Causal Inference and Treatment Effect Estimation using Stata

Enrique Pinzón

StataCorp LLC

Apri 17, 2024

Objective

- Provide a quick overview of the causal-inference tools available in Stata
- Discuss theory of estimators and present estimation and postestimation results
- Not a course or overview of causal inference
- Examples are presented to illustrate the tools

Road map

- Selection on observables
- Selection on unobservables

About us

- About me
- About you
- How I work and a basic Stata overview

STATA CONCEPTS

Factor variables

```
. sysuse auto, clear
(1978 automobile data)
. summarize i.foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
foreign					
Domestic	74	.7027027	.4601885	0	1
Foreign	74	.2972973	.4601885	0	1

Factor variables

```
. summarize i.rep78
```

Variable	Obs	Mean	Std. dev.	Min	Max
rep78					
1	69	.0289855	.1689948	0	1
2	69	.115942	.3225009	0	1
3	69	.4347826	.4993602	0	1
4	69	.2608696	.4423259	0	1
5	69	.1594203	.3687494	0	1

Factor variables

```
. summarize i.rep78#i.foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
rep78#foreign					
1#Domestic	69	.0289855	.1689948	0	1
1#Foreign	69	(empty)			
2#Domestic	69	.115942	.3225009	0	1
2#Foreign	69	(empty)			
3#Domestic	69	.3913043	.4916177	0	1
3#Foreign	69	.0434783	.2054251	0	1
4#Domestic	69	.1304348	.3392485	0	1
4#Foreign	69	.1304348	.3392485	0	1
5#Domestic	69	.0289855	.1689948	0	1
5#Foreign	69	.1304348	.3392485	0	1

```
. summarize rep78#foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
rep78#foreign					
1#Domestic	69	.0289855	.1689948	0	1
1#Foreign	69	(empty)			
2#Domestic	69	.115942	.3225009	0	1
2#Foreign	69	(empty)			
3#Domestic	69	.3913043	.4916177	0	1
3#Foreign	69	.0434783	.2054251	0	1
4#Domestic	69	.1304348	.3392485	0	1
4#Foreign	69	.1304348	.3392485	0	1
5#Domestic	69	.0289855	.1689948	0	1
5#Foreign	69	.1304348	.3392485	0	1

Factor variables

```
. summarize i.rep78##i.foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
rep78					
1	69	.0289855	.1689948	0	1
2	69	.115942	.3225009	0	1
3	69	.4347826	.4993602	0	1
4	69	.2608696	.4423259	0	1
5	69	.1594203	.3687494	0	1
foreign					
Domestic	74	.7027027	.4601885	0	1
Foreign	74	.2972973	.4601885	0	1
rep78# foreign					
1#Domestic	69	.0289855	.1689948	0	1
1#Foreign	69	(empty)			
2#Domestic	69	.115942	.3225009	0	1
2#Foreign	69	(empty)			
3#Domestic	69	.3913043	.4916177	0	1
3#Foreign	69	.0434783	.2054251	0	1
4#Domestic	69	.1304348	.3392485	0	1
4#Foreign	69	.1304348	.3392485	0	1
5#Domestic	69	.0289855	.1689948	0	1
5#Foreign	69	.1304348	.3392485	0	1

Factor variables

```
. summarize c.mpg#i.foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
foreign# c.mpg					
Domestic	74	13.93243	9.948265	0	34
Foreign	74	7.364865	11.93886	0	41

```
. summarize c.mpg##i.foreign
```

Variable	Obs	Mean	Std. dev.	Min	Max
mpg	74	21.2973	5.785503	12	41
foreign					
Domestic	74	.7027027	.4601885	0	1
Foreign	74	.2972973	.4601885	0	1
foreign# c.mpg					
Domestic	74	13.93243	9.948265	0	34
Foreign	74	7.364865	11.93886	0	41

```
. summarize c.mpg##c.mpg
```

Variable	Obs	Mean	Std. dev.	Min	Max
mpg	74	21.2973	5.785503	12	41
c.mpg#c.mpg	74	486.5946	282.0663	144	1681

margins

- $\widehat{\beta}$ is just the beginning
- Construct functions of $x\widehat{\beta}$ or sometimes $E(Y|X)$ (nonparametric)
- Based on Stata's `predict`

Probit model

- The prediction is $\widehat{P}(Y|X) = \Phi\left(X\widehat{\beta}\right)$ questions could be:
  1. Marginal effect `margins, dydx(xk)`

$$\frac{\partial \widehat{P}(Y|X)}{\partial X_{ik}} = \phi\left(X\widehat{\beta}\right)\widehat{\beta}_k$$

  2. Contrast or marginal effect of a discrete covariate `margins r.xd` or `margins, dydx(xd)` which is $\widehat{P}(Y|X_{-d}, X_{id} = 1) - \widehat{P}(Y|X_{-d}, X_{id} = 0)$

$$\Phi\left(X_{-d}\widehat{\beta}_{-d} + \widehat{\beta}_d\right) - \Phi\left(X_{-d}\widehat{\beta}_{-d}\right)$$

  3. Specific counterfactuals of covariates. What if `x5` is 2 for everyone in the population, or 3, ..., or 100. `margins, at(x5=(2(1)100))`. $\Phi\left(X_{-5}\widehat{\beta}_{-5} + 2 * \beta_5\right)$, or $\Phi\left(X_{-5}\widehat{\beta}_{-5} + 3 * \beta_5\right)$, or $\Phi\left(X_{-5}\widehat{\beta}_{-5} + 100 * \beta_5\right)$, or

SELECTION ON OBSERVABLES

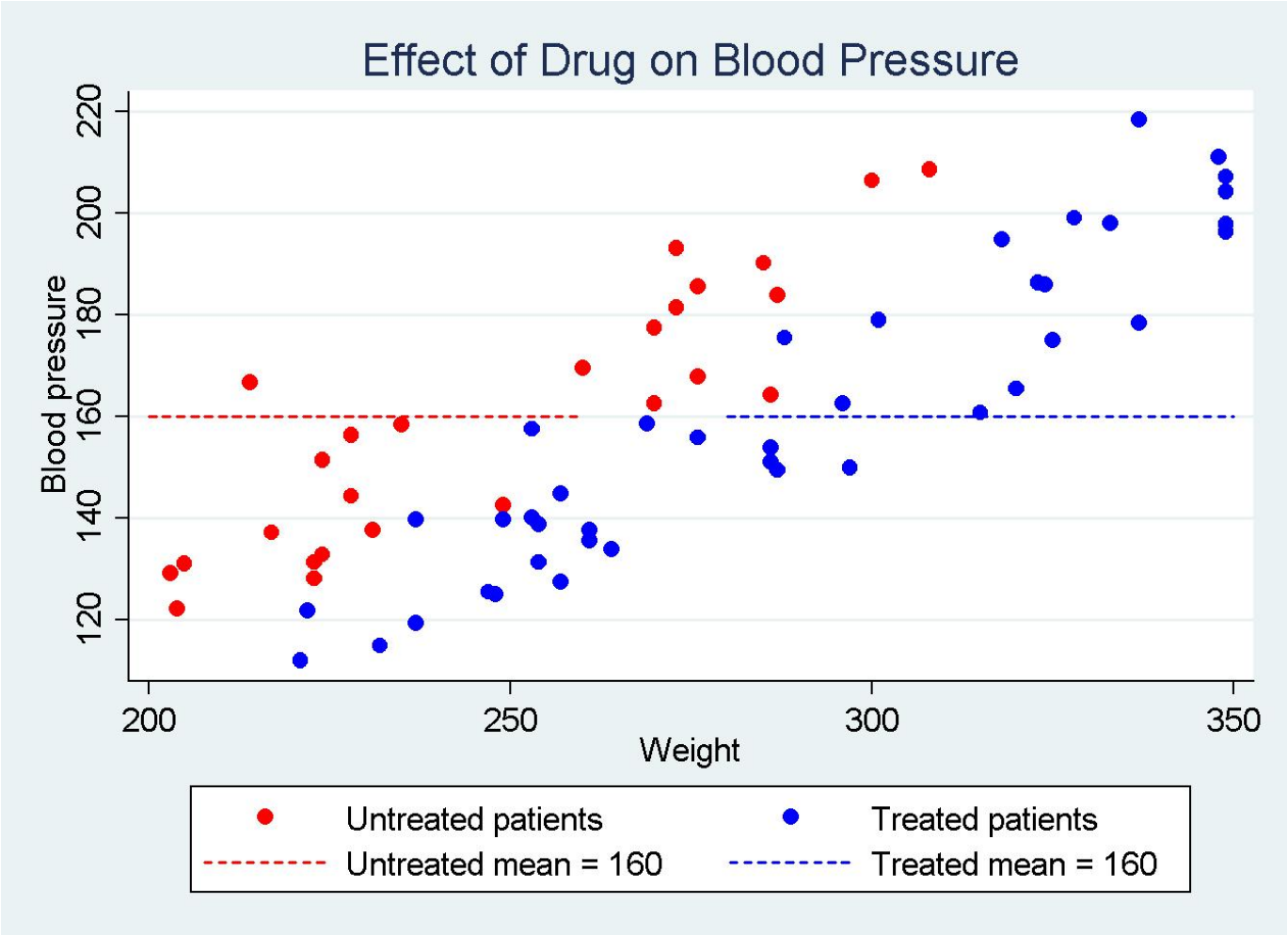
Motivation

We are interested in the outcomes of receiving a treatment in scenarios were researchers have observational data.

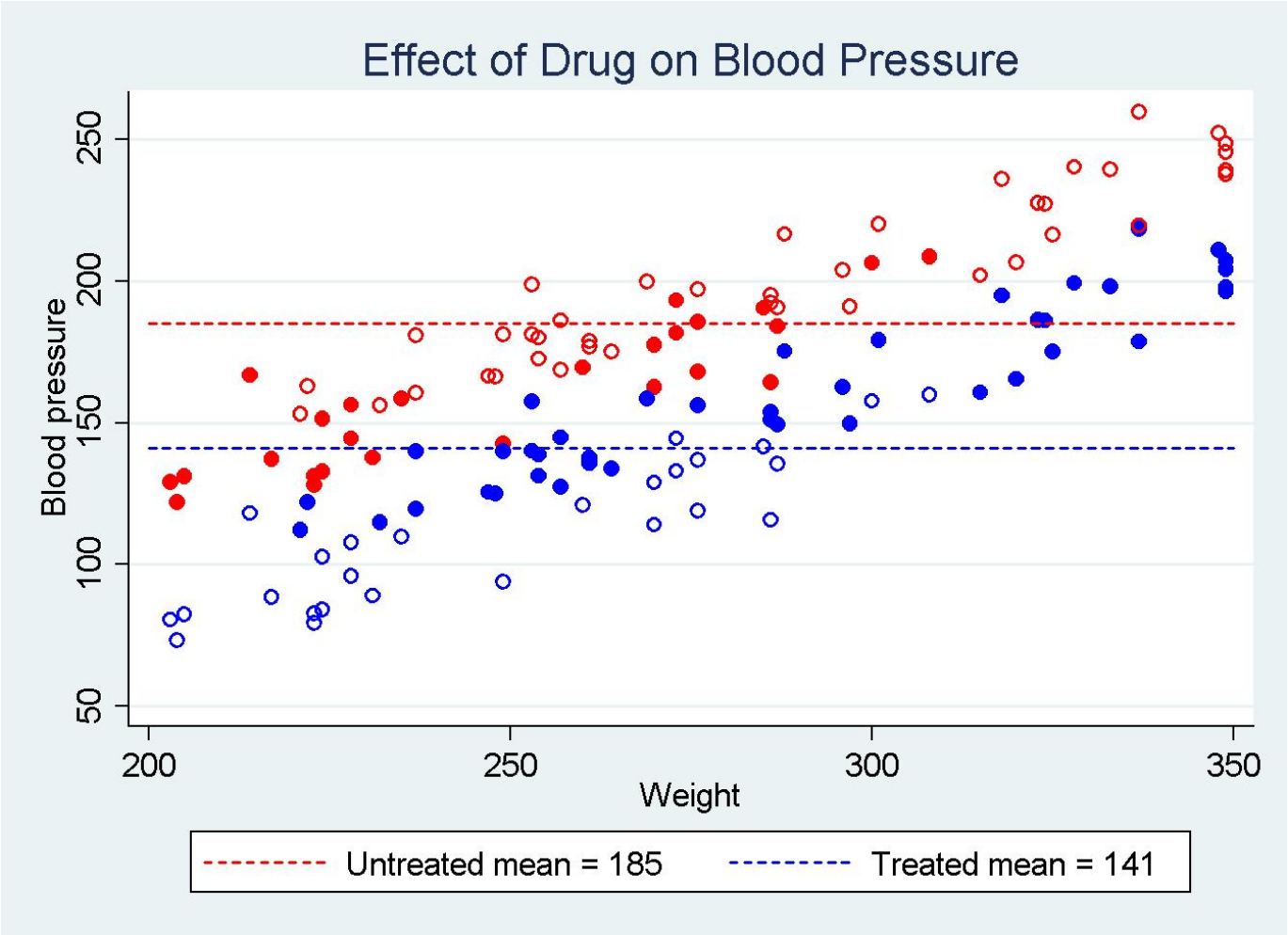
For instance:

- The impact on public education outcomes for schools that received a transfer and those that did not.
- Employment outcomes for individuals that participated in a job training program and those that did not.
- The effect on birth weight for babies of mothers that smoked relative to those of mothers that did not.

Observed Effect of Statin on Blood Presure



Potential Outcomes of Statin on Blood Presure



Observed and Unobserved Data

```
. list y treat y1 y0 te in 1/10, noobs sep(0)
```

y	treat	y1	y0	te
46.12323	0	93.2825	46.12323	47.15927
-28.48061	1	-28.48061	-14.79876	-13.68186
283.9639	1	283.9639	141.76	142.2039
92.78703	0	189.4475	92.78703	96.66045
-175.9654	0	-347.9302	-175.9654	-171.9648
-274.3629	1	-274.3629	-136.4753	-137.8875
-29.50582	0	-58.54392	-29.50582	-29.0381
-86.76294	1	-86.76294	-45.42468	-41.33826
-266.5881	1	-266.5881	-131.3949	-135.1931
196.3055	1	196.3055	98.41778	97.88767

Selection on observables approach to treatment effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- What can we do?
  - We can account for covariates that are correlated with the treatment (selection on observables)
  - We will think of the problem in terms of models that govern the treatment result and the outcome
- There will be cases were this approach will not suffice. We will get to it later.

Selection on observables approach to treatment effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- What can we do?
  - We can account for covariates that are correlated with the treatment (selection on observables)
  - We will think of the problem in terms of models that govern the treatment result and the outcome
- There will be cases were this approach will not suffice. We will get to it later.

Selection on observables approach to treatment effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- What can we do?
  - We can account for covariates that are correlated with the treatment (selection on observables)
  - We will think of the problem in terms of models that govern the treatment result and the outcome
- There will be cases were this approach will not suffice. We will get to it later.

Selection on observables approach to treatment effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- What can we do?
  - We can account for covariates that are correlated with the treatment (selection on observables)
  - We will think of the problem in terms of models that govern the treatment result and the outcome
- There will be cases were this approach will not suffice. We will get to it later.

Notation and Definitions

- The potential outcome is denoted by the random variable y_τ with $\tau \in \{0, 1, \ldots, K\}$. The potential realizations will be denoted by:
  - y_{0i} is the outcome individual i if they do not receive the treatment, where $i = 1 \ldots n$
  - y_{ki} is the potential outcome for individual i if they receive different discrete levels of the treatment, where $k = 1 \ldots K$
  - Usually people think about the binary case where there are only two levels y_{0i} and y_{1i}
- Potential outcome mean

$$POM = E(y_\tau)$$

- Average treatment effect

$$ATE = E(y_{ki} - y_{0i})$$

- Average treatment effect on the treated

$$ATET = E(y_{ki} - y_{0i} | \tau = k)$$

- From now on we will focus on binary treatments. All results are valid for multivariate treatments unless explicitly noted.

Assumptions

- We will be dealing with a cross-sectional random sample of n individuals
- **Overlap:**

$$0 < P(\tau_i = 1 | X_i = x) < 1$$

- **Conditional Independence**: Conditional on the covariates, X, the potential outcomes, y_0, y_1, and the treatment, τ, are independent

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$
\begin{aligned}
y_0 &= x\beta_0 + \varepsilon_0 \\
y_1 &= x\beta_1 + \varepsilon_1 \\
y &= \tau y_1 + (1-\tau)\, y_0
\end{aligned}
$$

TREATMENT MODEL:

$$
\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}
$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

Estimators

- Regression Adjustment (RA)
- Inverse Probability Weighting (IPW)
- Augmented Inverse Probability Weighting (AIPW)
- Inverse Probability Weighted Regression Adjustment (IPWRA)
- Nearest Neighbor Matching
- Propensity Score Matching

Data from Cattaneo (2010) Journal of Econometrics

`bweight`:	infant birth weight (grams)
`lbweight`:	1 if low birthweight baby
`mbsmoke`:	1 if mother smoked
`prenatal`:	trimester of first prenatal care visit
`fbaby`:	1if first baby
`mmarried`:	1 if mother married
`mage`:	mother's age
`fage`:	father's age
`alcohol`:	1 if alcohol consumed during pregnancy

- Sample of newborns from the United States from 1997

Data from Cattaneo (2010) Journal of Econometrics

`bweight`:	infant birth weight (grams)
`lbweight`:	1 if low birthweight baby
`mbsmoke`:	1 if mother smoked
`prenatal`:	trimester of first prenatal care visit
`fbaby`:	1if first baby
`mmarried`:	1 if mother married
`mage`:	mother's age
`fage`:	father's age
`alcohol`:	1 if alcohol consumed during pregnancy

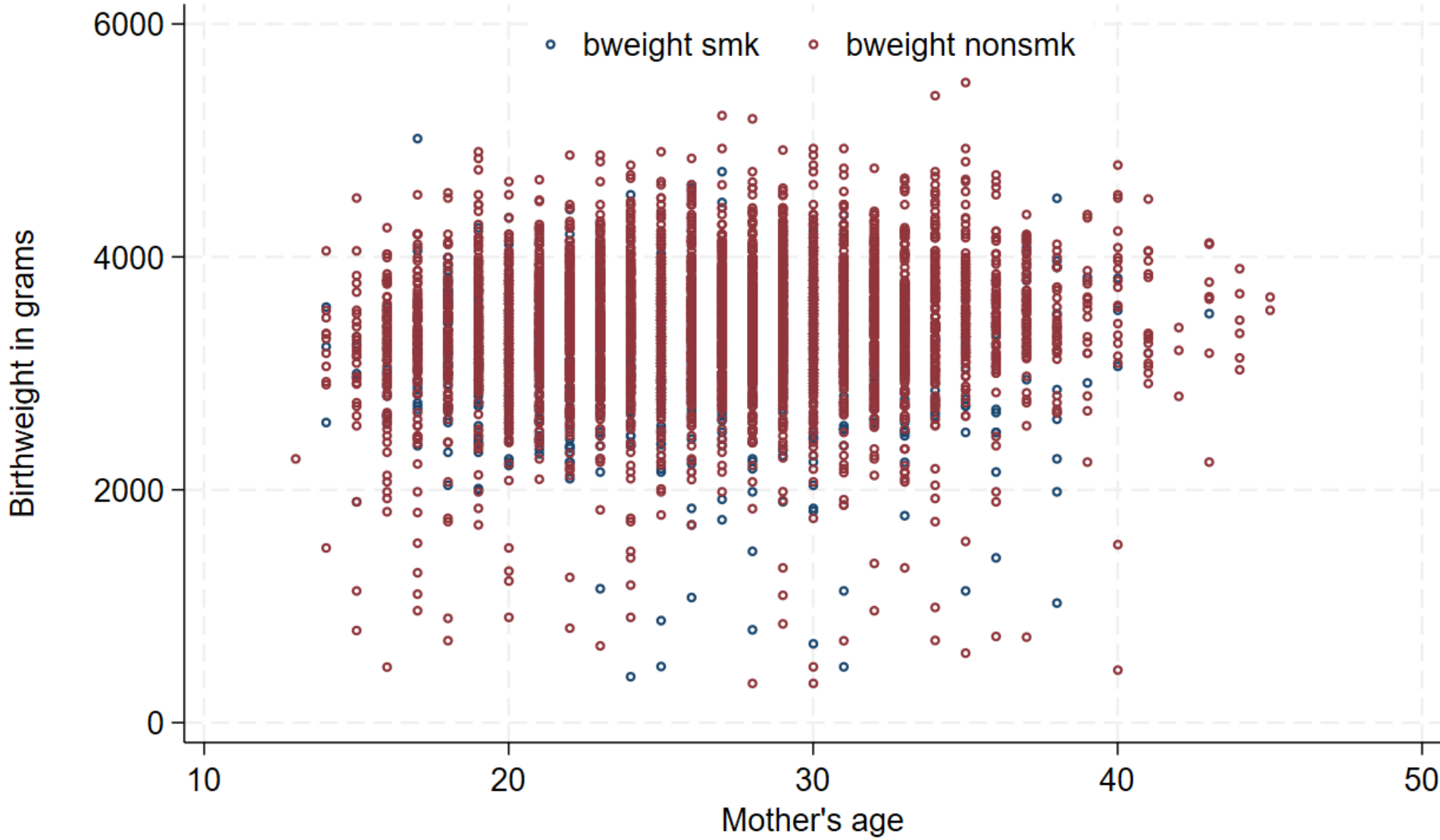
- Sample of newborns from the United States from 1997

Smoking data

