Heterogeneous differences-in-differences in Stata

Eduardo Garcia Echeverri

Colombian Stata Conference, 2023

Outline

1. Heterogeneity in treatment effects

2. Model setup

3. Estimation in Stata
   - Regression adjusted
   - Inverse-probability weighting
   - Augmented inverse-probability weighting
   - Extended two-way fixed effects

4. Aggregation of treatment effects

5. Conclusion

Why heterogeneous treatment effects?

Classic differences-in-differences: Treatment effects are obtained by estimating

$$y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_t + \gamma_g + \varepsilon_{it}$$

y_{it}: **Outcome** of interest.

D_{it}: Binary **treatment**.

γ_t: **Time** fixed effects

γ_g: **Group** fixed effects. Treatment happens at the **group level**.

$\beta_1 =$ ATT (average **treatment effect** on the treated)

Why heterogeneous treatment effects?

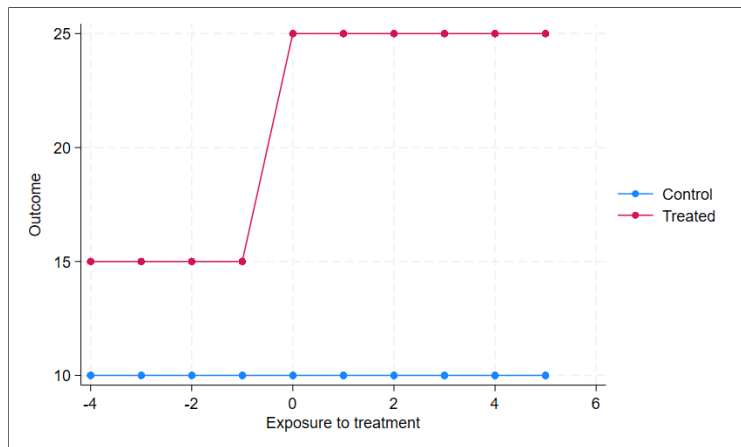**Classic differences-in-differences:** Treatment effects are obtained by estimating

$$y_{it} = \beta_0 + \beta_1 D_{it} + \gamma_t + \gamma_g + \varepsilon_{it}$$

Implicit assumptions:

- ATT is the same **irrespective of when** unit is treated.
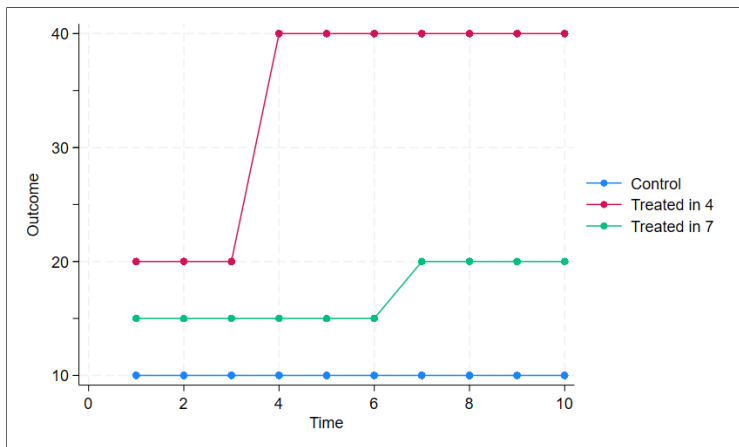- ATT is **constant** after unit is treated.

We are assuming **homogeneous treatment effects**.

Classic differences-in-differences

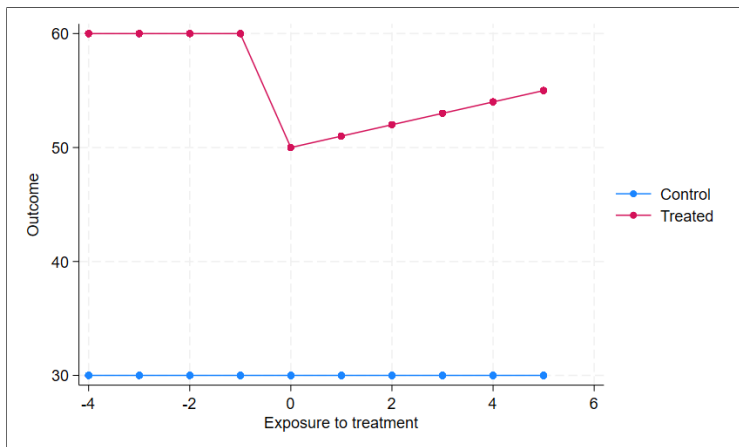$ATT = DID =$ difference in treated$-$difference in control $= 10-0 = 10$

