

Heterogeneous Difference-in-Differences in Stata

Di Liu

Principal Econometrician

Stata

Table of Contents

- 1 Heterogeneous treatment effects and TWFE
- 2 Example: minimum wage
- 3 Aggregate ATET's
- 4 Four estimators

Introduction

Setup

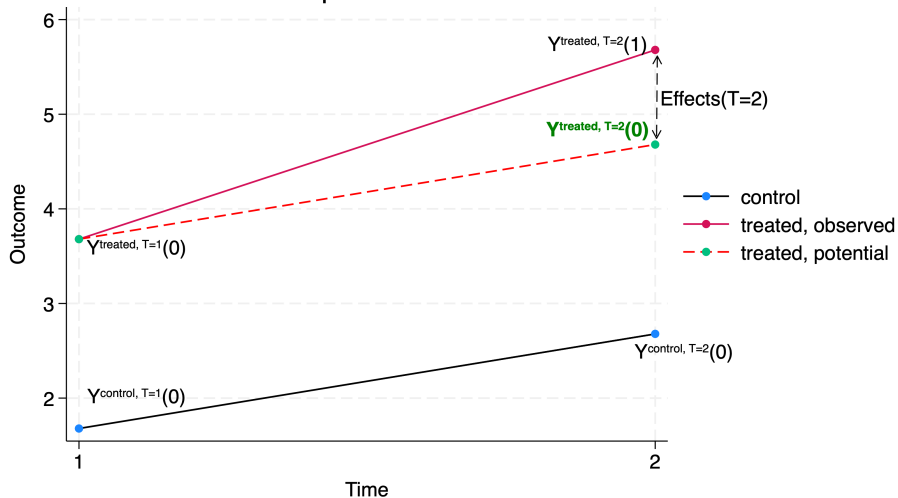
- Estimate treatment effects using panel data or repeated cross-section
- Treatments may start at different times
- Staggered treatment (once treated, always treated)

Group/Cohort: units in the same group start the treatment at the same time, different groups start treatments at different times

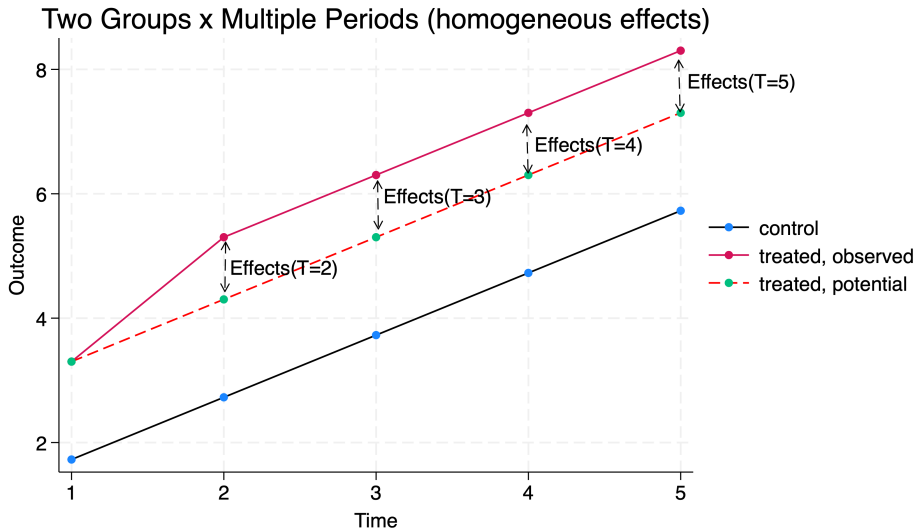
Objective: Estimate treatment effects for the treated groups