```
 . webuse cattaneo2, clear
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138154)
. describe *weight* m* alcohol f*
```

Variable name	Storage type	Display format	Value label	Variable label
bweight	int	%9.0g		Infant birthweight (grams)
lbweight	byte	%9.0g		1 if low birthweight baby
mmarried	byte	%11.0g	mmarried	1 if mother married
mhisp	byte	%9.0g		1 if mother hispanic
mage	byte	%9.0g		Mother´s age
medu	byte	%9.0g		Mother´s education attainment
monthslb	int	%9.0g		Months since last birth
msmoke	byte	%27.0g	smoke2	Cigarettes smoked during pregnancy
mbsmoke	byte	%9.0g	mbsmoke	1 if mother smoked
mrace	byte	%9.0g		1 if mother is white
alcohol	byte	%9.0g		1 if alcohol consumed during pregnancy
fhisp	byte	%9.0g		1 if father hispanic
foreign	byte	%9.0g		1 if mother born abroad
fage	byte	%9.0g		Father´s age
fedu	byte	%9.0g		Father´s education attainment
frace	byte	%9.0g		1 if father is white
fbaby	byte	%9.0g	YesNo	1 if first baby

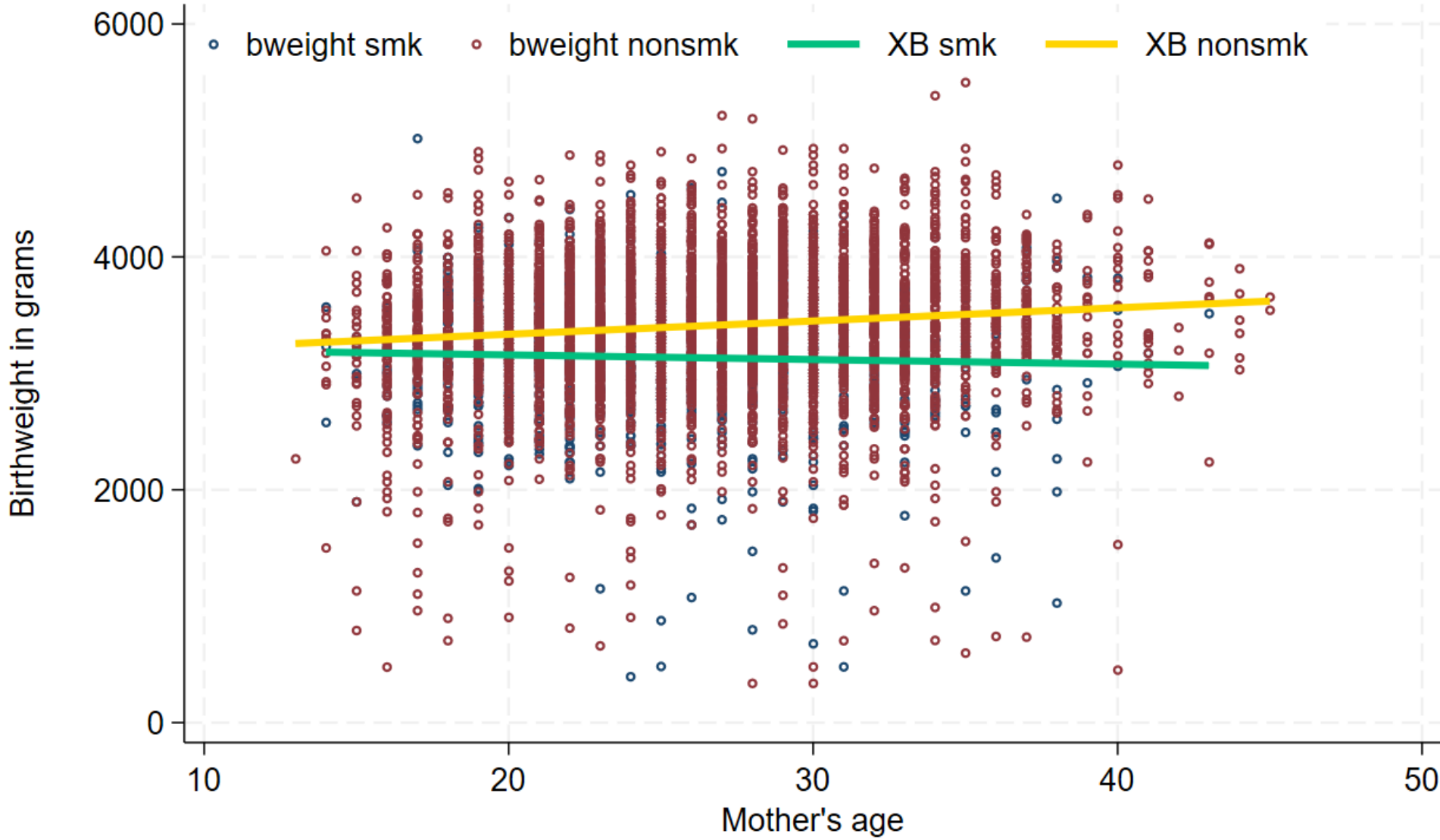
Effect of Smoking Mothers on Birthweight



Regression Adjustment (RA)

- We model the potential outcome and do not say anything about the treatment mechanism
- A conditional expectation is estimated for the treatment and control groups.
- The results from the estimations are used to compute POMs and thereafter ATEs, and ATETs.

RA graphically



Models for the Potential Outcome

Outcome Model	$E(y\|x,z,\tau)$
`linear`	$x\beta_\tau$
`logit`	$exp(x\beta_\tau)/\{1+exp(x\beta_\tau)\}$
`probit`	$\Phi(x\beta_\tau)$
`poisson`	$exp(x\beta_\tau)$
`hetprobit`	$\Phi(x\beta_\tau/exp(z\alpha_\tau))$

Regression adjustment estimation

```
. teffects ra (bweight i.prenatal1 i.mmarried mage i.fbaby) (mbsmoke)
Iteration 0:  EE criterion =  7.734e-24
Iteration 1:  EE criterion =  1.196e-25
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-239.6392	23.82402	-10.06	0.000	-286.3334	-192.945
POmean						
mbsmoke Nonsmoker	3403.242	9.525207	357.29	0.000	3384.573	3421.911

```
. estimates store linear
```

Regression adjustment estimation

```
. teffects ra (bweight i.prenatal1 i.mmarried mage i.fbaby) (mbsmoke), atet
Iteration 0:  EE criterion =  7.629e-24
Iteration 1:  EE criterion =  2.697e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
mbsmoke (Smoker vs Nonsmoker)	-223.3017	22.7422	-9.82	0.000	-267.8755	-178.7278
POmean						
mbsmoke Nonsmoker	3360.961	12.75749	263.45	0.000	3335.957	3385.966

Regression adjustment estimation

```
. teffects ra (bweight i.prenatal1 i.mmarried mage i.fbaby) (mbsmoke), pom
Iteration 0:  EE criterion =  7.734e-24
Iteration 1:  EE criterion =  2.850e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
mbsmoke						
Nonsmoker	3403.242	9.525207	357.29	0.000	3384.573	3421.911
Smoker	3163.603	21.86351	144.70	0.000	3120.751	3206.455

Regression adjustment estimation

```
. teffects ra (bweight i.prenatal1 i.mmarried mage i.fbaby, poisson) (mbsmoke)
Iteration 0:  EE criterion =  3.950e-17
Iteration 1:  EE criterion =  1.145e-23
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : Poisson
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-239.6669	23.83757	-10.05	0.000	-286.3877	-192.9462
POmean						
mbsmoke Nonsmoker	3403.178	9.526006	357.25	0.000	3384.508	3421.849

```
. estimates store exponential
```

Regression adjustment estimation

```
. teffects ra (lbweight i.prenatal1 i.mmarried mage i.fbaby, probit) (mbsmoke)
Iteration 0:  EE criterion =  1.018e-18
Iteration 1:  EE criterion =  5.723e-34
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : probit
Treatment model: none
```

lbweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	.0500546	.0118733	4.22	0.000	.0267833	.0733259
POmean						
mbsmoke Nonsmoker	.0517931	.003734	13.87	0.000	.0444745	.0591116

```
. estimates store probit
```

Regression adjustment estimation

```
. teffects ra (lbweight i.prenatal1 i.mmarried mage i.fbaby) (mbsmoke)
Iteration 0:  EE criterion =  1.589e-32
Iteration 1:  EE criterion =  4.225e-35
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

lbweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	.0496685	.0118577	4.19	0.000	.0264278	.0729092
POmean						
mbsmoke Nonsmoker	.0518917	.0037393	13.88	0.000	.0445629	.0592205

```
. estimates store linearp
```

Comparing results

```
. estimates table linear exponential, se keep(ATE:)
```

Variable	linear	exponent~l
mbsmoke (Smoker vs Nonsmoker)	-239.63921	-239.66693
	23.824021	23.837567

Legend: b/se

Comparing results

```
. estimates table probit linearp, se keep(ATE:)
```

Variable	probit	linearp
mbsmoke (Smoker vs Nonsmoker)	.05005461	.04966851
	.01187334	.01185771

Legend: b/se

Replay and `aequations`

```
. teffects, aequations
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

lbweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	.0496685	.0118577	4.19	0.000	.0264278	.0729092
POmean						
mbsmoke Nonsmoker	.0518917	.0037393	13.88	0.000	.0445629	.0592205
OME0						
prenatal1 Yes	-.0139963	.0115359	-1.21	0.225	-.0366063	.0086137
mmarried Married	-.0480361	.0108809	-4.41	0.000	-.0693622	-.02671
mage	-.0004581	.0008032	-0.57	0.568	-.0020323	.0011162
fbaby Yes	.0039126	.007564	0.52	0.605	-.0109125	.0187377
_cons	.1071458	.0222382	4.82	0.000	.0635597	.1507318
OME1						
prenatal1 Yes	.0180394	.0239757	0.75	0.452	-.0289521	.0650309
mmarried Married	-.0454804	.0233014	-1.95	0.051	-.0911502	.0001895
mage	.0032633	.0023316	1.40	0.162	-.0013066	.0078331
fbaby Yes	-.06742	.0216083	-3.12	0.002	-.1097716	-.0250684
_cons	.0619625	.0570132	1.09	0.277	-.0497812	.1737063

Replay and `aequations`

```
. teffects ra (lbweight i.prenatal1 i.mmarried mage i.fbaby) (mbsmoke),   ///
>         aequations
Iteration 0:  EE criterion =  1.589e-32
Iteration 1:  EE criterion =  4.225e-35
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

lbweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	.0496685	.0118577	4.19	0.000	.0264278	.0729092
POmean						
mbsmoke Nonsmoker	.0518917	.0037393	13.88	0.000	.0445629	.0592205
OME0						
prenatal1 Yes	-.0139963	.0115359	-1.21	0.225	-.0366063	.0086137
mmarried Married	-.0480361	.0108809	-4.41	0.000	-.0693622	-.02671
mage	-.0004581	.0008032	-0.57	0.568	-.0020323	.0011162
fbaby Yes	.0039126	.007564	0.52	0.605	-.0109125	.0187377
_cons	.1071458	.0222382	4.82	0.000	.0635597	.1507318
OME1						
prenatal1 Yes	.0180394	.0239757	0.75	0.452	-.0289521	.0650309
mmarried Married	-.0454804	.0233014	-1.95	0.051	-.0911502	.0001895
mage	.0032633	.0023316	1.40	0.162	-.0013066	.0078331
fbaby Yes	-.06742	.0216083	-3.12	0.002	-.1097716	-.0250684
_cons	.0619625	.0570132	1.09	0.277	-.0497812	.1737063

Manual/Heuristic computation of effects

```
. quietly regress bweight i.prenatal1 i.mmarried mage i.fbaby if mbsmoke==1
.
. predict double te1
(option xb assumed; fitted values)
.
. quietly regress bweight i.prenatal1 i.mmarried mage i.fbaby if mbsmoke==0
.
. predict double te0
(option xb assumed; fitted values)
.
. generate double teffect = te1 - te0
.
. summarize teffect
```

Variable	Obs	Mean	Std. dev.	Min	Max
teffect	4,642	-239.6392	99.00799	-488.4602	8.261835

Manual/Heuristic computation of effects

```
. quietly regress bweight ///
>         i.mbsmoke##(i.prenatal1 i.mmarried c.mage i.fbaby), vce(robust)
. estimates store reg
. margins r.mbsmoke, vce(unconditional) contrast(nowald) noestimcheck
Contrasts of predictive margins                         Number of obs = 4,642
Expression: Linear prediction, predict()
```

	Contrast	Unconditional std. err.	[95% conf.	interval]
mbsmoke (Smoker vs Nonsmoker)	-239.6392	23.84972	-286.396	-192.8824

```
.
. di  23.84972*sqrt((4632)/4642)
23.824017
```

GMM: Standard error computation

- System of equations estimation
- Accounts for uncertainty of estimated quantities used to fit a model
- An extension of a method of moments estimation
- For example let $y = X\beta + \varepsilon$ and:

$$
\begin{aligned}
E\left(X'\varepsilon\right) &= 0 \\
E\left(X'\underbrace{\left(y - X\beta\right)}_{e=y-X\beta}\right) &= 0 \\
E\left(X'y\right) &= E\left(X'X\right)\beta \\
E\left(X'X\right)^{-1}E\left(X'y\right) &= \beta
\end{aligned}
$$

- Replace expected values by sample means. Namely:

$$
\begin{aligned}
\left(\frac{X'X}{N}\right)^{-1}\left(\frac{X'y}{N}\right) &= \widehat{\beta} \\
\left(X'X\right)^{-1}\left(X'y\right) &= \widehat{\beta}
\end{aligned}
$$

GMM

- Key insight is to define $E(X'\varepsilon) = 0$
- X are instruments
- e are score equations from likelihood or error terms in linear case

GMM estimation

```
. local X i.mbsmoke##(i.prenatal1 i.mmarried c.mage i.fbaby)
. quietly gmm (bweight -{xb: `X´ _cons}),    ///
>         instruments(`X´) onestep quickderivatives
. estimates store gmm
```

GMM results

```
. estimates table gmm reg, eq(1)
```

Variable	gmm	reg
mbsmoke Smoker	24.423541	24.423541
prenatal1 Yes	64.408589	64.408589
mmarried Married	160.95126	160.95126
mage	2.5468279	2.5468279
fbaby Yes	-71.328598	-71.328598
mbsmoke# prenatal1 Smoker#Yes	-39.297259	-39.297259
mbsmoke# mmarried Smoker # Married	-27.289554	-27.289554
mbsmoke# c.mage Smoker	-9.917709	-9.917709
mbsmoke# fbaby Smoker#Yes	112.76851	112.76851
_cons	3202.7457	3202.7457

Inverse Probability Weighting (IPW)

- In contrast to RA estimators, IPW estimate models for the treatment
- We fit a model for the treatment and compute the probabilities of treatment
- We then compute a weighted average, using the inverse of the probability of being in each group.

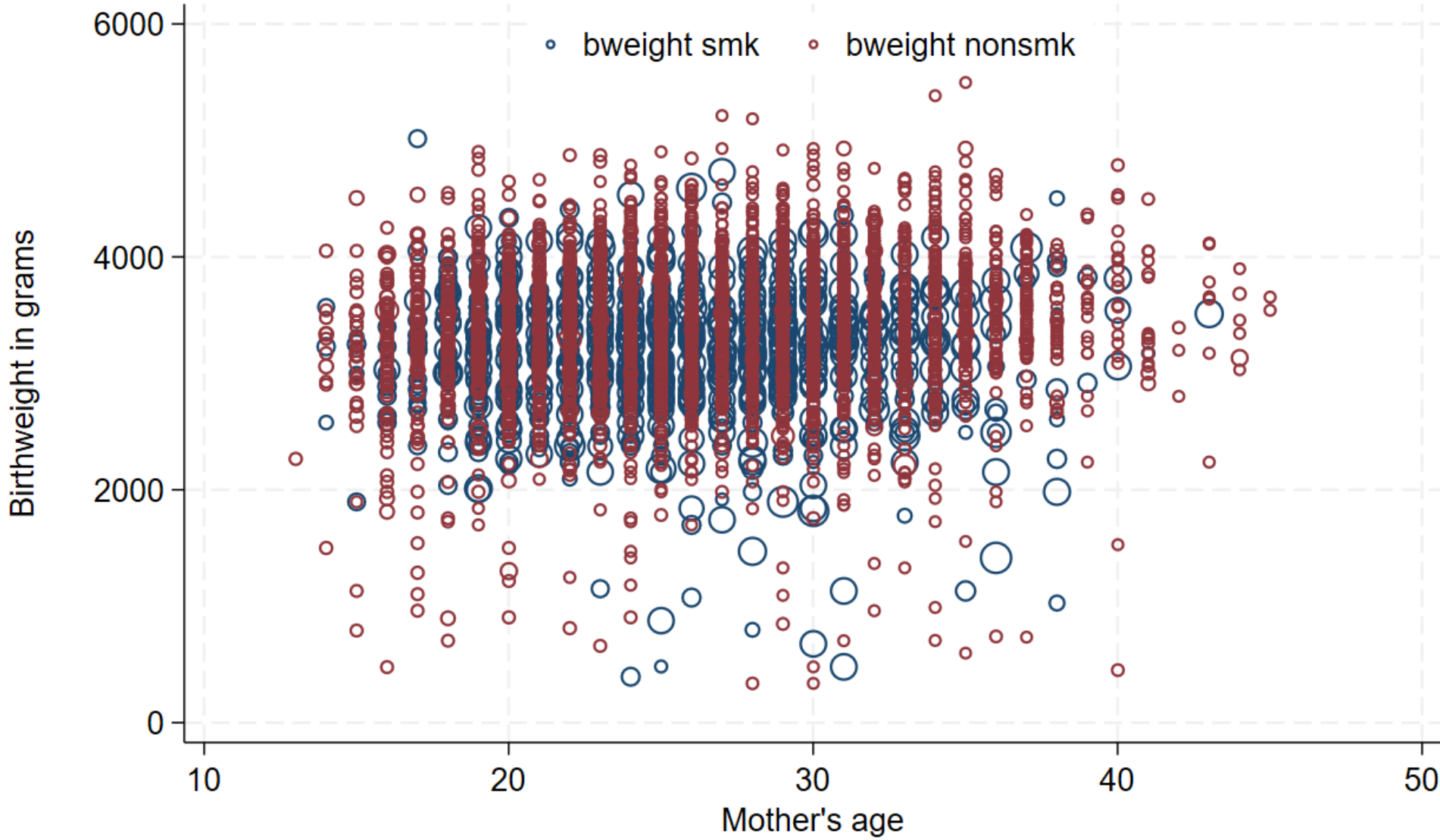
Obtaining weights

```
. logit mbsmoke i.mmarried i.alcohol mage fedu
Iteration 0:  Log likelihood = -2230.7484
Iteration 1:  Log likelihood =  -2063.901
Iteration 2:  Log likelihood = -2050.4497
Iteration 3:  Log likelihood = -2050.3995
Iteration 4:  Log likelihood = -2050.3995
Logistic regression                                     Number of obs =  4,642
                                                        LR chi2(4)    = 360.70
                                                        Prob > chi2   = 0.0000
Log likelihood = -2050.3995                             Pseudo R2     = 0.0808
```

mbsmoke	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
mmarried						
Married	-.883019	.0958196	-9.22	0.000	-1.070822	-.695216
1.alcohol	1.513532	.1779139	8.51	0.000	1.164827	1.862237
mage	-.0041619	.0077716	-0.54	0.592	-.0193939	.01107
fedu	-.072674	.010471	-6.94	0.000	-.0931968	-.0521512
_cons	-.030232	.1995288	-0.15	0.880	-.4213012	.3608373

```
. predict double prob
(option pr assumed; Pr(mbsmoke))
. generate double ps = 1/prob
. replace  ps        = 1/(1-prob) if mbsmoke==0
(3,778 real changes made)
```

