terrorism. In addition, we use the 1998–1999 truce as a natural experiment. We find that stocks of firms with a significant part of their business in the Basque Country showed a positive relative performance when truce became credible, and a negative relative performance at the end of the cease-fire. (*JEL* D74, G14, P16)

Political instability is believed to have strong adverse effects on economic prosperity. However, to date, the evidence on this matter is scarce, probably because it is difficult to know how economies would have evolved in absence of political conflicts.

This article investigates the economic impact of terrorist and political conflict, the Basque Country had dropped to the sixth position in per capita GDP.[1] During that period, terrorist activity by the Basque terrorist organization ETA resulted in almost 800 deaths. Basque entrepreneurs and corporations had been specific targets of violence and extortion (including assassina-

Synthetic Control Methods

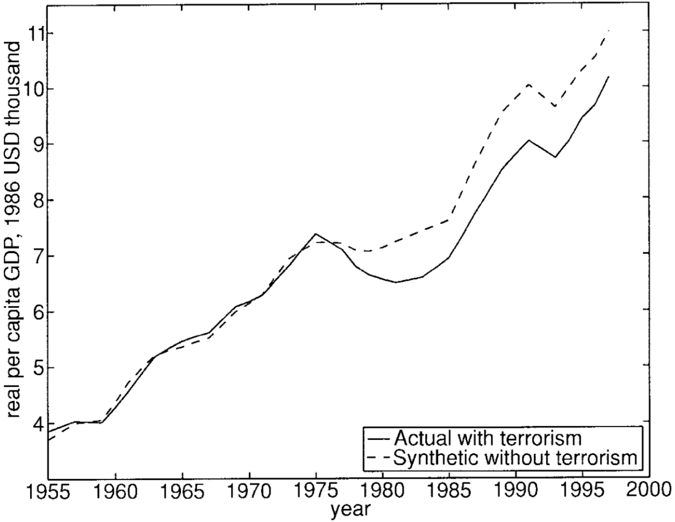
Table 3—Pre-Terrorism Characteristics, 1960's

	Basque Country (1)	Spain (2)	"Synthetic" Basque Country (3)
Real per capita GDP[a]	5,285.46	3,633.25	5,270.80
Investment ratio (percentage)[b]	24.65	21.79	21.58
Population density[c]	246.89	66.34	196.28
Sectoral shares (percentage)[d]			
Agriculture, forestry, and fishing	6.84	16.34	6.18
Energy and water	4.11	4.32	2.76
Industry	45.08	26.60	37.64
Construction and engineering	6.15	7.25	6.96
Marketable services	33.75	38.53	41.10
Nonmarketable services	4.07	6.97	5.37
Human capital (percentage)[e]			
Illiterates	3.32	11.66	7.65
Primary or without studies	85.97	80.15	82.33
High school	7.46	5.49	6.92
More than high school	3.26	2.70	3.10

Sources: Authors' computations from Matilde Mas et al. (1998) and Fundación BBV (1999).

[a] 1986 USD, average for 1960–1969.
[b] Gross Total Investment/GDP, average for 1964–1969.
[c] Persons per square kilometer, 1969.
[d] Percentages over total production, 1961–1969.
[e] Percentages over working-age population, 1964–1969.