Canonical DID

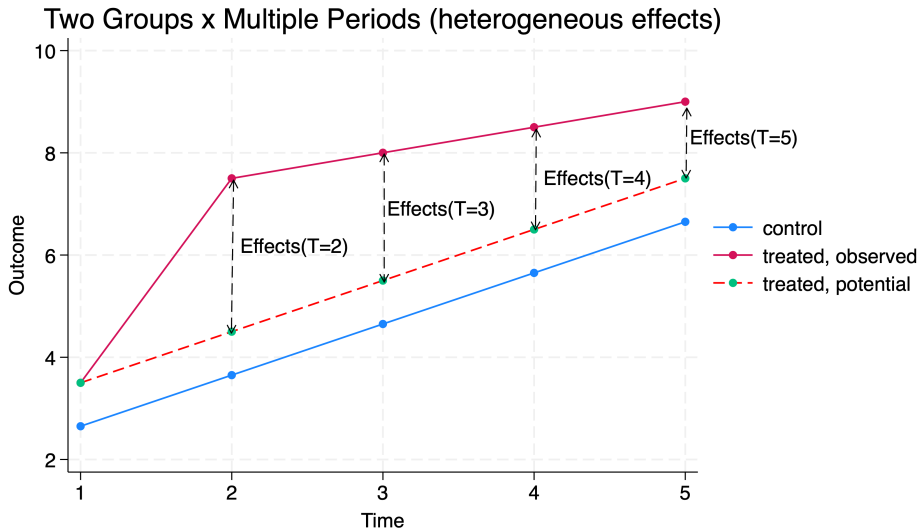
Two Groups x Two Periods



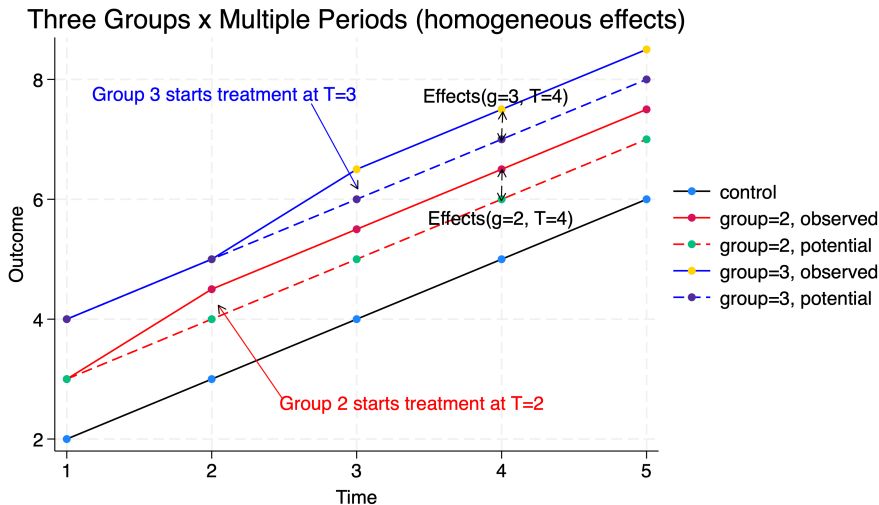
Homogeneous effects across time



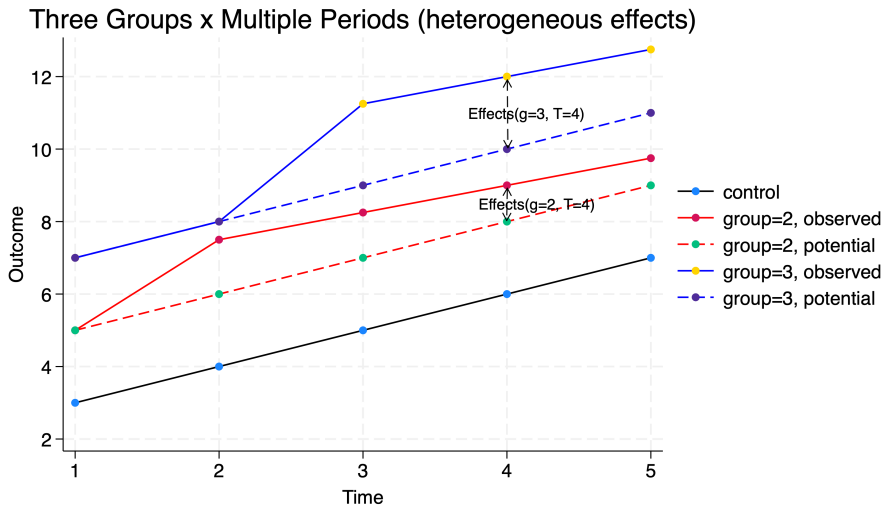
Heterogeneous effects across time



Homogeneous effects across groups and time



Heterogeneous effects across groups and time



Problematic TWFE

$$y_{it} = \theta_t + \eta_i + d_{it}\alpha + v_{it}$$

But,

$$\alpha = \sum w_k \text{Good_DID}_k + \sum w_j \text{Bad_DID}_j$$

- 1 Newly treated relative to the control group (good)
- 2 Newly treated relative to the not-yet treated group (good)
- 3 Newly treated relative to already treated group (bad)

When is TWFE good?