IPW graphically



Treatment Models

Treatment Model	$P(\tau \mid w, z)$
`logit`	$exp(w\gamma_\tau) / \{1 + exp(w\gamma_\tau)\}$
`probit`	$\Phi(w\gamma_\tau)$
`hetprobit`	$\Phi(w\gamma_\tau / \exp(z\theta_\tau))$

- Only the logit model is available for multivalued treatments

$$P(\tau \mid w) = \frac{exp(w\gamma_\tau)}{1 + \sum_{k=1}^{K} exp(w\gamma_k)}$$

Treatment Models

Treatment Model	$P(\tau \mid w, z)$
`logit`	$exp(w\gamma_\tau) / \{1 + exp(w\gamma_\tau)\}$
`probit`	$\Phi(w\gamma_\tau)$
`hetprobit`	$\Phi(w\gamma_\tau / \exp(z\theta_\tau))$

- Only the logit model is available for multivalued treatments

$$P(\tau \mid w) = \frac{exp(w\gamma_\tau)}{1 + \sum_{k=1}^{K} exp(w\gamma_k)}$$

IPW estimation

```
. teffects ipw (bweight) (mbsmoke i.mmarried c.mage##c.mage fbaby medu)
Iteration 0:  EE criterion =  1.714e-21
Iteration 1:  EE criterion =  4.748e-27
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-231.7203	25.17975	-9.20	0.000	-281.0717	-182.3689
POmean						
mbsmoke Nonsmoker	3403.527	9.576358	355.41	0.000	3384.757	3422.296

```
. estimates store orig
```

IPW estimation

```
. teffects ipw (bweight) (mbsmoke i.mmarried c.mage##c.mage fbaby medu, probit)
Iteration 0:  EE criterion =  4.621e-21
Iteration 1:  EE criterion =  8.328e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: probit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-230.6886	25.81524	-8.94	0.000	-281.2856	-180.0917
POmean						
mbsmoke Nonsmoker	3403.463	9.571369	355.59	0.000	3384.703	3422.222

IPW estimation

```
. teffects ipw (bweight) (mbsmoke i.mmarried c.mage##c.mage fbaby medu), atet
Iteration 0:  EE criterion =  1.714e-21
Iteration 1:  EE criterion =  4.924e-27
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
mbsmoke (Smoker vs Nonsmoker)	-225.6992	23.7133	-9.52	0.000	-272.1764	-179.222
POmean						
mbsmoke Nonsmoker	3363.359	14.28989	235.37	0.000	3335.351	3391.367

Balancing

- Inspects the validity of the model as a cuasi-experimental tool
- After appropriately weighted (controling for covariates) it is as if we had an experiment
- Raw data is not balanced, after estimation they should be balanced

`tebalance` subcommands

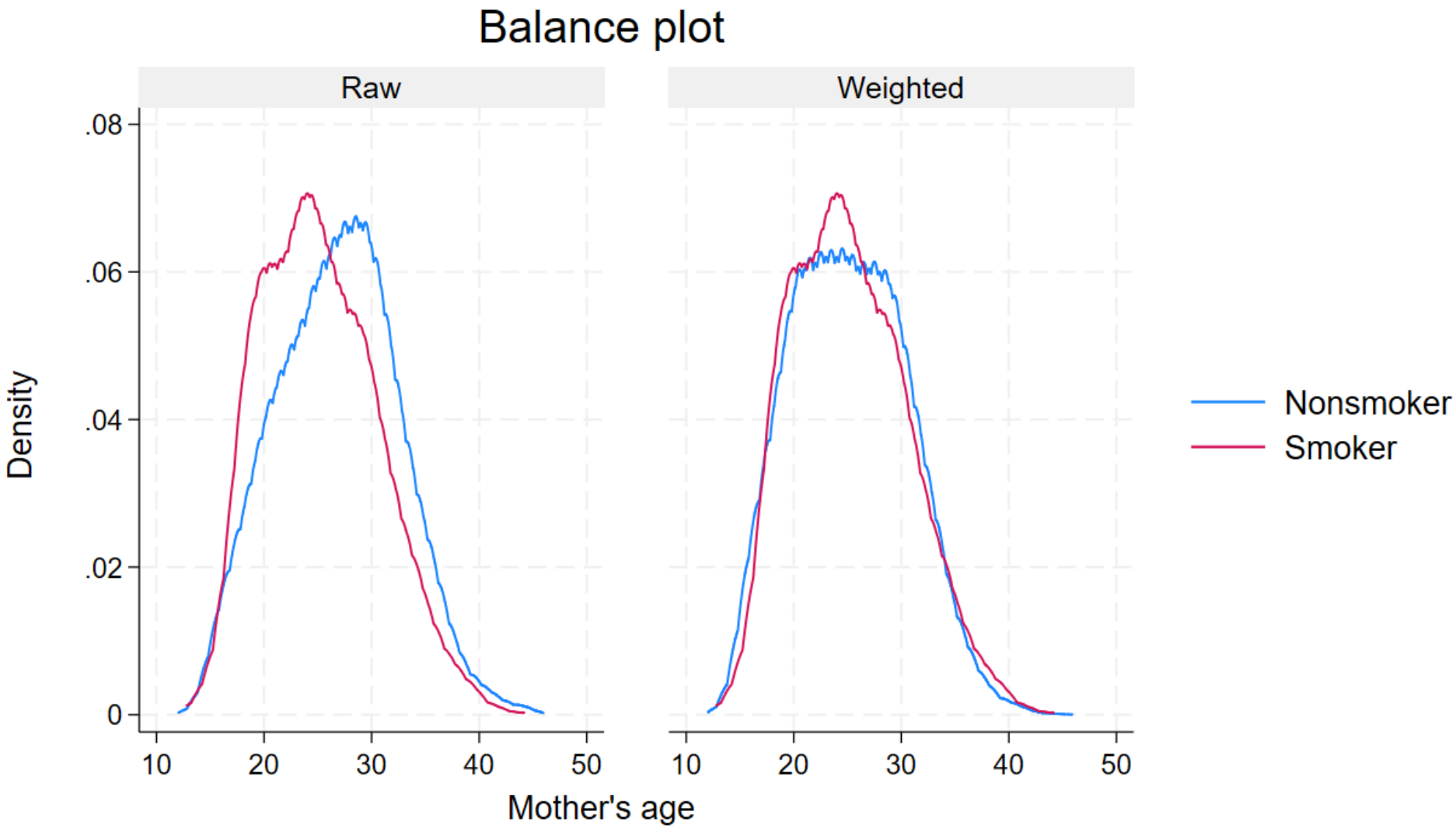
Subcommand	Description
`summarize`	means and standardized variances raw and balanced
`density`	kernel densities of raw and balanced
`box`	boxplot of balanced data
`overid`	overidentification test

tebalance summarize

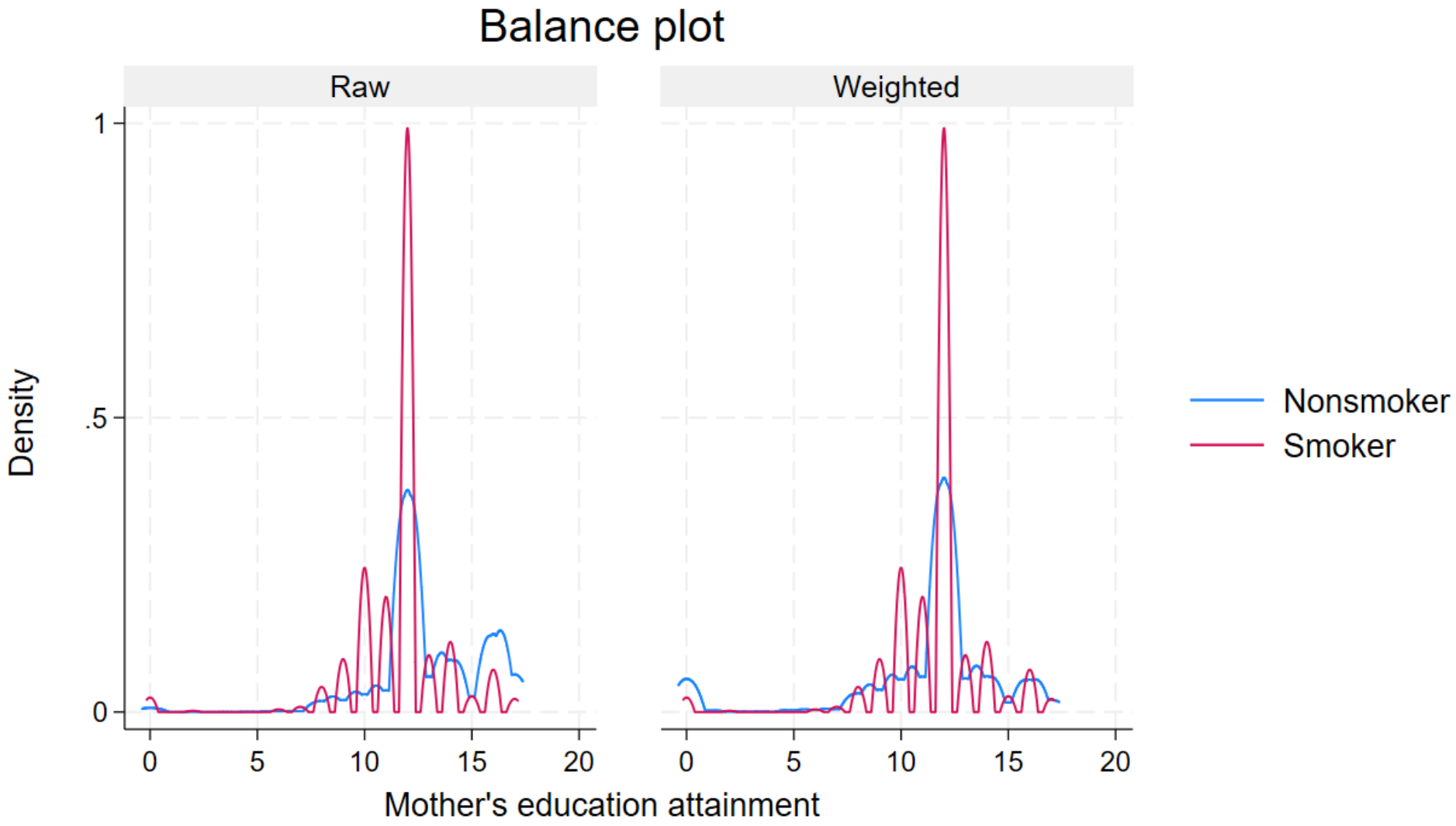
```
. tebalance summarize
Covariate balance summary
                            Raw      Weighted
------------------------------------------------
Number of obs =           4,642       4,642.0
Treated obs   =             864       2,296.2
Control obs   =           3,778       2,345.8
------------------------------------------------
```

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
mmarried				
Married	-.5953009	.0091224	1.335944	1.001066
mage	-.300179	.0053862	.8818025	.9803239
mage#				
mage	-.3028275	.0031547	.8274389	1.022628
fbaby	-.1663271	.006125	.9430944	1.003313
medu	-.5474357	.1350784	.7315846	.3510256

tebalance density mage



tebalance density medu



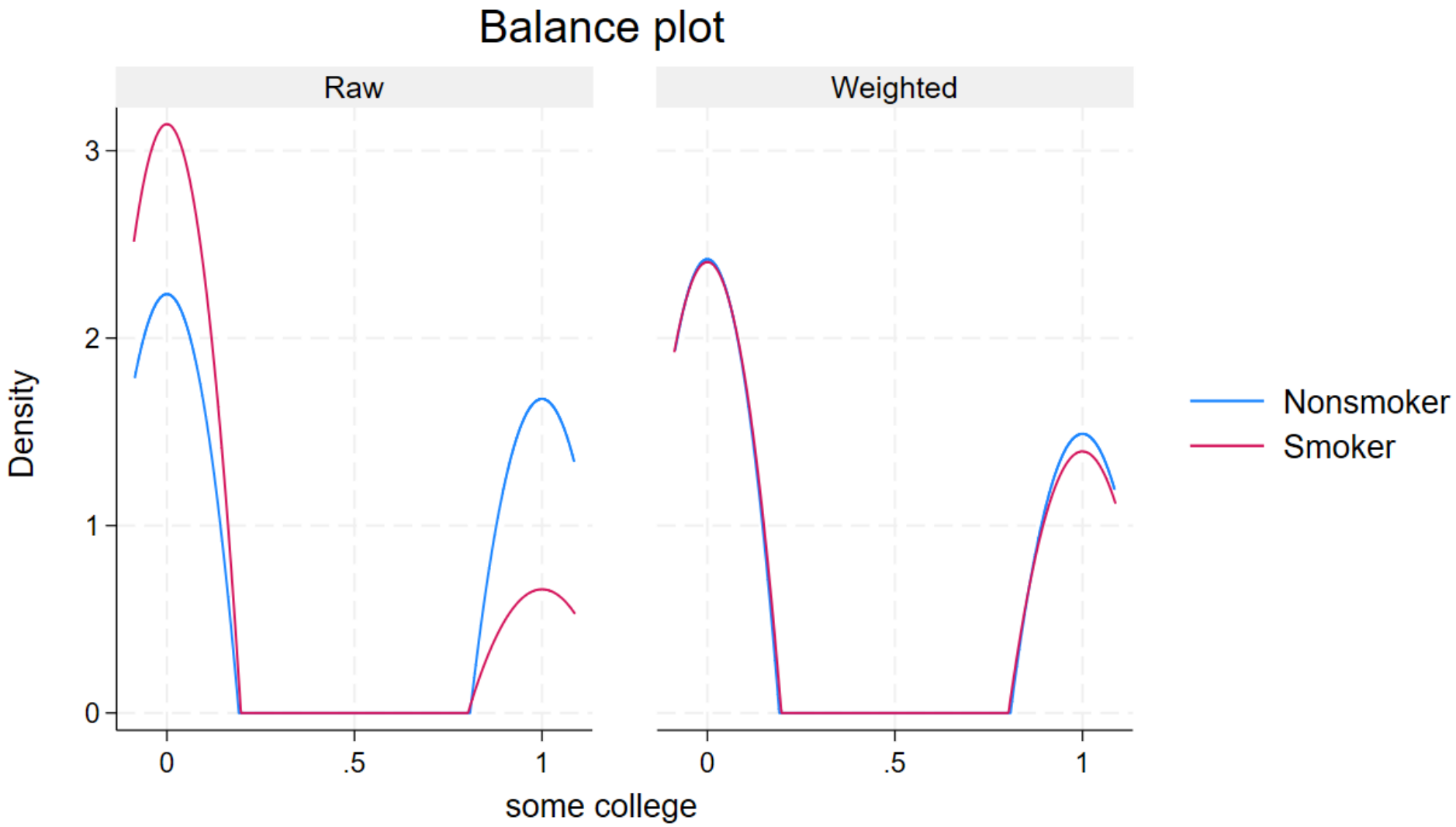
Let's try again

```
. generate mhigh = medu>12
. label var mhigh "some college"
. qui teffects ipw (bweight) (mbsmoke i.mmarried c.mage##c.mage fbaby i.mhigh)
. estimates store new
. tebalance summarize
Covariate balance summary
```

	Raw	Weighted
Number of obs =	4,642	4,642.0
Treated obs =	864	2,319.1
Control obs =	3,778	2,322.9

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
mmarried				
Married	-.5953009	-.0082226	1.335944	1.007121
mage	-.300179	-.0466757	.8818025	1.009618
mage#				
mage	-.3028275	-.0444132	.8274389	1.010684
fbaby	-.1663271	.0310087	.9430944	1.00675
mhigh				
1	-.5783293	-.0284375	.5863736	.9852987

tebalance density mhigh



Overidentification

- Back to GMM. Remember $E(X'e) = 0$
- Our `teffects ipw` model is exactly identified
- Let's add a set of moments. Weighted mean difference
- Test if the additional restrictions are equal to zero statistically
- Overidentification test

`tebalance overid` for original

```
. estimates restore orig
(results orig are active now)
. tebalance overid
Iteration 0:  Criterion =  .01623295
Iteration 1:  Criterion =  .01623417  (backed up)
Iteration 2:  Criterion =   .0175529
Iteration 3:  Criterion =  .01796254
Iteration 4:  Criterion =  .01869533
Iteration 5:  Criterion =  .01878694
Iteration 6:  Criterion =   .0187975
Iteration 7:  Criterion =  .01879778
Iteration 8:  Criterion =  .01879835
Iteration 9:  Criterion =  .01879855
Iteration 10: Criterion =  .01879856
Overidentification test for covariate balance
H0: Covariates are balanced
         chi2(6)      =  52.1956
         Prob > chi2  =   0.0000
```

tebalance overid for new

```
. estimates restore new
(results new are active now)
. tebalance overid
Iteration 0:  Criterion =  .05245216
Iteration 1:  Criterion =  .05250902  (backed up)
Iteration 2:  Criterion =  .05265924
Iteration 3:  Criterion =   .0547431
Iteration 4:  Criterion =  .05502277
Iteration 5:  Criterion =  .05513076
Iteration 6:  Criterion =  .05573705
Iteration 7:  Criterion =  .05608525
Iteration 8:  Criterion =  .05610047
Iteration 9:  Criterion =  .05610049
Iteration 10: Criterion =  .05610049
Overidentification test for covariate balance
H0: Covariates are balanced
         chi2(6)      =  16.8252
         Prob > chi2  =   0.0099
```

Doubly Robust Estimators

- Doubly robust estimators model both the treatment and the outcome model
- These models are interesting because they are consistent even if one of the models is misspecified
- Augmented Inverse Probability Weighting (AIPW) and Inverse Probability Weighted Regression Adjustment(IPWRA) have this property

Double Robust Estimators Inverse Probability Weighted Regression Adjustment (IPWRA)

- Estimate a treatment model and compute inverse-probability weights
- Use the estimated inverse-probability weights and fit weighted regression models of the outcome for each treatment level
- Compute the means of the treatment-specific predicted outcomes

Double Robust Estimators AIPW

- Estimate a treatment model and compute inverse-probability weights
- Estimate separate regression model of the outcome for each treatment level
  - We allow the outcome model to be estimated by nonlinear least squares or weighted nonlinear least squares
- Compute the weighted means of the treatment-specific predicted outcomes, where the weights are the inverse-probability weights computed in step.

Estimation

```
. teffects aipw (bweight i.prenatal1 i.mmarried mage i.fbaby)     ///
>                          (mbsmoke i.mmarried c.mage##c.mage fbaby medu)
Iteration 0:  EE criterion =  1.721e-21
Iteration 1:  EE criterion =  2.239e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-232.0409	25.66973	-9.04	0.000	-282.3527	-181.7292
POmean						
mbsmoke Nonsmoker	3403.457	9.570043	355.64	0.000	3384.7	3422.214

