Panel Methods (1): The Parametric Approach

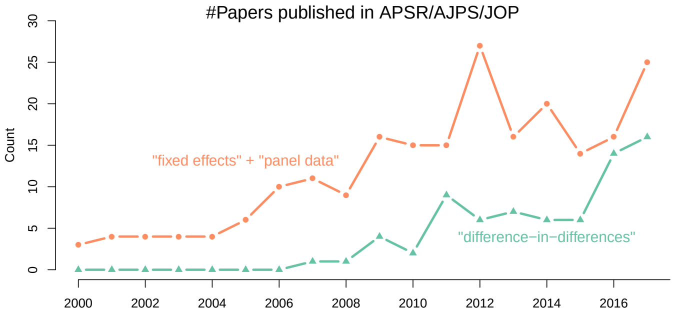
Northwestern Causal Inference Main Workshop

Yiqing Xu
Stanford University

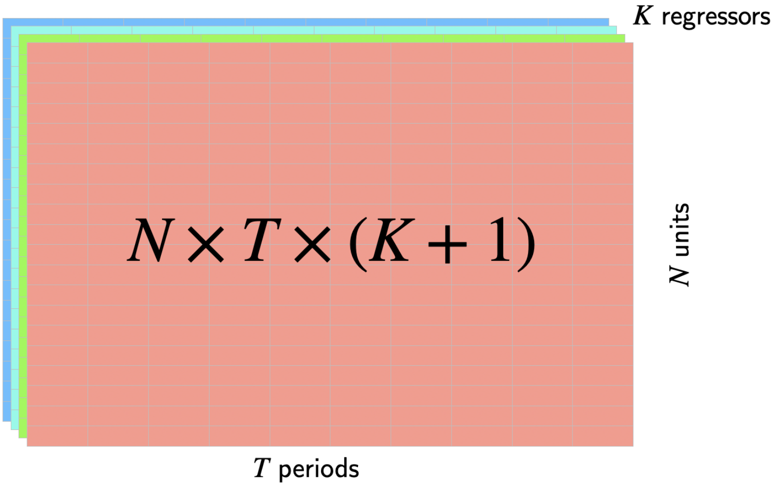
July 31st, 2024

Motivations

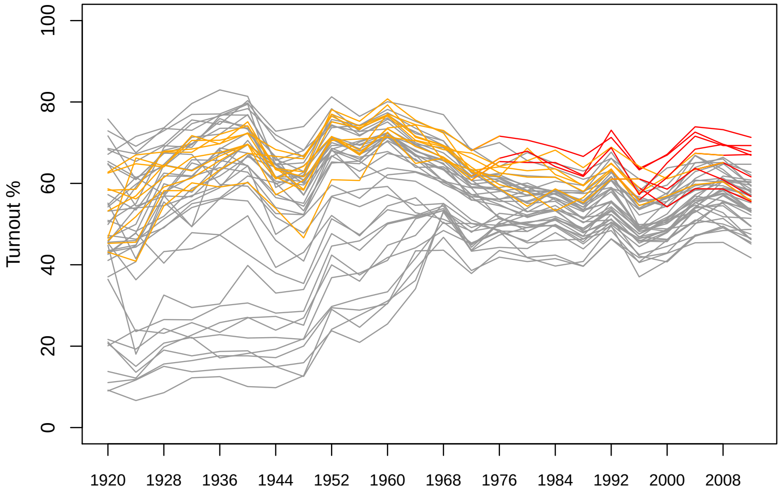
- Causal inference is challenging with observational data
- Panel data often come to the rescue (false hope?)
- In the social sciences, difference-in-differences (DiD) and two-way fixed effects (TWFE) are the most popular



What's a Panel Dataset?

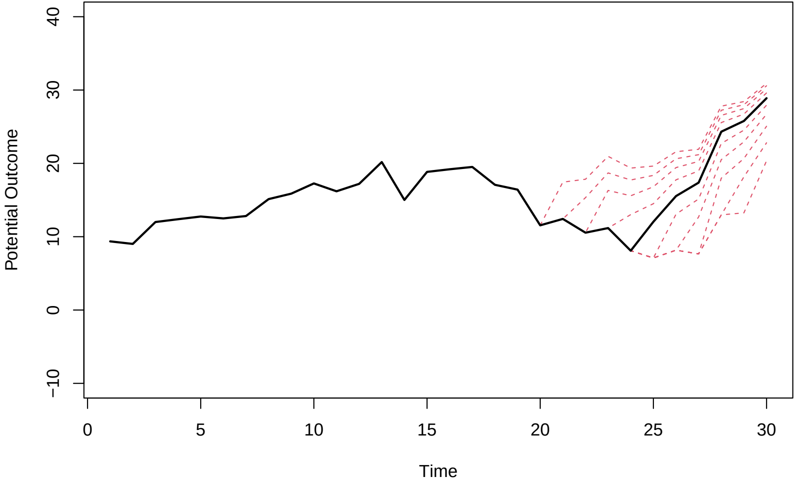


What's a Panel Dataset?



What's Causal Inference with Panel Data?

What would happen if a group of units experienced a set of counterfactual treatment histories?



What's Special about Panel Data?

- The fundamental problem of causal inference (Holland 1986)
$$\tau_i = Y_{1i} - Y_{i0}$$
- A statistical solution makes use of others' information
$$\text{e.g. } ATE = E[Y_1] - E[Y_0]$$
- A scientific solution exploits homogeneity or invariance assumptions

  e.g. The long-run growth rate of the US economy is 2.5%.

