

# Difference-in-differences models for causal inference in Stata

Eduardo Garcia Echeverri

APSA Annual Meeting and Exhibition, Vancouver, 2025

# Differences-in-differences (DID)

DID is one of the most venerable **methods** for **causal inference**.  
**Easy to implement** in **Stata**.

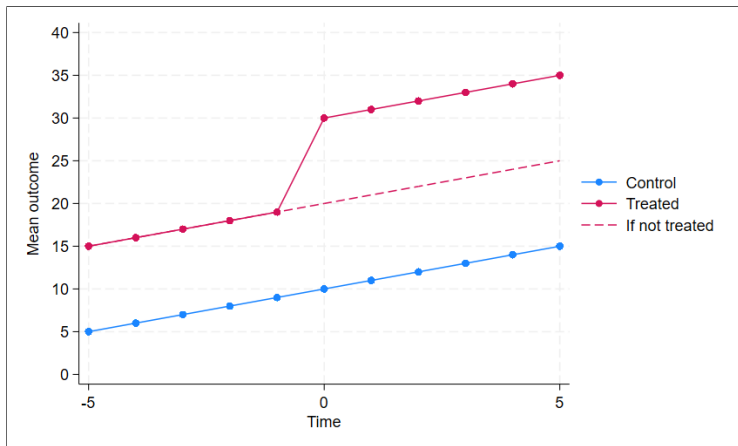
## Examples:

- Effects of **minimum wage** on **employment**.
- Do **tariffs** help or hinder **economic growth**?
- By **how much** does the presence of police **reduce crime**?

## DID compares

1. Average **difference over time** in outcome for the **treated**
2. Average **difference over time** in outcome for the **controls**

## In a nutshell...



$$ATT = DID = \text{difference treated} - \text{difference control} = 20 - 10 = 10$$

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion

# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion

# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion

## Example: *Card and Krueger (1994)*

In 1992, **New Jersey increased** the **minimum hourly wage**:

- \$4.25 → \$5.05

To study **effect on employment**, the authors surveyed a **panel**:

- **331 fast food restaurants** in New Jersey:
  - approx. 30% of workers were paid minimum wage.
- February 15th - March 4th (**before raise**)
- November 5th - December 31st (**after raise**)

As **controls**, the authors simultaneously surveyed a **second panel**:

- **79 fast food restaurants** in **eastern Pennsylvania**:
  - \$4.25 → \$4.25

## Example: *Card and Krueger (1994)*

**Table 1: Average number of employees at fast food restaurants**

	<i>New Jersey</i>	<i>Pennsylvania</i>	<b>Difference</b>
<i>Before increase</i>	20.43	23.38	2.95
<i>After increase</i>	20.90	21.10	0.20
<b>Difference</b>	0.47	-2.28	<b>2.75</b>

$\theta = \text{DID estimator of ATT} = \uparrow 2.75 \text{ employees}$



# DID estimator – Panel data

**Two-way fixed effects (FE) regression:**

$$y_{it} = \theta D_{it} + u_i + \nu_t + \varepsilon_{it}$$

where,

$y_{it}$ : **Employment** at restaurant  $i$  at time  $t$ .

$D_{it}$ : **Treatment** dummy.

- 1 if minimum wage is \$5.05;
- 0 otherwise.

$u_i$ : Restaurant  $i$  **fixed effect**.

$\nu_t$ : **Time fixed effect**.

$\theta$ : **ATT**

# DID estimator – Panel data

**The two-way FE regression**

$$y_{it} = \theta D_{it} + u_i + \nu_t + \varepsilon_{it}$$

is estimated as a **one-way FE regression with time dummies**

$$y_{it} = \underset{(1.34)}{2.75} D_{it} - \underset{(1.2)}{2.3} + u_i + \varepsilon_{it}$$

and **clustered standard errors** at the restaurant level.

# General DID model with controls – Panel data

**Generalized model** to account for more than two states/periods and to include **controls**:

$$y_{it} = \theta D_{it} + X'_{it}\beta + u_i + \nu_t + \varepsilon_{it}$$

Again, **estimation** can be done by:

1. **One-way FE regression with time dummies**, or
2. Linear regression using the **two-way transformed variables**:
  - $\{\ddot{y}_{it}, \ddot{D}_{it}, \ddot{X}_{it}\}$
  - $\ddot{z}_{it} = z_{it} - \bar{z}_i - \bar{z}_t + \bar{\bar{z}}$

**Standard errors should be clustered** at the unit level.

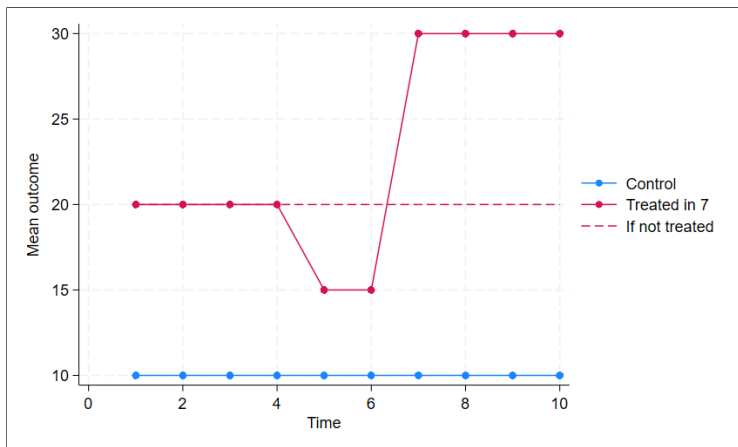
# Assumptions

So when does **DID** = **ATT**?