Synthetic Control Methods



Synthetic Control Methods

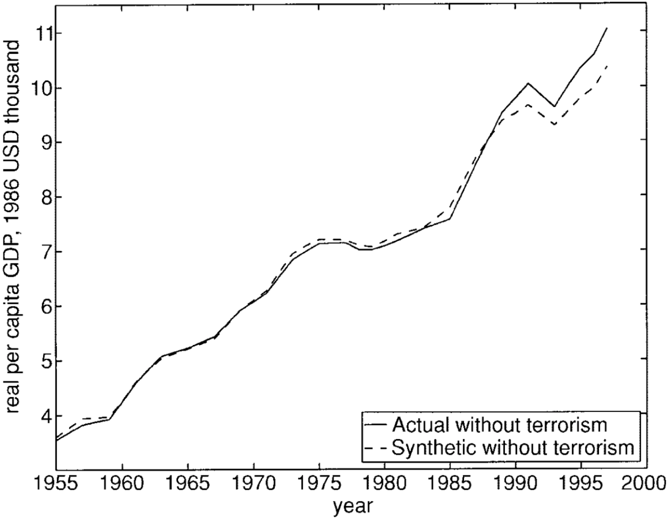


FIGURE 4. A "PLACEBO STUDY," PER CAPITA GDP FOR CATALONIA

Synthetic Control Methods

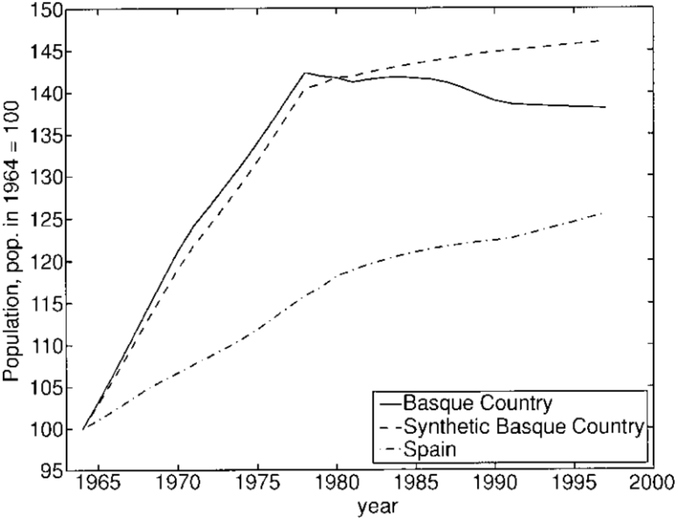


FIGURE 5. POPULATION

Synthetic Control Methods

Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Alberto ABADIE, Alexis DIAMOND, and Jens HAINMUELLER

Building on an idea in Abadie and Gardeazabal (2003), this article investigates the application of synthetic control methods to comparative case studies. We discuss the advantages of these methods and apply them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. We demonstrate that, following Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per-capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is very large, especially in situations where traditional regression methods are not appropriate.

KEY WORDS: Observational studies; Proposition 99; Tobacco control legislation; Treatment effects.

1. INTRODUCTION

Social scientists are often interested in the effects of events or policy interventions that take place at an aggregate level and affect aggregate entities, such as firms, schools, or geographic or administrative areas (countries, regions, cities, etc.). To estimate the effects of these events or interventions, researchers often use comparative case studies. In comparative case studies, researchers estimate the evolution of aggregate outcomes

Comparing the evolution of an aggregate outcome (e.g., state-level crime rate) between a unit affected by the event or intervention of interest and a set of unaffected units requires only aggregate data, which are often available. However, when data are not available at the same level of aggregation as the outcome of interest, information on a sample of disaggregated units can sometimes be used to estimate the aggregate outcomes of interest (like in Card 1990 and Card and Krueger 1994).

Synthetic Control Methods

Table 2. State weights in the synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

Synthetic Control Methods

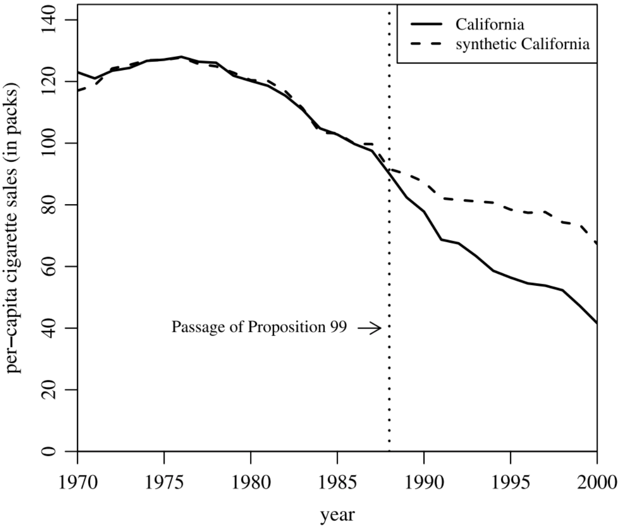


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Synthetic Control Methods

Comparative Politics and the Synthetic Control Method

Alberto Abadie Harvard University and NBER
Alexis Diamond International Finance Corporation
Jens Hainmueller Stanford University

In recent years, a widespread consensus has emerged about the necessity of establishing bridges between quantitative and qualitative approaches to empirical research in political science. In this article, we discuss the use of the synthetic control method as a way to bridge the quantitative/qualitative divide in comparative politics. The synthetic control method provides a systematic way to choose comparison units in comparative case studies. This systematization opens the door to precise quantitative inference in small-sample comparative studies, without precluding the application of qualitative approaches. Borrowing the expression from Sidney Tarrow, the synthetic control method allows researchers to put "qualitative flesh on quantitative bones." We illustrate the main ideas behind the synthetic control method by estimating the economic impact of the 1990 German reunification on West Germany.

Synthetic Control Methods

Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models

Yiqing Xu*

University of California, San Diego

Forthcoming, *Political Analysis*

ABSTRACT

Difference-in-differences (DID) is commonly used for causal inference in time-series cross-sectional data. It requires the assumption that the average outcomes of treated and control units would have followed parallel paths in the absence of treatment. In this paper, we propose a method that not only relaxes this often-violated assumption, but also unifies the synthetic control method (Abadie, Diamond and Hainmueller 2010) with linear fixed effects models under a simple framework, of which DID is a special case. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients. This method has several advantages. First, it allows the treatment to be correlated with unobserved unit and time heterogeneities under reasonable modelling assumptions. Second, it generalizes the synthetic control method to the case of multiple treated units and variable treatment periods, and improves efficiency and interpretability. Third, with a built-in cross-validation procedure, it avoids specification searches and thus is easy to implement. An empirical example of Election Day Registration and voter turnout in the United States is provided.