- Panel data allow us to construct treated counterfactuals using information from both the past and the others with the caveat that treatment assignment mechanism may be complicated
- Causal inference with panel data is also challenging because of
  - potentially (very) high dimensional treatment
  - limited knowledge about treatment assignment mechanisms
  - both cross-sectional and temporal interference (SUTVA violations)

Today's Plan

- First, introduce the traditional, **parametric** approach, covering models such as unit fixed effects models and two-way fixed effects models (TWFE)
- Investigate the difference-in-differences (DID) design, including its identification, estimation, its link to TWFE models, and common treats
- Discuss the synthetic control methods and its extensions
- [Advanced workshop] Discuss the pathologies of TWFE models and potential solutions

The Parametric Approach

- Panel Setup and Unobserved Confounding
- Fixed Effects (vs. Pooled OLS)
- Time Effects (and TWFE)
- Dynamic Treatment Effects
- Heterogeneous Treatment Effects

Panel Setup

Single unit:

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{it} \\ \vdots \\ y_{iT} \end{pmatrix}_{T\times 1} \qquad \mathbf{X}_i = \begin{pmatrix} x_{i,1,1} & x_{i,1,2} & x_{i,1,j} & \cdots & x_{i,1,K} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{i,t,1} & x_{i,t,2} & x_{i,t,j} & \cdots & x_{i,t,K} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{i,T,1} & x_{i,T,2} & x_{i,T,j} & \cdots & x_{i,T,K} \end{pmatrix}_{T\times K}$$

Panel with all units:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_i \\ \vdots \\ \mathbf{y}_N \end{pmatrix}_{NT\times 1} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_N \end{pmatrix}_{NT\times K}$$

Unobserved Effects Model: Farm Output

- For a randomly drawn cross-sectional unit i, the model is given by

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

  - y_{it}: output of farm i in year t
  - $\mathbf{x}_{it}$: $1 \times K$ vector of variable inputs for farm i in year t, such as labor, fertilizer, etc. plus an intercept
  - $\boldsymbol{\beta}$: $K \times 1$ vector of marginal effects of variable inputs, the causal quantity of interest
  - c_i: farm effect, i.e. the sum of all time-invariant inputs known to farmer i (but unobserved for the researcher), such as soil quality, managerial ability, etc.
    - often called: unobserved effect, unobserved heterogeneity, capturing unobserved confounding
  - ε_{it}: time-varying unobserved inputs, such as rainfall, unknown to the farmer at the time the decision on the variable inputs $\mathbf{x}_{it}$ is made
    - often called: idiosyncratic error
- What happens when we regress y_{it} on $\mathbf{x}_{it}$?

Pooled OLS

- When we ignore the panel structure and regress y_{it} on $\mathbf{x}_{it}$ we get

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

  with composite error $v_{it} \equiv c_i + \varepsilon_{it}$

- Main assumption to obtain consistent estimates for $\boldsymbol{\beta}$ is:
  - $E[v_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}] = E[v_{it}|\mathbf{x}_{it}] = 0$ for $t = 1, 2, ..., T$
    - $\mathbf{x}_{it}$ are strictly exogenous: the composite error v_{it} in each time period is uncorrelated with the past, current, and future regressors
    - But: labour input $\mathbf{x}_{it}$ likely depends on soil quality c_i and so we have omitted variable bias and $\hat{\boldsymbol{\beta}}$ is not consistent
  - No correlation between $\mathbf{x}_{it}$ and v_{it} $\Rightarrow$ no correlation between unobserved effect c_i and $\mathbf{x}_{it}$ for all t
    - Violations are common: whenever we omit a time-constant variable that is correlated with the regressors (unoberseved confounding)

Unobserved Effects Model: Program Evaluation

- Program evaluation model:

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

  - y_{it}: log wage of individual i in year t
  - $prog_{it}$: indicator coded 1 if individual i participants in training program at t and 0 otherwise
  - $\boldsymbol{\beta}$: effect of program
  - c_i: sum of all time-invariant unobserved characteristics that affect wages, such as ability, etc.
- What happens when we regress y_{it} on $prog_{it}$? $\hat{\boldsymbol{\beta}}$ not consistent since $prog_{it}$ is likely correlated with c_i (e.g. ability)
- Always ask: Is there a time-constant unobserved variable (c_i) that is correlated with the regressors? If yes, pooled OLS is problematic
- Additional problem: $v_{it} \equiv c_i + \varepsilon_{it}$ are serially correlated for same i since c_i is present in each t and thus pooled OLS standard errors are invalid

Fixed Effect Regression

- Our unobserved effects model is:

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

- If we have data on multiple time periods, we can think of c_i as fixed effects or "nuisance parameters" to be estimated

- OLS estimation with fixed effects yields:

$$(\widehat{\boldsymbol{\beta}}, \widehat{c}_1, \ldots, \widehat{c}_N) = \underset{\boldsymbol{b}, m_1, \ldots, m_N}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \mathbf{x}_{it}\boldsymbol{b} - m_i)^2$$

this amounts to including N farm dummies in regression of y_{it} on $\mathbf{x}_{it}$

Derivation: Fixed Effects Regression

$$(\widehat{\boldsymbol{\beta}}, \widehat{c}_1, \ldots, \widehat{c}_N) = \underset{\boldsymbol{b}, m_1, \ldots, m_N}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \mathbf{x}_{it}\boldsymbol{b} - m_i)^2$$

The first-order conditions (FOC) for this minimization problem are:

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}'(y_{it} - \mathbf{x}_{it}\widehat{\boldsymbol{\beta}} - \widehat{c}_i) = 0$$

and

$$\sum_{t=1}^{T} (y_{it} - \mathbf{x}_{it}\widehat{\boldsymbol{\beta}} - \widehat{c}_i) = 0$$

for $i = 1, \ldots, N$.

Derivation: Fixed Effects Regression

Therefore, for $i = 1, \ldots, N$,

$$\widehat{c}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}\widehat{\boldsymbol{\beta}}) = \bar{y}_i - \bar{\mathbf{x}}_i\widehat{\boldsymbol{\beta}},$$

where

$$\bar{\mathbf{x}}_i \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}, \qquad \bar{y}_i \equiv \frac{1}{T}\sum_{t=1}^{T}y_{it}.$$

Plug this result into the first FOC to obtain:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(y_{it} - \bar{y}_i)\right)$$

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{y}_{it}\right)$$

with time-demeaned variables $\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$.

Fixed Effects Regression

Running a regression with the time-demeaned variables $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$ and $\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i$ is numerically equivalent to a regression of y_{it} on $\mathbf{x}_{it}$ and unit specific dummy variables.

Fixed effects estimator is often called the within estimator because it only uses the time variation within each cross-sectional unit.