## Assumptions:

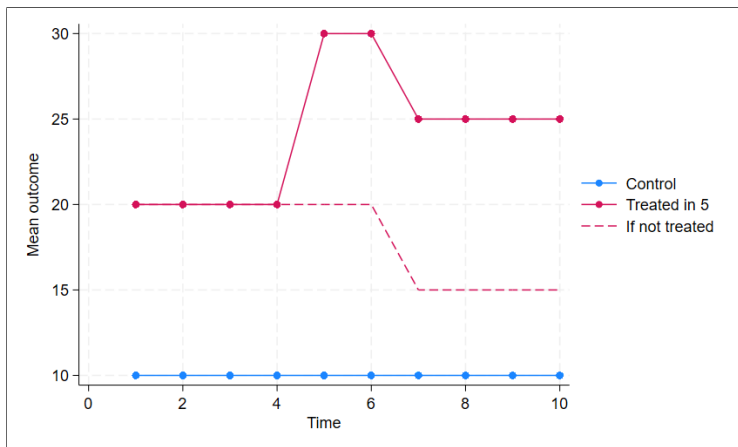
1.  $y_{it} = \theta D_{it} + X'_{it}\beta + u_i + \nu_t + \varepsilon_{it}$
2.  $\mathbb{E}[\ddot{Z}_{it}\ddot{Z}'_{it}] > 0$ , where  $\ddot{Z}_{it} = (\ddot{D}_{it}, \ddot{X}'_{it})$
3.  $\mathbb{E}[X_{it}\varepsilon_{is}] = 0$  for all  $t$  and  $s$ .
4. Random variables  $D_{it}$  and  $\varepsilon_{it}$  are **independent conditional on**  $X_{i1}, X_{i2}, \dots, X_{iT}$ .

# Implications of assumption 4: No anticipatory effects



$$DID = (30 - 15) - 0 = 30 - 15 = 15 \neq 10 = ATT$$

# Implications of assumption 4: Parallel trends if not treated



$$DID = (25 - 20) - 0 = 5 - 0 = 5 \neq 10 = ATT$$

# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion

## DID commands in Stata

**Stata** has two commands to **fit classical DID models**:

1. `xtdidregress` for **panel data**
2. `didregress` for **repeated cross-sectional data**.



## Example: xtdidregress

*Moser et al. (2012)* study the effects of **compulsory licensing\*** on **domestic inventions**.

\*Allowing firms to **produce foreign inventions without consent** of patent owners.

### Possible effects:

- ↑: Access to foreign technology can **inspire new innovations**.
- ↓: **Discourage domestic inventions**.

## Trading with the enemy act (TWEA)

In WW1, the **USA enacted the TWEA**. American companies were granted **compulsory licensing** over German-owned patents.

**Data:** Panel of **industry subclasses**, spanning from **1875-1939**.

**outcome:** uspatents, **US patents issued** in subclass.

**control:** fpatents, **foreign non-TWEA patents** issued in subclass.

**treat:** gotpatents, 1 if subclass got a **TWEA patent**.

## Data – Moser et al. (2012)

fpatents[17536]

	year	uspatents	fpatents	class	gotpatent
17509	1898	0	1	17550 None	
17510	1899	0	0	17550 None	
17511	1900	0	0	17550 None	
17512	1901	0	0	17550 None	
17513	1902	0	0	17550 None	
17514	1903	0	0	17550 None	
17515	1904	0	1	17550 None	
17516	1905	0	0	17550 None	
17517	1906	0	0	17550 None	
17518	1907	0	0	17550 None	
17519	1908	0	0	17550 None	
17520	1909	0	0	17550 None	
17521	1910	0	0	17550 None	
17522	1911	0	0	17550 None	
17523	1912	0	0	17550 None	
17524	1913	0	0	17550 None	
17525	1914	0	1	17550 None	
17526	1915	0	0	17550 None	
17527	1916	0	0	17550 None	
17528	1917	0	0	17550 None	
17529	1918	0	0	17550 None	
17530	1919	0	0	17550 Patent	
17531	1920	0	0	17550 Patent	
17532	1921	0	0	17550 Patent	
17533	1922	0	0	17550 Patent	
17534	1923	0	0	17550 Patent	
17535	1924	0	1	17550 Patent	
17536	1925	0	1	17550 Patent	
17537	1926	1	0	17550 Patent	
17538	1927	1	0	17550 Patent	
17539	1928	1	2	17550 Patent	
17540	1929	0	2	17550 Patent	
17541	1930	1	2	17550 Patent	