- 1 There are **only two periods**
- 2 The treatment effects are **homogenous across both groups and time**

Overview of heterogeneous DID in Stata 18

Estimation:

- 1 `xthdidregress` and `hdidregress` for panel data and repeated cross-section data
- 2 Four estimators: `ra`, `ipw`, `aipw` in Callaway and Sant'Anna (2021) and `twfe` in Wooldridge (2021)

Post-estimation:

- 1 `estat atetplot`: visualize ATETs
- 2 `estat aggregation`: aggregate ATETs along different dimensions
- 3 `estat ptrends`: pre-treatment parallel trend tests
- 4 `estat sci`: simultaneous CI for RA, IPW, and AIPW estimators

Table of Contents

- 1 Heterogeneous treatment effects and TWFE
- 2 Example: minimum wage**
- 3 Aggregate ATET's
- 4 Four estimators

Increasing minimum wage and young employment

- **Outcome:** county-level employment for young workers
- **Treatment:** minimum wage restrictions introduced by State government; see Callaway and Sant'Anna (2021)
- **Multiple periods:** 2002 - 2007 (6 years)
- **Multiple treatment timings:** 2004, 2006, 2007

```
xthdidregress aipw
```

Define covariates

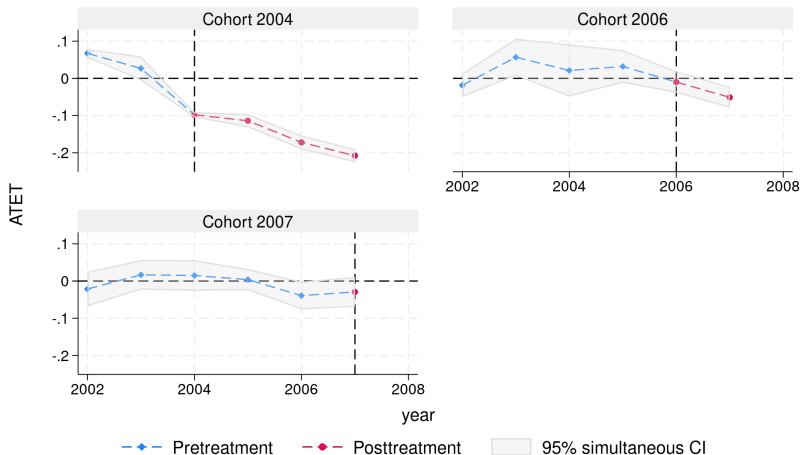
```
global covars i.region pop medinc white hs pov ///  
              c.pop#c.pop c.medinc#c.medinc
```

Use AIPW estimator

```
xthdidregress aipw (lemp $covars) (treat $covars), group(state)
```

- Adding covariates for conditional parallel trend
- There are 18 ATET(g,t)'s (6 years \times 3 cohorts)
- Standard errors are adjusted by clusters of state

estat atetplot, sci



- Specify option `sci` for simultaneous confidence intervals
- For cohorts 2004 and 2006, minimum wage restriction decreases the employment rate for young workers

Table of Contents

- 1 Heterogeneous treatment effects and TWFE
- 2 Example: minimum wage
- 3 Aggregate ATET's**
- 4 Four estimators

Summarize ATET(g,t)'s

- How do the ATETs vary with the length of **exposure to the treatment**? (event study)
- How do the ATETs vary with **cohorts**? (does start treatment earlier matter?)
- How do the ATETs vary with **time**? (Good year vs. lousy year)
- Overall ATETs across time and cohorts