The regression with the time-demeaned variables is consistent for $\boldsymbol{\beta}$ even when $Cov[\mathbf{x}_{it}, c_i] \neq 0$, because time-demeaning eliminates the unobserved effects:

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it} \\
\bar{y}_i &= \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{\varepsilon}_i \\
\hline
(y_{it} - \bar{y}_i) &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (c_i - c_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \\
\ddot{y}_{it} &= \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\varepsilon}_{it}
\end{aligned}
$$

Fixed Effects Regression: Main Results

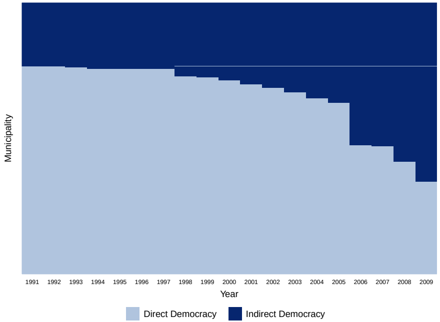
- Identification assumptions:
  1. $E[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i] = 0, \; t = 1, 2, ..., T$
     - regressors are strictly exogenous conditional on the unobserved effect
     - allows $\mathbf{x}_{it}$ to be arbitrarily related to c_i
  2. $rank(\ddot{\mathbf{X}}) = K < NT$ and $E(\ddot{\mathbf{x}}_i ? \ddot{\mathbf{x}}_i)$ is p.d. and finite
     - regressors vary over time for at least some i and are not collinear
- Fixed effects estimator:
  1. Demean and regress $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$ (need to correct degrees of freedom)
  2. Regress y_{it} on $\mathbf{x}_{it}$ and unit dummies (dummy variable regression)
  3. Regress y_{it} on $\mathbf{x}_{it}$ with canned fixed effects routine
     - Stata: `reghdfe y x , a(PanelID) cl(PanelID)`
- Properties (under assumptions 1-2):
  - $\hat{\boldsymbol{\beta}}_{FE}$ is consistent: $\operatorname*{plim}_{N \to \infty} \hat{\boldsymbol{\beta}}_{FE,N} = \boldsymbol{\beta}$
  - $\hat{\boldsymbol{\beta}}_{FE}$ also unbiased conditional on $\mathbf{X}$

Fixed Effects Regression: Main Issues

- Inference:
  - Standard errors have to be "clustered" by panel unit (e.g. farm) to allow correlation in the ε_{it}'s for the same i.
  - Yields valid inference as long as number of clusters is reasonably large
- Typically we care about $\boldsymbol{\beta}$, but unit fixed effects c_i could be of interest
  - $\widehat{c_i}$ from dummy variable regression is unbiased but not consistent for c_i (based on fixed T and $N \to \infty$)
  - `reghdfe` demeans the data before running the regression and therefore does not estimate $\widehat{c_i}$ in the first run, but you can recover them using `predict`

Example: Direct Democracy and Naturalizations

- Do minorities fare worse under direct democracy than under representative democracy?
- Hainmueller and Hangartner (2016, AJPS) examine data on naturalization requests of immigrants in Switzerland, where municipalities vote on naturalization applications in:
  - referendums (direct democracy)
  - elected municipality councils (representative democracy)
- Annual panel data from 1,400 municipalities for the 1991-2009 period
  - y_{it}: naturalization rate = $\#$ naturalizations$_{it}$ / eligible foreign population $_{it-1}$
  - x_{it}: 1 if municipality used representative democracy, 0 if municipality used direct democracy in year t



Naturalization Panel Data

```
. des muniID muni_name year nat_rate repdem

               storage  display     value
variable name    type   format      label      variable label
------------------------------------------------------------------------------------
muniID          float   %8.0g                  municipality code
muni_name       str43   %43s                   municipality name
year            float   %ty                    year
nat_rate        float   %9.0g                  naturalization rate (percent)
repdem          float   %9.0g                  1 representative democracy, 0 direct
```

Panel Data Long Format

```
. list muniID muni_name year nat_rate repdem in 31/40
```

	muniID	muni_name	year	nat_rate	repdem
31.	2	Affoltern A.A.	2002	4.638365	0
32.	2	Affoltern A.A.	2003	4.844814	0
33.	2	Affoltern A.A.	2004	5.621302	0
34.	2	Affoltern A.A.	2005	4.387827	0
35.	2	Affoltern A.A.	2006	8.115358	1
36.	2	Affoltern A.A.	2007	7.067371	1
37.	2	Affoltern A.A.	2008	8.977719	1
38.	2	Affoltern A.A.	2009	6.119704	1
39.	3	Bonstetten	1991	.8333334	0
40.	3	Bonstetten	1992	.8403362	0

Pooled OLS

```
. reg nat_rate repdem

      Source |       SS           df       MS      Number of obs   =      4655
-------------+----------------------------------   F(1, 4653)      =    376.17
       Model |  5958.63488         1  5958.63488   Prob > F        =    0.0000
    Residual |  73705.2336     4,653  15.8403683   R-squared       =    0.0748
-------------+----------------------------------   Adj R-squared   =    0.0746
       Total |  79663.8685     4,654  17.1172902   Root MSE        =      3.98

------------------------------------------------------------------------------
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |   2.503318     .12907    19.40   0.000     2.250279    2.756356
       _cons |   2.222683   .0690427    32.19   0.000     2.087326    2.358039
------------------------------------------------------------------------------
```

Decompose Within and Between Variation

```
. tsset  muniID year ,  yearly
       panel variable:  muniID (strongly balanced)
        time variable:  year, 1991 to 2009
                delta:  1 year

. xtsum  nat_rate

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
nat_rate overall |  2.938992    4.137305          0   24.13793 |     N =    4655
         between |              1.622939          0   7.567746 |     n =     245
         within  |              3.807039  -3.711323   24.80134 |     T =      19
```

Time-Demeaning for Fixed Effects: $y_{it} \rightarrow \ddot{y}_{it}$

```
. * get municipality means
. egen means_nat_rate = mean(nat_rate) , by(muniID)

. * compute deviations from means
. gen  dm_nat_rate = nat_rate - means_nat_rate

. list muniID muni_name year nat_rate means_nat_rate dm_nat_rate in 20/40 ,ab(20)
```

	muniID	muni_name	year	nat_rate	means_nat_rate	dm_nat_rate
20.	2	Affoltern A.A.	1991	.2173913	3.595932	-3.37854
21.	2	Affoltern A.A.	1992	.9473684	3.595932	-2.648563
22.	2	Affoltern A.A.	1993	1.04712	3.595932	-2.548811
23.	2	Affoltern A.A.	1994	.8342023	3.595932	-2.761729
24.	2	Affoltern A.A.	1995	2.002002	3.595932	-1.59393
25.	2	Affoltern A.A.	1996	1.7769	3.595932	-1.819031
26.	2	Affoltern A.A.	1997	1.862745	3.595932	-1.733186
27.	2	Affoltern A.A.	1998	2.054155	3.595932	-1.541776
28.	2	Affoltern A.A.	1999	2.402135	3.595932	-1.193796