Variables

Filter variables here

Name	Label	Type	Format	Value label
<input checked="" type="checkbox"/> year	Year	int	%9.0g	
<input checked="" type="checkbox"/> uspatents	Number of US patents	byte	%9.0g	
<input checked="" type="checkbox"/> fpatents	Number of foreign pate...	byte	%9.0g	
<input checked="" type="checkbox"/> classid	Class ID	float	%9.0g	
<input checked="" type="checkbox"/> gotpatent	Subclass got patent pos...	byte	%9.0g	gp

Variables Snapshots

Properties

Variables

Name	gotpatent
Label	Subclass got patent post 1918
Type	byte
Format	%9.0g
Value label	gp
Notes	

Data

Frame	default
Filename	patents.dta
Label	Excerpt from Moser and Voena (2012)
Notes	
Variables	5
Observations	471,120
Size	4.04M
Memory	64M

Ready

Vars: 5 Order: Dataset Obs: 471,120 Filter: Off Mode: Browse CAP NUM

## Fitting a DID model for panel data

We want to **estimate**  $\theta$  (ATT) in the **linear model**:

$$\text{uspatents}_{it} = \beta_0 + \theta \text{gotpatent}_{it} + \beta_1 \text{fpatent}_{it} + u_i + \nu_t + \varepsilon_{it}$$

We simply **type in Stata**,

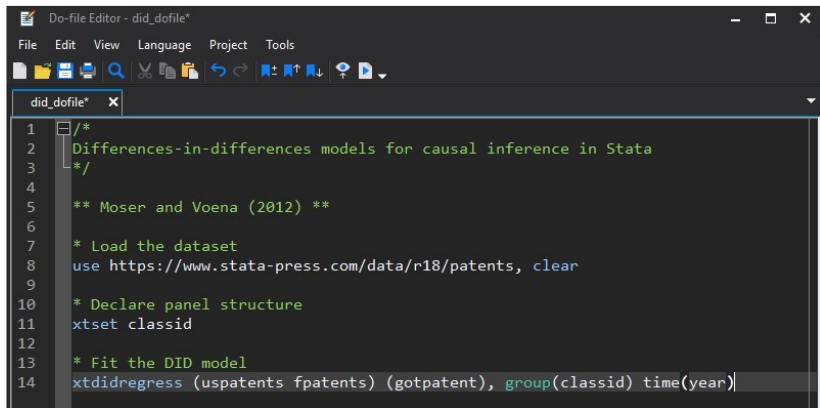
```
xtdidregress (uspatents fpatents) (gotpatent),  
group(classid) time(year)
```

# Syntax – xtdidregress

*DID for panel data*

```
xtdidregress (ovar omvarlist) (tvar [, continuous]) [if] [in] [weight],  
             group(groupvars) [time(timevar) options]
```

## Full code



The screenshot shows a Stata Do-file Editor window titled "did\_dofile\*". The window has a menu bar with "File", "Edit", "View", "Language", "Project", and "Tools". Below the menu bar is a toolbar with various icons. The main editing area shows the following code:

```
1 /*  
2 Differences-in-differences models for causal inference in Stata  
3 */  
4  
5 ** Moser and Voena (2012) **  
6  
7 * Load the dataset  
8 use https://www.stata-press.com/data/r18/patents, clear  
9  
10 * Declare panel structure  
11 xtset classid  
12  
13 * Fit the DID model  
14 xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)
```

# Output – Header

Time variable: year

Control: gotpatent = 0

Treatment: gotpatent = 1

	Control	Treatment
Group		
classid	6912	336
Time		
Minimum	1875	1919
Maximum	1875	1919

# Output – Results

Difference-in-differences regression

Number of obs = 471,120

Data type: Longitudinal

(Std. err. adjusted for 7,248 clusters in `classid`)

uspatents	Robust		t	P> t	[95% conf. interval]	
	Coefficient	std. err.				
ATET						
gotpatent						
(Patent vs None)	.150516	.0356081	4.23	0.000	.0807137	.2203183

Note: ATET estimate adjusted for covariates, panel effects, and time effects.



# Display coefficients for controls and time fixed effects

```
keep if year>=1917 & year<=1919
xtddidregress (uspatents fpatents) (gotpatent), group(classid) ///
time(year) aequations
```

Difference-in-differences regression		Number of obs = 21,744					
Data type: Longitudinal							
		(Std. err. adjusted for 7,248 clusters in classid)					
uspatents		Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ATET							
gotpatent							
(Patent vs None)		.0574715	.0446729	1.29	0.198	-.0301004	.1450434
Controls							
fpatents		.0974619	.0475782	2.05	0.041	.0041947	.1907291
year							
1918		-.0306437	.0116726	-2.63	0.009	-.0535255	-.0077619
1919		-.0516186	.0124995	-4.13	0.000	-.0761213	-.0271159
_cons		.4410144	.0075294	58.57	0.000	.4262545	.4557742

Note: ATET estimate adjusted for covariates, panel effects, and time effects.

## What if we had few clusters?

**Cluster-robust SE's and CI's** can have **low coverage** when

- Number of clusters is **small**;
- Size of clusters is **highly heterogeneous**.

**Solution: Wild cluster bootstrap** – *Cameron et al. (2008)*

Bootstrap samples are created by **perturbing the original data** with **random noise**.