Group (cohort) heterogeneity

$$ATT_{\text{red}} = 20 \qquad ATT_{\text{green}} = 5$$

Time heterogeneity

$$ATT_{t=0} = -10 \qquad ATT_{t=5} = -5$$

Time-cohort heterogeneity
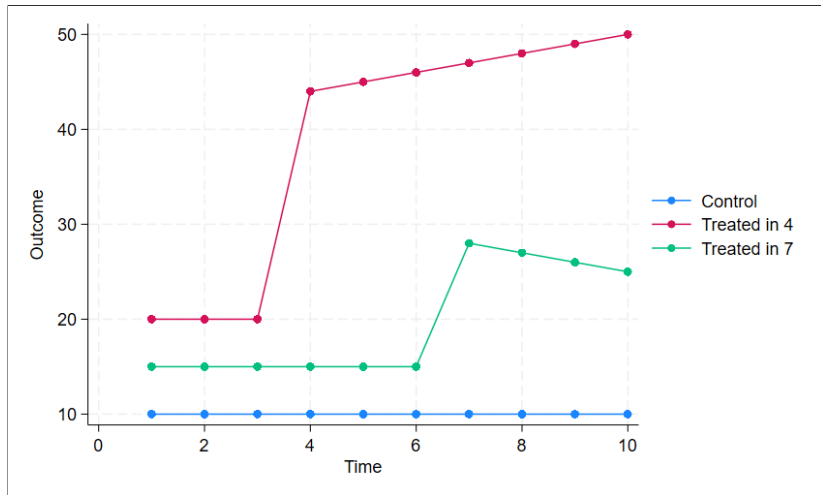
Heterogeneous DID

A **growing literature** has emerged to **estimate heterogeneous ATTs**:

- Callaway & Sant'Anna (2021), Wooldridge (2021), Chaisemartin and D'Haultfoeuille (2020)...

... and to **diagnose/understand** treatment effect heterogeneity:

- Borusyak, Jaravel, and Spiess (2018), Goodman-Bacon (2021)...

Many of these features have been **incorporated into Stata 18**.

Outline

The heterogeneous DID model

Panel (or **repeated cross-sectional** data) with $\{1, 2 \ldots, T\}$ periods:

   t: a specific **time period**.

D_{it}: 1 if **unit is treated** in period t, 0 otherwise.
   - **Irreversible treatment**: Once treated, unit remains treated.
   - No unit is treated at $t = 1$.

G_i: **Group** of unit i. When did i **first receive** treatment?
   - $G_i = 5$ if unit i **first received treatment** in $t = 5$.
   - $G_i = \infty$ if unit i **never received treatment**.

X_i: Time-invariant **controls** for unit i.

Potential and observed outcomes

$Y_{i,t}(0)$: **potential outcome** of unit i at time t if it is **never treated**.
- If $G_i = \infty$, then $Y_{i,t}(0)$ is **observed**.
- If $G_i \neq \infty$, then $Y_{i,t}(0)$ is **unobserved**.

$Y_{i,t}(g)$: **potential outcome** of unit i at time t if it had been **first treated at time** g.
- If $G_i = g$, then $Y_{i,t}(g)$ is **observed**.
- If $G_i \neq g$, then $Y_{i,t}(g)$ is **unobserved**.

$Y_{i,t}$: **observed outcome** in the data.
- $Y_{i,t} = Y_{i,t}(0)$ when $G_i = \infty$.
- $Y_{i,t} = Y_{i,t}(g)$ when $G_i = g$.

Heterogeneous Treatment Effects

Group-time average treatment effects on the treated:

$$ATT(g, t) = \mathbb{E}\big[Y_{i,t}(g) - Y_{i,t}(0)|G_i = g\big]$$

In **group** g and **time** t, what was the **average effect of being treated?**

Up to $(T-1)^2$ different ATTs $\Rightarrow$ rich **heterogeneity!**

Problem: ATT are based on **unobservables** $\Rightarrow$ **Assumptions**

Assumption 1: No anticipatory effect

Before treatment happens (for $t < g$),

$$\mathbb{E}\big[Y_{i,t}(g)|X, G_i = g\big] = \mathbb{E}\big[Y_{i,t}(0)|X, G_i = g\big]$$

Outcome **doesn't respond in anticipation** to the treatment.