Fixed Effects Regression with Demeaned Data

```
. egen means_repdem   = mean(repdem) , by(muniID)

. gen  dm_repdem   = repdem   - means_repdem

.
. * regression with demeaned data
. reg dm_nat_rate dm_repdem , cl(muniID)

Linear regression                                      Number of obs =     4655
                                                       F(  1,    244) =   265.18
                                                       Prob > F      =   0.0000
                                                       R-squared     =   0.1052
                                                       Root MSE      =   3.6017

                             (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
 dm_nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   dm_repdem |     3.0228   .1856244    16.28   0.000     2.657169    3.388431
       _cons |   6.65e-10   5.81e-09     0.11   0.909    -1.08e-08    1.21e-08
------------------------------------------------------------------------------
```

Fixed Effects Regression with Canned Routine

```
. reghdfe nat_rate repdem, a(muniID) cl(muniID)
(MWFE estimator converged in 1 iterations)

HDFE Linear regression                            Number of obs   =      4,655
Absorbing 1 HDFE group                            F(   1,    244) =     265.18
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.2423
                                                  Adj R-squared   =     0.2002
                                                  Within R-sq.    =     0.1052
Number of clusters (muniID)  =        245         Root MSE        =     3.7000

                              (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |     3.0228   .1856244    16.28   0.000     2.657169    3.388431
       _cons |   2.074036   .0531153    39.05   0.000     1.969413    2.178659
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
      muniID |       245         245           0    *|
-----------------------------------------------------+
* = FE nested within cluster; treated as redundant for DoF computation
```

Fixed Effects Regression with Dummies

```
. reg nat_rate repdem i.muniID, cl(muniID)

Linear regression                               Number of obs  =      4655
                                                F(  0,    244) =         .
                                                Prob > F       =         .
                                                R-squared      =    0.2423
                                                Root MSE       =       3.7

                             (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |     3.0228   .1906916    15.85   0.000     2.647188    3.398412
             |
      muniID |
          2  |   1.367365   5.17e-14  2.6e+13   0.000     1.367365    1.367365
          3  |   1.292252   5.17e-14  2.5e+13   0.000     1.292252    1.292252
          9  |   1.284652   5.17e-14  2.5e+13   0.000     1.284652    1.284652
         10  |   1.271783   5.17e-14  2.5e+13   0.000     1.271783    1.271783
         13  |   .3265469   5.17e-14  6.3e+12   0.000     .3265469    .3265469
```

Problems that Fixed Effects Do Not Solve

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

- 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 fewer murders
     - can include time-varying controls, but avoid post-treatment bias
  2. Simultaneity & feedback
     - 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...

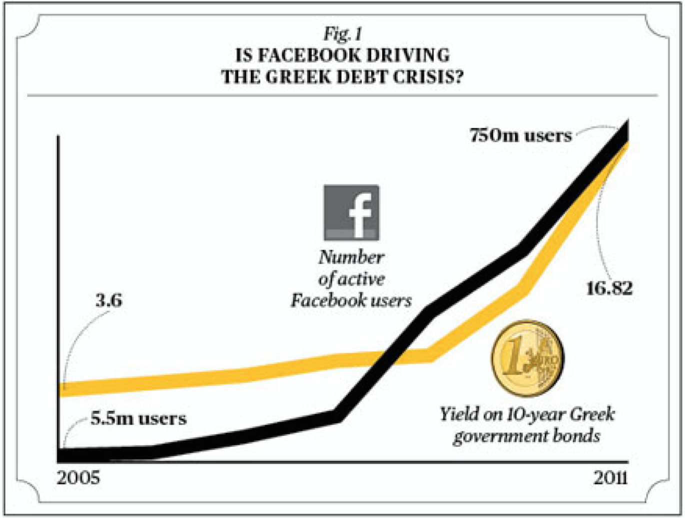
Summary: Fixed Effects, Pooled OLS (and Random Effects)

Assm. 1 Regressors are strictly exogenous conditional on the time-invariant unobservables
Assm. 2 Regressors are uncorrelated with time-varying & time-invariant unobservables

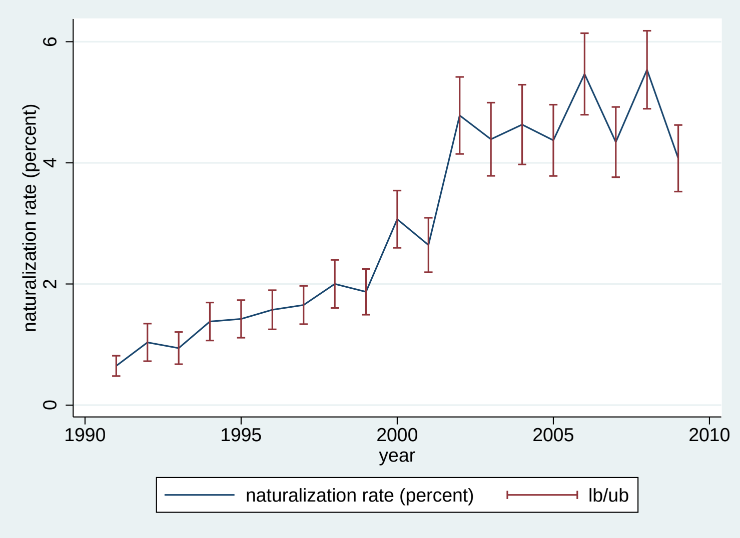
- Results:
    - Fixed effects estimator is consistent given Assm. 1, but rules out time-invariant regressors
    - Pooled OLS (and random effects) are consistent under Assum. 2, and allow for time-invariant regressors
    - Specification tests (e.g., Hausman test) are not gonna cut it

- Assum. 2 is strong, so fixed effects models are typically more credible
    - One main advantage of panel data is to account for time-invariant unobserved confounding

Common Shocks or Causation?