- Stata **documentation**:
  - <https://www.stata.com/manuals/rwildbootstrap.pdf>

**Implementation:** `wildbootstrap` option.

## wildbootstrap option

```

xtdidregress (uspatents fpatents) (gotpatent) in 1/2000, ///
group(classid) time(year) wildbootstrap

```

DID with wild-cluster bootstrap inference

Number of obs = 2,000

No. of clusters = 31

Replications = 1,000

Data type: Longitudinal

Error weight: rademacher

uspatents	Coefficient	t	P> t	[95% conf. interval]	
ATET gotpatent (Patent vs None)	-.0842576	-0.19	0.872	-1.455363	.9171657

Note: ATET estimate adjusted for covariates, panel effects, and time effects.

## wildbootstrap option

```

xtdidregress (uspatents fpatents) (gotpatent) in 1/2000, ///
group(classid) time(year)

```

```

Difference-in-differences regression          Number of obs = 2,000
Data type: Longitudinal

              (Std. err. adjusted for 31 clusters in classid)

```

		Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ATET							
gotpatent							
(Patent vs None)		-.0842576	.4350006	-0.19	0.848	-.9726473	.8041321

```

Note: ATET estimate adjusted for covariates, panel effects, and time effects.

```

## DID model – Repeated cross-sectional data

The DID **model changes slightly**

$$y_{ist} = \theta D_{st} + X'_{ist}\beta + \gamma_s + \gamma_t + \varepsilon_{it}$$

**s: group-level index** (treatment is administered at this level)

Estimation of  $\theta$  is done by **Frisch-Waugh-Lowell theorem**:

1. **Regress**  $y_{ist}$ ,  $D_{st}$ ,  $X_{ist}$ , and the time period dummies **on the group-level dummies**;
2. **Regress the residuals** of  $y_{ist}$  **on the residuals** of  $D_{st}$ ,  $X_{ist}$ , and the time period dummies.

## Example: didregress

We want to study the **effect of a new procedure** for admitting patients to hospitals on their **satisfaction**.

**Repeated cross-sectional** data:

- 7368 patients admitted to 46 hospitals;
- Spanning from January to July 2021.

**Variables:**

**outcome:** satis, **patient satisfaction** score (0-10)

**treat:** procedure, 1 if **new admission procedure** is implemented.

# Data – Patient admissions

	hospital	frequency	month	procedure	satis					
1	1	High	July	New	4.106527					
2	1	Medium	March	Old	3.319475					
3	1	Very high	February	Old	3.41172					
4	1	Medium	April	New	3.004025					
5	1	Low	March	Old	3.11072					
6	1	Low	July	New	2.882164					
7	1	Medium	April	New	4.410257					
8	1	Very high	January	Old	2.902885					
9	1	Medium	March	Old	3.050226					
10	1	Medium	January	Old	2.438842					
11	1	Very high	January	Old	3.333466					
12	1	Low	April	New	2.166077					
13	1	Medium	June	New	3.837375					

# Example: didregress

```
didregress (satis) (procedure), group(hospital) time(month)
```

Difference-in-differences regression

Number of obs = 7,368

Data type: Repeated cross-sectional

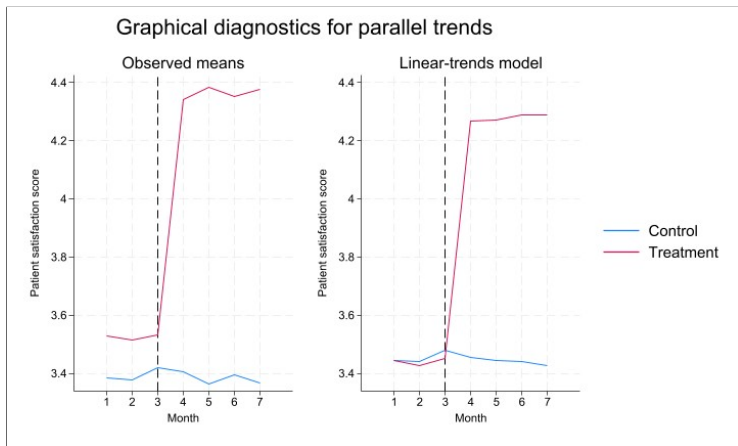
(Std. err. adjusted for 46 clusters in hospital)

satis	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ATET procedure (New vs Old)	.8479879	.0321121	26.41	0.000	.7833108	.912665

Note: ATET estimate adjusted for group effects and time effects.



# Model diagnostics – Pretreatment parallel trends



## Model diagnostics – Testing pretreatment parallel trends

The graph indicates **parallel trends prior to policy implementation**. We can **formally test** using `estat ptrends`:

```
. estat ptrends

Parallel-trends test (pretreatment time period)
H0: Linear trends are parallel

F(1, 45) = 0.55
Prob > F = 0.4615
```

## Model diagnostics – Testing anticipatory effects

We can also **test for parallel trends** and **no anticipatory effects** with a **Granger-type** causality model.

- Augment model with dummies that indicate future treatment
- Jointly test if their coefficients are 0.

```
. estat granger  
  
Granger causality test  
H0: No effect in anticipation of treatment  
  
F(2, 45) = 0.33  
Prob > F = 0.7239
```

## Difference-in-difference-in-differences

Could this result be explained by **unobserved factors** that affect both the **frequency of visits** and **patient satisfaction**?