Simulated data

```
. cscript
-------------------------------------------------------------------------------BEGIN
. set obs 5000
Number of observations (_N) was 0, now 5,000.
. set seed 111
.
. generate x1 = rbeta(2,5)
. generate x2 = (rchi2(5)-5)/sqrt(2*5)
. generate a  = int(rbeta(2,3)*3)
. generate e  = rnormal()
. generate e0 = rnormal(0, 2)
. generate e1 = (rchi2(5)-5)/sqrt(2*5)
. generate t  = .5*(1-x1 + x2 -a/2) + e > 0
. generate y0 = -.5 - x1*x2 + x1*a - x1 + x2 + a + e0
. generate y1 = .5 - x1*x2 + x1*a - x1 + x2 + a + e1
. generate y  = t*y1 + (1-t)*y0
```

Simulated data estimation

```
. quietly teffects ra (y c.x1#c.x2 i.a) (t)
. estimates store ra
. quietly teffects ipw (y) (t x1)
. estimates store ipw
. quietly teffects aipw (y c.x1#c.x2 i.a) (t x1 x2 i.a)
. estimates store aipw_ipw
. quietly teffects aipw (y c.x1##(c.x2 i.a)) (t x1)
. estimates store aipw_ra
. estimates table ra ipw aipw*, keep(ATE:) t se
```

Variable	ra	ipw	aipw_ipw	aipw_ra
t				
(1 vs 0)	1.2205215	1.3360659	1.0763364	1.0613425
	.05149324	.05525988	.05263483	.05114926
	23.70	24.18	20.45	20.75

Legend: b/se/t

Nearest Neighbor Matching

- Can be understood as an outcome model within our framework
- Matches the closest individuals in terms of covariates
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Nearest Neighbor Matching

- Can be understood as an outcome model within our framework
- Matches the closest individuals in terms of covariates
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Nearest Neighbor Matching

- Can be understood as an outcome model within our framework
- Matches the closest individuals in terms of covariates
- Matching is done with replacement and ties are allowed
- Bias adjustment with more than one covariate
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Nearest Neighbor Matching

- Can be understood as an outcome model within our framework
- Matches the closest individuals in terms of covariates
- Matching is done with replacement and ties are allowed
- Bias adjustment with more than one covariate
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Nearest neighbor estimation

```
. teffects nnmatch (bweight mage) (mbsmoke)
Treatment-effects estimation                   Number of obs      =      4,642
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Distance metric: Mahalanobis                               max =        260
```

bweight	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-274.8308	22.70001	-12.11	0.000	-319.322	-230.3396

Nearest neighbor estimation: matches

```
. teffects nnmatch (bweight mage) (mbsmoke), generate(nei)
Treatment-effects estimation                   Number of obs      =      4,642
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Distance metric: Mahalanobis                               max =        260
```

bweight	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-274.8308	22.70001	-12.11	0.000	-319.322	-230.3396

```
. list bweight mbsmoke nei1-nei5 in 1/10
```

	bweight	mbsmoke	nei1	nei2	nei3	nei4	nei5
1.	3459	Nonsmoker	60	84	185	244	351
2.	3260	Nonsmoker	317	372	381	573	581
3.	3572	Nonsmoker	196	433	602	646	808
4.	2948	Nonsmoker	11	140	175	201	276
5.	2410	Nonsmoker	317	372	381	573	581
6.	3147	Nonsmoker	20	133	278	303	320
7.	3799	Nonsmoker	20	133	278	303	320
8.	3629	Nonsmoker	60	84	185	244	351
9.	2835	Nonsmoker	144	255	395	560	684
10.	3880	Nonsmoker	173	284	452	458	621

Propensity Score Matching

- Can be classified within the class of treatment models
- Estimate the treatment probabilities (propensity scores)
- Assign values to unobserved outcomes based on observed ones with similar propensity scores
- Estimate ATE
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Propensity Score Matching

- Can be classified within the class of treatment models
- Estimate the treatment probabilities (propensity scores)
- Assign values to unobserved outcomes based on observed ones with similar propensity scores
- Estimate ATE
- These estimators are nondifferentiable therefore the bootstrap is not allowed
- These estimators do not allow for multivalued treatments

Propensity score matching: Estimation

```
. teffects psmatch (bweight) (mbsmoke i.mmarried c.mage##c.mage i.fbaby medu)
Treatment-effects estimation                   Number of obs      =      4,642
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Treatment model: logit                                     max =         74
```

bweight	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-210.9683	32.021	-6.59	0.000	-273.7284	-148.2083

Summarize: Selection on observables

- We model treatment and outcome
- After controlling for covariates it is as if treatment were assigned randomly
- We have parametric, semiparametric, and doubly robust estimators

SELECTION ON UNOBSERVABLES: Difference in Differences

Difference-in-differences (DID)

- One of the most popular causal effects estimators (1855)
- Understand the effect of a treatment on an outcome for the treated group
  - Subsidy on productivity
  - A drug on cholesterol levels
  - An after-school program on GPA
- How is it different from other treatment effects estimators?
  - Observational data for repeated cross-sectional and panel data
  - Identification does not depend on controlling for covariates
  - Identification hinges on control for group and time unobservable characteristics
- Estimate of causal effect of a treatment controlling for unobservables

Difference-in-differences (DID)

- One of the most popular causal effects estimators (1855)
- Understand the effect of a treatment on an outcome for the treated group
  - Subsidy on productivity
  - A drug on cholesterol levels
  - An after-school program on GPA
- How is it different from other treatment effects estimators?
  - Observational data for repeated cross-sectional and panel data
  - Identification does not depend on controlling for covariates
  - Identification hinges on control for group and time unobservable characteristics
- Estimate of causal effect of a treatment controlling for unobservables

Difference-in-differences (DID)

- One of the most popular causal effects estimators (1855)
- Understand the effect of a treatment on an outcome for the treated group
  - Subsidy on productivity
  - A drug on cholesterol levels
  - An after-school program on GPA
- How is it different from other treatment effects estimators?
  - Observational data for repeated cross-sectional and panel data
  - Identification does not depend on controlling for covariates
  - Identification hinges on control for group and time unobservable characteristics
- Estimate of causal effect of a treatment controlling for unobservables

Difference-in-differences (DID)

- One of the most popular causal effects estimators (1855)
- Understand the effect of a treatment on an outcome for the treated group
  - Subsidy on productivity
  - A drug on cholesterol levels
  - An after-school program on GPA
- How is it different from other treatment effects estimators?
  - Observational data for repeated cross-sectional and panel data
  - Identification does not depend on controlling for covariates
  - Identification hinges on control for group and time unobservable characteristics
- Estimate of causal effect of a treatment controlling for unobservables

Stata implementation

- Two-way fixed effects also known as generalized DID (default)
- Allows 2x2 design
- Provides a wide range of standard errors
- Provides diagnostics and tests
- Binary or continuous treatment
- Difference-in-difference-in-differences (DDD) with group and time interactions
- Caveats
  - Treatment effects are homogeneous
  - Standard error literature is large and growing
  - Diagnostics and tests
  - https://friosavila.github.io/playingwithstata/index.html

Stata implementation

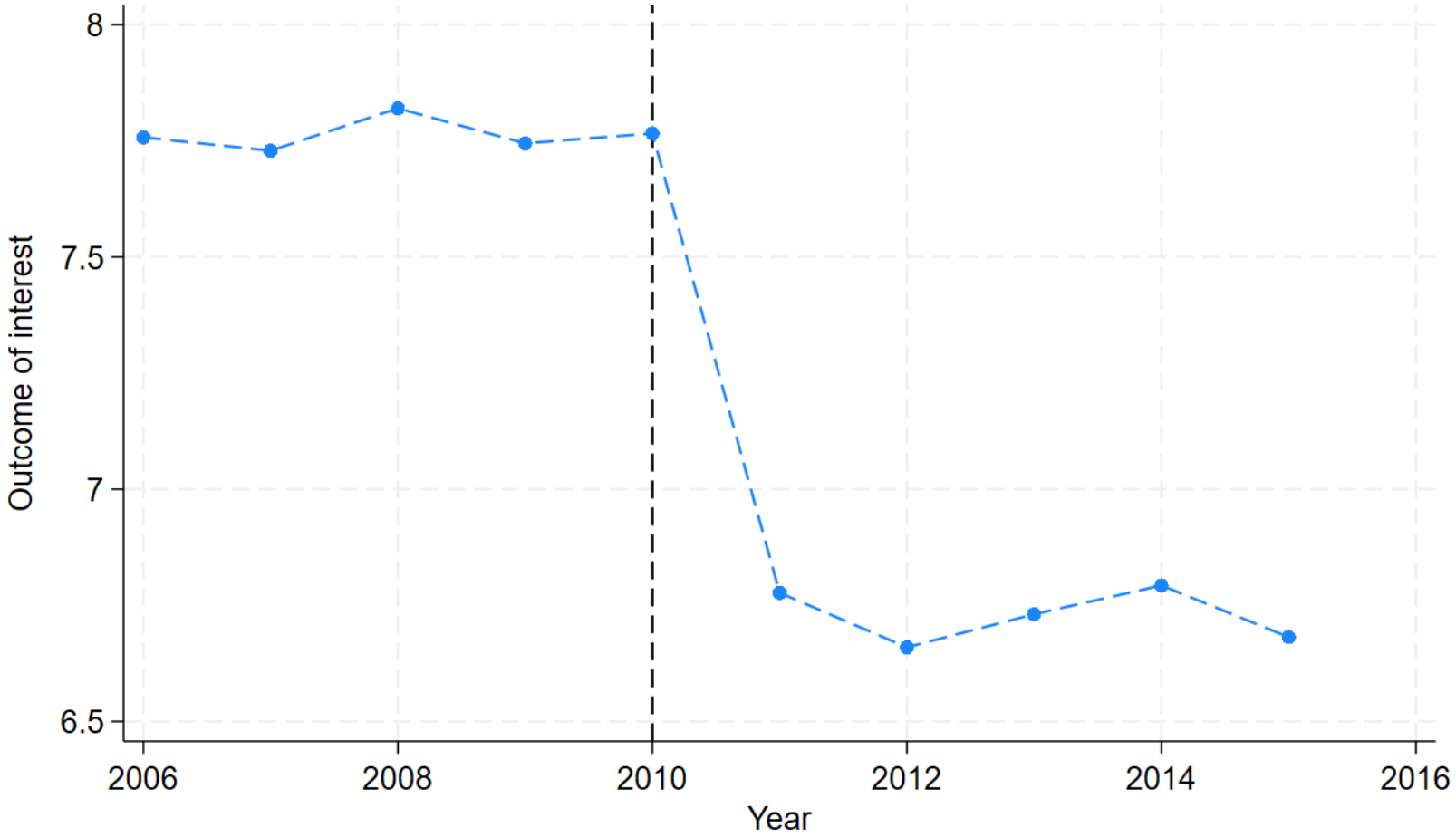
- Two-way fixed effects also known as generalized DID (default)
- Allows 2x2 design
- Provides a wide range of standard errors
- Provides diagnostics and tests
- Binary or continuous treatment
- Difference-in-difference-in-differences (DDD) with group and time interactions
- Caveats
  - Treatment effects are homogeneous
  - Standard error literature is large and growing
  - Diagnostics and tests
  - https://friosavila.github.io/playingwithstata/index.html

Outline

- Basic concepts
- Stata examples

Basic Concepts

Treated group



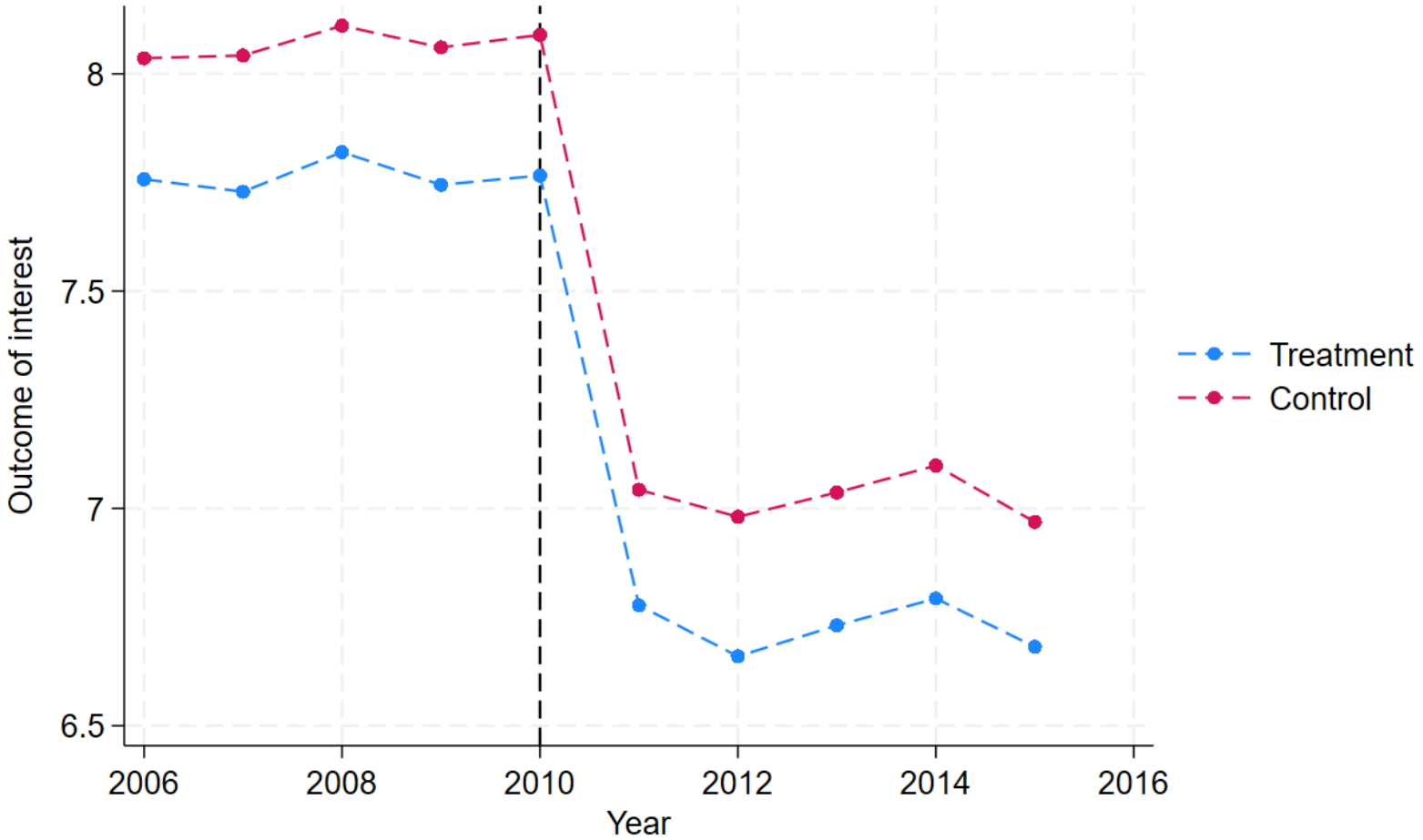
What have we learned

- Clearly there is a change in the outcome after treatment for the treated
- Is it causal?
  - Time specific effects. Another policy. Covid-19.
  - Group unobservable characteristics correlated to treatment. Jargon.
- What can we do?
  - Control for time-specific effects
  - Control for group-specific unobservables (fixed-effects)
  - Use a causal-inference framework

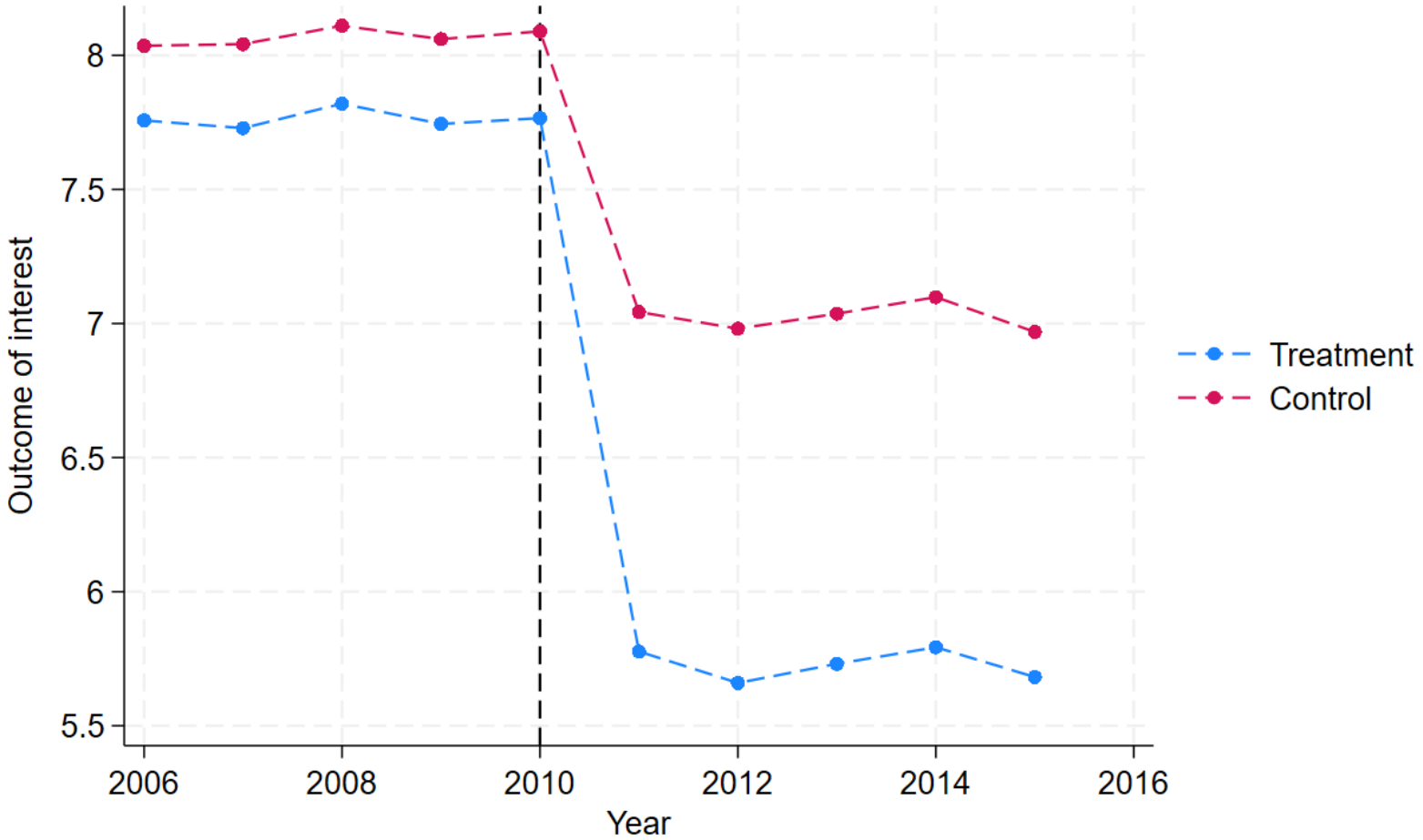
What have we learned

- Clearly there is a change in the outcome after treatment for the treated
- Is it causal?
  - Time specific effects. Another policy. Covid-19.
  - Group unobservable characteristics correlated to treatment. Jargon.
- What can we do?
  - Control for time-specific effects
  - Control for group-specific unobservables (fixed-effects)
  - Use a causal-inference framework

Graphical representation I



Graphical representation II



Card and Krueger (1994)

- Intervention: Increase in the minimum wage
- Group: New Jersey and Pennsylvania
- Outcome: Employment

Linear Framework: Card and Krueger (1994)

- Individuals (i) in a state (s) at two time period $t \in \{0, 1\}$
- Potential outcomes (for now no covariates):

$$\begin{aligned} E\left(y_{ist0}|s,t\right) &= \lambda_t + \gamma_s \\ E\left(y_{ist1}|s,t\right) &= \lambda_t + \gamma_s + \beta \end{aligned}$$