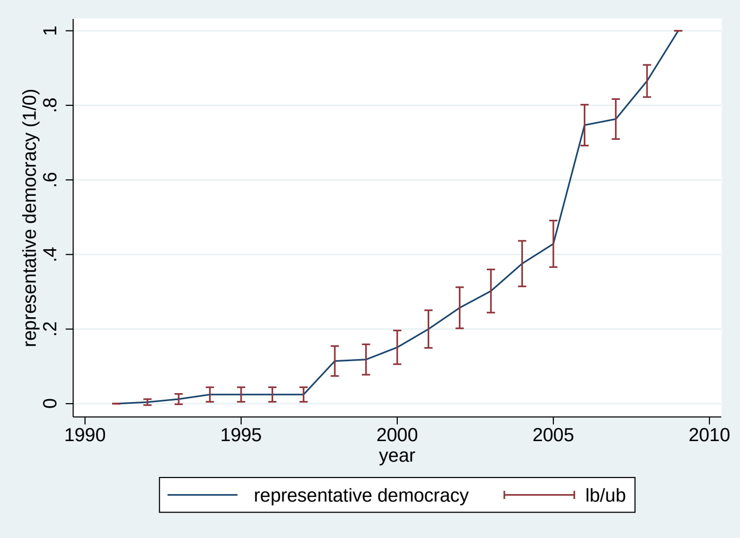


Naturalization Rate Over Time



`xtgraph nat_rate`

Representative Democracy Over Time




```
xtgraph repdem
```

Adding Time Effects

- Reconsider our unobserved effects model:

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

  - Fixed effects assumption: $E[\varepsilon_{it}|\mathbf{x}_i, c_i] = 0,\ t = 1, 2, ..., T$: regressors are strictly exogenous conditional on the unobserved effect
  - Typical violation: Common shocks that affect all units in the same way and are correlated with $\mathbf{x}_{it}$.
    - Trends in farming technology or climate affect productivity
    - Trends in immigration inflows affect naturalization rates
  - We can allow for such common shocks by including time effects into the model

Fixed Effects Regression: Adding Time Effects

- Linear time trend:

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

  - Linear time trend common to all units

- Time fixed effects:

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

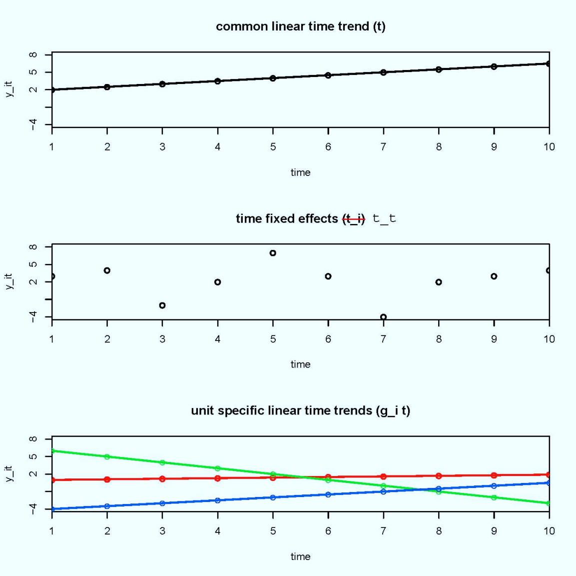
  - Common shock in each time period
  - This is called the two-way fixed effects model
  - Referred to as "generalized difference-in-differences"

- Unit specific linear time trends:

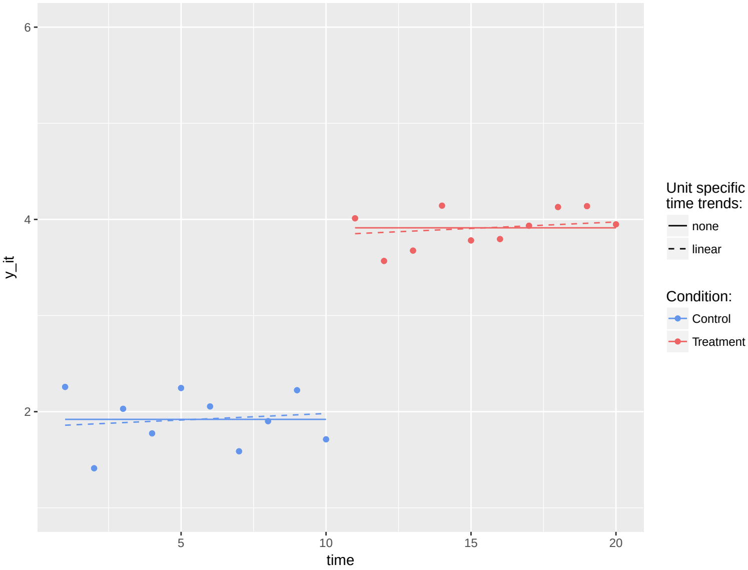
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + g_i \cdot t + t_t + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

  - Linear time trends that vary by unit

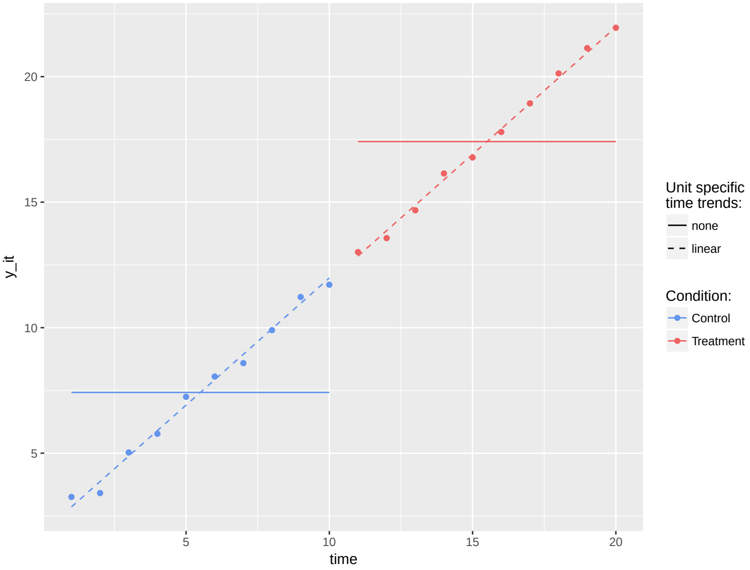
Modeling Time Effects



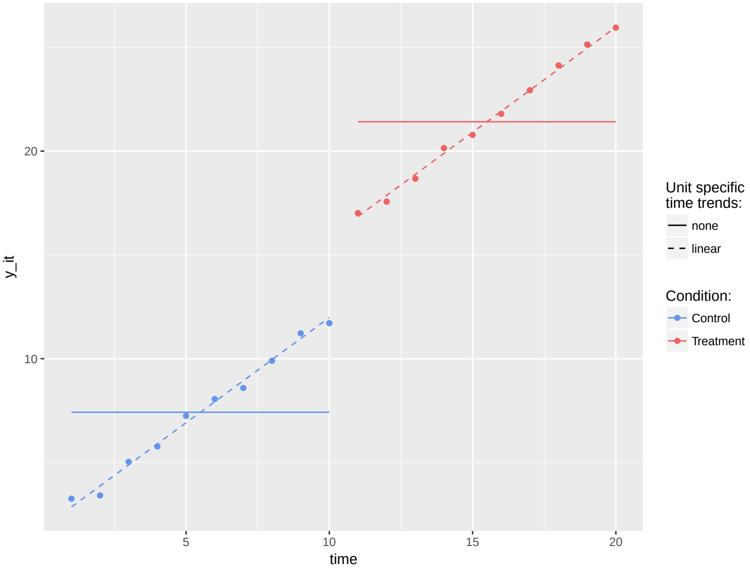
Unit Specific Time Trends



Unit Specific Time Trends



Unit Specific Time Trends



Unit Specific Time Trends Often Eliminate "Results"