We can express the aggregations as a weighted mean of all ATETs

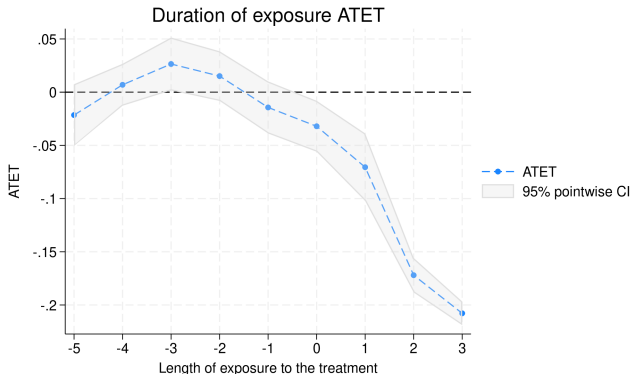
$$\theta = \sum_{g \in \mathbf{G}} \sum_{t=2}^T \underbrace{w(g, t)}_{\text{weight}} ATET(g, t)$$

Event study

- Let $e = t - g$ be the length of exposure to the treatment.

$$\theta(\mathbf{e}) = \sum_{g \in \mathbf{G}} \underbrace{\mathbb{1}(g + e \leq T) \Pr(G = g | g + e \leq T)}_{\text{proportions used to estimate ATET}(g, g+e)} \text{ATET}(g, g + e)$$

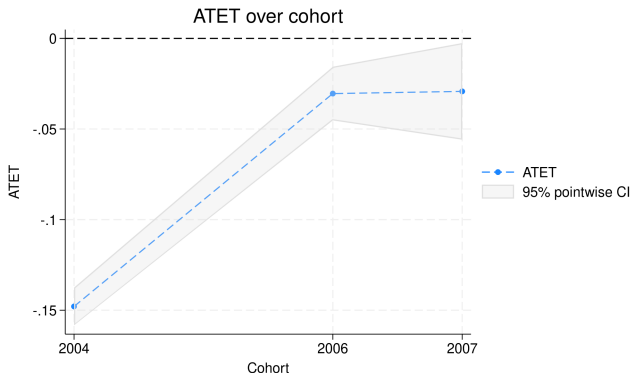
- estat aggregation, **dynamic** graph



ATEs over cohort

$$\theta(\mathbf{g}) = \sum_{t=g}^T \underbrace{\Pr(G = g | G = g, t \geq g)}_{\text{proportions used to estimate post-treatment ATE}(g,t)} \text{ATE}(g, t)$$

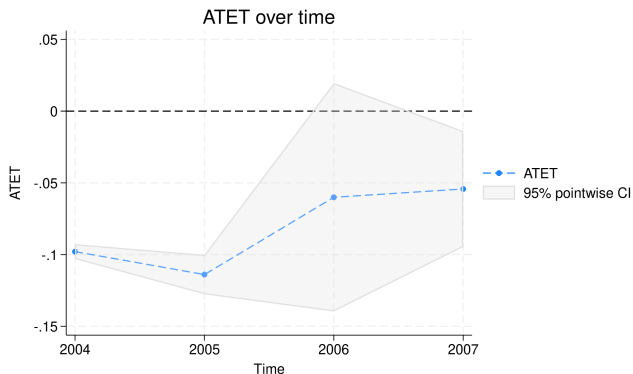
- estat aggregation, **cohort** graph



ATETs across time

$$\theta(\mathbf{t}) = \sum_{g \in \mathbf{G}} \mathbb{1}(t \geq g) \Pr(G = g | G \leq t) \text{ATET}(g, t)$$

- estat aggregation, **time** graph



Overall aggregations

- A single number to summarize ATET's

$$\theta = \frac{1}{\kappa} \sum_{g \in \mathbf{G}} \sum_{t=2}^T \mathbb{1}(t \geq g) \Pr(G = g | G \leq T) ATET(g, t)$$

- estat aggregation, **overall**

```
. estat aggreg, overall
```

```
Overall ATET
```

```
Number of obs = 15,988
```

```
(Std. err. adjusted for 29 clusters in state)
```

lemp	ATET	Robust std. err.	z	P> z	[95% conf. interval]	
treat (1 vs 0)	-.062811	.0256879	-2.45	0.014	-.1131582	-.0124637