- λ_t is a time effect
- γ_s is a state effect
- y_{ist1} is only observed if state s at time t receives the treatment, an increase in minimum wage, $D_{st} = 1$
- y_{ist0} is only observed if state s at time t does not receive the treatment, $D_{st} = 0$

Linear Framework: Card and Krueger (1994)

- Individuals (i) in a state (s) at two time period $t \in \{0, 1\}$
- Potential outcomes (for now no covariates):

$$\begin{aligned} E\left(y_{ist0}|s,t\right) &= \lambda_t + \gamma_s \\ E\left(y_{ist1}|s,t\right) &= \lambda_t + \gamma_s + \beta \end{aligned}$$

- λ_t is a time effect
- γ_s is a state effect
- y_{ist1} is only observed if state s at time t receives the treatment, an increase in minimum wage, $D_{st} = 1$
- y_{ist0} is only observed if state s at time t does not receive the treatment, $D_{st} = 0$

Linear Framework: Card and Krueger (1994)

- Individuals (i) in a state (s) at two time period $t \in \{0, 1\}$
- Potential outcomes (for now no covariates):

$$\begin{aligned} E\left(y_{ist0}|s,t\right) &= \lambda_t + \gamma_s \\ E\left(y_{ist1}|s,t\right) &= \lambda_t + \gamma_s + \beta \end{aligned}$$

- λ_t is a time effect
- γ_s is a state effect
- y_{ist1} is only observed if state s at time t receives the treatment, an increase in minimum wage, $D_{st} = 1$
- y_{ist0} is only observed if state s at time t does not receive the treatment, $D_{st} = 0$

Card and Krueger (1994) continued

- New Jersey increased minimum wage in April (treatment)
- Neighboring Pennsylvania did not (control)
- Before wage change in February:

$$\begin{aligned} E\left(y_{ist}|PA, Feb\right) &= \lambda_{Feb} + \gamma_{PA} \\ E\left(y_{ist}|NJ, Feb\right) &= \lambda_{Feb} + \gamma_{NJ} \\ E\left(y_{ist}|NJ, Feb\right) - E\left(y_{ist}|PA, Feb\right) &= \gamma_{NJ} - \gamma_{PA} \end{aligned}$$

- The model assumes a common time trend and differing state effects
- Differencing eliminates unobserved time effects

Card and Krueger (1994) continued

- New Jersey increased minimum wage in April (treatment)
- Neighboring Pennsylvania did not (control)
- Before wage change in February:

$$\begin{aligned} E\left(y_{ist}|PA, Feb\right) &= \lambda_{Feb} + \gamma_{PA} \\ E\left(y_{ist}|NJ, Feb\right) &= \lambda_{Feb} + \gamma_{NJ} \\ E\left(y_{ist}|NJ, Feb\right) - E\left(y_{ist}|PA, Feb\right) &= \gamma_{NJ} - \gamma_{PA} \end{aligned}$$

- The model assumes a common time trend and differing state effects
- Differencing eliminates unobserved time effects

Card and Krueger (1994) continued

- After the minimum wage change, in November:

$$E\left(y_{ist}|NJ, Nov\right) - E\left(y_{ist}|PA, Nov\right) = \gamma_{NJ} - \gamma_{PA} + \beta$$

- Difference-in-differences looks at differences before and after the policy:

$$\left[E\left(y_{ist}|., Nov\right) - E\left(y_{ist}|., Nov\right)\right] - \left[E\left(y_{ist}|., Feb\right) - E\left(y_{ist}|., Feb\right)\right]$$

- The difference in these two differences is β
- It is also the average treatment effect on the treated (ATT)

Card and Krueger (1994) continued

- After the minimum wage change, in November:

$$E\left(y_{ist}|NJ, Nov\right) - E\left(y_{ist}|PA, Nov\right) = \gamma_{NJ} - \gamma_{PA} + \beta$$

- Difference-in-differences looks at differences before and after the policy:

$$\left[E\left(y_{ist}|., Nov\right) - E\left(y_{ist}|., Nov\right)\right] - \left[E\left(y_{ist}|., Feb\right) - E\left(y_{ist}|., Feb\right)\right]$$

- The difference in these two differences is β
- It is also the average treatment effect on the treated (ATT)

Parallel trends

- y_{ist0} potential outcome of not being treated
- $D_{st} \equiv D$ if group s was treated at time t, $D \in \{0, 1\}$
- s and t are $\in \{0, 1\}$
- At $t = 0$ no one is treated
- Parallel trends:

$$\underbrace{E\left(y_{ist0} | s = 1, D = 1, t = 1\right)}_{\text{potential outcome of treated in group } s = 1 \text{ had they remained untreated at } t = 1}$$

$$E\left(y_{ist0} | s = 1, D = 1, t = 1\right) - E\left(y_{ist0} | s = 1, D = 1, t = 0\right) =$$
$$E\left(y_{ist0} | s = 0, D = 1, t = 1\right) - E\left(y_{ist0} | s = 0, D = 1, t = 0\right)$$

- Could be conditional on covariates

Parallel trends

- y_{ist0} potential outcome of not being treated
- $D_{st} \equiv D$ if group s was treated at time t, $D \in \{0, 1\}$
- s and t are $\in \{0, 1\}$
- At $t = 0$ no one is treated
- Parallel trends:

$$\underbrace{E\left(y_{ist0} | s = 1, D = 1, t = 1\right)}_{\text{potential outcome of treated in group } s = 1 \text{ had they remained untreated at } t = 1}$$

$$E\left(y_{ist0} | s = 1, D = 1, t = 1\right) - E\left(y_{ist0} | s = 1, D = 1, t = 0\right) =$$
$$E\left(y_{ist0} | s = 0, D = 1, t = 1\right) - E\left(y_{ist0} | s = 0, D = 1, t = 0\right)$$

- Could be conditional on covariates

Observed Outcome and Estimating equation

$$
\begin{aligned}
E\left(y_{ist}|s,t\right) &= D_{st}E\left(y_{ist1}|s,t\right)+\left(1-D_{st}\right)E\left(y_{ist0}|s,t\right)\\
E\left(y_{ist}|s,t\right) &= D_{st}\left(\lambda_t+\gamma_s+\beta\right)+\left(1-D_{st}\right)\left(\lambda_t+\gamma_s\right)\\
E\left(y_{ist}|s,t\right) &= \lambda_t+\gamma_s+D_{st}\beta
\end{aligned}
$$

- This suggests fitting a regression model with a dummy variable D_{st}
- The specification could have regressors

Generalized DID or two-way fixed effects

$$y_{ist} = \gamma_s + \gamma_t + D_{st}\beta + \varepsilon_{ist}$$

- D_{st} is an observation level indicator of treatment $D_{st} \in \{0, 1\}$
- In panel data if individuals are nested in s individual effect absorb state effects
- You may include covariates in the specification above

2 x 2 specification DID

$$y_{its} = \gamma_{\mathbb{1}treated} + \gamma_{\mathbb{1}post} + \mathbb{1}treated \times \mathbb{1}post\beta + \varepsilon_{its}$$

- Works when all units are treated at the same time (balanced)
- This model is nested in the generalized DID
  - $\mathbb{1}treated$ is a linear combination of the group dummies
  - $\mathbb{1}post$ is a linear combination of the time dummies
- This model assumes all post periods and all treatment groups are equivalent.

Alternative specifications

- D_{st} is not binary but continuous (intensity of treatment)
- Differences occur between two groups (differencing two group unobservables)
- DDD or triple differences. It incorporates unobservables from two control groups.
  - Number of parameters is large
  - Identification is more challenging

Standard error computation

Treatment occurs at the group level, state, county, country, etc. and time

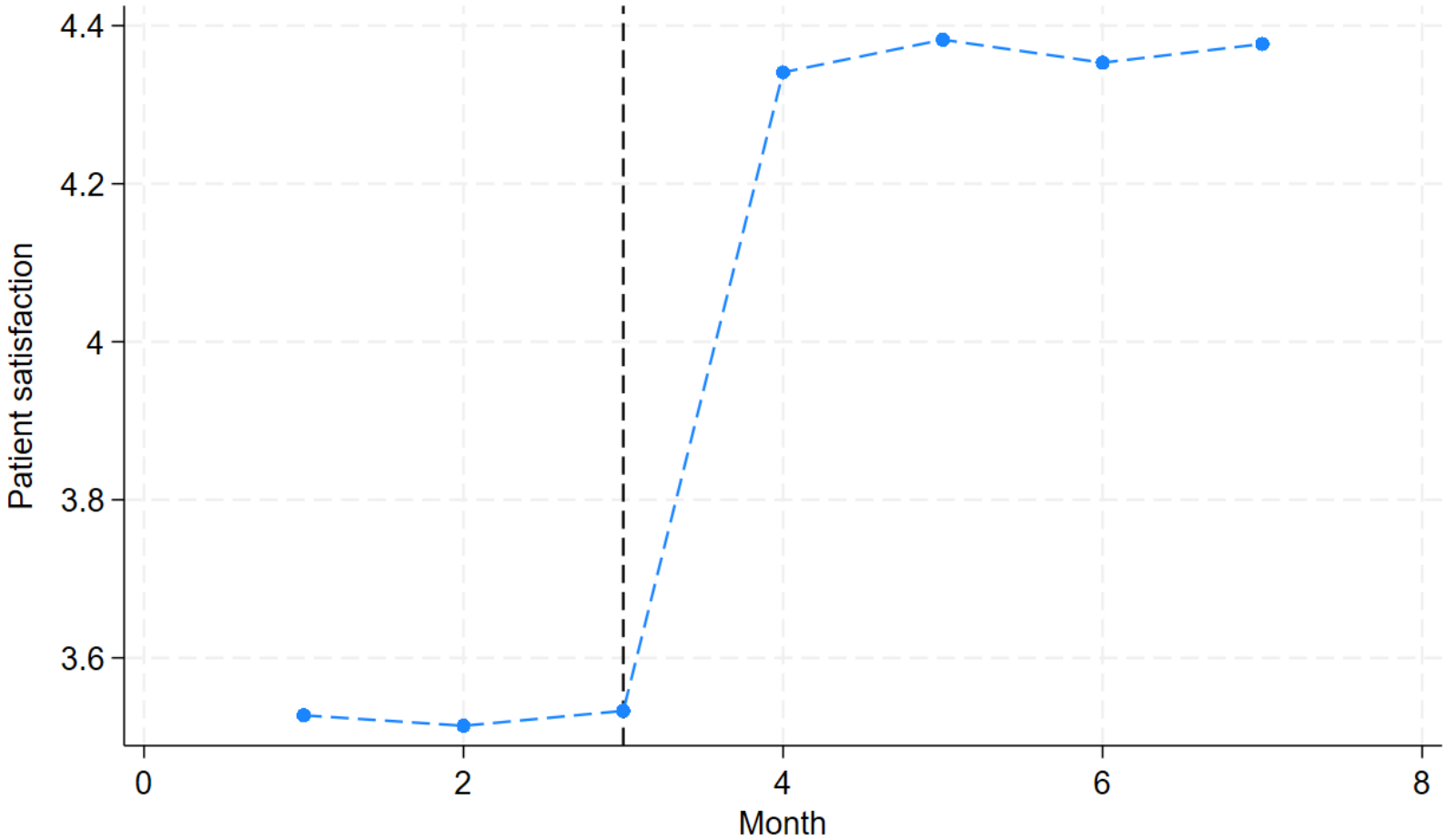
- Cluster at the group level Bertrand, Dufflo, Mullainathan (2004)
- Few number of elements in the group:
  - Donald and Lang (2007) aggregation and other aggregation methods
  - Wild-cluster bootstrap
  - Bias-corrected standard errors with Bell and McCaffrey (2002) degrees of freedom adjustment
  - Other suggestions

Stata Examples

Artificial data

```
 . webuse hospdd, clear
(Artificial hospital admission procedure data)
. describe
Contains data from https://www.stata-press.com/data/r18/hospdd.dta
 Observations:         7,368                  Artificial hospital admission
                                                procedure data
    Variables:             5                  7 Mar 2023 19:52
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
hospital        byte    %9.0g                 Hospital ID
frequency       byte    %9.0g      size       Hospital visit frequency
month           byte    %8.0g      mnth       Month
procedure       byte    %9.0g      pol        Admission procedure
satis           float   %9.0g                 Patient satisfaction score
-------------------------------------------------------------------------------
Sorted by: hospital
```

Graphical representation III



Estimation

```
. didregress (satis) (procedure), group(hospital) time(month)
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

```
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

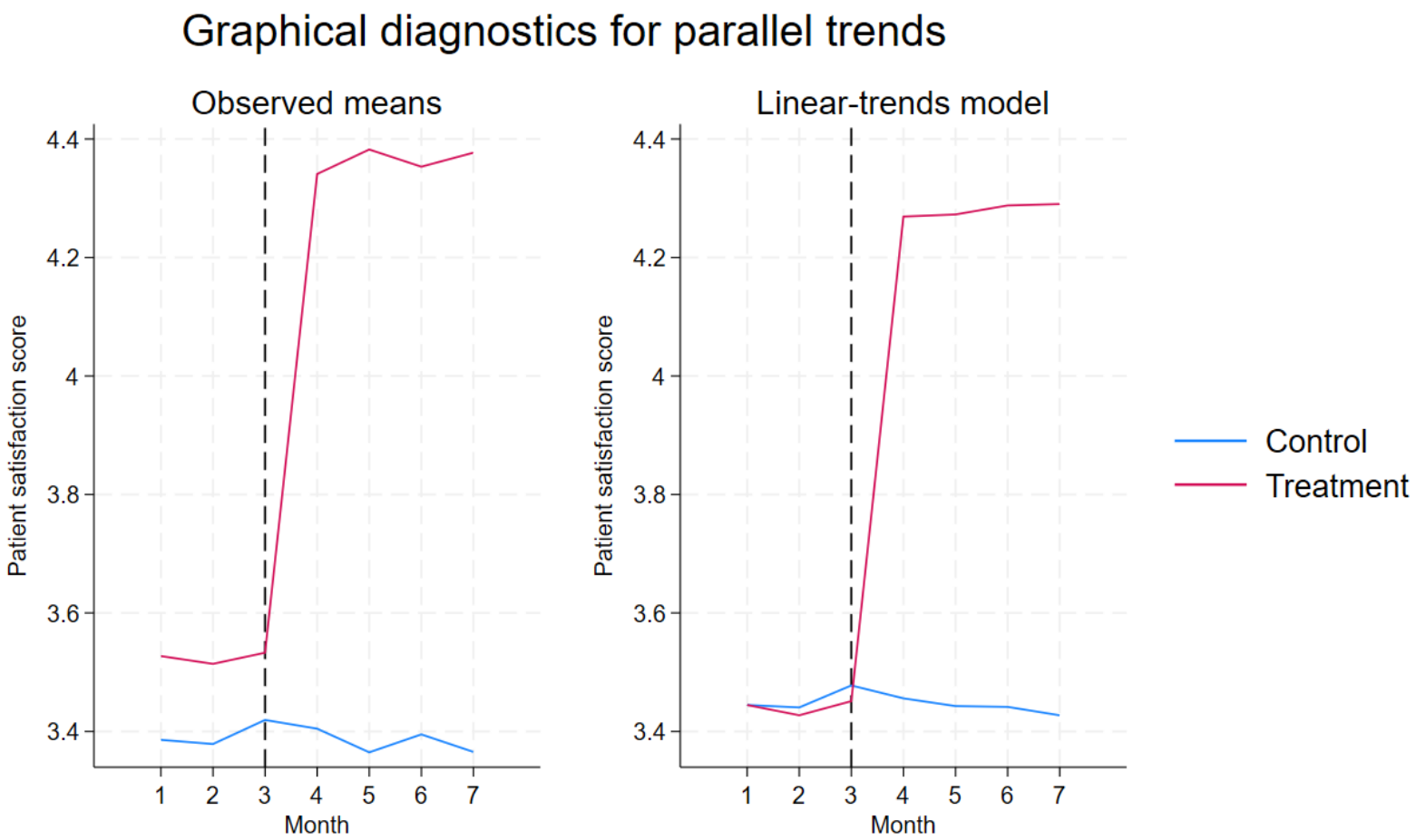
```
Note: ATET estimate adjusted for group effects and time effects.
```

Diagnostic plots

```
estat trendplot
```

- First plot: Mean of the outcome for treated and untreated units
- Second plot: Trend of treated and control groups (group interacted with time)

Diagnostic plots



Tests: `estat ptrends`

```
. estat ptrends
Parallel-trends test (pretreatment time period)
H0: Linear trends are parallel
F(1, 45) =   0.55
Prob > F = 0.4615
```

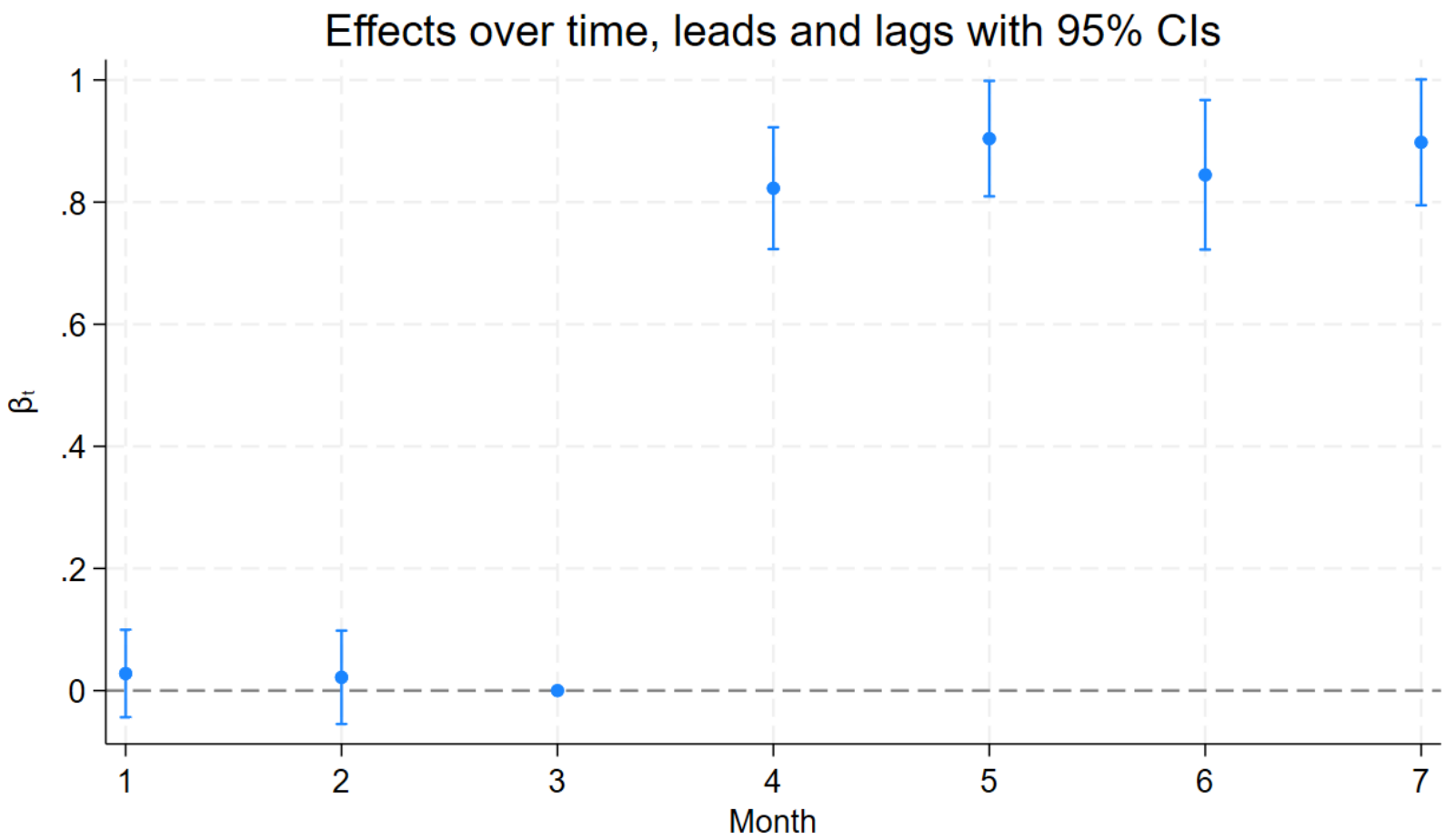
- Augmented model with trends for treated vs. control group before and after treatment. Test if the pretreatment trends are parallel.