Anticipatory effects **bias** ATT estimation. For example:

y_{it}: **Sales** by firms.

D_{it}: 1 if county **reduces sales tax**, 0 otherwise.

- Firms may **sell less today** to grow inventories and **sell more later**.

Anticipatory effects bias DID

$$DID = \color{red}{(30 - 15)} - \color{blue}{0} = \color{red}{30} - \color{blue}{15} = 15 \neq 10 = ATT$$

Assumption 2: Parallel trends with never-treated

After treatment happens $(t \geq g)$,

$$\mathbb{E}\big[Y_{i,t}(0) - Y_{i,t-1}(0)|X, G_i = g\big] = \mathbb{E}\big[Y_{i,t}(0) - Y_{i,t-1}(0)|X, G_i = \infty\big]$$

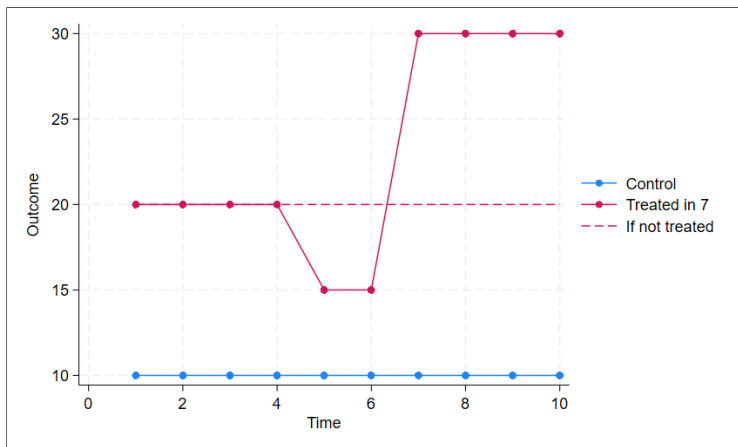If group **had not been treated**, outcome would move as in the **never treated group**.

Violations of this assumption **bias** ATT estimation. For example:

y_{it}: **Sales** by firms.

D_{it}: 1 if county **reduces sales tax**, 0 otherwise.

- **After** decreasing the sales tax, some treated counties **increase the corporate income tax**.

Non-parallel trends effects bias DID
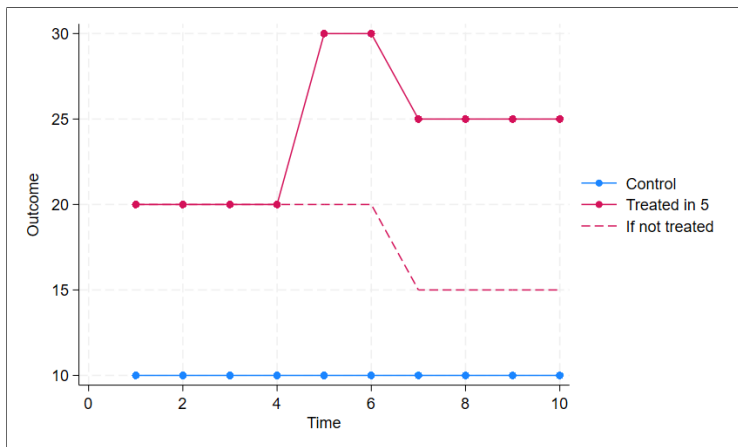
$$DID = (25 - 20) - 0 = 5 - 0 = 5 \neq 10 = ATT$$

Identification – Callaway, Sant'Anna (2021)

Theorem: Given some **technical conditions**, if **assumptions 1 and 2** hold

$$\Rightarrow ATT(g,t) \text{ can be } \textbf{estimated from the data}.$$

Result also holds if **parallel trends** with **not-yet treated groups**.

Option `controlgroup` lets you **choose the control group**

Outline

Commands hdidregress and xthdidregress

Estimate ATTs that vary over group/cohort and over time:

- hdidregress for **cross-sectional data**
- xthdidregress for **panel data**.

