Microeconometrics

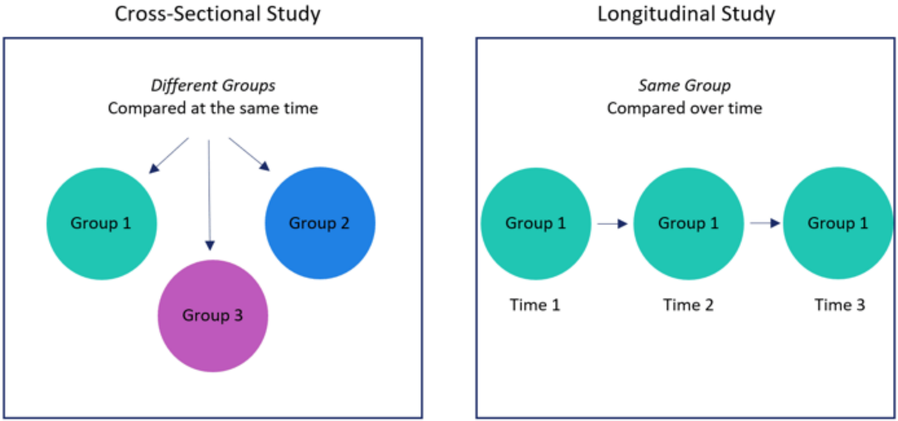
Linear panel data methods

Alex Armand
alex.armand@novasbe.pt

Summary of today's class

1. Panel data setup
2. Assumptions needed in panel data models
3. Estimation methods and interpretation
    1. Pooled OLS (POLS)
    2. Fixed Effects (FE)
    3. First Differencing (FD)
    4. Random Effects (RE)
4. Applications

From cross section to panel data



Random sampling with panel data

- Random sampling across i with fixed time periods T

$$\{(x_{it}, y_{it}) : t = 1, ..., T\}$$

- Two cases:
  1. **Panel is balanced**: each individual i is observed in all periods T
     - We will assume this case for this topic
  2. **Panel is unbalanced**: don't observe some i for some of the periods t
     - Trickier because we must know why we are missing some time periods for some units
     - Selection models later on in the course

General specification

$$y_{it} = \mathrm{z}_i\delta + \mathrm{w}_{it}\gamma + \mathrm{g}_t\theta + v_{it}$$

- $\mathrm{z}_i \Rightarrow$ set of time-constant observed variables
- $\mathrm{w}_{it} \Rightarrow$ set of time-varying observed variables
- $\mathrm{g}_t \Rightarrow$ vector of aggregate time effects

Common feature: **decomposition of the error term**

$$v_{it} = c_i + u_{it}$$

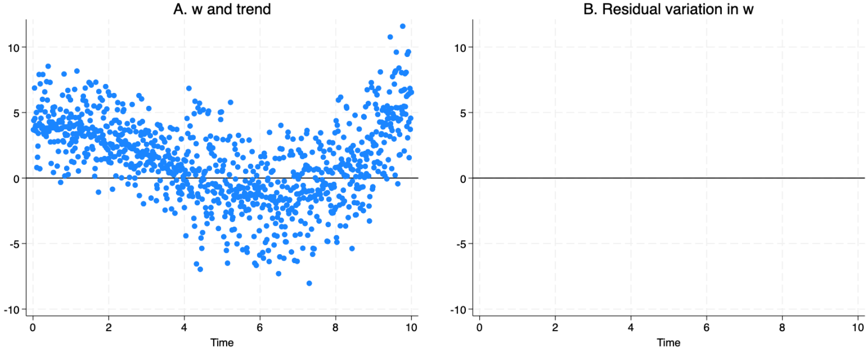
1. $c_i \Rightarrow$ **time-invariant unobservable characteristics**
2. $u_{it} \Rightarrow$ **idiosyncratic errors**

Aggregate time effects

- Time effects remove trends in y_{it} and w_{it} (focus on residual variation)
- **How to choose time effects**?
  - Depends on how much residual variation you want to use $\Rightarrow$ allow for variation within each time group
- **Example**: assume w_{it} varies daily over 10 years $\Rightarrow$ create some variables capturing time and space
  - $time_t \Rightarrow$ measure the time t continuously (each year is 1 unit)
  - $year_t \Rightarrow$ measure the year of t
  - $d_{j,t} \Rightarrow$ dummy equal to 1 if time t belongs to year j
  - $days_{j,t} \Rightarrow$ dummy equal to 1 if time t belongs to group j in which days are groups of 3 days
  - $reg_i \Rightarrow$ dummy equal to 1 if the region of the respondent is region 1, 0 otherwise

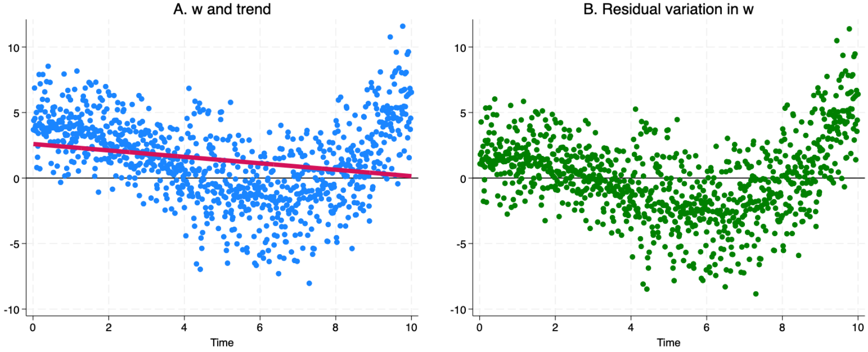
Aggregate time effects – an example

$$\text{no time effects} \Rightarrow \quad y_{it} = w_{it}\gamma + v_{it}$$



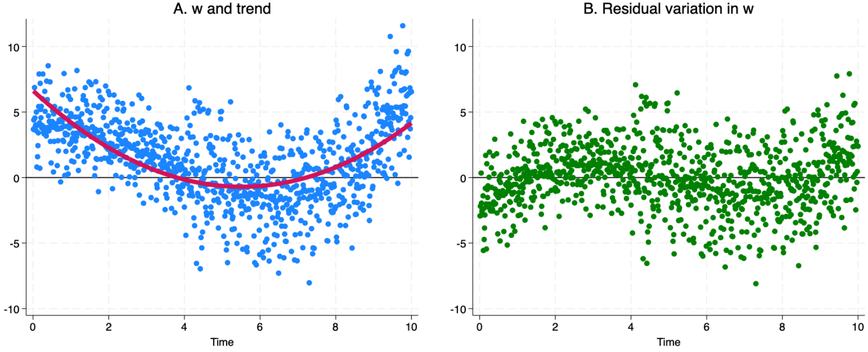
Aggregate time effects – an example

linear trend in time $\Rightarrow \quad y_{it} = \mathrm{w}_{it}\gamma + \theta_1 time_t + v_{it}$



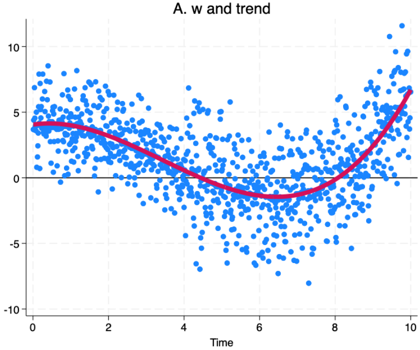
Aggregate time effects – an example

quadratic trend in time $\Rightarrow$ $\quad y_{it} = \mathrm{w}_{it}\gamma + \theta_1 time_t + \theta_2 time_t^2 + v_{it}$