- frequency: 1 “Low”, 2 “Medium”, 3 “High”, 4 “Very high”

What if parallel trends only hold for patients **with same frequency**?

DID model is **misspecified**  $\Rightarrow$  **ATT estimation is biased**.

- Use a **DDD model instead**.

# DDD model

We **expand the DID model**

$$y_{ist} = \theta D_{sgt} + X'_{isgt} \beta + \gamma_s + \gamma_t + \gamma_g + \gamma_t \gamma_g + \gamma_s \gamma_g + \varepsilon_{it}$$

where  $g$  is a **group-level index**.

**Remark:** Treatment is now administered at the  $(s, g)$  level.

So, in our example, index  $g$  is defined by the values in `frequency`

## Fitting a DDD model

First, a **new treatment identifier**:

- `hightreat = (procedure == 1)*(frequency >= 3)`

Effect of the **new procedure and visiting the hospital frequently** on patient **satisfaction**.

To fit the model, we simply add `frequency` in `group()`.

- `didregress (satis) (hightrt), ///`  
`group(hospital frequency) time(month)`

# Fitting a DDD model

Triple-differences regression  
Data type: Repeated cross-sectional

Number of obs = 7,368

(Std. err. adjusted for 46 clusters in hospital)

		Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ATET	satis						
	hightrt (1 vs 0)	.764154	.0402603	18.98	0.000	.6830655	.8452425

Note: ATET estimate adjusted for group effects, time effects, and group- and time-effects interactions.

# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion



# Outline

## 1 Classical DID and DDD

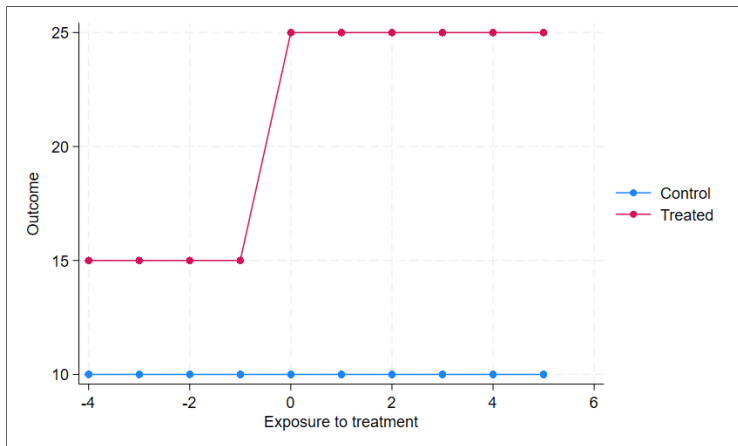
- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

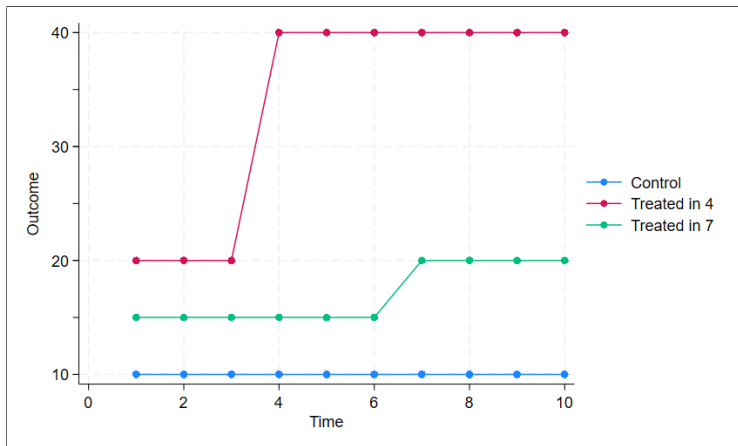
## 3 Conclusion

# Classic differences-in-differences



$$ATT = DID = \text{difference in treated} - \text{difference in control} = 10 - 0 = 10$$