Tests: `estat granger`

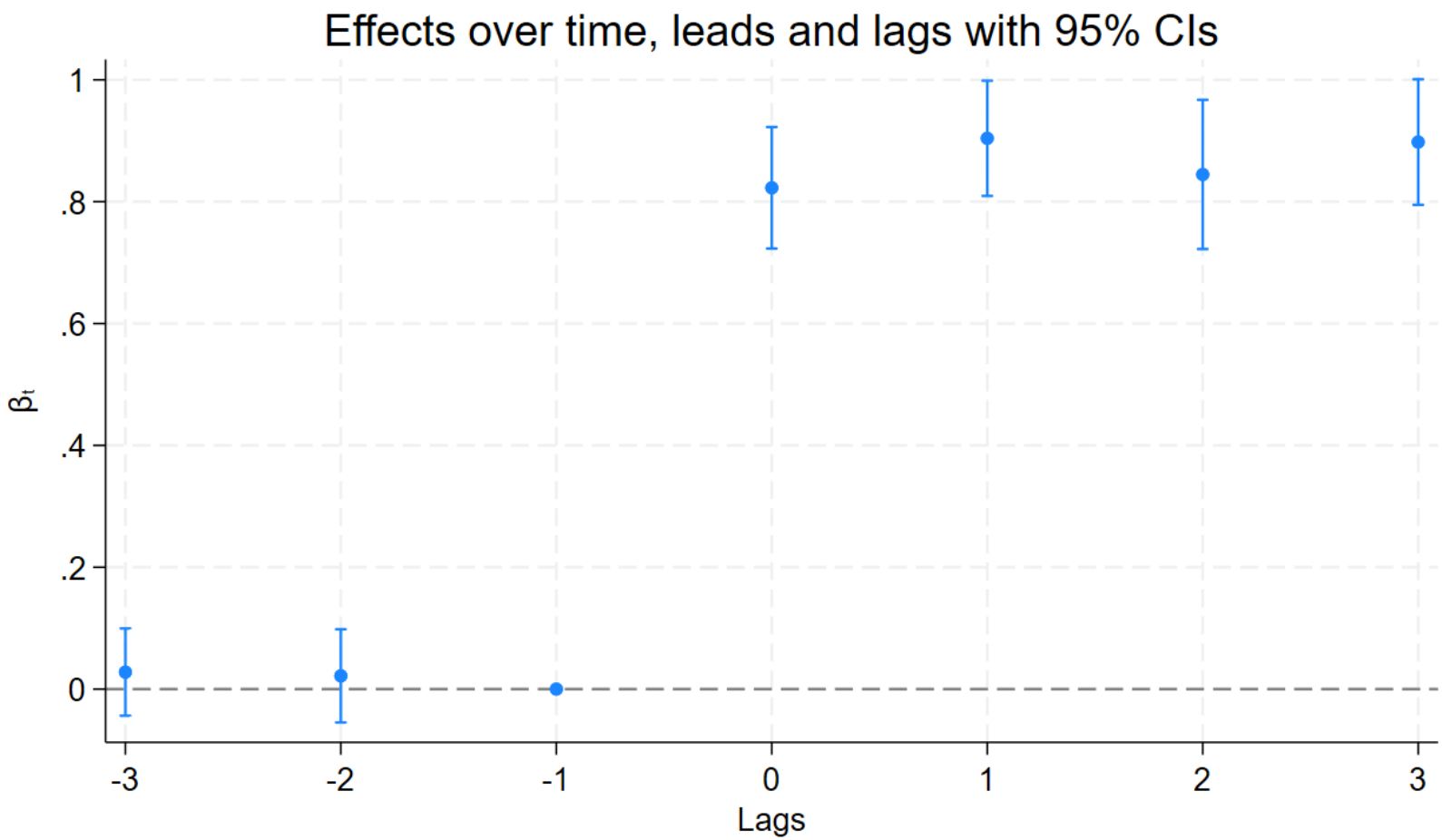
```
. estat granger
Granger causality test
H0: No effect in anticipation of treatment
F(2, 45) =   0.33
Prob > F = 0.7239
```

- Augment the model to include dummies as if treatment had occurred in the past. Test coefficients jointly.

estat grangerplot



estat grangerplot, lagview



estat grangerplot (results)

```
. estat grangerplot, nodraw verbose
Linear regression, absorbing indicators                 Number of obs     =      7,368
Absorbed variable: hospital                             No. of categories =         46
                                                        F(12, 45)         =      94.68
                                                        Prob > F          =     0.0000
                                                        R-squared         =     0.5334
                                                        Adj R-squared     =     0.5298
                                                        Root MSE          =     0.7240
                                          (Std. err. adjusted for 46 clusters in hospital)
------------------------------------------------------------------------------
             |               Robust
       satis |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       month |
    February |   -.007044   .0263953    -0.27   0.791    -.0602068    .0461188
       March |    .0335696  .0255925     1.31   0.196    -.0179764    .0851156
       April |    .0187852  .0250623     0.75   0.457    -.0316927    .0692632
         May |   -.0211152  .0269569    -0.78   0.438    -.0754092    .0331788
        June |    .0091208  .0179016     0.51   0.613     -.026935    .0451766
        July |   -.0203444  .0306266    -0.66   0.510    -.0820296    .0413407
             |
       _lead3 |    .027897   .035569     0.78   0.437    -.0437426    .0995367
       _lead2 |    .0217322  .0380076    0.57   0.570     -.054819    .0982833
        _lag0 |    .8228153  .0494933   16.62   0.000     .7231307    .9224999
        _lag1 |    .9040498  .0469682   19.25   0.000     .8094511    .9986486
        _lag2 |     .844724  .0608006   13.89   0.000     .7222654    .9671826
        _lag3 |    .8978885  .0511588   17.55   0.000     .7948494    1.000928
        _cons |    3.433074  .0198449  173.00   0.000     3.393104    3.473044
------------------------------------------------------------------------------
```

A 2 × 2 specification

- Create dummy variables for treated group and post time period
- Tell `didregress` not to include group and time effects
- Add dummies to the outcome equation

```
. bysort hospital: egen treated = mean(procedure)
. replace treated = 1 if treated>0
(3,064 real changes made)
. generate post = 0
. replace post = 1 if month>3
(3,684 real changes made)
```

A 2 × 2 specification

- Create dummy variables for treated group and post time period
- Tell `didregress` not to include group and time effects
- Add dummies to the outcome equation

```
. bysort hospital: egen treated = mean(procedure)
. replace treated = 1 if treated>0
(3,064 real changes made)
. generate post = 0
. replace post = 1 if month>3
(3,684 real changes made)
```

A 2 × 2 specification

```
. didregress (satis treated post) (procedure),    ///
>         group(hospital) time(month) nogteffects
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

```
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	.0320051	26.50	0.000	.7835263	.9124494

```
Note: ATET estimate adjusted for covariates.
```

Difference-in-difference-in-differences DDD

- Augmented DID
- Selection on unobservables provides identification
- What if there are unobservables that vary at the group and time level
- Find a new group not exposed to treatment but exposed to the problematic time-varying confounder
- Subtract the effect of that group from the original DID
- In our example think about individual's frequency of visit affecting satisfaction

Difference-in-difference-in-differences DDD

- Augmented DID
- Selection on unobservables provides identification
- What if there are unobservables that vary at the group and time level
- Find a new group not exposed to treatment but exposed to the problematic time-varying confounder
- Subtract the effect of that group from the original DID
- In our example think about individual's frequency of visit affecting satisfaction

Difference-in-difference-in-differences DDD

- Augmented DID
- Selection on unobservables provides identification
- What if there are unobservables that vary at the group and time level
- Find a new group not exposed to treatment but exposed to the problematic time-varying confounder
- Subtract the effect of that group from the original DID
- In our example think about individual's frequency of visit affecting satisfaction

DDD preparing my data

```
. generate hightrt = procedure==1 & (frequency==3 | frequency==4)
. label define trt 0 "Untreated" 1 "Treated"
. label values hightrt trt
```

DDD estimation

```
. didregress (satis) (hightrt), group(hospital frequency) time(month)
  (output omitted)
Treatment and time information
Time variable: month
Control:       hightrt = 0
Treatment:     hightrt = 1
```

	Control	Treatment
Group		
hospital	28	18
frequency	2	2
Time		
Minimum	1	4
Maximum	1	4

```
Triple-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						
hightrt (Treated vs Untreated)	.764154	.0402603	18.98	0.000	.6830655	.8452425

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

Other estimation alternatives

- `didregress (y x1 ... xk) (c, continuous), ...`
- `didregress (y ...) (d...), group(g1 g2)`
- `xtdidregress (y x1 ... xk) (d), group(`*`groupvar`*`) time(`*`timevar`*`)`

Standard error considerations

- Default standard errors are cluster robust standard errors at the group level BDM (2004)
- `didregress` is equivalent to `areg` considers group fixed effects as regressors in the degrees of freedom adjustment
- `xtdidregress` is equivalent to `xtreg` does not consider group fixed effects as regressors
- When the number of elements per groups (states, counties, countries) is small cluster robust standard errors do not work well. Alternatives are:
  - Wild cluster bootstrap
  - Bias corrected standard errors
  - Aggregation methods

Wild-cluster bootstrap

- Covariates remain the same across iteration
- We impose the null hypothesis of $ATET = 0$
- What changes is the weights given to residuals at each iteration $\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$, $\tilde{\beta}$, and $\tilde{\varepsilon} = \widehat{\varepsilon} * w$
- No standard errors are computed (rely on normal approximation)
- P-values and confidence intervals are computed
- Algorithm computes p-values and then solves a bisection-algorithm to get CI
- Problem to find CI upper bound and CI lower bound are two separate optimization problems
- Speed is substantially improved

Wild-cluster bootstrap

- Covariates remain the same across iteration
- We impose the null hypothesis of $ATET = 0$
- What changes is the weights given to residuals at each iteration $\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$, $\tilde{\beta}$, and $\tilde{\varepsilon} = \widehat{\varepsilon} * w$
- No standard errors are computed (rely on normal approximation)
- P-values and confidence intervals are computed
- Algorithm computes p-values and then solves a bisection-algorithm to get CI
- Problem to find CI upper bound and CI lower bound are two separate optimization problems
- Speed is substantially improved

Error weights

Error weight	Formula
`rademacher`	-1 with pr 0.5 and 1 with pr 0.5
`mammen`	$1-\phi$ with pr $\phi/\sqrt{(5)}$, ϕ otherwise, $\phi=(1+\sqrt{5})/2$
`webb`	-$\sqrt{3/2}$, -$\sqrt{2/2}$, -$\sqrt{1/2}$, $\sqrt{1/2}$, $\sqrt{2/2}$, $\sqrt{3/2}$ pr $1/6$
`normal`	standard normal
`gamma`	shape parameter 4 scale parameter $1/2$

Wildbootstrap I

```
. didregress (satis) (procedure), ///
>         group(hospital) time(month) wildbootstrap(rseed(111))
Performing 1,000 replications for p-value for constraint
  procedure = 0 ...
Computing confidence interval for procedure
  Lower bound: ..... done (5)
  Upper bound: .........10.. done (12)
  (output omitted)
```

Wildbootstrap II

```
. didregress (satis) (procedure), ///
>         group(hospital) time(month) wildbootstrap(rseed(111))
Performing 1,000 replications for p-value for constraint
  procedure = 0 ...
Computing confidence interval for procedure
  Lower bound: ..... done (5)
  Upper bound: .........10.. done (12)
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

```
DID with wild-cluster bootstrap inference          Number of obs = 7,368
                                                   Replications  = 1,000
Data type:    Repeated cross-sectional
Error weight: rademacher
```

satis	Coefficient	t	P>|t|	[95% conf.	interval]
ATET					
procedure (New vs Old)	.8479879	26.41	0.000	.7806237	.9157614

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

Bias-corrected standard errors

- Cluster generalization of HC2 (scale residuals inverse of square of diagonals from projection matrix)
- Bell and McCaffrey (2002) suggest a degrees of freedom adjustment (per parameter)

Bias-corrected SEs

```
. didregress (satis) (procedure), group(hospital) time(month) vce(hc2)
Computing degrees of freedom ...
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

Difference-in-differences regression — Number of obs = 7,368
Number of clusters = 46

Data type: Repeated cross-sectional

satis	Coefficient	Robust HC2 std. err.	t	P>|t|	[95% conf.	interval]
ATET						
procedure (New vs Old)	.8479879	.0325552	26.05	0.000	.7819941	.9139816

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

Degrees of freedom adjustment

```
. mat list r(table)
r(table)[9,9]
              ATET:   Controls:   Controls:   Controls:   Controls:   Controls:
             r1vs0.         1b.          2.          3.          4.          5.
          procedure       month       month       month       month       month
     b    .84798786           0  -.00960766   .02196858  -.00328387  -.00940274
    se    .03255515           .   .01836738   .01817606   .02210113   .02325151
     t    26.047731           .  -.52308262   1.2086544  -.14858393  -.40439255
pvalue    3.558e-25           .   .60348306   .23310851   .88254581   .68783978
    ll     .7819941           .  -.04660147  -.01463989  -.04779783  -.05623368
    ul    .91398163           .   .02738615   .05857705   .04123009    .0374282
    df    36.496106          45          45          45          45          45
  crit    2.0271372   2.0141034   2.0141034   2.0141034   2.0141034   2.0141034
 eform            0           0           0           0           0           0
           Controls:   Controls:   Controls:
                  6.          7.
              month       month       _cons
     b  -.00383754  -.01119415    3.444675
    se   .01906173    .0230133   .01140018
     t   -.2013216  -.48642083   302.15965
pvalue   .84135438   .62902945   4.517e-76
    ll  -.04222984   -.0575453   3.4217139
    ul   .03455476   .03515701   3.4676362
    df          45          45          45
  crit   2.0141034   2.0141034   2.0141034
 eform            0           0           0
```

Aggregation methods

$$
\begin{aligned}
y_{its} &= \gamma_s + \gamma_t + z_{1ist}\beta_1 + z_{2st}\beta_2 + D_{st}\delta + \varepsilon_{ist} \\
y_{ist} &= z_{1ist}\beta_2 + C_{st} + \varepsilon_{ist} \\
\widehat{C}_{st} &= z_{2st}\beta_2 + D_{st}\delta + \nu_{st}
\end{aligned}
$$

- Obtain $\widehat{C}_{st}$
- Aggregate at the s, t level and regress
  - `dlang, constant`: regress $\widehat{C}_{st}$ on z_{2st}, D_{st} and time and group fixed effects, degrees of freedom are a function of the number of observations in the collapsed sample and the number of regressors in z_{2st} and D_{st}.
  - `standard`: regress $\widehat{C}_{st}$ on z_{2st}, D_{st}
  - `dlang, varying`: $\widehat{C}_{st}$ is the constant of a regression of each group defined by st, i.e. β_1 is not constant but varying.

aggregate(dlang)

```
. didregress (satis) (procedure), group(hospital) time(month) aggregate(dlang)
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

```
Difference-in-differences regression                    Number of obs = 322
Data type:   Repeated cross-sectional
Aggregation: DonaldLang
```

satis	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ATET						
procedure (New vs Old)	.8500467	.0255727	33.24	0.000	.7997311	.9003623

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

aggregate(standard)

```
. didregress `specs´, group(hospital) time(month) aggregate(standard) vce(hc2)
Computing degrees of freedom ...
Treatment and time information
Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

```
Difference-in-differences regression                 Number of obs    = 322
                                                     No. of clusters =  46