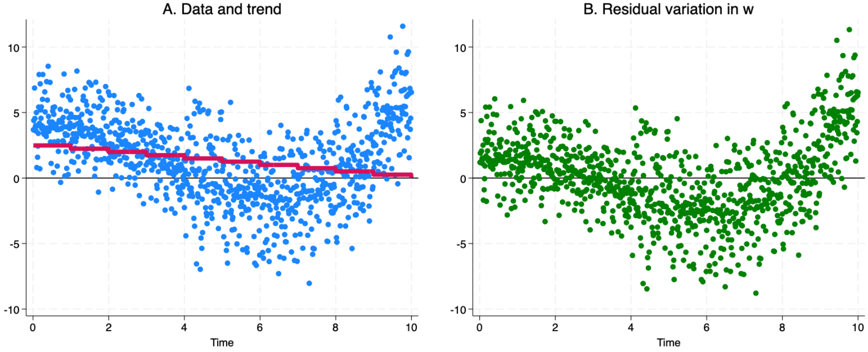
Aggregate time effects – an example

cubic trend in time $\Rightarrow$ $\quad y_{it} = \mathrm{w}_{it}\gamma + \theta_1 time_t + \theta_2 time_t^2 + \theta_3 time_t^3 + v_{it}$



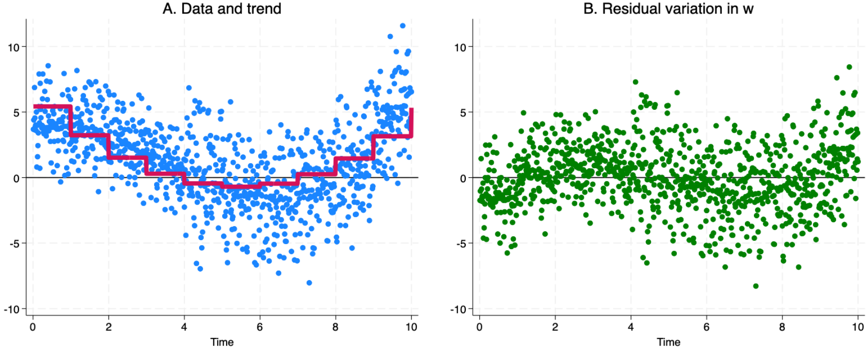
Aggregate time effects – an example

linear trend in year $\Rightarrow \quad y_{it} = \mathrm{w}_{it}\gamma + \theta_1 year_t + v_{it}$



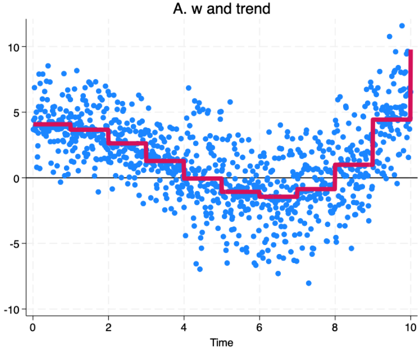
Aggregate time effects – an example

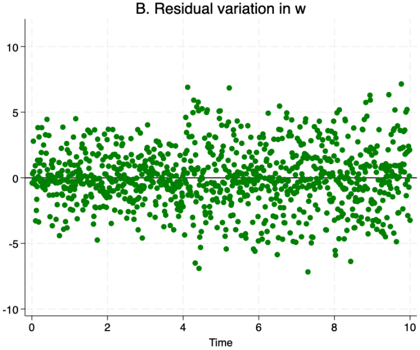
quadratic trend in year $\Rightarrow$ $\quad y_{it} = \mathrm{w}_{it}\gamma + \theta_1 year_t + \theta_2 year_t^2 + v_{it}$



Aggregate time effects – an example

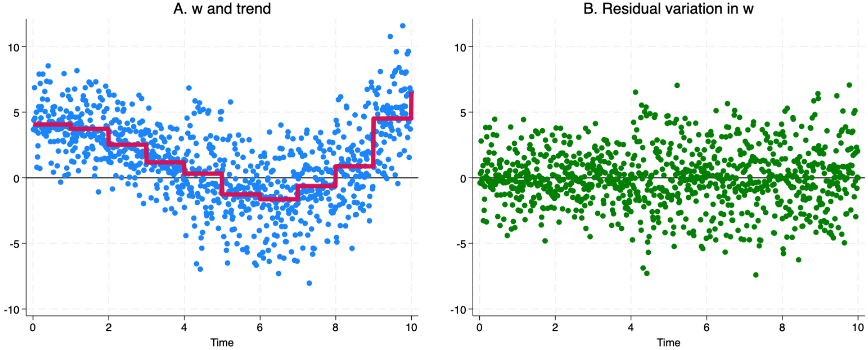
cubic trend in year $\Rightarrow$ $y_{it} = \mathrm{w}_{it}\gamma + \theta_1 year_t + \theta_2 year_t^2 + \theta_3 year_t^3 + v_{it}$





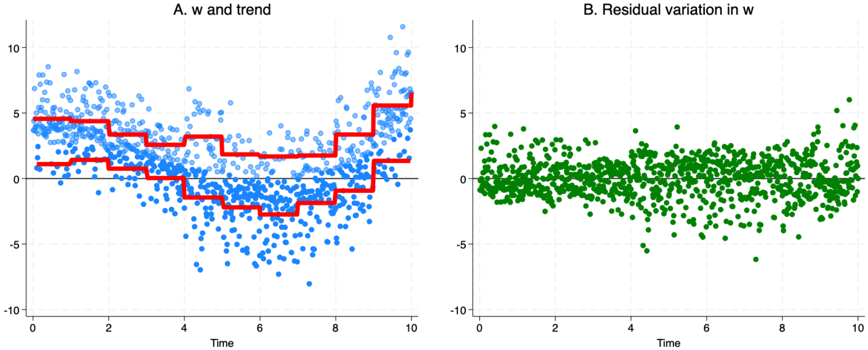
Aggregate time effects – an example

non-linear trend in year $\Rightarrow$ $$y_{it} = w_{it}\gamma + \sum_{j=2}^{10} \theta_j d_{j,t} + v_{it}$$



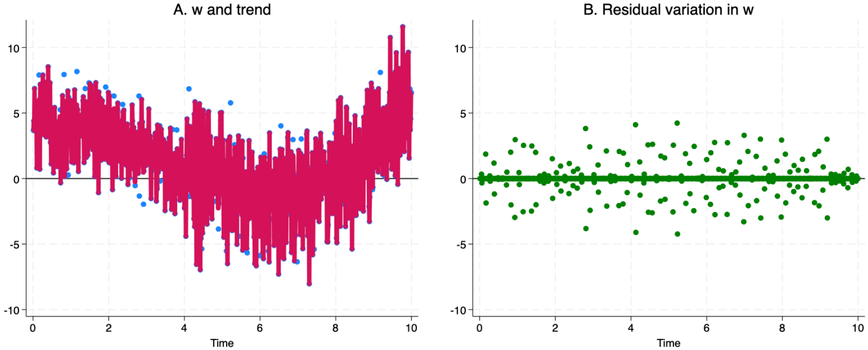
Aggregate time effects – an example

(local) non-linear trend in year $\Rightarrow \quad y_{it} = \mathrm{w}_{it}\gamma + \sum_{j=2}^{10} \theta_j d_{j,t} \cdot reg_i + v_{it}$



Aggregate time effects – an example

non-linear trend in group of days $\Rightarrow \quad y_{it} = \mathrm{w}_{it}\gamma + \sum_{j=2}^{10} \theta_j days_{j,t} + v_{it}$



Consequences of decomposing the error term

- **Correlation structure of the error term**: v_{it} is almost certainly serially correlated

$$
\begin{aligned}
corr(v_{i,t}, v_{i,t-1}) &= corr(c_i + u_{i,t}, c_i + u_{i,t-1}) \\
&= \sigma_c^2 + corr(c_i, u_{i,t-1}) + corr(c_i, u_{i,t}) \\
&\quad + corr(u_{i,t}, u_{i,t-1})
\end{aligned}
$$