Both commands come with **four estimators**:

- Callaway, Sant'Anna (2021):
  - **Regression adjusted**
  - Inverse-**probability weighting**
  - **Augmented** inverse-probability weighting
- Wooldridge (2021):
  - **Extended** two-way fixed effects

Regression adjusted estimator

Syntax:

xthdidregress ra (*ovar* [*omvarlist*]) (*tvar*) [*if*] [*in*] [*weight*],
<u>group</u>(*groupvar*) [*options*]

ovar: **continuous outcome** of interest

omvarlist: **covariates** in the **outcome model**

tvar: binary **treatment**

groupvar: categorical variable indicating **group level at which treatment occurs. Required.**

The RA estimator

$$ATT(g, t) = \mathbb{E}\left[\frac{G_g}{\mathbb{E}\left[G_g\right]} \left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X) \right) \right]$$

$m_{g,t}^{nev}(X)$: **Difference in the control group** conditional on X.

- $m_{g,t}^{nev}(X) = \mathbb{E}\left[Y_t - Y_{g-1} | X, G_i = \infty\right]$

The term in orange is the **difference in the differences** between group g and the control group.

Heuristically

Algorithm:

1. keep if time is t or $g-1$
2. keep if cohort is g or C
3. generate $\Delta Y = Y_t - Y_{g-1}$
4. regress ΔY on X for the group C and predict $\hat{m}_{g,t}^{nev}(X)$
5. generate $\widehat{TE} = \Delta Y - \hat{m}_{g,t}^{nev}(X)$
6. summarize $\widehat{TE}$ if cohort is g
7. Repeat for each g and t.

Influence function approach in Callaway & Sant'Anna (2021).

Example: the RA estimator in Stata

Question: How is the **number of registrations of a dog breed** in the American Kennel Club affected by that **dog breed being the protagonist** in a movie?

Data

```
. webuse akc, clear
(Fictional dog breed and AKC registration data)


. describe

Contains data from https://www.stata-press.com/data/r18/akc.dta
 Observations:          1,410              Fictional dog breed and AKC registration data
    Variables:              5              1 Feb 2023 14:23

Variable      Storage   Display    Value
   name        type     format     label      Variable label

year            int     %10.0g                Year
breed           int     %34.0g     Breed      Dog breed
movie           byte    %9.0g                 Was a movie protagonist
best            byte    %9.0g                 Won best in show in past 10 years
registered      int     %9.0g                 Number of AKC registrations

Sorted by: breed
```

Data

Staggered treatment

```
. tabulate year movie

               Was a movie
               protagonist
    Year          0        1         Total

    2031        141        0           141
    2032        141        0           141
    2033        141        0           141
    2034        137        4           141
    2035        137        4           141
    2036        134        7           141
    2037        119       22           141
    2038        119       22           141
    2039        119       22           141
    2040        119       22           141

    Total     1,307      103         1,410
```

Output 1

```
. xtset breed year

Panel variable: breed (strongly balanced)
 Time variable: year, 2031 to 2040
         Delta: 1 unit

.
. xthdidregress ra (registered best) (movie), group(breed)
note: variable _did_cohort, containing cohort indicators formed by treatment variable movie and group variable breed, was added to the dataset.

Computing ATET for each cohort and time:
Cohort 2034 (9): ......... done
Cohort 2036 (9): ......... done
Cohort 2037 (9): ......... done

Treatment and time information

Time variable: year
Time interval: 2031 to 2040
Control:       _did_cohort = 0
Treatment:     _did_cohort > 0
```

	_did_cohort
Number of cohorts	4
Number of obs	
Never treated	1190
2034	40
2036	30
2037	150

Output 2

```
Heterogeneous-treatment-effects regression        Number of obs     =  1,410
                                                  Number of panels  =    141

Estimator:        Regression adjustment
Panel variable:   breed
Treatment level:  breed
Control group:    Never treated

                             (Std. err. adjusted for 141 clusters in breed)
```