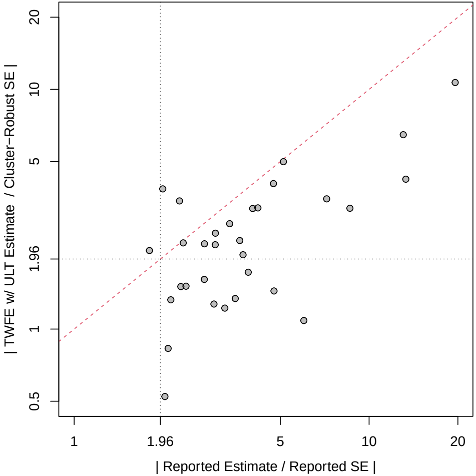
TABLE 5.2.3
Estimated effects of labor regulation on the performance of firms in Indian states

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	−.186	−.185	−.104	.0002
	(.064)	(.051)	(.039)	(.020)
Log development expenditure per capita		.240	.184	.241
		(.128)	(.119)	(.106)
Log installed electricity capacity per capita		.089	.082	.023
		(.061)	(.054)	(.033)
Log state population		.720	0.310	−1.419
		(.96)	(1.192)	(2.326)
Congress majority			−.0009	.020
			(.01)	(.010)
Hard left majority			−.050	−.007
			(.017)	(.009)
Janata majority			.008	−.020
			(.026)	(.033)
Regional majority			.006	.026
			(.009)	(.023)
State-specific trends	No	No	No	Yes
Adjusted R^2	.93	.93	.94	.95

Notes: Adapted from Besley and Burgess (2004), table IV. The table reports regression DD estimates of the effects of labor regulation on productivity. The

"labor regulation increased in states where output was declining anyway"

A survey of political science papers published in top journals



Fixed Effects Regression: Linear Time Trend

```
. reghdfe nat_rate repdem year, a(muniID) cl(muniID)
(MWFE estimator converged in 1 iterations)

HDFE Linear regression                            Number of obs   =      4,655
Absorbing 1 HDFE group                            F(   2,    244) =     247.57
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.2891
                                                  Adj R-squared   =     0.2494
                                                  Within R-sq.    =     0.1604
Number of clusters (muniID)  =        245         Root MSE        =     3.5844

                              (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |   .8247928   .2590615     3.18   0.002     .3145106    1.335075
        year |   .2313692   .0171752    13.47   0.000     .1975386    .2651997
       _cons |  -460.0353   34.2984    -13.41   0.000     -527.594   -392.4766
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
      muniID |       245         245           0    *|
-----------------------------------------------------+
* = FE nested within cluster; treated as redundant for DoF computation
```

Fixed Effects Regression: Year Fixed Effects

```
. reghdfe nat_rate repdem, a(muniID year) cl(muniID)
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                            Number of obs   =      4,655
Absorbing 2 HDFE groups                           F(   1,    244) =      15.76
Statistics robust to heteroskedasticity           Prob > F        =     0.0001
                                                  R-squared       =     0.3129
                                                  Adj R-squared   =     0.2717
                                                  Within R-sq.    =     0.0081
Number of clusters (muniID)  =        245         Root MSE        =     3.5308

                              (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |   1.203658   .3031499     3.97   0.000     .6065335    1.800783
       _cons |   2.594572   .0867445    29.91   0.000     2.423709    2.765436
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-----------------------------------------------------+
 Absorbed FE | Categories  - Redundant  = Num. Coefs |
-------------+---------------------------------------|
      muniID |       245         245           0    *|
        year |        19           1          18     |
-----------------------------------------------------+
* = FE nested within cluster; treated as redundant for DoF computation
```

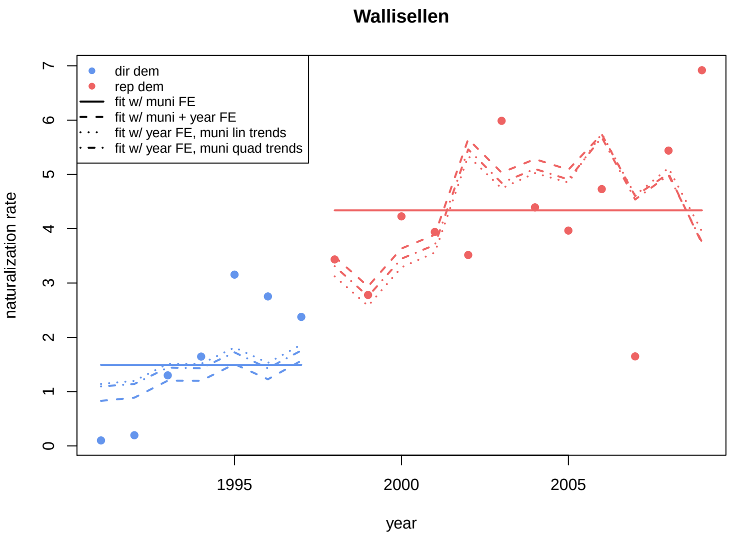
Fixed Effects Regression: Unit Specific Time Trends

```
. reghdfe nat_rate repdem, a(muniID year muniID#c.year) cl(muniID)
(MWFE estimator converged in 3 iterations)

HDFE Linear regression                            Number of obs     =      4,655
Absorbing 3 HDFE groups                           F(   1,     244)  =       9.33
Statistics robust to heteroskedasticity           Prob > F          =     0.0025
                                                  R-squared         =     0.3777
                                                  Adj R-squared     =     0.3014
                                                  Within R-sq.      =     0.0049
Number of clusters (muniID)  =        245         Root MSE          =     3.4581

                               (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |   .9865241   .3229047     3.06   0.002     .3504877    1.622561
       _cons |   2.656704   .0923972    28.75   0.000     2.474706    2.838702
------------------------------------------------------------------------------

Absorbed degrees of freedom:
-------------------------------------------------------+
   Absorbed FE | Categories  - Redundant  = Num. Coefs |
---------------+---------------------------------------|
        muniID |       245         245           0    *|
          year |        19           1          18     |
 muniID#c.year |       245           0         245    ?|
-------------------------------------------------------+
? = number of redundant parameters may be higher
* = FE nested within cluster; treated as redundant for DoF computation
```