- We require **two types of assumptions concerning errors**:
  1. relationship between covariates and c_i
  2. relationship between covariates and u_{it}

Relationship between covariates and c_i

1. **Fixed effect**: no restrictions on the relationship between c_i and x_{it}
2. **Random effect**:

$$Cov(x_{it}, c_i) = 0, \quad t = 1, ..., T$$

3. **Correlated random effects**: we model the relationship between c_i and x_{it}

Relationship between covariates and u_{it}

1. **Contemporaneous exogeneity**

$$E(u_{it}|\mathrm{x}_{it}, c_i) = 0$$

2. **Strict exogeneity**

$$E(u_{it}|\mathrm{x}_{i1}, ..., \mathrm{x}_{iT}, c_i) = 0$$

3. **Sequential exogeneity**

$$E(u_{it}|\mathrm{x}_{it}, \mathrm{x}_{i,t-1}, ..., \mathrm{x}_{i1}, c_i) = 0$$

Estimation

Consider the simple model:

$$\begin{aligned} y_{it} &= x_{it}\beta + v_{it} \\ v_{it} &= c_i + u_{it} \end{aligned}$$

1. **Pooled OLS (POLS)**
2. Fixed Effects (FE)
3. First Differencing (FD)
4. Random Effects (RE)

Pooled OLS (POLS)

- Same as OLS but with stacked observations
- Consistency ensured by
  1. **Rank condition**
  2. **Orthogonality:** $E(x'_{it} v_{it}) = 0$

$$\begin{aligned} & E(x'_{it} c_i) &= 0 \\ \text{Contemporaneous exogeneity:} \quad & E(x'_{it} u_{it}) &= 0, \; t = 1, ..., T \end{aligned}$$

- Decomposition of error term $\Rightarrow$ inference robust to **serial correlation**

Estimation

Consider the simple model:

$$
\begin{aligned}
y_{it} &= x_{it}\beta + v_{it} \\
v_{it} &= c_i + u_{it}
\end{aligned}
$$

1. Pooled OLS (POLS)
2. **Fixed Effects (FE)**
3. First Differencing (FD)
4. Random Effects (RE)

Fixed Effects (FE)

Obtain **FE estimator** in 2 steps:

1. **Between equation ($\bar{y}_i$)**: average the equation across $t \rightarrow$ cross-section

$$\bar{y}_i \equiv \frac{\sum_{t=1}^{T} y_{it}}{T} = \frac{\sum_{t=1}^{T} \mathrm{x}_{it}\beta + c_i + u_{it}}{T}$$

$$= \bar{\mathrm{x}}_i\beta + c_i + \bar{u}_i$$

2. **Within equation ($\ddot{y}_{it}$)**: subtract the between equation from the original equation to eliminate c_i

$$\ddot{y}_{it} \equiv (y_{it} - \bar{y}_i) = (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\beta + u_{it} - \bar{u}_i$$

$$= \ddot{\mathrm{x}}_{it}\beta + \ddot{u}_{it}$$

Fixed Effects (FE)

FE estimator uses pooled OLS for $\ddot{y}_{it} = \ddot{\mathrm{x}}_{it}\beta + \ddot{u}_{it}$

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

Conditions for identification

1. **Orthogonality**: $E(\ddot{\mathrm{x}}_{it}'\ddot{u}_{it}) = 0$, t=1,...,T
2. **Rank**: $rank\ E(\ddot{\mathrm{x}}_{it}'\ddot{\mathrm{x}}_{it}) = K$
   - Condition is **violated** if x_{it} has elements that:
     - do not vary over time
     - change over time in the same way → $\ddot{\mathrm{x}}_{it}$ transformation introduces constant terms

FE and orthogonality

- The key condition for consistency is

$$E(\ddot{\mathrm{x}}_{it}' \ddot{u}_{it}) = E(\ddot{\mathrm{x}}_{it}' u_{it}) = E[(\mathrm{x}_{it} - \bar{\mathrm{x}}_i)' u_{it}] = 0$$

  - The first step is possible because $\bar{u}_i$ is equal to 0

- Includes the following conditions:

$$\begin{aligned} E(\mathrm{x}_{it}' u_{it}) &= 0 \\ E(\bar{\mathrm{x}}_i' u_{it}) &= E\left[\left(\frac{\sum_{s=1}^{T} \mathrm{x}_{is}}{T}\right)' u_{it}\right] = \frac{\sum_{s=1}^{T} E(\mathrm{x}_{is}' u_{it})}{T} = 0 \end{aligned}$$

- **Strict exogeneity!**

EXAMPLE: union status and wages

Impact of being in the **union on log-wages** using different panels for US

- Two approaches:
  1. **POLS**: $y_i = \alpha + \gamma union_i + x_i \beta + v_i$
  2. **FE**: $y_{it} = \alpha + \gamma union_{i,t} + x_{it} \beta + c_i + u_{it}$
- Compare estimates – **what do we learn**?

	Dep. variable: Log-wage	
Dataset	**POLS**	**FE**
Union status in ...		
May CPS, 1974–75	0.19	0.09
LSYM, 1970–78	0.28	0.19
Michigan PSID, 1970–79	0.23	0.14

Note. Adapted from Freeman (1984), "Longitudinal Analyses of the Effects of Trade Union," *Journal of Labor Economics* Vol. 2, No. 1, pp. 1-26. Estimates are calculated using the surveys listed at left. The cross-section estimates include controls for demographic and human capital variables.

APPLICATION: market concentration and airfares

Measure the impact of market concentration on airfares

- $N = 1,149$ U.S. air routes and the years 1997 through 2000
- y_{it} is $\log(fare_{it})$ and the key explanatory variable is $concen_{it}$, the concentration ratio for route i.
- Other covariates are year dummies and the time-constant variables $\log(dist_i)$ and $[\log(dist_i)]^2$.
- Note that what we call c_i Stata refers to as u_i.

Dataset

```
. use airfare

. tab year

1997, 1998, |
 1999, 2000 |      Freq.     Percent        Cum.
------------+-----------------------------------
       1997 |      1,149       25.00       25.00
       1998 |      1,149       25.00       50.00
       1999 |      1,149       25.00       75.00
       2000 |      1,149       25.00      100.00
------------+-----------------------------------
      Total |      4,596      100.00

. sum fare concen dist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        fare |      4596    178.7968    74.88151         37        522
      concen |      4596    .6101149     .196435      .1605          1
        dist |      4596     989.745    611.8315         95       2724
```

POLS estimates (with “wrong” standard errors)

```
. reg lfare concen ldist ldistsq y98 y99 y00

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =  523.18
       Model |  355.453858     6  59.2423096           Prob > F      =  0.0000
    Residual |  519.640516  4589  .113236112           R-squared     =  0.4062
-------------+------------------------------           Adj R-squared =  0.4054
       Total |  875.094374  4595  .190444913           Root MSE      =  .33651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203   .0300691    11.98   0.000     .3011705    .4190702
       ldist |  -.9016004    .128273    -7.03   0.000    -1.153077   -.6501235
     ldistsq |   .1030196   .0097255    10.59   0.000     .0839529    .1220863
         y98 |   .0211244   .0140419     1.50   0.133    -.0064046    .0486533
         y99 |   .0378496   .0140413     2.70   0.007      .010322    .0653772
         y00 |     .09987   .0140432     7.11   0.000     .0723385    .1274015
       _cons |   6.209258   .4206247    14.76   0.000     5.384631    7.033884
------------------------------------------------------------------------------
```