Keywords: causal inference, TSCS data, difference-in-differences, synthetic control method, interactive fixed effects, factor analysis

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1 Differencing Models

2 Fixed Effects

3 Random Effects

4 (Almost) Twenty Questions

5 Fun with Comparative Case Studies

6 Fun with Music Lab

7 Concluding Thoughts for the Course

And now a very special Fun With

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. Concluding Thoughts for the Course

1. Differencing Models
2. Fixed Effects
3. Random Effects
4. (Almost) Twenty Questions
5. Fun with Comparative Case Studies
6. Fun with Music Lab
7. **Concluding Thoughts for the Course**

Where are you?

Where are you?

You've been given a powerful set of tools

Your New Weapons

Your New Weapons

- Basic probability theory
  - Probability axioms, random variables, marginal and conditional probability, building a probability model
  - Expected value, variances, independence
  - CDF and PDF (discrete and continuous)

Your New Weapons

- Basic probability theory
  - Probability axioms, random variables, marginal and conditional probability, building a probability model
  - Expected value, variances, independence
  - CDF and PDF (discrete and continuous)
- Properties of Estimators
  - Bias, Efficiency, Consistency
  - Central limit theorem

Your New Weapons

- Basic probability theory
  - Probability axioms, random variables, marginal and conditional probability, building a probability model
  - Expected value, variances, independence
  - CDF and PDF (discrete and continuous)
- Properties of Estimators
  - Bias, Efficiency, Consistency
  - Central limit theorem
- Univariate Inference
  - Interval estimation (normal and non-normal Population)
  - Confidence intervals, hypothesis tests, p-values
  - Practical versus statistical significance

Your New Weapons

Your New Weapons

- Simple Regression
  - regression to approximate the conditional expectation function
  - idea of conditioning
  - kernel and loess regressions
  - OLS estimator for bivariate regression
  - Variance decomposition, goodness of fit, interpretation of estimates, transformations

Your New Weapons

- Simple Regression
  - regression to approximate the conditional expectation function
  - idea of conditioning
  - kernel and loess regressions
  - OLS estimator for bivariate regression
  - Variance decomposition, goodness of fit, interpretation of estimates, transformations
- Multiple Regression
  - OLS estimator for multiple regression
  - Regression assumptions
  - Properties: Bias, Efficiency, Consistency
  - Standard errors, testing, p-values, and confidence intervals
  - Polynomials, Interactions, Dummy Variables
  - F-tests
  - Matrix notation

Your New Weapons

Your New Weapons

- Diagnosing and Fixing Regression Problems
  - Non-normality
  - Outliers, leverage, and influence points, Robust Regression
  - Non-linearities and GAMs
  - Heteroscedasticity and Clustering

Your New Weapons

- Diagnosing and Fixing Regression Problems
  - Non-normality
  - Outliers, leverage, and influence points, Robust Regression
  - Non-linearities and GAMs
  - Heteroscedasticity and Clustering
- Causal Inference
  - Frameworks: potential outcomes and DAGs
  - Measured Confounding
  - Unmeasured Confounding
  - Methods for repeated data

Your New Weapons

- Diagnosing and Fixing Regression Problems
  - Non-normality
  - Outliers, leverage, and influence points, Robust Regression
  - Non-linearities and GAMs
  - Heteroscedasticity and Clustering
- Causal Inference
  - Frameworks: potential outcomes and DAGs
  - Measured Confounding
  - Unmeasured Confounding
  - Methods for repeated data
- And you learned how to use `R`: you're not afraid of trying something new!

Using these Tools

Using these Tools

So, Admiral Ackbar, now that you've learned how to run these regressions we can just use them blindly, right?


IT'S A TRAP!

Beyond Linear Regressions

You need more training

Beyond Linear Regressions

Beyond Linear Regressions

- SOC504: with me again!
  we move from guided replication to replication and extension on your own.

Beyond Linear Regressions

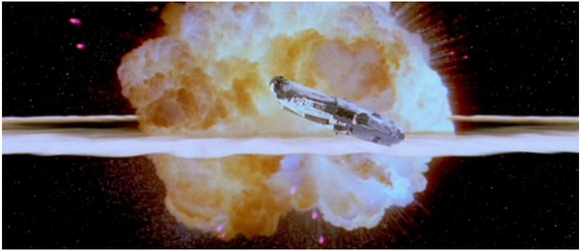
- SOC504: with me again!
  we move from guided replication to replication and extension on your own.
- Social Networks (Graduate or Undergraduate) with Matt Salganik
  fun with social network analysis!

Beyond Linear Regressions

- SOC504: with me again!
  we move from guided replication to replication and extension on your own.
- Social Networks (Graduate or Undergraduate) with Matt Salganik
  fun with social network analysis!

Thanks!

Thanks so much for an amazing semester.



Fill out your evaluations!