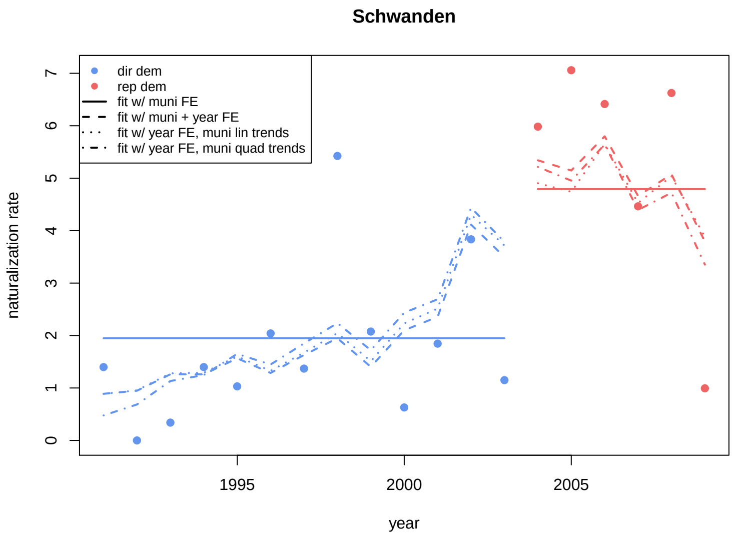
Unit Specific Quadratic Time Trends

```
. g year2 = year^2

. reghdfe nat_rate repdem, a(muniID year muniID#c.year muniID#c.year2) cl(muniID)
(MWFE estimator converged in 3 iterations)

HDFE Linear regression                            Number of obs   =      4,655
Absorbing 4 HDFE groups                           F(   1,     244) =       8.81
Statistics robust to heteroskedasticity           Prob > F        =     0.0033
                                                  R-squared       =     0.3777
                                                  Adj R-squared   =     0.2575
                                                  Within R-sq.    =     0.0049
Number of clusters (muniID)   =        245        Root MSE        =     3.5650

                              (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |   .9862918   .3322911     2.97   0.003     .3317666    1.640817
       _cons |    2.65677   .0950831    27.94   0.000     2.469482    2.844059
------------------------------------------------------------------------------

Absorbed degrees of freedom:
---------------------------------------------------------+
      Absorbed FE | Categories  - Redundant  = Num. Coefs |
------------------+--------------------------------------|
           muniID |       245         245           0    *|
             year |        19           1          18     |
     muniID#c.year|       245           0         245    ?|
    muniID#c.year2|       245           0         245    ?|
---------------------------------------------------------+
? = number of redundant parameters may be higher
* = FE nested within cluster; treated as redundant for DoF computation
```

Fixed Effects Regression: Unit Specific Time Trends



Fixed Effects Regression: Unit Specific Time Trends



Distributed Lag Model

$$y_{it} = x_{it}\boldsymbol{\beta}_0 + x_{it-1}\boldsymbol{\beta}_1 + x_{it-2}\boldsymbol{\beta}_2 + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Model recognizes that effect of change in x may occur with a lag
  - effect of new tax credit for children on fertility rate

- Interpretation of coefficients:
  - Consider temporary increase in x_{it} from level m to $m+1$ at t, which lasts only one period

$$\begin{aligned}
y_{t-1} &= m\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i \\
y_t &= (m+1)\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i \\
y_{t+1} &= m\boldsymbol{\beta}_0 + (m+1)\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i \\
y_{t+2} &= m\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + (m+1)\boldsymbol{\beta}_2 + c_i \\
y_{t+3} &= m\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i
\end{aligned}$$

  - $\boldsymbol{\beta}_0 = y_t - y_{t-1}$ immediate change in y due to temporary one-unit increase in x (impact propensity)
  - $\boldsymbol{\beta}_1 = y_{t+1} - y_{t-1}$ change in y one period after temporary one-unit increase in x
  - $\boldsymbol{\beta}_2 = y_{t+2} - y_{t-1}$ change in y two periods after temporary one-unit increase in x
  - $y_{t+3} = y_{t-1}$ change in y is zero three periods after temporary one-unit increase in x

Lagged Effects of Direct Democracy

```
. reghdfe nat_rate repdem L1.repdem L2.repdem L3.repdem, a(muniID year) cl(muniID)
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                          Number of obs     =      3,920
Absorbing 2 HDFE groups                         F(   4,     244)  =       5.61
Statistics robust to heteroskedasticity         Prob > F          =     0.0002
                                                R-squared         =     0.3078
                                                Adj R-squared     =     0.2580
                                                Within R-sq.      =     0.0124
Number of clusters (muniID)  =        245       Root MSE          =     3.7157

                             (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |
         --. |   .6364802   .3593924     1.77   0.078    -.0714272    1.344388
         L1. |   1.201065   .4233731     2.84   0.005      .367133    2.034998
         L2. |  -.1648692   .4697434    -0.35   0.726    -1.090139    .7604003
         L3. |  -.5245206   .4109918    -1.28   0.203    -1.334065    .2850239
             |
       _cons |   2.906479   .1150951    25.25   0.000     2.679772    3.133186
------------------------------------------------------------------------------
```

Long-run Effect of Direct Democracy

```
. lincom repdem + L1.repdem + L2.repdem + L3.repdem

 ( 1)  repdem + L.repdem + L2.repdem + L3.repdem = 0
```

nat_rate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	1.294485	.4426322	2.92	0.004	.4226175	2.166353