POLS estimates (with serial correlation)

```
. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

                                    (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203     .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004    .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196    .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244    .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496    .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987    .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258    .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------
```

FE estimates

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe cluster(id)

                              (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328    .004163     5.48   0.000     .0146649    .0310007
         y99 |   .0363819   .0051275     7.10   0.000     .0263215    .0464422
         y00 |   .0977717   .0055054    17.76   0.000     .0869698    .1085735
       _cons |   4.953331   .0296765   166.91   0.000     4.895104    5.011557
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

FE estimates: non-linear effects

- Let the effect of concen depend on route distance.

```
. sum ldist if y00

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       ldist |      1149    6.696482    .6595331   4.553877   7.909857

. gen ldistconcen = (ldist - 6.7)*concen
```

FE estimates: non-linear effects

```
. xtreg lfare concen ldistconcen y98 y99 y00, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .1652538   .0482782     3.42   0.001     .0705304    .2599771
 ldistconcen |  -.2498619   .0828545    -3.02   0.003    -.4124251   -.0872987
         y98 |   .0230874   .0041459     5.57   0.000      .014953    .0312218
         y99 |   .0355923   .0051452     6.92   0.000     .0254972    .0456874
         y00 |   .0975745   .0054655    17.85   0.000     .0868511    .1082979
       _cons |    4.93797   .0317998   155.28   0.000     4.875578    5.000362
-------------+----------------------------------------------------------------
     sigma_u |  .50598296
     sigma_e |  .10605257
         rho |  .95791776   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From FE to TWFE

Many applications of FE models use **Two Way Fixed Effects (TWFE)**

- Consider this example:

$$y_{it} = \alpha d_{it} + \tau_t + c_i + u_{it}$$

  - d_{it} is the time-varying variable of interest
  - τ_t and c_i are the time-specific FE and the individual-specific FE

- τ_t and c_i can also be captured by adding indicator variables for each period and individual
  - Replicates FE but with $\uparrow$ computation (many parameters to estimate)

APPLICATION: the costs of low birth weight

How can we identify the **causal impact of low birth weight**?

- Cross-sectional regression would fail orthogonality

$$h_i = \alpha + \beta bw_i + X_i\gamma + \epsilon_i$$

  - bw_i is birth weight of child i
  - h_i is an indicator of health

- We cannot use **panel data** since you are born only once
- Can we use panel data methods applied to non-panel data?
  - Almond et al. QJE (2005) $\Rightarrow$ effect of birth weight on child health applying FE estimator to non-panel data – how?

APPLICATION: the costs of low birth weight

How can we identify the **causal impact of low birth weight**?

- Cross-sectional regression would fail orthogonality

$$h_i = \alpha + \beta bw_i + X_i \gamma + \epsilon_i$$

  - bw_i is birth weight of child i
  - h_i is an indicator of health

- We cannot use **panel data** since you are born only once
- Can we use panel data methods applied to non-panel data?
  - Almond et al. QJE (2005) ⇒ effect of birth weight on child health applying FE estimator to non-panel data – how?