Table of Contents

- 1 Heterogeneous treatment effects and TWFE
- 2 Example: minimum wage
- 3 Aggregate ATET's
- 4 Four estimators

Potential outcome framework

Some notations:

- $y_{i,t}(g)$ is unit i 's potential outcome at time t if it starts treatment at time g
- $y_{i,t}(\infty)$ is unit i 's potential outcome at time t if it is never treated
- G_i indicates unit i 's cohort (when the treatment starts), and it is one element in $\mathbf{G} = \{2, \dots, G, \infty\}$
- $y_{i,t}$ is unit i 's observed outcome at time t

$$y_{i,t} = \underbrace{\mathbb{1}(t < G_i)y_{i,t}(\infty)}_{\text{before treatment}} + \underbrace{\mathbb{1}(t \geq G_i)y_{i,t}(G_i)}_{\text{after treatment}}$$

Heterogeneous ATETs

$$ATET(g, t) = \mathbf{E}[y_{i,t}(g) - y_{i,t}(\infty) | G_i = g]$$

Remarks

- $ATET(g, t)$ is a function of two arguments: cohort g and time t
- $ATETs$ can be heterogeneous over cohorts, across time, across both time and cohorts
- Objective: consistently estimate $ATETs$ and summarize them

Key assumptions

- Observe I.I.D samples of $\{y_{i,t}, \mathbf{x}_{i,t}, \mathbf{z}_{i,t}, d_{i,t}\}_{i=1, t=1}^{i=N, t=T}$, where $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$ are covariates, and $d_{i,t}$ is observational level treatment indicator
- No one is treated in the first period
- No anticipation in pre-treatment periods $t < g$

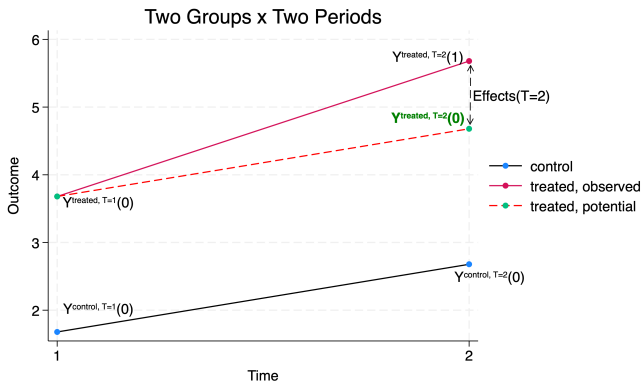
$$\mathbf{E}[y_{i,t}(g)|\mathbf{x}, G_i = g] = \mathbf{E}[y_{i,t}(\infty)|\mathbf{x}, G_i = g]$$

- Conditional parallel trend

$$\mathbf{E}[y_{i,t}(\infty) - y_{i,t-1}(\infty)|\mathbf{x}, G_i = g] = \mathbf{E}[y_{i,t}(\infty) - y_{i,t-1}(\infty)|\mathbf{x}, G_i = \infty]$$

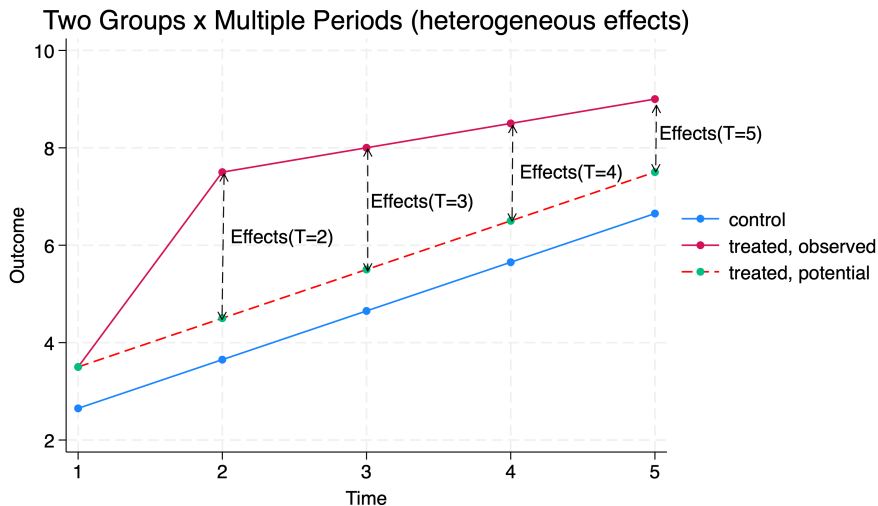
- Overlap assumption for propensity scores