Cohort	ATET	Robust std. err.	z	P>|z|	[95% conf. interval]	
2034						
year						
2032	-254.8927	266.1024	-0.96	0.338	-776.4439	266.6584
2033	-257.5329	217.9389	-1.18	0.237	-684.6852	169.6194
2034	701.1318	127.0935	5.52	0.000	452.0331	950.2304
2035	1099.044	282.0704	3.90	0.000	546.196	1651.892
2036	1367.632	225.8702	6.05	0.000	924.9343	1810.329
2037	2008.294	237.2396	8.47	0.000	1543.313	2473.275
2038	2472.624	278.2949	8.88	0.000	1927.176	3018.072
2039	2689.615	504.3324	5.33	0.000	1701.142	3678.088
2040	3110.97	568.916	5.47	0.000	1995.915	4226.025
2036						
year						
2032	216.0259	122.9107	1.76	0.079	-24.87472	456.9265
2033	-172.5154	372.0776	-0.46	0.643	-901.7741	556.7433
2034	-218.0495	504.5267	-0.43	0.666	-1206.904	770.8045
2035	621.033	156.1306	3.98	0.000	315.0227	927.0434
2036	999.0781	180.1055	5.55	0.000	646.0779	1352.078
2037	1003.333	250.5916	4.00	0.000	512.1829	1494.484
2038	1556.669	451.6914	3.45	0.001	671.3697	2441.967
2039	2590.674	662.6979	3.91	0.000	1291.81	3889.538
2040	2225.712	486.9917	4.57	0.000	1271.225	3180.198

Output 3

2037						
year						
2032	-114.582	160.0972	-0.72	0.474	-428.3668	199.2028
2033	-127.9856	183.3941	-0.70	0.485	-487.4315	231.4603
2034	33.40901	168.0312	0.20	0.842	-295.9262	362.7442
2035	130.3495	166.2261	0.78	0.433	-195.4477	456.1468
2036	-10.48288	167.5059	-0.06	0.950	-338.7884	317.8226
2037	1717.016	268.5592	6.39	0.000	1190.65	2243.383
2038	2086.798	278.0215	7.51	0.000	1541.886	2631.71
2039	2473.611	268.186	9.22	0.000	1947.976	2999.246
2040	2835.117	378.6699	7.49	0.000	2092.938	3577.296

Note: ATET computed using covariates.

Graphical representation: `estat atetplot`
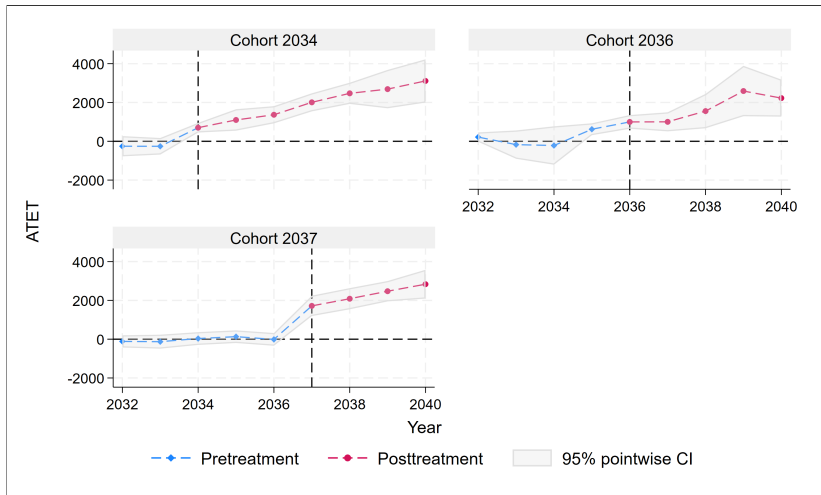
Simultaneous confidence intervals: `estat atetplot, sci`

Heterogeneous differences-in-differences in Stata
└─ Estimation in Stata
  └─ Inverse-probability weighting
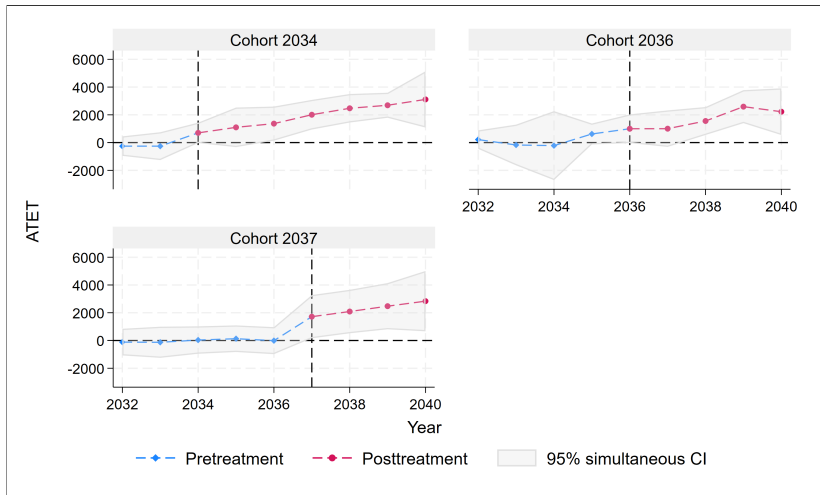
Inverse-probability weighting estimator