Twins to identify the causal effect (Almond et al. QJE 2005)

- Standard identification problem:

$$h_{ij} = \alpha + bw_{ij}\beta + X_i'\gamma + a_i + \varepsilon_{ij}, \quad (1)$$

where h_{ij} is the underlying health of newborn j for mother i, bw_{ij} is birth weight, X_i is a vector of mother-specific observable determinants of health (e.g., race, age, education), a_i reflects mother-specific unobservable determinants of health (e.g., genetic factors), and ε_{ij} represents other newborn-specific idiosyncratic factors, assumed to be independent of all observable and unobservable factors.

- Using twins, we can apply FE to eliminate a_i
  - Time dimension is replaced by the number of siblings $\Rightarrow$ observe multiple birth weight per mother

Comparison of estimates across POLS and FE

OLS is over-estimating the impact on hospital costs – why?

TABLE III
POOLED OLS AND TWINS FIXED EFFECTS ESTIMATES OF THE EFFECT OF BIRTH WEIGHT

Birth weight coefficient	Including congenital anomalies		Excluding congenital anomalies	
	Pooled OLS	Fixed effects	Pooled OLS	Fixed effects
Hospital costs	−29.95	−4.93	—	—
(in 2000 dollars)	(0.84)	(0.44)	—	—
	[−0.506]	[−0.083]	—	—
Adj. R^2	0.256	0.796	—	—
Sample size	44,410	44,410	—	—
Mortality, 1-year	−0.1168	−0.0222	−0.1069	−0.0082
(per 1000 births)	(0.0016)	(0.0016)	(0.0017)	(0.0012)
	[−0.412]	[−0.078]	[−0.377]	[−0.029]
Adj. R^2	0.169	0.585	0.164	0.629
Sample size	189,036	189,036	183,727	183,727

Comparison of estimates across POLS and FE

Similar when looking at dummy variables instead of bw_{ij}

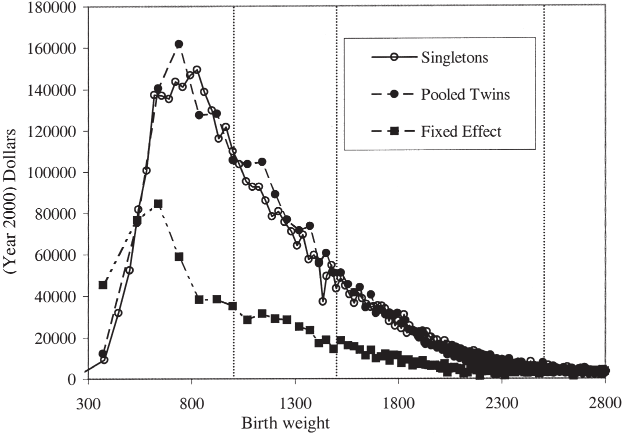


FIGURE Ia
Hospital Costs and Birth Weight
Note: 1995–2000 NY/NJ Hospital Discharge Microdata.

Another TWFE example: difference-in-differences

Example: introduction of a policy restricted to some regions of the country

Time	$t = 0$	$t = 1$
Regions affected	No region	Northern Mexico + Veracruz

Difference-in-Differences (DiD)

A random sample of individuals across the country is observed in two periods $t = 0, 1$

$$y_{it} = \beta + \alpha_i d_{it} + v_{it}$$

$$d_{it} = \begin{cases} 1 & \text{if affected by policy at time } t \\ 0 & \text{otherwise} \end{cases}$$

- DiD estimator applies a standard **TWFE error decomposition**

$$v_{it} = \tau_t + c_i + u_{it}$$

  - c_i: unobservable individual fixed effect
  - τ_t: aggregate time effect
  - u_{it}: idiosyncratic error term

Identification

Compute $E[y_{it}|d_i, t]$ for all groups and for each period $t = 0, 1$

- $E[y_{it}|d_i, t]$ for $d_i = 1$:

$$\begin{cases} \beta + E[c_i|d_i = 1] + \tau_{t_0} + E[u_{it_0}|d_i = 1] & \text{if } t = 0 \\ \beta + E[\alpha_i|d_i = 1] + E[c_i|d_i = 1] + \tau_{t_1} + E[u_{it_1}|d_i = 1] & \text{if } t = 1 \end{cases}$$

- $E[y_{it}|d_i, t]$ for $d_i = 0$:

$$\begin{cases} \beta + E[c_i|d_i = 0] + \tau_{t_0} + E[u_{it_0}|d_i = 0] & \text{if } t = 0 \\ \beta + E[c_i|d_i = 0] + \tau_{t_1} + E[u_{it_1}|d_i = 0] & \text{if } t = 1 \end{cases}$$

Identification

Assumption

Randomization hypothesis holds in first differences:

$$E[v_{it_1} - v_{it_0} | d_i = 1] = E[v_{it_1} - v_{it_0} | d_i = 0] = E[v_{it_1} - v_{it_0}]$$

- $E[y_{it} | d_i, t]$ for $d_i = 1$:

$$\begin{cases} \beta + E[c_i | d_i = 1] + \tau_{t_0} + E[u_{it_0}] & \text{if } t = 0 \\ \beta + E[\alpha_i | d_i = 1] + E[c_i | d_i = 1] + \tau_{t_1} + E[u_{it_1}] & \text{if } t = 1 \end{cases}$$

- $E[y_{it} | d_i, t]$ for $d_i = 0$:

$$\begin{cases} \beta + E[c_i | d_i = 0] + \tau_{t_0} + E[u_{it_0}] & \text{if } t = 0 \\ \beta + E[c_i | d_i = 0] + \tau_{t_1} + E[u_{it_1}] & \text{if } t = 1 \end{cases}$$

Identification: first and second differences

- **1st**: cancels out time-invariant characteristics