Data type:    Repeated cross-sectional
Aggregation:  Standard
```

satis	Coefficient	Robust HC2 std. err.	t	P>|t|	[95% conf.	interval]
ATET						
procedure (New vs Old)	.8500467	.0329513	25.80	0.000	.7832444	.916849

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

Selection on unobservables: Summarizing

- DID and DDD estimation for cross-sectional and panel-data
- Graphical diagnostics and tests to validate identification strategy
- Standard errors for situations with the number of groups is small
- Just a first step from which we will build

Context: Treatment-effects estimation in Stata

The effect of a treatment or exposure on an outcome

- Average treatment effect (ATE) and average treatment effect on the treated (ATET)
- `teffects`: cross-sectional data selection on observables
- `didregress` and `xtdidregress`: repeated-measures selection on unobservables
- Estimation, inference, visualization, diagnostics, and tests
- One treatment effect. That assumes that the treatment does not change over groups or time. Homogeneous treatment effects
- New: Heterogeneous treatment effects for selection on unobservables. Heterogeneous Difference in differences (DID). Multiple ATETs

Context: Treatment-effects estimation in Stata

The effect of a treatment or exposure on an outcome

- Average treatment effect (ATE) and average treatment effect on the treated (ATET)
- `teffects`: cross-sectional data selection on observables
- `didregress` and `xtdidregress`: repeated-measures selection on unobservables
- Estimation, inference, visualization, diagnostics, and tests
- One treatment effect. That assumes that the treatment does not change over groups or time. Homogeneous treatment effects
- New: Heterogeneous treatment effects for selection on unobservables. Heterogeneous Difference in differences (DID). Multiple ATETs

Context: Treatment-effects estimation in Stata

The effect of a treatment or exposure on an outcome

- Average treatment effect (ATE) and average treatment effect on the treated (ATET)
- `teffects`: cross-sectional data selection on observables
- `didregress` and `xtdidregress`: repeated-measures selection on unobservables
- Estimation, inference, visualization, diagnostics, and tests
- One treatment effect. That assumes that the treatment does not change over groups or time. Homogeneous treatment effects
- New: Heterogeneous treatment effects for selection on unobservables. Heterogeneous Difference in differences (DID). Multiple ATETs

Context: Treatment-effects estimation in Stata

The effect of a treatment or exposure on an outcome

- Average treatment effect (ATE) and average treatment effect on the treated (ATET)
- `teffects`: cross-sectional data selection on observables
- `didregress` and `xtdidregress`: repeated-measures selection on unobservables
- Estimation, inference, visualization, diagnostics, and tests
- One treatment effect. That assumes that the treatment does not change over groups or time. Homogeneous treatment effects
- New: Heterogeneous treatment effects for selection on unobservables. Heterogeneous Difference in differences (DID). Multiple ATETs

Starting point: Why heterogeneous treatment effects?

- DID with different groups treated at different times
- With multiple treatment times the ATET for DID was obtained via

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

- β_1 is the ATET
- A generalization of the well understood 2 by 2 model.
- Two-way fixed-effects (TWFE) was criticized in the last six years
- Why:
  1. Homogeneity (well understood)
  2. Hidden cost of generalizing of 2 by 2

Starting point: Why heterogeneous treatment effects?

- DID with different groups treated at different times
- With multiple treatment times the ATET for DID was obtained via

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

- β_1 is the ATET
- A generalization of the well understood 2 by 2 model.
- Two-way fixed-effects (TWFE) was criticized in the last six years
- Why:
  1. Homogeneity (well understood)
  2. Hidden cost of generalizing of 2 by 2

Starting point: Why heterogeneous treatment effects?

- DID with different groups treated at different times
- With multiple treatment times the ATET for DID was obtained via

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

- β_1 is the ATET
- A generalization of the well understood 2 by 2 model.
- Two-way fixed-effects (TWFE) was criticized in the last six years
- Why:
  1. Homogeneity (well understood)
  2. Hidden cost of generalizing of 2 by 2

What happens when we go beyond the 2 by 2 model?

- ATET might be different for groups treated at different times
- Within groups, ATET might change over time
- If this is the case:

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

- Weighted average
  - Borusyak, Jaravel, and Spiess (2018)
  - de Chaisemartin and D'Haultfoeuille (2020)
  - Goodman-Bacon (2021) (`estat bdecomp`)

What happens when we go beyond the 2 by 2 model?

- ATET might be different for groups treated at different times
- Within groups, ATET might change over time
- If this is the case:

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

- Weighted average
  - Borusyak, Jaravel, and Spiess (2018)
  - de Chaisemartin and D'Haultfoeuille (2020)
  - Goodman-Bacon (2021) (`estat bdecomp`)

Bacon decomposition

- β_1 is a weighted average of (2 by 2) estimates

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

- 2 by 2 estimates using:
  1. Never treated groups
  2. Early treated groups
  3. Later treated groups
- Some of the comparisons are valid comparisons. Some are not.
- Validity: satisfying a parallel-trends assumption.

Bacon decomposition

- β_1 is a weighted average of (2 by 2) estimates

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

- 2 by 2 estimates using:
  1. Never treated groups
  2. Early treated groups
  3. Later treated groups
- Some of the comparisons are valid comparisons. Some are not.
- Validity: satisfying a parallel-trends assumption.

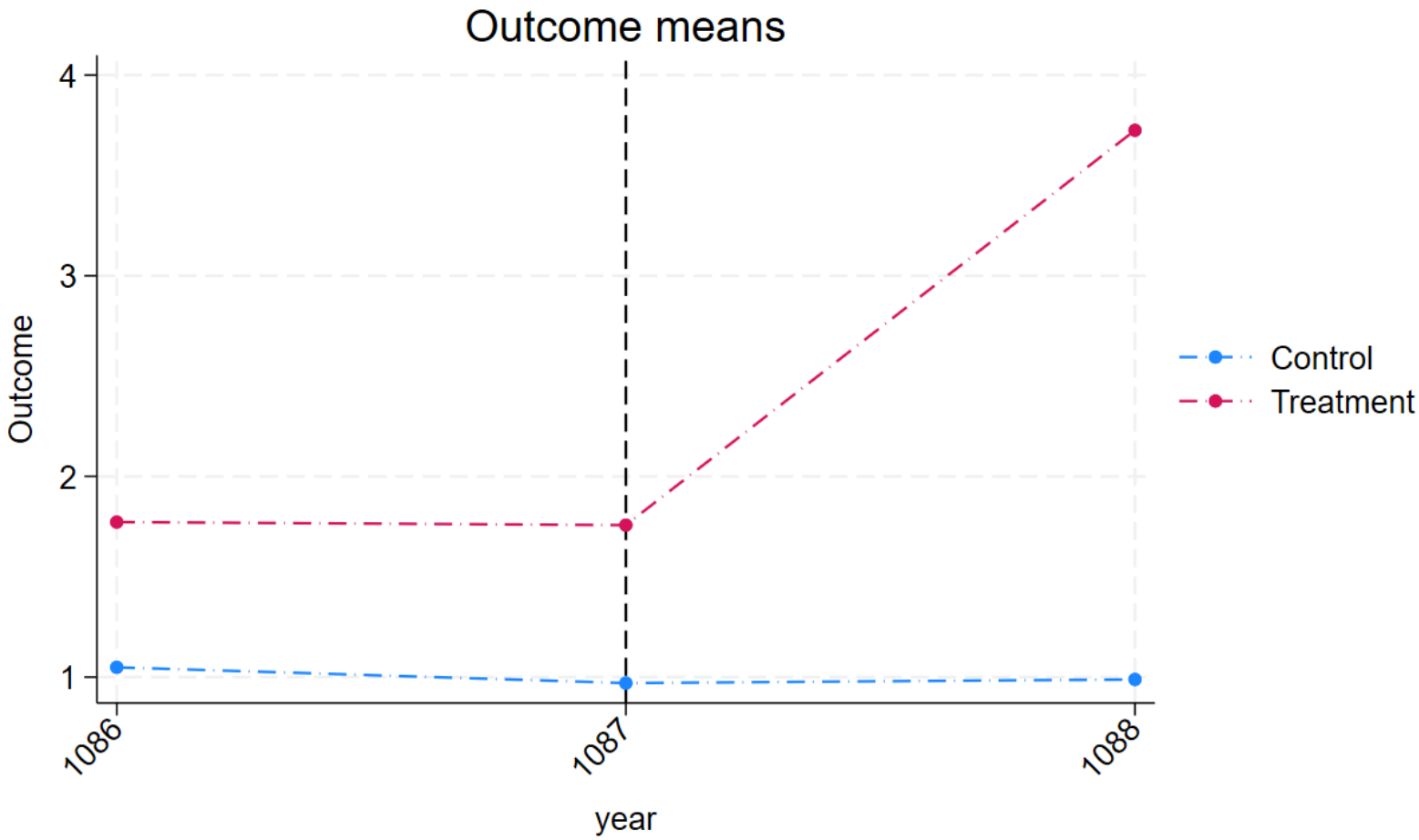
Let's see it work

- What a 2 by 2 model looks like
- What a DID model with multiple treatment times looks like graphically
  - homogeneous treatment
  - heterogeneous treatment
- What the Bacon decomposition tells us
  - homogeneous treatment
  - heterogeneous treatment
- Build our understanding of heterogeneous DID

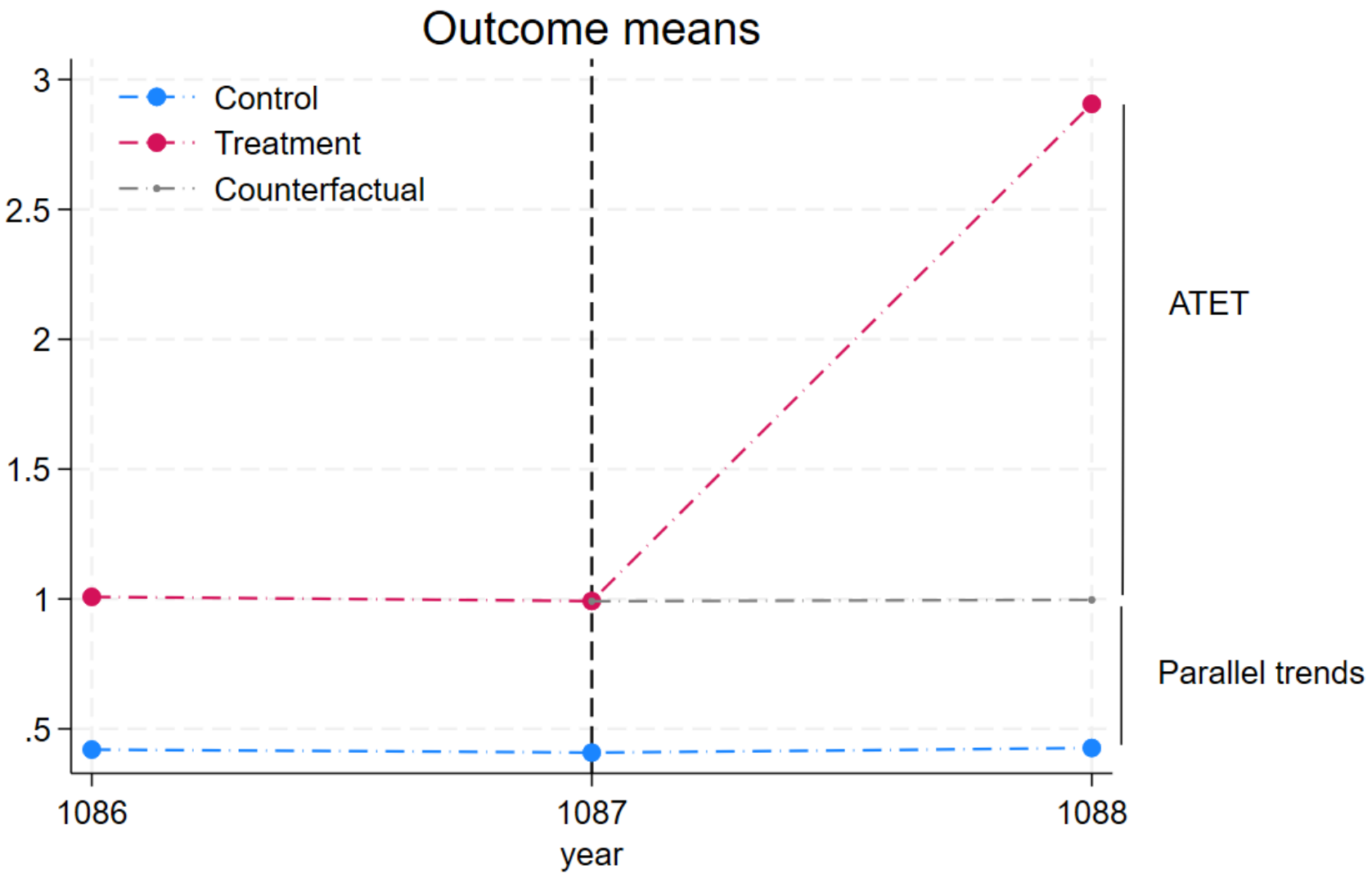
Let's see it work

- What a 2 by 2 model looks like
- What a DID model with multiple treatment times looks like graphically
  - homogeneous treatment
  - heterogeneous treatment
- What the Bacon decomposition tells us
  - homogeneous treatment
  - heterogeneous treatment
- Build our understanding of heterogeneous DID

2 by 2 framework: `estat trendplots`



2 by 2 framework: Parallel(Common)-trends assumption



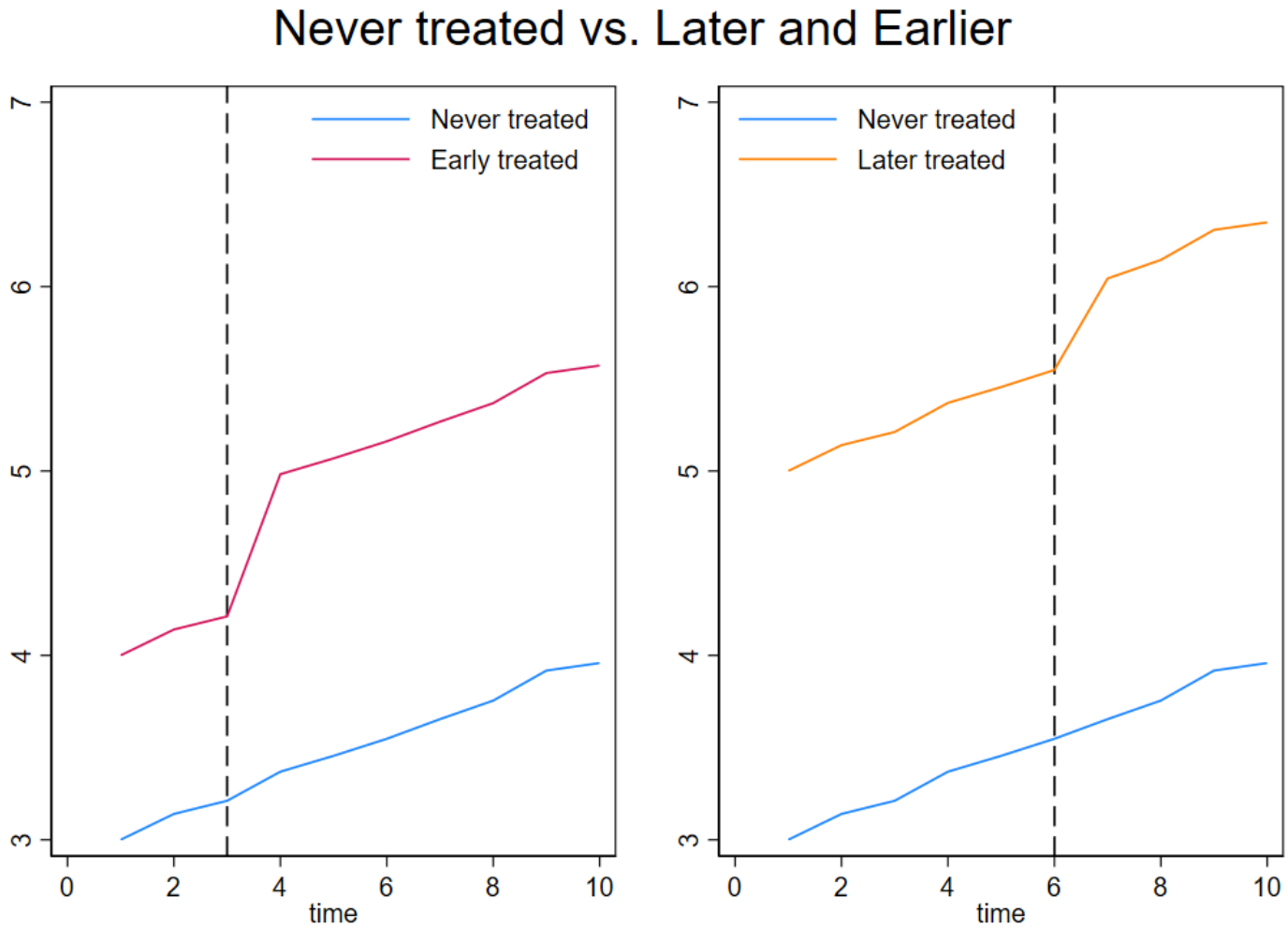
Homogeneous treatment with multiple treatment times I

- 10 time periods
- Treatment occurs for some groups at time 3 (Earlier treated)
- For others at time 6 (Later treated)
- Some units remain untreated
- What are the comparisons used by TWFE

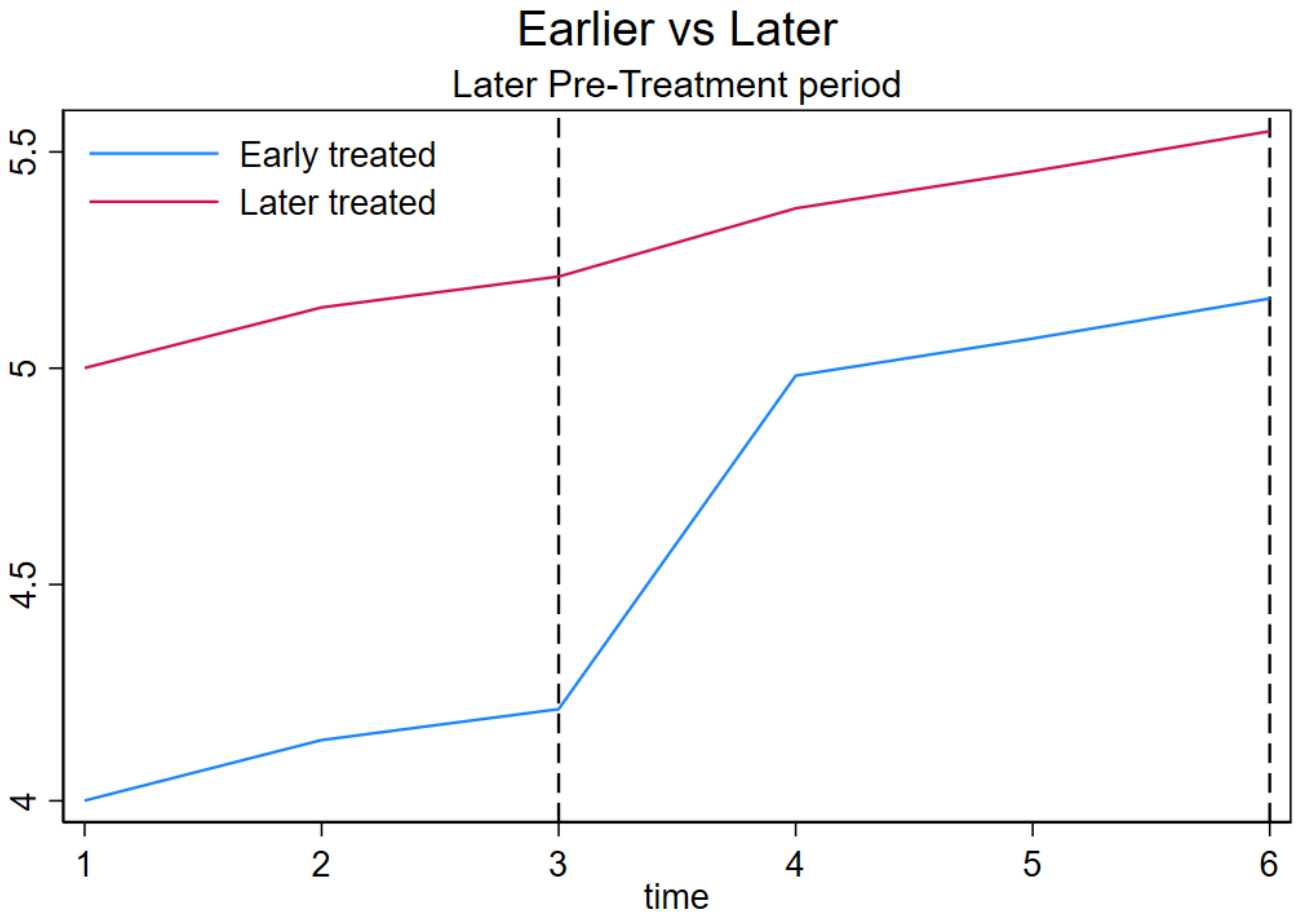
Homogeneous treatment with multiple treatment times I

- 10 time periods
- Treatment occurs for some groups at time 3 (Earlier treated)
- For others at time 6 (Later treated)
- Some units remain untreated
- What are the comparisons used by TWFE

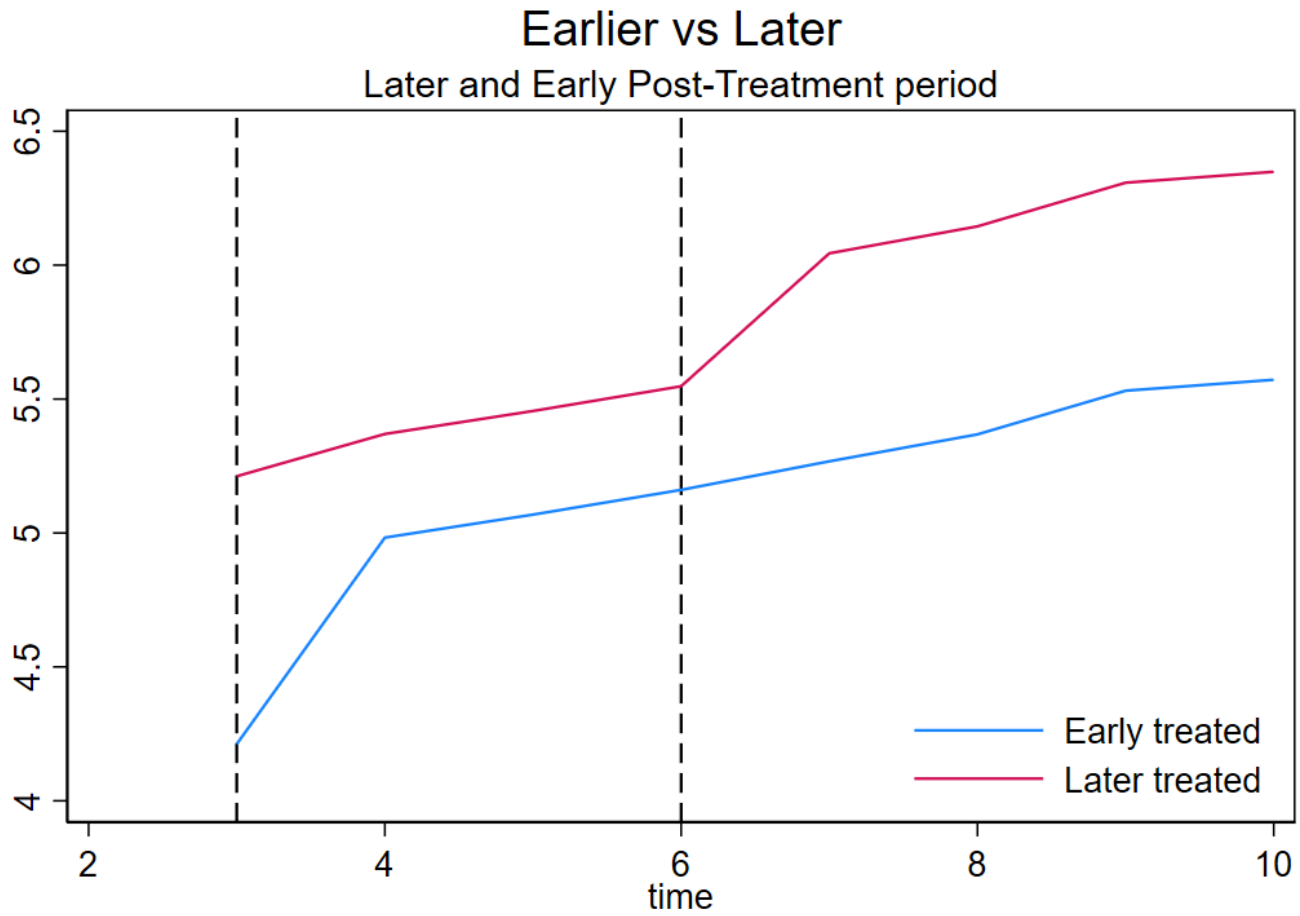
Homogeneous treatment with multiple treatment times I



Homogeneous treatment with multiple treatment times II



Homogeneous treatment with multiple treatment times III



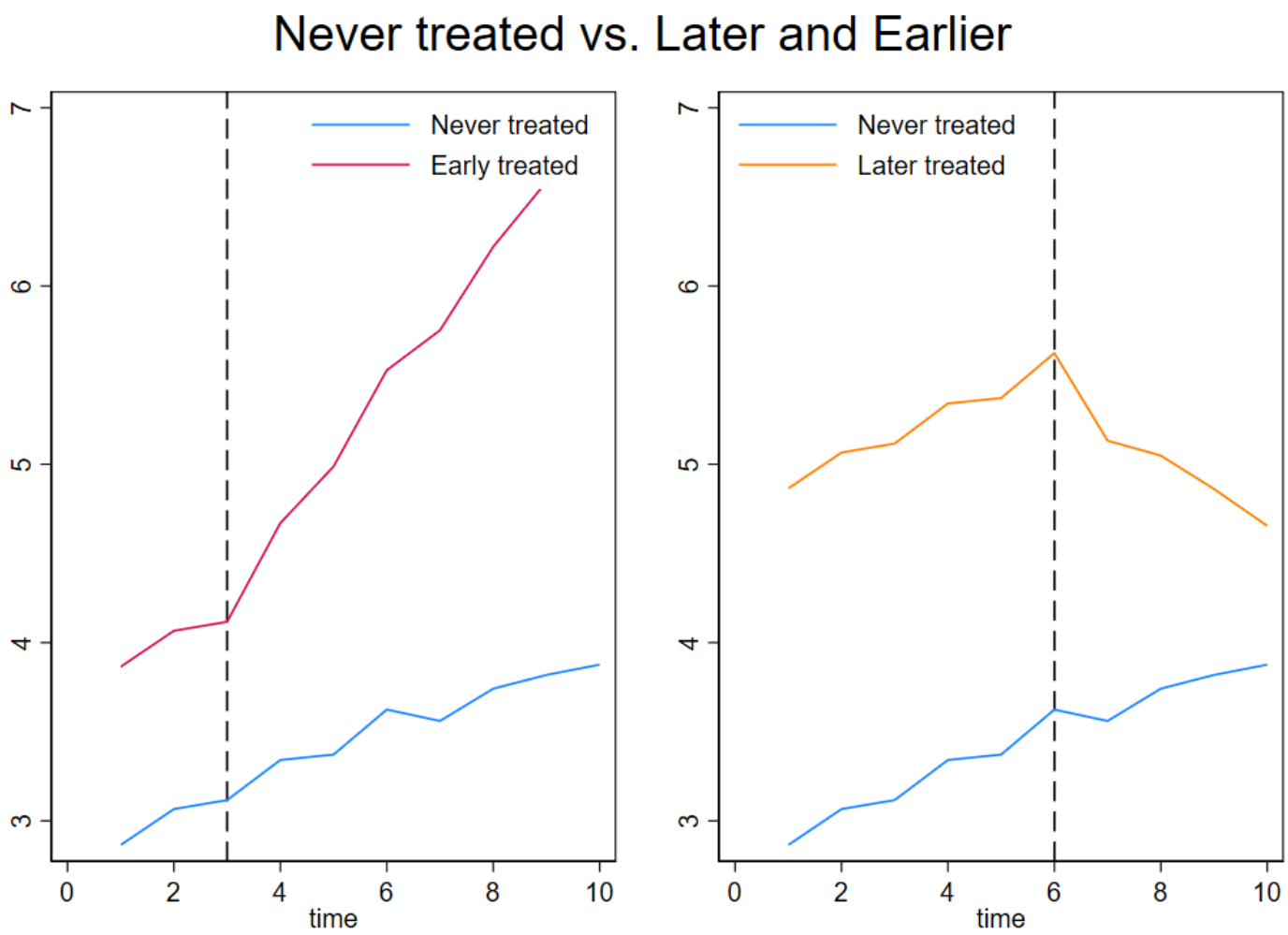
Homogeneous treatment: `estat bdecomp`