Syntax:

xthdidregress ipw (*ovar*) (*tvar* [*tmvarlist*]) [*if*] [*in*] [*weight*],
<u>group</u>(*groupvar*) [*options*]

ovar: **continuous outcome** of interest

tmvarlist: **covariates** in the **treatment model**

tvar: binary **treatment**

groupvar: categorical variable indicating **group level at which treatment occurs. Required.**

Heterogeneous differences-in-differences in Stata
└─Estimation in Stata
　└─Inverse-probability weighting

The IPW estimator

$$ATT(g, t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\frac{p_g(X)}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1}\right)\right]$$

$p_g(X)$: **Probability of being in group g** given X and given that observation is either in g or C.

  • Generalized **propensity score**.

The term in orange is the **inverse-probability weight**.

Heuristically

Algorithm:

1. keep if time is t or $g-1$
2. keep if cohort is g or C
3. generate $\Delta Y = Y_t - Y_{g-1}$
4. logit $D_{i,t}$ on X and predict $\hat{p}_g(X)$
5. generate inverse-probability weights
6. summarize ΔY if cohort is g using weights in 5.
7. Repeat for each g and t.

Influence function approach in Callaway & Sant'Anna (2021).

Heterogeneous differences-in-differences in Stata
└─ Estimation in Stata
  └─ Augmented inverse-probability weighting

Augmented inverse-probability weighting estimator

Syntax:

xthdidregress aipw (*ovar* [*omvarlist*]) (*tvar* [*tmvarlist*]) [*if*] [*in*] [*weight*],
<u>group</u>(*groupvar*) [*options*]

ovar: **continuous outcome** of interest

omvarlist: **covariates** in the **outcome model**

tmvarlist: **covariates** in the **treatment model**

tvar: binary **treatment**

groupvar: categorical variable indicating **group level at which treatment occurs. Required.**

The AIPW estimator

$$ATT(g, t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\frac{p_g(X)}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-1} - m_{g,t}^{nev}(X)\right)\right]$$

$p_g(X)$: **Probability of being in group g** given X and given that observation is either in g or C.

$m_{g,t}^{nev}(X)$: **Difference in the control group** conditional on X.

Inverse-probability weight in orange. **Augmented** term in violet

Doubly robust

Heterogeneous differences-in-differences in Stata
└─Estimation in Stata
   └─Augmented inverse-probability weighting

Heuristically

Algorithm:

1. keep if time is t or $g-1$
2. keep if cohort is g or C
3. generate $\Delta Y = Y_t - Y_{g-1}$
4. logit $D_{i,t}$ on X and predict $\hat{p}_g(X)$
5. generate inverse-probability weights
6. regress ΔY on X for the group C and predict $\hat{m}_{g,t}^{nev}(X)$
7. generate $\widehat{TE} = \Delta Y - \hat{m}_{g,t}^{nev}(X)$
8. summarize $\widehat{TE}$ if cohort is g using weights in 5
9. Repeat for each g and t.

Influence function approach in Callaway & Sant'Anna (2021).

Heterogeneous differences-in-differences in Stata
└─ Estimation in Stata
   └─ Extended two-way fixed effects

Extended two-way fixed effects estimator

Syntax:

xthdidregress twfe (*ovar* [*omvarlist*]) (*tvar*) [*if*] [*in*] [*weight*],
<u>group</u>(*groupvar*) [*options*]

ovar: **continuous outcome** of interest

omvarlist: **covariates** in the **outcome model**

tvar: binary **treatment**

groupvar: categorical variable indicating **group level at which treatment occurs. Required.**

The TWFE estimator

Consider the **extended two-way fixed effects regression**:

$$Y_{it'} = \eta + \sum_{g=q}^{T} \alpha_g G_{ig} + \sum_{t=q}^{T} \gamma_t f_t + \sum_{g=q}^{T} \sum_{t=q}^{T} \tau_{g,t} D_{it} G_{ig} f_t + \varepsilon_{it'}$$

q: **first treatment** period

f_t: 1 if $t' = t$, 0 otherwise.