$$\begin{aligned}\Delta y_T &= E[y_{it}|d_i = 1, t = 1] - E[y_{it}|d_i = 1, t = 0] \\ &= E[\alpha_i|d_i = 1] + (\tau_{t_1} - \tau_{t_0}) + E[u_{it_1} - u_{it_0}]\end{aligned}$$

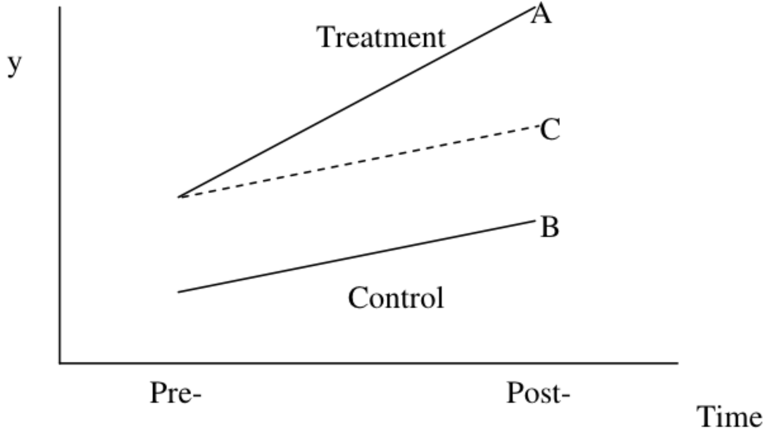
$$\begin{aligned}\Delta y_C &= E[y_{it}|d_i = 0, t = 1] - E[y_{it}|d_i = 0, t = 0] \\ &= (\tau_{t_1} - \tau_{t_0}) + E[u_{it_1} - u_{it_0}]\end{aligned}$$

- **2nd**: cancels out time effects and identifies ATT

$$\Delta y_T - \Delta y_C = E[\alpha_i|d_i = 1] = \alpha^{ATT}$$

What is DiD indentifying?

Excess outcome change for the treated as compared to the non-treated



Estimating DiD

- Use two dummy variables ⇒ TWFE
  1. **Group identifier** (d_i): 1 if individual i lives in affected regions (independently from time), 0 otherwise
  2. **Post** (T_t): 1 if observation refers to $t = 1$, 0 otherwise

- Estimate the following with **OLS**:

$$y_{it} = \alpha_0 + \{\alpha_1 d_i + \alpha_2 T_t + \alpha_3 d_i \cdot T_t\} + \mathrm{X_{it}}\beta + v_{it}$$

  - α_1 ⇒ difference in means across all periods between affected and non-affected regions
  - α_2 ⇒ difference in means between $t = 0$ and $t = 1$
  - α_3 ⇒ **DiD estimate**

Violations

1. **Ashenfelter's dip**: $u_{it_1} - u_{it_0}$ is not unrelated to d

$$E[\alpha^{DiD}] = \alpha^{ATT} + E\left[u_{it_1} - u_{it_0} | d_{it_1} = 1\right] + \\ -E\left[u_{it_1} - u_{it_0} | d_{it_1} = 0\right]$$

   - *Example*: enrolment in a training programme is more likely if a temporary dip in earnings occurs just before the programme

2. **Lack of common trend**: aggregate time effects are different

$$E[\alpha^{DiD}] = \alpha^{ATT} + (\tau_{1,t_1} - \tau_{1,t_0}) - (\tau_{0,t_1} - \tau_{0,t_0})$$

3. **Composition changes**: effects driven by before-after changes in the composition of respondents

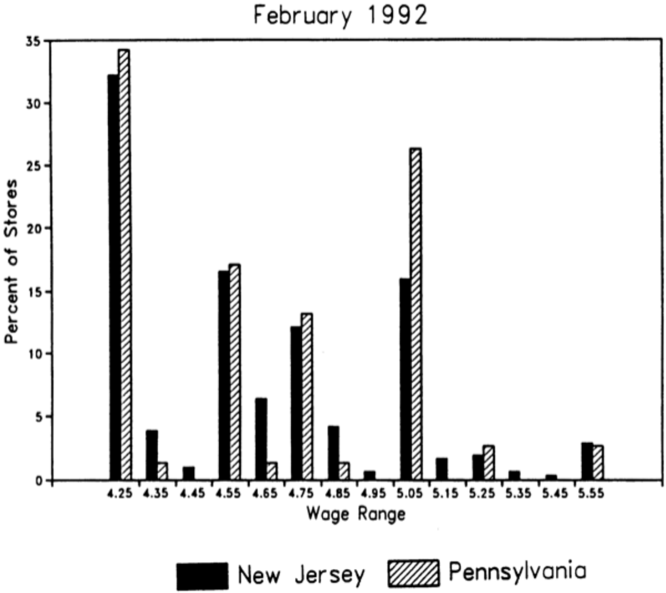
APPLICATION: Card and Krueger (1994)

What is the effect of minimum wage on employment?

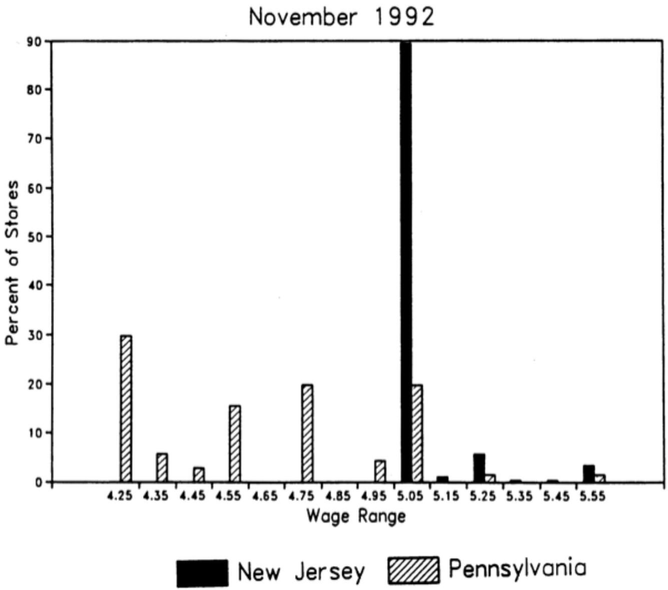
> *On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above $5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment.* (*JEL* J30, J23)

- Minimum wage change in one state (NJ) between $t = 0$ and $t = 1$
- Compare with neighbouring state (PA) for comparable employment
- **Assumptions are valid?**

Wage range at t = 0 (pre)   t = 1 (post)



Wage range at t = 0 (pre) t = 1 (post)



DiD estimates of the effect of minimum wage

- **Columns (i) and (ii)**: 1st differences
- **Column (iii)**: 2nd difference → ATT estimate corresponding to α_3 in this OLS regression

$$y_{it} = \alpha_0 + \{\alpha_1 NJ_i + \alpha_2 Post_t + \alpha_3 NJ_i \cdot Post_t\} + \mathrm{X_{it}}\beta + v_{it}$$

Outcome measure	Mean change in outcome NJ (i)	PA (ii)	NJ − PA (iii)
Store Characteristics:			
1. Fraction full-time workers[c] (percentage)	2.64 (1.71)	−4.65 (3.80)	7.29 (4.17)
2. Number of hours open per weekday	−0.00 (0.06)	0.11 (0.08)	−0.11 (0.10)
3. Number of cash registers	−0.04 (0.04)	0.13 (0.10)	−0.17 (0.11)
4. Number of cash registers open at 11:00 A.M.	−0.03 (0.05)	−0.20 (0.08)	0.17 (0.10)

TWFE and parallel trends

- Consider 2x2 setting ⇒ DiD formula:

$$\alpha^{2x2} = \{E[Y_T|Post] - E[Y_T|Pre]\} - \{E[Y_C|Post] - E[Y_C|Pre]\}$$

  - Y_T (group affected) and Y_C (group not affected)

- Rewrite it using potential outcomes

$$\alpha^{2x2} = \underbrace{\{E[Y_T^1|Post] - E[Y_T^0|Post]\}}_{\text{ATT}} + \underbrace{\{E[Y_T^0|Post] - E[Y_T^0|Pre]\} - \{E[Y_C^0|Post] - E[Y_C^0|Pre]\}}_{\text{Non-parallel trends bias (i.e., selection bias)}}$$

- To cancel out the second term ⇒ **parallel trends**

Special case: staggered rollout

Intervention d_{it} is adopted at multiple times t across multiple groups

Time	$t = 0$	$t = 1$	t =2
Regions affected	No region	Northern Mexico + Veracruz	Northern Mexico + Veracruz + Baja California + Pacific Coast

Special case: staggered rollout