Canonical DID



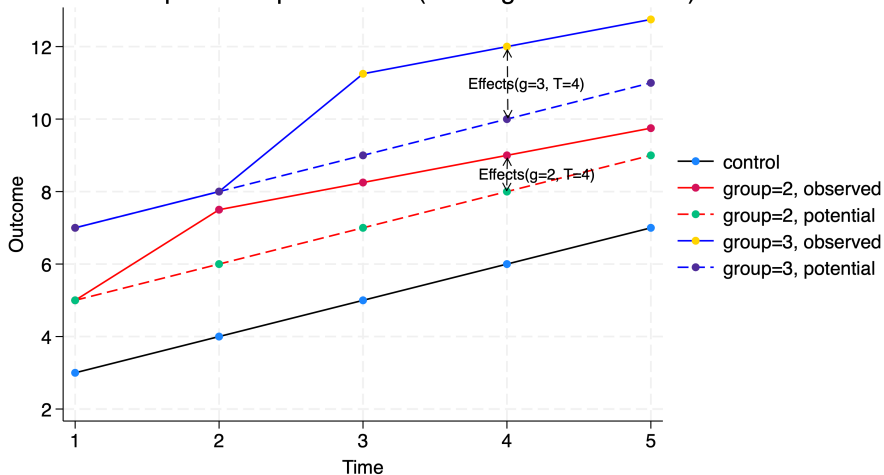
$$\begin{aligned}
 \text{Effects} &= [Y^{\text{treated}, T=2(1)} - Y^{\text{treated}, T=1(0)}] - [Y^{\text{control}, T=2(0)} - Y^{\text{control}, T=1(0)}] \\
 &= \underbrace{[Y^{\text{treated}, T=2} - Y^{\text{treated}, T=1}]}_{\text{treated differences}} - \underbrace{[Y^{\text{control}, T=2} - Y^{\text{control}, T=1}]}_{\text{untreated differences}}
 \end{aligned}$$

Heterogenous DID: benchmark time $g - 1$



ATE(g, t) is reduced to 2×2 DID

Three Groups x Multiple Periods (heterogeneous effects)



Regression adjustment (RA)

$$\begin{aligned} ATET(g, t) &= \mathbf{E} \left[\frac{K_g}{\mathbf{E}(K_g)} (y_t - y_{g-1} - m_{g,t}) \right] \\ &= \underbrace{\mathbf{E} [y_t - y_{g-1} | K_g = 1]}_{\text{treated differences}} - \underbrace{\mathbf{E} [m_{g,t}(\mathbf{x}) | K_g = 1]}_{\text{untreated differences}} \end{aligned}$$

where

- $K_g = \mathbb{1}(G_i = g)$ and $m_{g,t}(\mathbf{x}) = \mathbf{E}(y_t - y_{g-1} | \mathbf{x}, G_i = \infty)$
- It is 2×2 difference-in-differences (two groups \times two periods)
- Benchmark time: one period before treatment ($g - 1$)
- Benchmark group: never-treated group ($G_i = \infty$)

In Stata, we type

- `xthdidregress ra $\underbrace{(y \quad \mathbf{x})}_{m_{g,t}(\mathbf{x})} (d), \text{group}(id)$`