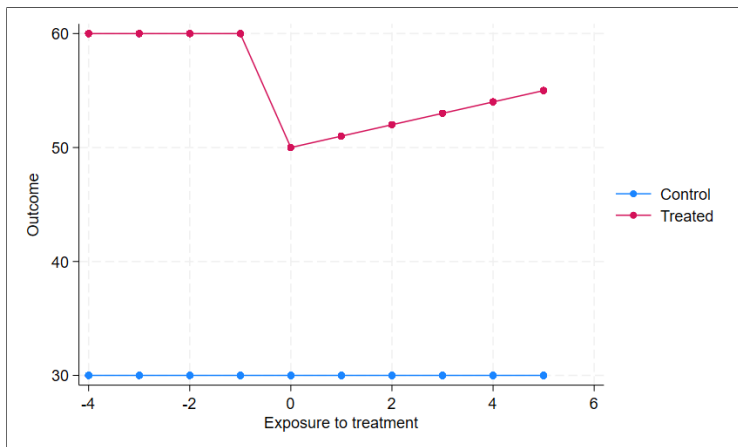
# Cohort heterogeneity



$$ATT_{\text{red}} = 20$$

$$ATT_{\text{green}} = 5$$

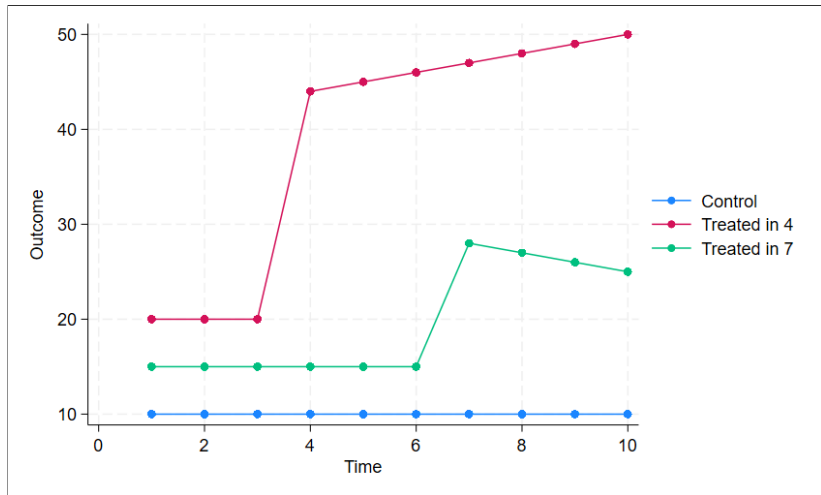
# Time heterogeneity



$$ATT_{t=0} = -10$$

$$ATT_{t=5} = -5$$

# Time-cohort heterogeneity



# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- **Model setup and assumptions**
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 Conclusion

# The heterogeneous DID model

**Panel** or **repeated cross-sectional** data with  $\{1, \dots, T\}$  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:

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

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:** ATTs are based on **unobservables**  $\Rightarrow$  **Assumptions**

# Assumption 1: No anticipatory effect

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

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

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

Anticipatory effects **bias** ATT estimation.

## Assumption 2: Parallel trends with never-treated

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

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

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

Violations of this assumption **bias** ATT estimation.

# Outline

## 1 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- **Estimation and model diagnostics**
- Aggregation of treatment effects

## 3 Conclusion

# 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],  
group(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

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

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

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

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  $\Delta Y$  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  $t$ .



## 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 name	Storage type	Display format	Value 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

Data Editor (Browse) - [akc]

File Edit View Data Tools

year[1] 2031

	year	breed	movie	best	registered
1	2031	Affenpinscher	0	0	1053
2	2032	Affenpinscher	0	0	1340
3	2033	Affenpinscher	0	0	1180
4	2034	Affenpinscher	0	0	1602
5	2035	Affenpinscher	0	0	934
6	2036	Affenpinscher	0	0	497
7	2037	Affenpinscher	0	0	1395
8	2038	Affenpinscher	0	0	1056
9	2039	Affenpinscher	0	0	1063
10	2040	Affenpinscher	0	0	1166
11	2031	Afghan Hound	0	0	1341
12	2032	Afghan Hound	0	0	1398
13	2033	Afghan Hound	0	0	1544
14	2034	Afghan Hound	0	0	791
15	2035	Afghan Hound	0	0	531
16	2036	Afghan Hound	0	0	643
17	2037	Afghan Hound	0	0	392
18	2038	Afghan Hound	0	0	887
19	2039	Afghan Hound	0	0	889
20	2040	Afghan Hound	0	0	1215
21	2031	Airedale Terrier	0	0	483
22	2032	Airedale Terrier	0	0	1196
23	2033	Airedale Terrier	0	0	1596
24	2034	Airedale Terrier	0	0	1625
25	2035	Airedale Terrier	0	0	1300
26	2036	Airedale Terrier	0	0	1114
27	2037	Airedale Terrier	0	0	897
28	2038	Airedale Terrier	0	0	467
29	2039	Airedale Terrier	0	0	1702
30	2040	Airedale Terrier	0	0	1628
31	2031	Akita	0	0	1276
32	2032	Akita	0	0	350
33	2033	Akita	0	0	909

**Variables**

Filter variables here

<input checked="" type="checkbox"/>	Name	Label	Type	Format	Value label
<input checked="" type="checkbox"/>	year	Year	int	%10.0g	
<input checked="" type="checkbox"/>	breed	Dog breed	int	%34.0g	Breed
<input checked="" type="checkbox"/>	movie	Was a movie protagonist	byte	%9.0g	
<input checked="" type="checkbox"/>	best	Won best in show in pa...	byte	%9.0g	
<input checked="" type="checkbox"/>	registered	Number of AKC registra...	int	%9.0g	

**Variables** | **Snapshots**

**Properties**

**Variables**

Name	year
Label	Year
Type	int
Format	%10.0g
Value label	
Notes	

**Data**

Frame	default
Filename	akc.dta
Label	Fictional dog breed and AKC registrati...
Notes	
Variables	5
Observations	1,410
Size	11.07K

Ready

Vars: 5 Order: Dataset Obs: 1,410 Filter: Off Mode: Browse CAP: NUM

# Staggered treatment

```
. tabulate year movie
```

Year	Was a movie protagonist		Total
	0	1	
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