- DiD extends to **TWFE** $\Rightarrow$

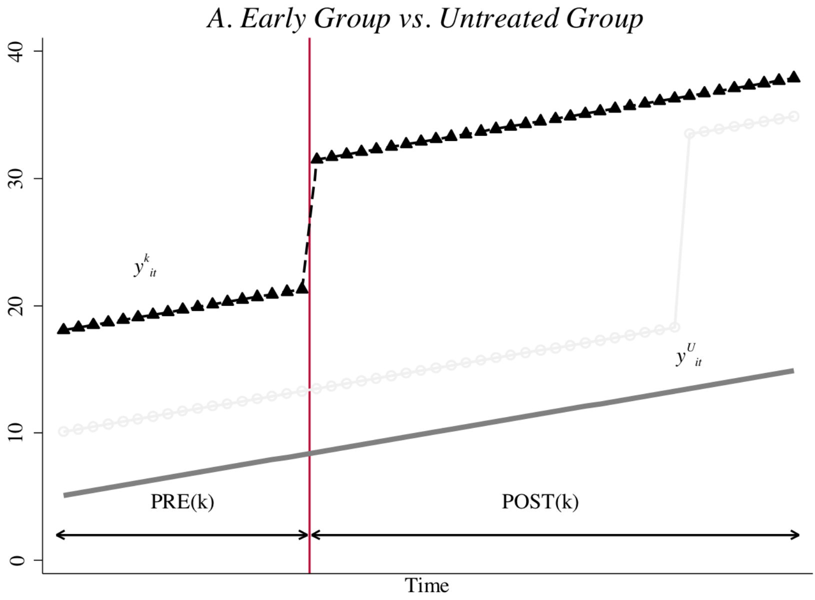
$$Y_{it} = \alpha d_{it} + \tau_t + \sigma_i + \epsilon_{it} \tag{1}$$

- The design in the example considers three different groups:
  1. Regions that are **never affected**: U
  2. Regions **affected early**: k
  3. Regions **affected later**: l
- Many comparisons are possible – Example: DiD for k and U

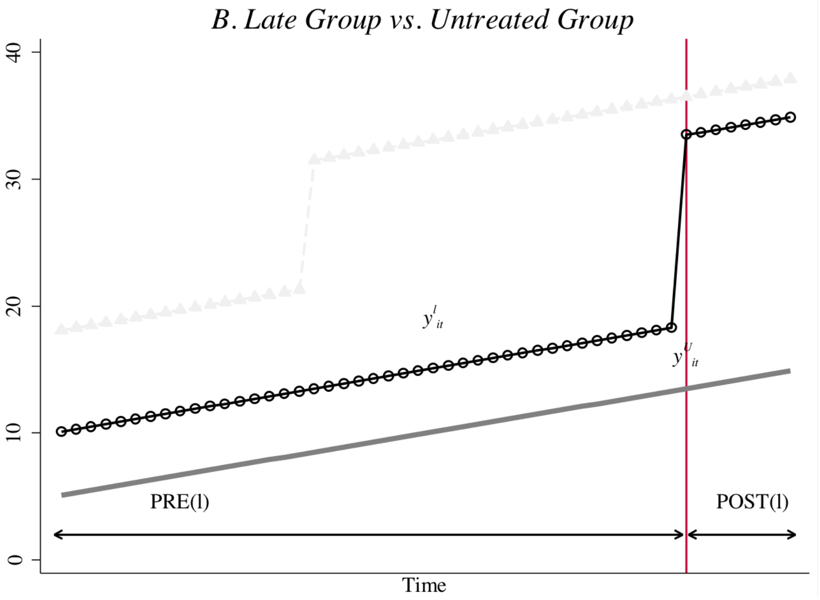
$$\delta^{2x2} = [\bar{y}_k^{post(k)} - \bar{y}_k^{pre(k)}] - [\bar{y}_U^{post(k)} - \bar{y}_U^{pre(k)}]$$

- Under what conditions will α equal the ATT?

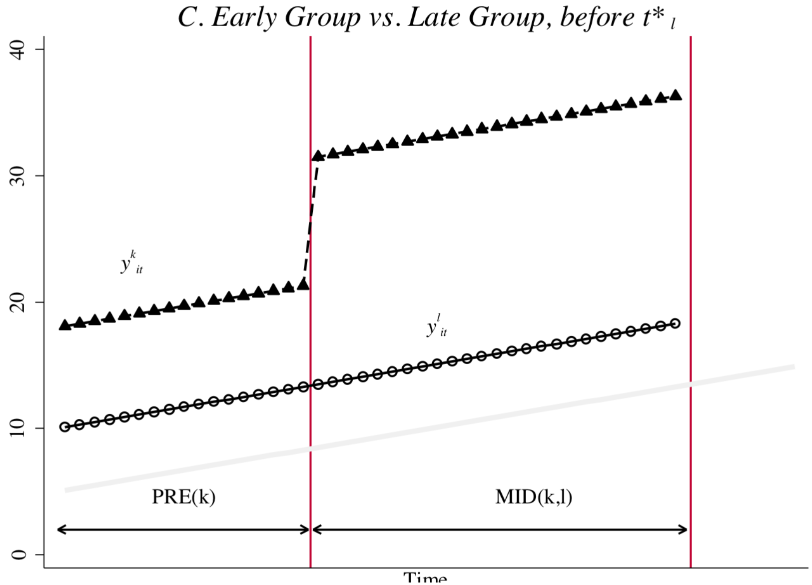
TWFE: early versus never affected



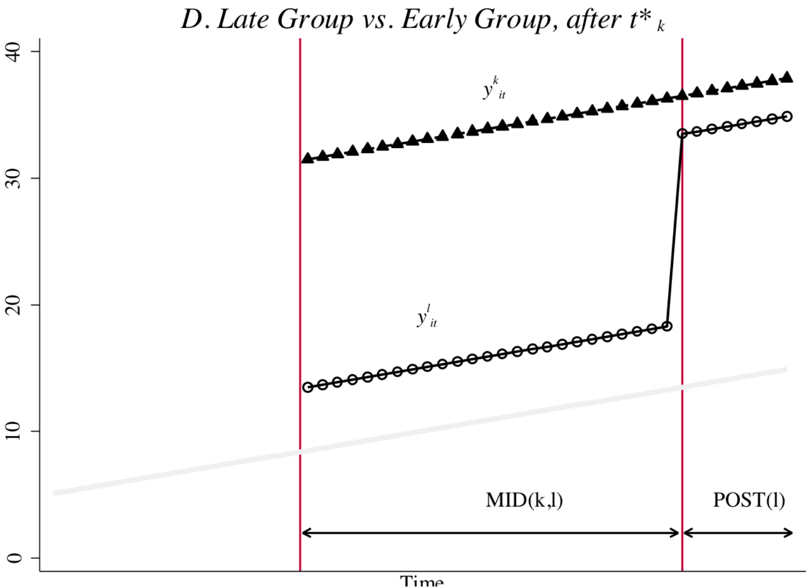
TWFE: late verus never affected



TWFE: early versus late before t^*



TWFE: early versus late after t^*



Goodman-Bacon decomposition

Consider the following 2x2 comparisons:

- $\delta_{kU}^{2x2} \Rightarrow$ affected to never affected
- $\delta_{kl}^{2x2,k} \Rightarrow$ affected early to affected later before t^*
- $\delta_{lk}^{2x2,l} \Rightarrow$ affected early to affected later after t^*

Goodman-Bacon decomposition: α in equation (1) calculates a **weighted average of all these 2x2 comparisons**

$$\tilde{\delta}^{TWFE} = \sum_{k \neq U} s_{kU} \tilde{\delta}_{kU}^{2x2} + \sum_{k \neq U} \sum_{l > k} s_{kl} \left[\mu_{kl} \delta_{kl}^{2x2,k} + (1 - \mu_{kl}) \delta_{lk}^{2x2,l} \right]$$

- All weights are positive
  - s captures variances
  - μ captures the relevance of comparisons between affected groups

Goodman-Bacon decomposition

- Substitute potential outcomes for all the Y values

$$\delta_{kU}^{2x2} = ATT_k Post + \Delta Y_k^0(Post(k), Pre) - \Delta Y_U^0(Post(k), Pre)$$

$$\delta_{kl}^{2x2} = ATT_k(MID) + \Delta Y_k^0(MID, Pre) - \Delta Y_l^0(MID, Pre)$$

$$\begin{aligned}\delta_{lk}^{2x2} = {} & ATT_l(Post(l)) + \Delta Y_l^0(Post(l), MID) - \Delta Y_k^0(Post(l), MID) \\ & -(ATT_k(Post) - ATT_k(MID))\end{aligned}$$

- Key takeaways:
  - All terms are a composition of ATT and **parallel trend biases**
  - δ_{lk}^{2x2} includes an additional term $\Rightarrow$ **heterogeneity bias**

Interpreting TWFE estimates

The estimator $\tilde{\delta}^{TWFE}$ is the sum of three components:

$$\tilde{\delta}^{TWFE} = VWATT + VWPT - \Delta ATT$$

1. ATT (*VW* stands for variance weighted)
2. Differences in parallel trends
3. Evolution of ATT over time
   - $\Delta ATT = 0$: ATT is constant over time $\Rightarrow$ fine for TWFE
   - $\Delta ATT > 0$: ATT is dynamic and introduces a bias $\Rightarrow$ attenuation bias or even change in sign!

Estimation

Consider the simple model:

$$\begin{aligned} y_{it} &= \mathrm{x}_{it}\beta + v_{it} \\ v_{it} &= c_i + u_{it} \end{aligned}$$

1. Pooled OLS (POLS)
2. Fixed Effects (FE)
3. **First Differencing (FD)**
4. Random Effects (RE)

First-differencing (FD) estimation

- Removes c_i by **differencing adjacent observations**:

$$\Delta y_{it} = \Delta \mathrm{x}_{it}\beta + \Delta u_{it}, \quad t = 2, ..., T \textbf{ (we lose t = 1)}$$

- For consistency:
  1. **Orthogonality**: $E(u_{it}|\mathrm{x}_i, c_i) = 0$, $t = 1, ..., T$
  2. **Rank**: $rank\ E(\Delta \mathrm{X}_i' \Delta \mathrm{X}_i) = K$
- FE and FD are the same when $T = 2$

FE estimates

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe cluster(id)

                                (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328    .004163     5.48   0.000     .0146649    .0310007
         y99 |   .0363819   .0051275     7.10   0.000     .0263215    .0464422
         y00 |   .0977717   .0055054    17.76   0.000     .0869698    .1085735
       _cons |   4.953331   .0296765   166.91   0.000     4.895104    5.011557
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

FD estimates