Inverse probability weighting (IPW)

$$ATET(g, t) = \mathbf{E} \left[\left(\frac{K_g}{\mathbf{E}(K_g)} - \frac{\frac{p_g(\mathbf{z})K_\infty}{1-p_g(\mathbf{z})}}{\mathbf{E} \left[\frac{p_g(\mathbf{z})K_\infty}{1-p_g(\mathbf{z})} \right]} \right) (Y_t - Y_{g-1}) \right]$$

where

- $p_g(\mathbf{z}) = \Pr(K_g = 1 | \mathbf{z}, K_g + K_\infty = 1) = \frac{\Pr(K_g=1|\mathbf{z})}{\Pr(K_g+K_\infty=1|\mathbf{z})}$
- $\frac{p_g(\mathbf{z})}{1-p_g(\mathbf{z})} = \frac{\Pr(K_g=1|\mathbf{z})}{\Pr(K_\infty=1|\mathbf{z})}$. Thus, in the benchmark group (never treated), attach more weights to observations that are more probably observed in the cohort g
- We estimate $p_g(\mathbf{z})$ by a logit regression

In Stata, we type

- `xthdidregress ipw (y) $\underbrace{(d \ \mathbf{z})}_{p_g(\mathbf{z})}$, group(id)`

Augmented inverse probability weighting (AIPW)

$$ATE_T(g, t) = \mathbf{E} \left[\underbrace{\left(\frac{K_g}{\mathbf{E}(K_g)} - \frac{\frac{\rho_g(\mathbf{z})K_\infty}{1-\rho_g(\mathbf{z})}}{\mathbf{E} \left[\frac{\rho_g(\mathbf{z})K_\infty}{1-\rho_g(\mathbf{z})} \right]} \right)}_{\text{IPW}} (Y_t - Y_{g-1} - \overbrace{m_{g,t}(\mathbf{x})}^{\text{augmented term}}) \right]$$

- AIPW is **doubly robust**: only one of the outcome model or the treatment model needs to be correctly specified

In Stata, we type

- `xthdidregress aipw $\underbrace{(y \ x)}_{m_{g,t}(\mathbf{x})} \underbrace{(d \ \mathbf{z})}_{\rho_g(\mathbf{z})}, \text{group}(\text{id})$`

TWFE

Traditional TWFE:

$$y_{it} = \theta_t + \eta_i + d_{it}\alpha + v_{it}$$

TWFE in Wooldridge (2021):

$$y_{it} = \theta_t + \eta_i + \sum_{g \in \mathbf{G}} \sum_{s=g}^T \alpha_{g,t} \mathbb{1}(G_i = g, t = s) + v_{it}$$

With covariates \mathbf{x} , add full interactions with θ_t , η_i , and $\mathbb{1}(G_i = g, t = s)$.

In Stata, we type

- `xthdidregress twfe` $\underbrace{(y \ x)}_{\text{TWFE outcome}}$ (d), group(id)

Summary

- We illustrate the heterogeneous treatment effects across time and groups
- Traditional TWFE is inconsistent
- `xthdidregress` and `hdidregress` implements four estimators to remedy the heterogeneity issues
- `estat atetplot` visualize the heterogeneity at the group-time level (zoom in)
- `estat aggregation` estimate and visualize the heterogeneity at a higher level (zoom out)

References

- Callaway, B., and P. H. Sant'Anna. 2021. Difference-in-Differences with multiple time periods. *Journal of Econometrics* 225: 200–230.
- Wooldridge, J. M. 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. URL <https://ssrn.com/abstract=3931952>.

Appendix

Contents

1	xthdidregress	2
2	estat aggregation, dynamic	3
3	estat aggregation, cohort	3
4	estat aggregation, time	4

1 xthdidregress

```
. global covars i.region pop medinc white hs pov c.pop#c.pop c.medinc#c.medinc
. xthdidregress aipw (lemp $covars) (treat $covars), group(state)
note: variable _did_cohort, containing cohort indicators formed by treatment
      variable treat and group variable state, was added to the dataset.
```

Computing ATET for each cohort and time:

```
Cohort 2004 (6): ..... done
Cohort 2006 (6): ..... done
Cohort 2007 (6): ..... done
```

Treatment and time information

```
Time variable: year
Time interval: 2001 to 2007
Control:      _did_cohort = 0
Treatment:    _did_cohort > 0
```

	_did_cohort
Number of cohorts	4
Number of obs	
Never treated	9639
2004	700
2006	1561
2007	4088

```
Heterogeneous-treatment-effects regression      Number of obs   = 15,988
                                                    Number of panels =    29
```

```
Estimator:      Augmented IPW
Panel variable:  countyreal
Treatment level: state
Control group:   Never treated
```