$\tau_{g,t} = ATT(g,t)$

Remarks:

- **Covariates** would enter **fully interacted** in the model.

Outline

1. Heterogeneity in treatment effects

2. Model setup

3. Estimation in Stata
   - Regression adjusted
   - Inverse-probability weighting
   - Augmented inverse-probability weighting
   - Extended two-way fixed effects

4. **Aggregation of treatment effects**

5. Conclusion

Aggregating treatment effects

You might be interested in exploring **heterogeneity just by**:

- **Cohort**
- **Time**
- **Exposure** to treatment (event studies)
- Even **no heterogeneity** at all

Some **post-estimation tools** come handy in this case.

Suppose you've just fitted **a heterogeneous DID model**:

xthdidregress ra (registered best) (movie), group(breed)

Overall aggregation – `estat aggregation, overall`

```
Overall ATET                                              Number of obs = 1,410

                                   (Std. err. adjusted for 141 clusters in breed)

                          Robust
  registered      ATET   std. err.      z    P>|z|     [95% conf. interval]

       movie
    (1 vs 0)   2093.318   122.5752   17.08   0.000     1853.075    2333.561
```

Aggregation by cohort

```
. estat aggregation, cohort graph sci

ATET over cohort                          Number of obs = 1,410
                                          Replications  =   999

                   (Std. err. adjusted for 141 clusters in breed)
```
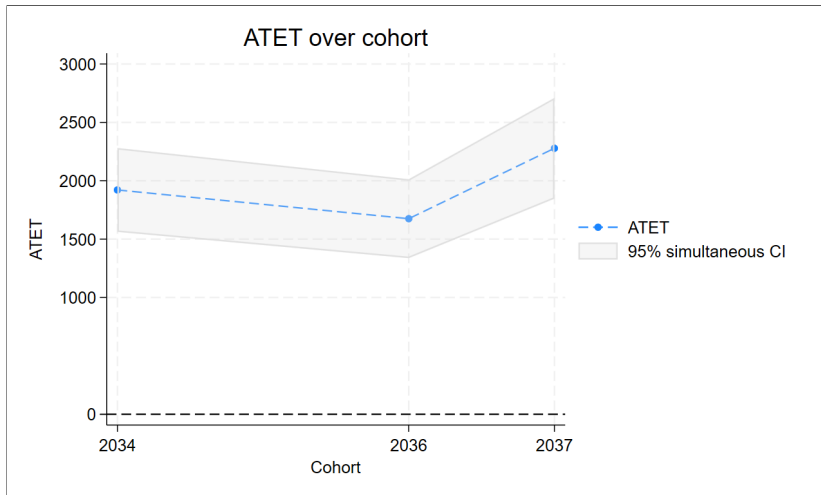
	Observed	Bootstrap	Simultaneous	
Cohort	ATET	std. err.	[95% conf. interval]	
2034	1921.33	135.3652	1561.16	2281.5
2036	1675.093	120.9415	1353.3	1996.886
2037	2278.136	175.554	1811.034	2745.238

```
Note: Simultaneous confidence intervals provide inference
      across all aggregations simultaneously.
```

Aggregation by cohort – Graph

Aggregation by time

```
. estat aggregation, time graph sci

ATET over time                              Number of obs = 1,410
                                            Replications  =   999

                    (Std. err. adjusted for 141 clusters in breed)

                        Observed   Bootstrap        Simultaneous
            Time            ATET    std. err.    [95% conf. interval]

            2034        701.1318    120.8096     388.4994    1013.764
            2035        1099.044    263.6189     416.8482     1781.24
            2036          1209.68   172.2839     763.8421    1655.518
            2037        1672.655    205.9913     1139.589    2205.722
            2038        2084.658    216.9237     1523.301    2646.015
            2039        2528.847    219.2507     1961.468    3096.227
            2040        2802.171    287.8548     2057.258    3547.085

Note: Simultaneous confidence intervals provide inference
      across all aggregations simultaneously.
```
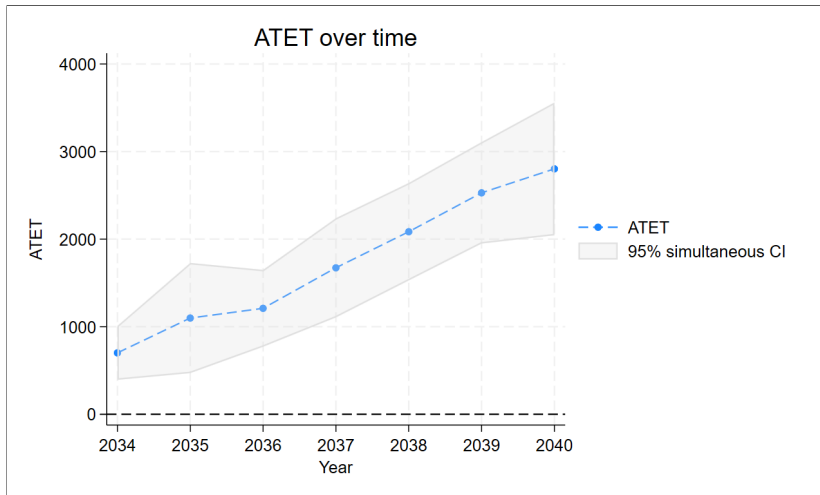
Aggregation by time – Graph