```
. reg D.(lfare concen y98 y99 y00), nocons tsscons cluster(id)

Linear regression                                      Number of obs =    3447
                                                       F(  3,  1148) =   26.29
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0382
                                                       Root MSE      =  .12508

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     D.lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |
         D1. |   .1759764   .0430367     4.09   0.000     .0915371    .2604158
             |
         y98 |
         D1. |   .0227692   .0041573     5.48   0.000     .0146124     .030926
             |
         y99 |
         D1. |   .0364365    .005153     7.07   0.000      .026326    .0465469
             |
         y00 |
         D1. |   .0978497   .0055468    17.64   0.000     .0869666    .1087328
------------------------------------------------------------------------------
```

- All estimates are similar to FE

Estimation

Consider the simple model:

$$
\begin{aligned}
y_{it} &= x_{it}\beta + v_{it} \\
v_{it} &= c_i + u_{it}
\end{aligned}
$$

1. Pooled OLS (POLS)
2. Fixed Effects (FE)
3. First Differencing (FD)
4. **Random Effects (RE)**

Random effects and Generalized Least Squares

Random effects make use of **Generalized Least Squares (GLS)**

- Exploits the correlation structure across multiple equations:
  1. *unconditional* variances across equations are different
  2. *unconditional* covariances across equations are non-zero

- **Equations can represent time** $\Rightarrow$ in the population, each period $1, .., T$ is one equation

$$\mathrm{y}_i = \mathrm{X}_i \beta + \mathrm{u}_i$$

  - y_i is $T \times 1$
  - X_i is $T \times K$
  - u_i is $T \times 1$

Variance-covariance (VCV) matrix of u_i

Unconditional variance-covariance matrix plays a key role in GLS:

$$W \equiv E(\mathrm{u}_i \mathrm{u}_i') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1G} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1G} & \sigma_{2G} & \cdots & \sigma_G^2 \end{pmatrix}$$

- **Properties:** $W^{-1/2}$ is a symmetric nonsingular matrix such that
  - $W^{-1/2}W^{-1/2} = W^{-1}$
  - $W^{-1/2}WW^{-1/2} = \mathrm{I}_G$

GLS transformation and estimator

If W is **known**, we can multiply y_i by $W^{-1/2}$

$$\begin{aligned} W^{-1/2}\mathrm{y}_i &= W^{-1/2}\mathrm{X}_i\beta + W^{-1/2}\mathrm{u}_i \\ \mathrm{y}_i^* &= \mathrm{X}_i^*\beta + \mathrm{u}_i^* \qquad (2) \end{aligned}$$

- Why? $\Rightarrow$ remove correlations in errors

$$E(\mathrm{u}_i^*\mathrm{u}_i^{*\prime}) = W^{-1/2}E(\mathrm{u}_i\mathrm{u}_i')W^{-1/2} = \mathrm{I}_G$$

- **GLS estimator** is the POLS estimator of equation (2)

$$\begin{aligned} \beta^{GLS} &= \left(\sum_{i=1}^N \mathrm{X}_i^{*\prime}\mathrm{X}_i^*\right)^{-1}\left(\sum_{i=1}^N \mathrm{X}_i^{*\prime}\mathrm{y}_i^*\right) \\ &= \left(\sum_{i=1}^N \mathrm{X}_i' W^{-1}\mathrm{X}_i\right)^{-1}\left(\sum_{i=1}^N \mathrm{X}_i' W^{-1}\mathrm{y}_i\right) \end{aligned}$$

GLS and assumptions

$$y_i^* = X_i^* \beta + u_i^* \tag{3}$$

1. **Orthogonality**: $E(X_i' W^{-1} u_i) = 0$
2. **Rank**: rank $E(X_i' W^{-1} X_i) = K$

If W is **not known** $\Rightarrow$ **feasible GLS procedure**

1. Assume initial $W \Rightarrow$ apply GLS
2. Use results to estimate $W \Rightarrow$ apply GLS
3. Iterate step 2 until GLS estimates are stable

Random Effects

For a random draw i from the population

$$
\begin{aligned}
y_{it} &= \mathrm{x}_{it}\beta + v_{it} \\
v_{it} &= c_i + u_{it}
\end{aligned}
$$

- **Orthogonality**

$$
\begin{aligned}
E(u_{it}|\mathrm{x}_{i1}, \mathrm{x}_{i2}, ..., \mathrm{x}_{iT}, c_i) &= 0, \ t = 1, ..., T \\
E(c_i|\mathrm{x}_{i1}, \mathrm{x}_{i2}, ..., \mathrm{x}_{iT}) &= E(c_i)
\end{aligned}
$$

- **Rank**
  - W is nonsingular and $E(\mathrm{X}_i' W^{-1} \mathrm{X}_i)$ nonsingular.
  - non-singularity means in practice that the matrices are invertible.

EXAMPLE: W without serial correlation

- Consider the simplest case:

$$\begin{aligned} Var(u_{it}) &= \sigma_u^2,\ t = 1, ..., T \\ Cov(u_{it}, u_{is}) &= 0,\ t \neq s \end{aligned}$$

- From the error decomposition, we can write:

$$Var(v_{it}) = Var(c_i + u_{it}) = \sigma_c^2 + \sigma_u^2$$

$$\begin{aligned} Cov(v_{it}, v_{is}) &= Cov(c_i + u_{it}, c_i + u_{is}) \\ &= Var(c_i) + Cov(c_i, u_{is}) + Cov(u_{it}, c_i) + Cov(u_{it}, u_{is}) \\ &= \sigma_c^2 \end{aligned}$$

EXAMPLE: W without serial correlation

- The matrix W can be written as:

$$W = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \sigma_c^2 \\ \sigma_c^2 & \cdots & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 \\ \sigma_c^2 & \cdots & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$

- **FGLS requires the initial** $W \Rightarrow$ only 2 parameters
  1. σ_u^2
  2. σ_c^2

Estimator of σ_u^2 and σ_c^2

Starting point $\Rightarrow$ POLS $\Rightarrow$ use residuals v_{it} to estimate both

1. How to estimate σ_u^2?
   - Since $v_{it} - \bar{v}_i = u_{it} - \bar{u}_i$, we can write:

$$\begin{aligned} Var(v_{it} - \bar{v}_i) &= Var(u_{it} - \bar{u}_i) = \sigma_u^2 + \frac{\sigma_u^2}{T} - 2Cov(u_{it}, \bar{u}_i) \\ &= \sigma_u^2 + \frac{\sigma_u^2}{T} - 2\frac{\sigma_u^2}{T} \\ &= \sigma_u^2 \frac{T-1}{T} \\ \frac{T}{T-1} Var(v_{it} - \bar{v}_i) &= \sigma_u^2 \end{aligned}$$

2. How to estimate σ_c^2?
   - Easy to compute because $Cov(v_{it}, v_{is}) = \sigma_c^2$ (see previous slide)
   - **Breusch-Pagan test**: test whether $H_0 : \sigma_c^2 = 0$

RE estimator

1. Use initial $\hat{W}$
2. Apply FGLS estimator as RE estimator:

$$\widehat{Avar}(\hat{\beta}_{RE}) = \left(\sum_{i=1}^{N} X_i'\hat{W}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\hat{W}^{-1}\hat{v}_i\hat{v}_i'\hat{W}^{-1}X_i\right)\left(\sum_{i=1}^{N} X_i'\hat{W}^{-1}X_i\right)^{-1}$$

3. Use new residuals to compute a new estimate of W
4. Repeat steps 2–3 until convergence

Application: market concentration and airfares

- $N = 1,149$ U.S. air routes and the years 1997 through 2000
- $y_{it} \Rightarrow \log(fare_{it})$
- key explanatory variable $\Rightarrow$ $concen_{it}$ (concentration ratio for route i)

```
. use airfare

. tab year

1997, 1998, |
 1999, 2000 |      Freq.     Percent        Cum.
------------+-----------------------------------
       1997 |      1,149       25.00       25.00
       1998 |      1,149       25.00       50.00
       1999 |      1,149       25.00       75.00
       2000 |      1,149       25.00      100.00
------------+-----------------------------------
      Total |      4,596      100.00

. sum fare concen dist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        fare |      4596    178.7968    74.88151         37        522
      concen |      4596    .6101149     .196435      .1605          1
        dist |      4596     989.745    611.8315         95       2724
```

POLS estimates

```
. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

                          (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203    .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004   .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196   .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244   .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496   .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987   .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258   .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------
```

FE estimates

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe cluster(id)

                               (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328    .004163     5.48   0.000     .0146649    .0310007
         y99 |   .0363819   .0051275     7.10   0.000     .0263215    .0464422
         y00 |   .0977717   .0055054    17.76   0.000     .0869698    .1085735
       _cons |   4.953331   .0296765   166.91   0.000     4.895104    5.011557
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Note that what we call c_i Stata refers to as u_i.

RE estimates

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, re cluster(id)

                                (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2089935   .0422459     4.95   0.000      .126193    .2917939
       ldist |  -.8520921   .2720902    -3.13   0.002    -1.385379   -.3188051
     ldistsq |   .0974604   .0201417     4.84   0.000     .0579833    .1369375
         y98 |   .0224743   .0041461     5.42   0.000      .014348    .0306005
         y99 |   .0366898   .0051318     7.15   0.000     .0266317     .046748
         y00 |    .098212   .0055241    17.78   0.000     .0873849     .109039
       _cons |   6.222005   .9144067     6.80   0.000     4.429801    8.014209
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Notice that the RE and POLS coefficients on the time-constant distance variables are pretty similar, something that often occurs.

RE estimates: omitting distance

```
. xtreg lfare concen y98 y99 y00, re cluster(id)

Random-effects GLS regression                   Number of obs      =      4596
Group variable: id                              Number of groups   =      1149


                                 (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .0468181   .0427562     1.09   0.274    -.0369826    .1306188
         y98 |   .0239229   .0041907     5.71   0.000     .0157093    .0321364
         y99 |   .0354453   .0051678     6.86   0.000     .0253167     .045574
         y00 |   .0964328   .0055197    17.47   0.000     .0856144    .1072511
       _cons |   5.028086   .0285248   176.27   0.000     4.972178    5.083993
-------------+----------------------------------------------------------------
     sigma_u |  .40942871
     sigma_e |  .10651186
         rho |  .93661309   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Estimate is much smaller than FE estimate
    - Can be very **harmful to omit time-constant variables in RE**