(Std. err. adjusted for 29 clusters in state)

Cohort	ATET	Robust			[95% conf. interval]	
		std. err.	z	P> z		
2004						
year						
2002	.0672458	.0061125	11.00	0.000	.0552655	.079226
2003	.0266718	.0122508	2.18	0.029	.0026608	.0506829
2004	-.0979371	.002649	-36.97	0.000	-.103129	-.0927451
2005	-.1139248	.0070092	-16.25	0.000	-.1276627	-.1001869
2006	-.1719979	.0082852	-20.76	0.000	-.1882366	-.1557592
2007	-.2078132	.0056814	-36.58	0.000	-.2189485	-.196678
2006						
year						
2002	-.0186685	.0105915	-1.76	0.078	-.0394274	.0020904
2003	.056737	.0181748	3.12	0.002	.0211151	.0923589
2004	.0212315	.0363779	0.58	0.559	-.0500679	.092531
2005	.0319911	.0158191	2.02	0.043	.0009863	.0629959
2006	-.009851	.0117487	-0.84	0.402	-.0328781	.013176
2007	-.0510452	.0092241	-5.53	0.000	-.069124	-.0329664

2007							
year							
2002	-.0215125	.014779	-1.46	0.145	-.0504788	.0074538	
2003	.0167167	.0132905	1.26	0.208	-.0093322	.0427655	
2004	.0149363	.0133763	1.12	0.264	-.0112809	.0411534	
2005	.0038453	.0092391	0.42	0.677	-.014263	.0219537	
2006	-.0390546	.0114977	-3.40	0.001	-.0615896	-.0165196	
2007	-.0292338	.0136042	-2.15	0.032	-.0558976	-.00257	

Note: ATET computed using covariates.

2 estat aggregation, dynamic

```
. estat aggreg, dynamic graph(name(d1))
```

Duration of exposure ATET

Number of obs = 15,988

(Std. err. adjusted for 29 clusters in state)

Exposure	Robust		z	P> z	[95% conf. interval]	
	ATET	std. err.				
-5	-.0215125	.014779	-1.46	0.145	-.0504788	.0074538
-4	.0069386	.0100519	0.69	0.490	-.0127627	.0266399
-3	.0264872	.0126915	2.09	0.037	.0016122	.0513621
-2	.0151101	.0118987	1.27	0.204	-.0082109	.0384311
-1	-.0143403	.0124878	-1.15	0.251	-.0388159	.0101353
0	-.032043	.0122219	-2.62	0.009	-.0559975	-.0080885
1	-.0705126	.0161956	-4.35	0.000	-.1022553	-.0387699
2	-.1719979	.0082852	-20.76	0.000	-.1882366	-.1557592
3	-.2078132	.0056814	-36.58	0.000	-.2189485	-.196678

Note: Exposure is the number of periods since the first treatment time.

3 estat aggregation, cohort

```
. estat aggreg, cohort graph(name(c1))
```

ATET over cohort

Number of obs = 15,988

(Std. err. adjusted for 29 clusters in state)

Cohort	Robust		z	P> z	[95% conf. interval]	
	ATET	std. err.				
2004	-.1479183	.0053113	-27.85	0.000	-.1583283	-.1375082
2006	-.0304481	.0075561	-4.03	0.000	-.0452578	-.0156384
2007	-.0292338	.0136042	-2.15	0.032	-.0558976	-.00257

4 estat aggregation, time

```
. estat aggreg, time graph(name(t1))
```

ATET over time

Number of obs = 15,988

(Std. err. adjusted for 29 clusters in state)

Time	ATET	Robust std. err.	z	P> z	[95% conf. interval]	
2004	-.0979371	.002649	-36.97	0.000	-.103129	-.0927451
2005	-.1139248	.0070092	-16.25	0.000	-.1276627	-.1001869
2006	-.0600513	.0406199	-1.48	0.139	-.1396648	.0195622
2007	-.0542855	.0206191	-2.63	0.008	-.0946981	-.0138728