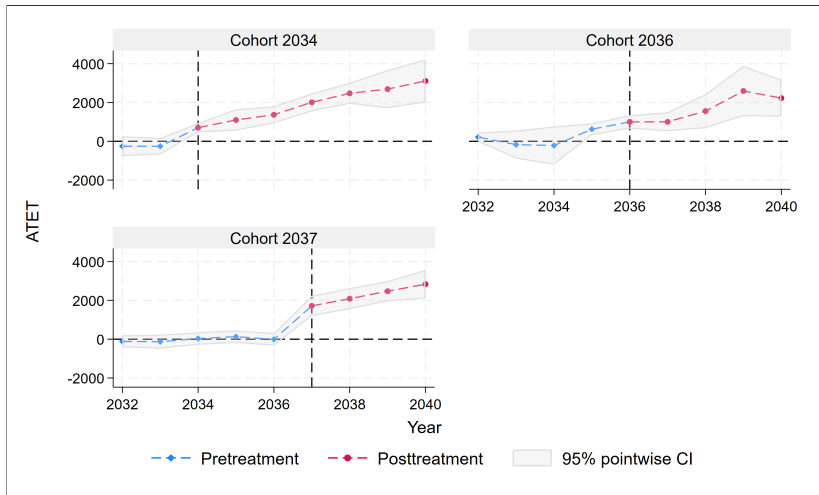
Heterogeneous-treatment-effects regression					Number of obs	= 1,410
					Number of panels	= 141
Estimator:	Regression adjustment					
Panel variable:	breed					
Treatment level:	breed					
Control group:	Never treated					
(Std. err. adjusted for 141 clusters in breed)						
		Robust				
Cohort	ATET	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

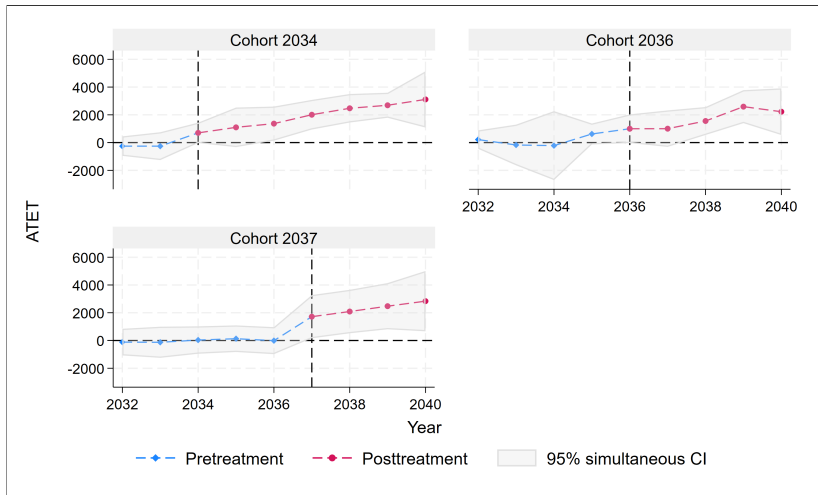
Note: ATET computed using covariates.

# Graphical representation: estat atetplot





# Simultaneous confidence intervals: estat atetplot, sci



## No anticipatory effects test – estat ptrends

Parallel-trends test (pretreatment time period)

H0: Treatment effects in all the pretreatment periods are zero

```
chi2(11) = 57.68  
Prob > chi2 = 0.0000
```

## Change the control group – controlgroup()

Sometimes all units are eventually treated. We need to use as controls the not yet treated.

```
xthdidregress ra (registered best) (movie), group(breed)  
controlgroup(notyet)
```

Heterogeneous-treatment-effects regression		Number of obs = <b>1,410</b>
		Number of panels = <b>141</b>
Estimator:	Regression adjustment	
Panel variable:	<b>breed</b>	
Treatment level:	<b>breed</b>	
Control group:	Not yet treated	

output omitted

# Inverse-probability weighting estimator

## Syntax:

```
xthdidregress ipw (ovar) (tvar [tmvarlist]) [if] [in] [weight],  
group(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.**

# The IPW estimator

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

$p_g(X)$ : **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**.

# Augmented inverse-probability weighting estimator

## Syntax:

```
xthdidregress aipw (ovar [omvarlist]) (tvar [tmvarlist]) [if] [in] [weight],  
group(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