Aggregation by exposure

```
.  estat aggregation, dynamic graph

Duration of exposure ATET                              Number of obs = 1,410

                                   (Std. err. adjusted for 141 clusters in breed)

                          Robust
   Exposure       ATET    std. err.      z     P>|z|      [95% conf. interval]

         -5    -114.582   160.0972     -0.72    0.474     -428.3668    199.2028
         -4   -70.65034   156.3185     -0.45    0.651      -377.029    235.7283
         -3   -.9117242   153.0999     -0.01    0.995      -300.982    299.1585
         -2    12.79653   144.8216      0.09    0.930     -271.0486    296.6417
         -1    30.71473   132.8508      0.23    0.817      -229.668    291.0975
          0    1434.409   206.3277      6.95    0.000      1030.014    1838.804
          1    1759.461   224.0229      7.85    0.000      1320.385    2198.538
          2    2147.486    221.903      9.68    0.000      1712.564    2582.408
          3    2651.452   284.8928      9.31    0.000      2093.073    3209.832
          4    2366.805   267.4253      8.85    0.000      1842.661    2890.949
          5    2689.615   504.3324      5.33    0.000      1701.142    3678.088
          6     3110.97    568.916      5.47    0.000      1995.915    4226.025

Note: Exposure is the number of periods since the first treatment time.
```
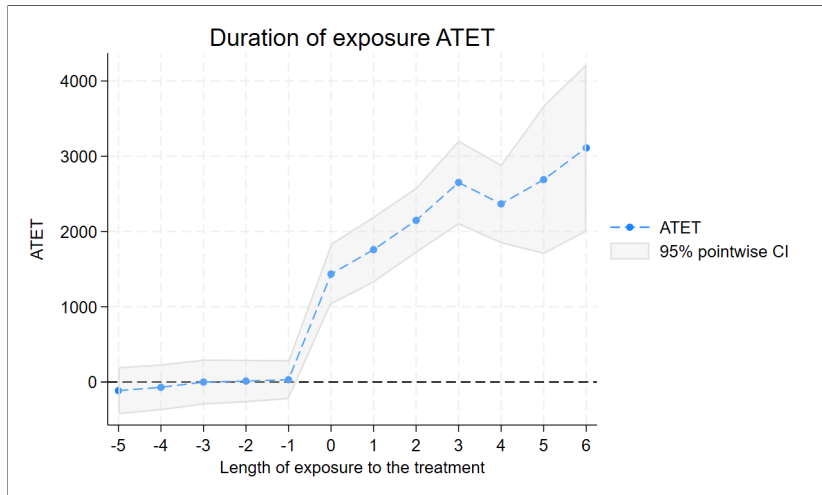
Aggregation by exposure – Graph

Outline

1 Heterogeneity in treatment effects

2 Model setup

3 Estimation in Stata
- Regression adjusted
- Inverse-probability weighting
- Augmented inverse-probability weighting
- Extended two-way fixed effects

4 Aggregation of treatment effects

5 Conclusion

Conclusion

1. **Heterogeneous DID is a powerful tool** to better understand treatment effects.

2. **Easy to implement** in Stata 18:
   - xthdidregress for **panel data**
   - hdidregress for **repeated cross section**
   - Results displayed as **tables or graphs**.

3. **Treatment effects** can be **aggregated by**:
   - Cohort,
   - Time
   - Exposure to treatment
   - Overall

Thank you!

Questions?