Lags and Leads Model

$$y_{it} = x_{it+1}\boldsymbol{\beta}_{-1} + x_{it}\boldsymbol{\beta}_0 + x_{it-1}\boldsymbol{\beta}_1 + x_{it-2}\boldsymbol{\beta}_2 + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Can use estimate of $\boldsymbol{\beta}_{-1}$ to test for anticipation effects
  - Consider temporary increase in x_{it} from level m to $m+1$ at t
    - $y_{t-2} = \boldsymbol{\beta}_{-1}m + m\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i$
    - $y_{t-1} = \boldsymbol{\beta}_{-1}(m+1) + m\boldsymbol{\beta}_0 + m\boldsymbol{\beta}_1 + m\boldsymbol{\beta}_2 + c_i$
- Anticipation effect: $\boldsymbol{\beta}_{-1} = y_{t-1} - y_{t-2}$ change in y in period $t-1$, the period before the temporary one-unit increase in x
- Placebo test: if x causes y, but y does not cause x, then $\boldsymbol{\beta}_{-1}$ should be close to zero

Leads and Lags

```
. reghdfe nat_rate F1.repdem repdem L1.repdem L2.repdem L3.repdem, a(muniID year) cl(muniID)
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                            Number of obs   =      3,675
Absorbing 2 HDFE groups                           F(   5,     244) =       3.76
Statistics robust to heteroskedasticity           Prob > F        =     0.0027
                                                  R-squared       =     0.3215
                                                  Adj R-squared   =     0.2692
                                                  Within R-sq.    =     0.0121
Number of clusters (muniID)  =        245         Root MSE        =     3.6808

                              (Std. Err. adjusted for 245 clusters in muniID)
------------------------------------------------------------------------------
             |               Robust
    nat_rate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      repdem |
        F1.  |   .1707685   .3212906     0.53   0.596    -.4620886    .8036255
        --.  |   .6975731   .4397095     1.59   0.114    -.1685376    1.563684
        L1.  |   .8723962   .4619322     1.89   0.060    -.0374873     1.78228
        L2.  |    .014941   .4583628     0.03   0.974    -.8879119    .9177939
        L3.  |  -.2904252   .4108244    -0.71   0.480     -1.09964    .5187895
             |
       _cons |   2.838663   .1187582    23.90   0.000     2.604741    3.072585
------------------------------------------------------------------------------
```

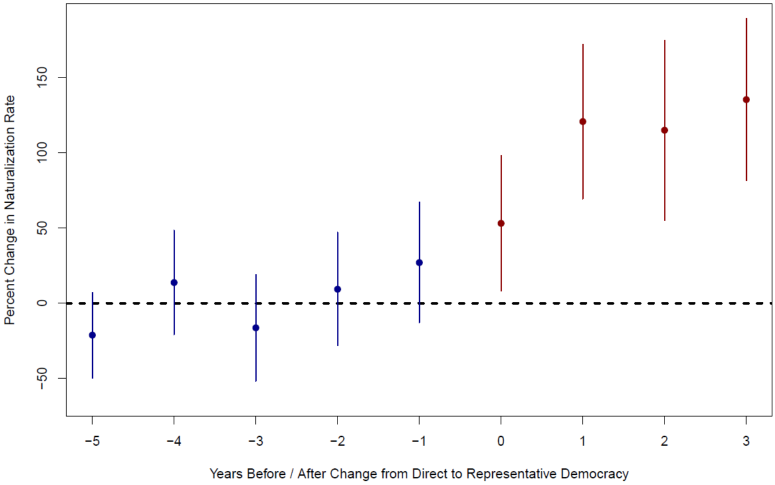
The Autor Test

- Let D_{it} be a binary indicator coded 1 if unit i switched from control to treatment between t and $t-1$; 0 otherwise
  - Lags: D_{it-1}: unit switched between $t-1$ and $t-2$
  - Leads: D_{it+1}: unit switches between $t+1$ and t
- Include lags and leads into the fixed effects model:

$$y_{it} = D_{it+2}\boldsymbol{\beta}_{-2} + D_{it+1}\boldsymbol{\beta}_{-1} + D_{it}\boldsymbol{\beta}_{0} + D_{it-1}\boldsymbol{\beta}_{1} + D_{it-2}\boldsymbol{\beta}_{2} + c_i + \varepsilon_{it}$$

- Interpretation of coefficients:
  - Leads $\boldsymbol{\beta}_{-1}$, $\boldsymbol{\beta}_{-2}$, etc. test for anticipation effects
  - Switch $\boldsymbol{\beta}_0$ tests for immediate effect
  - Lags $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, etc. test for long-run effects
    - highest lag dummy can be coded 1 for all post-switch years

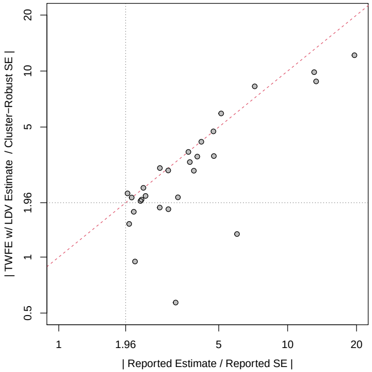
Dynamic Effect of Switching to Representative Democracy



Lagged Dependent Variables (LDVs)

$$y_{it} = \alpha\, y_{it-1} + c_i + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- y_{it} could be capital stock of firm i at time t, and α the capital depreciation rate
- Models with unit fixed effects and lagged y do not produce consistent estimators (Nickell bias)
  - after taking first differences to eliminate c_i, the differenced residual $\Delta\varepsilon_{it}$ is correlated with the lagged dependent variable Δy_{it-1} by construction
- We might use past levels y_{it-2} as an instrument for Δy_{it-1} and apply a general method of moment (GMM) estimator, but this requires strong assumptions (e.g. no serial correlation in ε_{it})
- When $T \to \infty$, the Nickell bias goes to 0; controlling for LDVs is fine as long as they are not correlated with treatment assignment
- However, many existing studies do not survive adding one LDV



Heterogeneous Treatment Effects

- So far we have assumed that the treatment effect is constant across units
- We can allow for heterogeneous treatment effects by including interaction of treatment with other regressors

$$y_{it} = treat_{it}\alpha_0 + (treat_{it} \cdot x_{it})\alpha_1 + x_{it}\boldsymbol{\beta} + c_i + t + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

- Often the treatment is interacted with a time-invariant regressor:

$$y_{it} = treat_{it}\alpha_0 + (treat_{it} \cdot x_i)\alpha_1 + c_i + t + \varepsilon_{it}, \qquad t = 1, 2, ..., T$$

  - Note: The lower order term on the time-invariant x_i is collinear with the fixed effects and drops out

Heterogeneous Effect of Direct Democracy