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

$p_g(X)$ : **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**

# Extended two-way fixed effects estimator

## Syntax:

```
xthdidregress twfe (ovar [omvarlist]) (tvar) [if] [in] [weight],  
group(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 Classical DID and DDD

- Model setup and assumptions
- Estimation, model diagnostics, and tests

## 2 Heterogeneous DID models

- Treatment effect heterogeneity
- Model setup and assumptions
- Estimation and model diagnostics
- Aggregation of treatment effects

## 3 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)						
registered	ATET	Robust 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

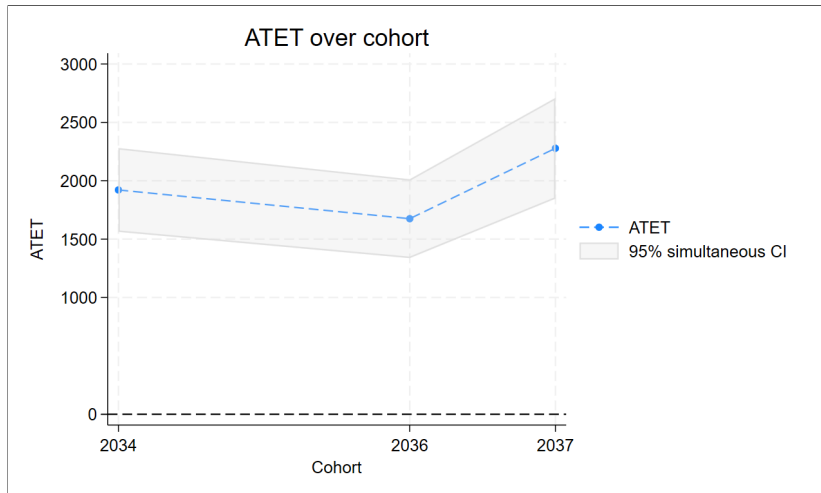
Replications = 999

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

Cohort	Observed ATET	Bootstrap std. err.	Simultaneous [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

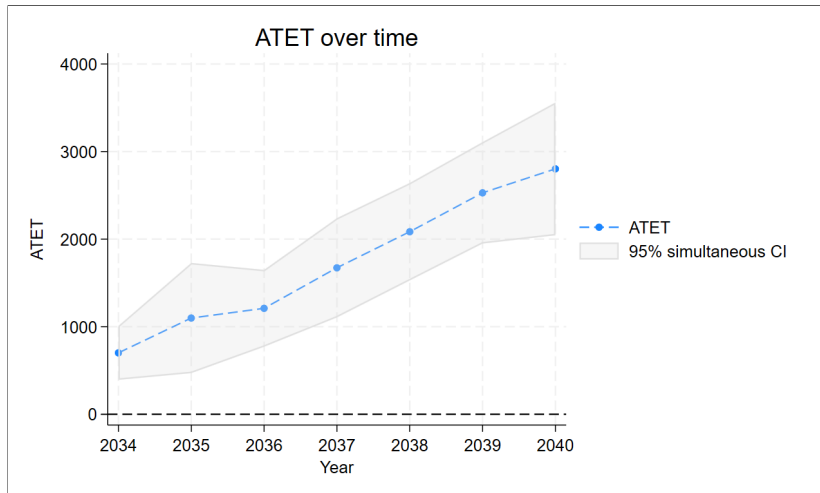
Replications = 999

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

Time	Observed ATET	Bootstrap std. err.	Simultaneous [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

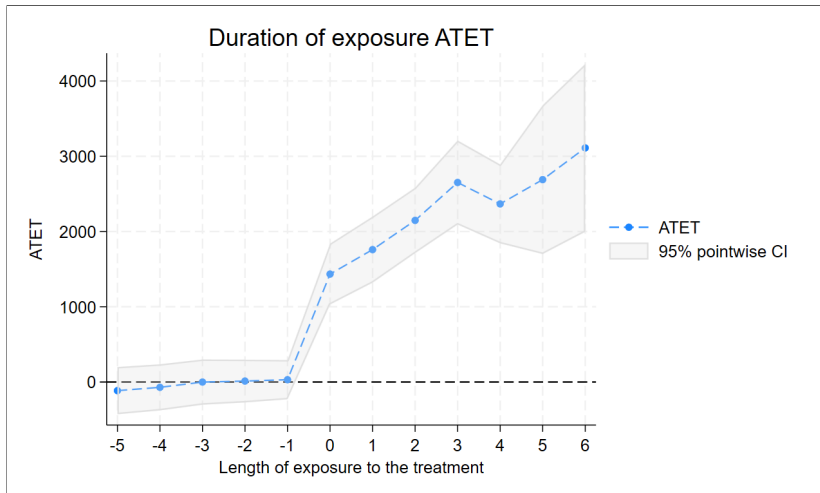
Number of obs = 1,410

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

Exposure	ATET	Robust 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 Classical DID and DDD
  - Model setup and assumptions
  - Estimation, model diagnostics, and tests
- 2 Heterogeneous DID models
  - Treatment effect heterogeneity
  - Model setup and assumptions
  - Estimation and model diagnostics
  - Aggregation of treatment effects
- 3 Conclusion

# Conclusion

1. **DID is a powerful tool** to better understand treatment effects.
2. **Easy to implement** in Stata.
3. Classic DID:
  - `xtdidregress` for **panel data**
  - `didregress` for **repeated cross section**
4. Heterogeneous DID:
  - `xthdidregress` for **panel data**
  - `hdidregress` for **repeated cross section**
  - **Treatment effects** can be **aggregated by**:
    - Cohort, time, exposure to treatment, or overall.

Questions?

Thank you!

# References

1. Callaway, B., and P. H.C.Sant'Anna.2021.Difference-in-differences with multiple time periods. Journal of Econometrics.
2. Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. Working paper, Department of Economics, Michigan State University.
3. Cameron,A.C., J. B. Gelbach, and D. L. Miller. 2008. Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics.
4. Moser, P., and A. Voena. 2012. Compulsory licensing: Evidence from the Trading with the Enemy Act. American Economic Review.