```
. estat bdecomp, summaryonly
DID treatment-effect decomposition
ATET = 4.041694                                  Number of obs      = 100,000
                                                 Number of groups   =  10,000
                                                 Number of cohorts  =       3
```

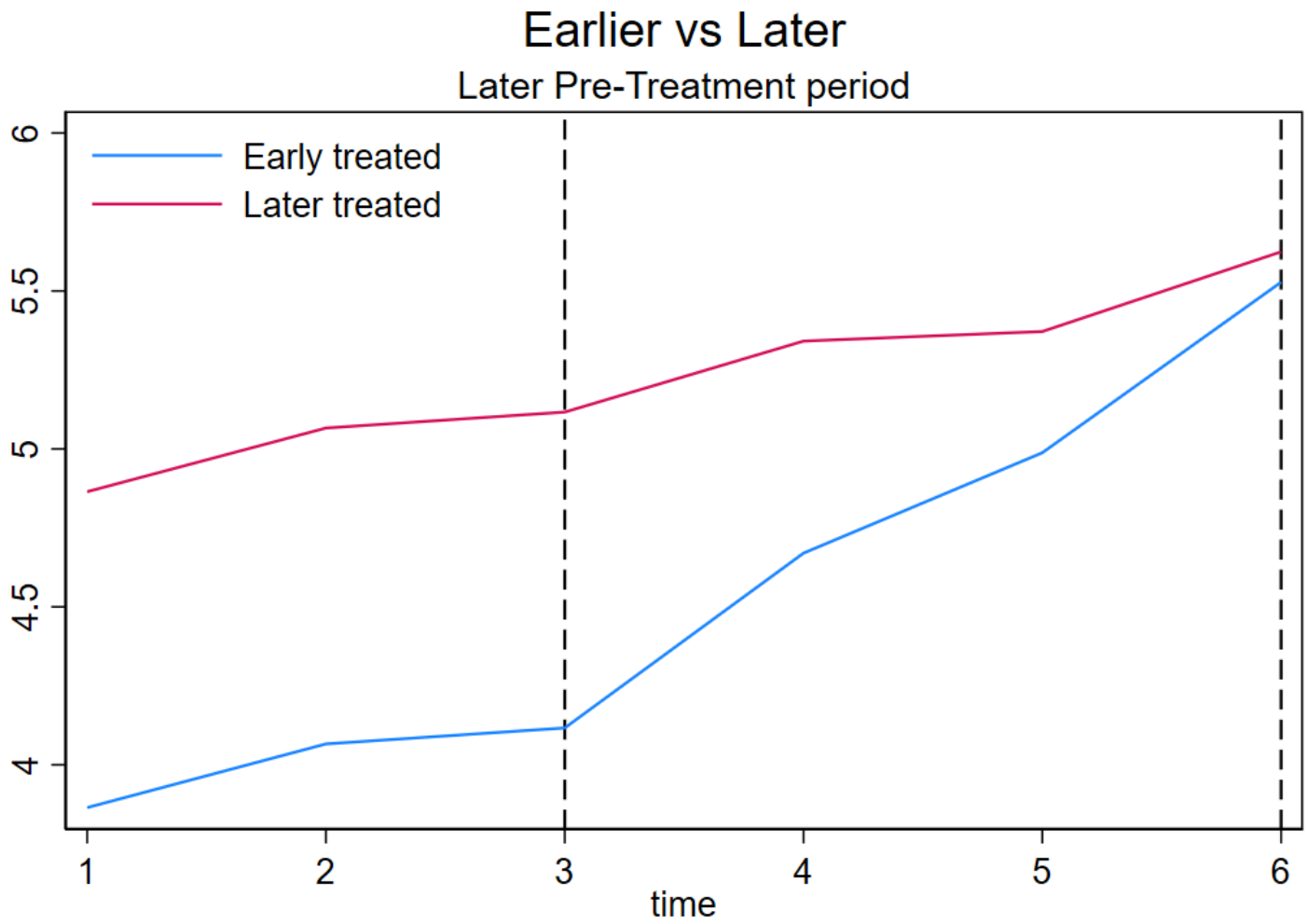
ATET decomposition summary	ATET component	Weight
Treated vs never treated	4.0435917	0.865605
Treated earlier vs later	4.0245287	0.057598
Treated later vs earlier	4.0331799	0.076797

Note: Number of cohorts includes never treated.
Note: The ATET reported by **xtdidregress** is a weighted average of the ATET components. If any component is substantially different from the ATET reported by **xtdidregress** and the weight is large, consider accounting for treatment-effect heterogeneity by using **xthdidregress**.

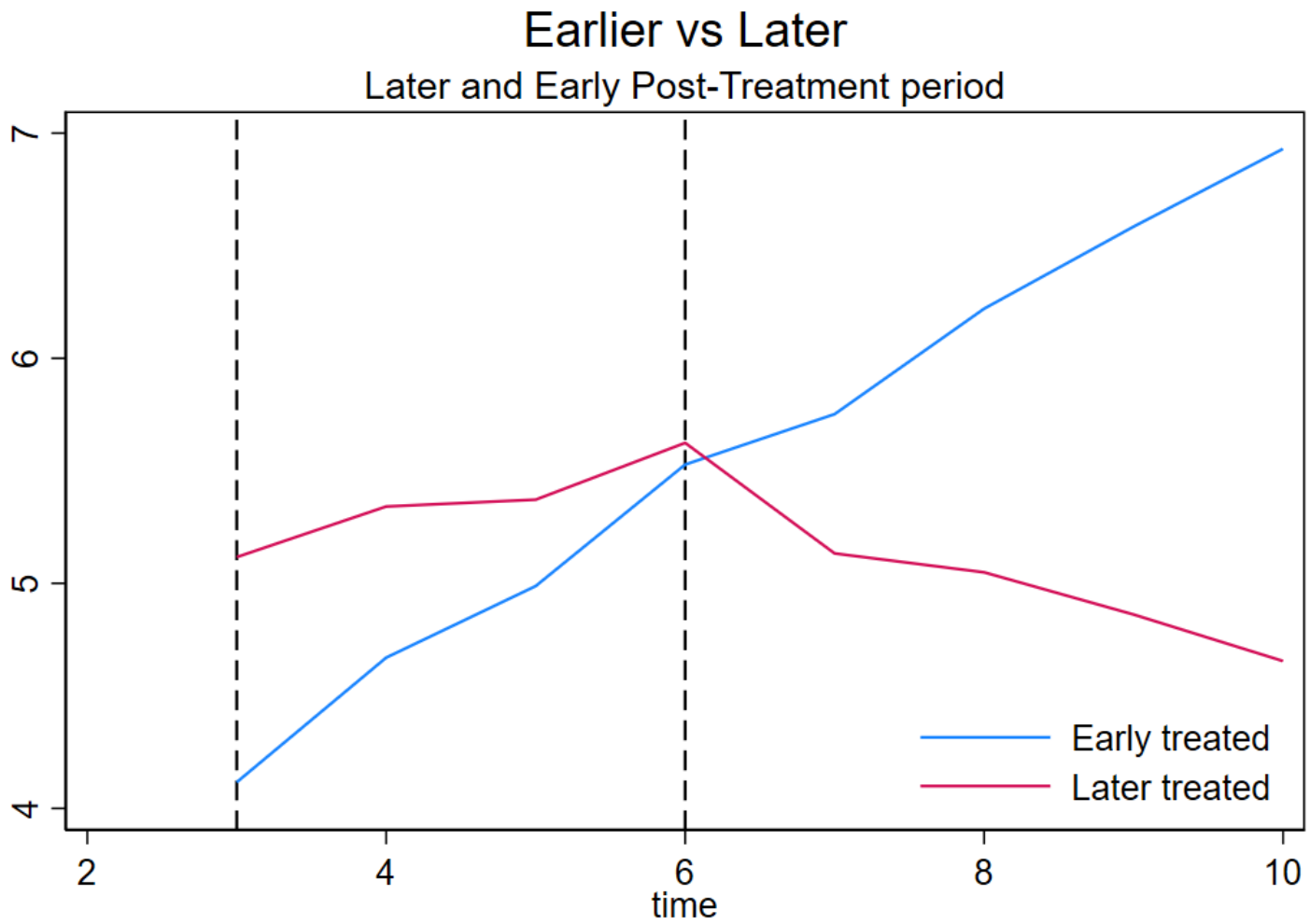
Heterogeneous treatment effect I



Heterogeneous treatment effect II



Heterogeneous treatment effect III



Heterogeneous treatment: `estat bdecomp`

```
. estat bdecomp, summaryonly
DID treatment-effect decomposition
ATET = 1.389688                                 Number of obs      = 50,000
                                                Number of groups   =  5,000
                                                Number of cohorts  =      3
```

ATET decomposition summary	ATET component	Weight
Treated vs never treated	1.9728088	0.860939
Treated earlier vs later	4.3441312	0.059597
Treated later vs earlier	-7.1439268	0.079463

Note: Number of cohorts includes never treated.
Note: The ATET reported by **xtdidregress** is a weighted average of the ATET components. If any component is substantially different from the ATET reported by **xtdidregress** and the weight is large, consider accounting for treatment-effect heterogeneity by using **xthdidregress**.

What would the aggregated effect be with “good” comparison groups?

```
. estat aggregation
Overall ATET                                    Number of obs = 50,000
                                (Std. err. adjusted for 50 clusters in state)
```

y	ATET	Robust std. err.	z	P>|z|	[95% conf.	interval]
tr (1 vs 0)	3.654175	.962046	3.80	0.000	1.768599	5.53975

What have we learned?

- With multiple treatment times, traditional DID assumes homogeneity
- 2 by 2 comparisons are well defined, but not all are useful
  - Comparison group matters
  - Time of comparison matters
- With multiple treatment times, it is important to assess heterogeneity
- If we suspect heterogeneity or do not want to assume homogeneity, use `hdidregress` and `xthdidregress`

What have we learned?

- With multiple treatment times, traditional DID assumes homogeneity
- 2 by 2 comparisons are well defined, but not all are useful
  - Comparison group matters
  - Time of comparison matters
- With multiple treatment times, it is important to assess heterogeneity
- If we suspect heterogeneity or do not want to assume homogeneity, use `hdidregress` and `xthdidregress`

Heterogeneous treatment effects approaches

- Solution 1: Callaway and Sant'Anna (2021)
  - Use valid comparison groups
  - Split the problem into 2 by 2 comparisons
  - Compute $ATET(g,t)$, where g is the cohort and t is the time
- Solution 2: Wooldridge (2021)
  - (Do not blame the messenger) Include heterogeneity in your regression
  - Add the adequate interactions with cohort and time
  - Compute $ATET(g,t)$, where g is the cohort and t is the time
  - Computing $ATET(g)$ or $ATET(t)$ is also possible
- Other solutions exist. Good surveys are Roth et al. (2022) and de Chaisemartin and D'Haultfoeuille (forthcoming)

Framework I

$$Y_{it} = Y_{it}(0) + \sum_{g=2}^{T} [Y_{it}(g) - Y_{it}(0)] G_{ig}$$

- Y_{it} observed outcome
- $Y_{it}(0)$ potential outcome of not being treated
- G_{ig} is an indicator for treatment group
- g is the time at which a group of individuals is treated (cohort)
- $Y_{it}(g)$ potential outcome for cohort g

Framework II

- Treatment is staggered
- Parallel trends
- No anticipation
- Overlap

Callaway and Sant'Anna

- Regression adjustment (RA)
- Inverse-probability weighting (IPW)
- Augmented inverse-probability weighting (AIPW)

RA

$$ATET(g,t) = E\left\{\frac{G_g}{E\left(G_g\right)}\left[Y_t - Y_{g-1} - m_{gt}\left(X\right)\right]\right\}$$

- $m_{gt}\left(X\right) = E\left(Y_t - Y_{g-1}|X, C = 1\right)$
- $C = 1$ is the never treated group ($G_g = \infty$)

RA

$$ATET(g,t) = E\left\{ \frac{G_g}{E\left(G_g\right)} \left[Y_t - Y_{g-1} - m_{gt}\left(X\right) \right] \right\}$$

- $ATET(g,t)$ is calculated using two groups: g and $C = 1$, never treated
- Outcomes are computed for two points in time
- 2 by 2 idea
- This is done for all g and all t
- We could have other 2 by 2 comparisons, i.e, using the not yet treated
- Identification assumptions are the same but need to hold for each 2 by 2

RA

$$ATET(g,t) = E\left\{ \frac{G_g}{E\left(G_g\right)} \left[Y_t - Y_{g-1} - m_{gt}\left(X\right) \right] \right\}$$

- $ATET(g,t)$ is calculated using two groups: g and $C = 1$, never treated
- Outcomes are computed for two points in time
- 2 by 2 idea
- This is done for all g and all t
- We could have other 2 by 2 comparisons, i.e, using the not yet treated
- Identification assumptions are the same but need to hold for each 2 by 2

Heuristically

1. `keep` if time is t or $g-1$
2. `keep` if cohort is g or cohort is $C=1$
3. `generate` $Y_t - Y_{g-1} \equiv \Delta Y$
4. `regress` ΔY on X for $C=1$ and `predict`, $\widehat{m}_{gt}(X)$
5. `generate` $\Delta Y - \widehat{m}_{gt}(X) = \widehat{TE}$
6. `summarize` $\widehat{TE}$ if cohort is g

- This is done for each g and t
- Doing it for all is a GMM problem, Callaway and Sant'Anna use influence functions.

IPW

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

- $p_g(X) = P(G_g = 1|X, G_g + C = 1)$, i.e., conditional on the sample we keep
- Steps are similar to before with the additional computation of $\widehat{p}_g(X)$ and the quotient $\frac{\frac{\widehat{p}_g(X)}{1-\widehat{p}_g(X)}}{\widehat{E}\left[\frac{\widehat{p}_g(X)}{1-\widehat{p}_g(X)}\right]}$

AIPW

$$ATET(g,t) = E\left\{\left(\frac{G_g}{E\left(G_g\right)} - \frac{\frac{p_g(X_1)}{1-p_g(X_1)}}{E\left[\frac{p_g(X_1)}{1-p_g(X_1)}\right]}\right)\left[Y_t - Y_{g-1} - m_{gt}\left(X_2\right)\right]\right\}$$

- Notice $m_{gt}\left(X_2\right)$. Emphasizes we could have different covariates
- AIPW is doubly robust. You may incorrectly specify at least one of $m_{gt}(X_2)$ or $p_g(X_1)$ and still recover $ATET(g,t)$

Wooldridge

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- q is the first treatment time and $q \ldots T$ the post period
- f_s is 1 if the time period is s and 0 otherwise
- We have group and time effects α_g and α_s
- Heterogeneity is captured by interacting group and time effects
- $\tau_{gs} \equiv ATET(g,t)$
- We use all our data and do not partition them
- If I have covariates, they enter fully interacted
- Extended two-way fixed effects

Wooldridge

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- q is the first treatment time and $q \ldots T$ the post period
- f_s is 1 if the time period is s and 0 otherwise
- We have group and time effects α_g and α_s
- Heterogeneity is captured by interacting group and time effects
- $\tau_{gs} \equiv ATET(g,t)$
- We use all our data and do not partition them
- If I have covariates, they enter fully interacted
- Extended two-way fixed effects

Wooldridge

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- q is the first treatment time and $q \ldots T$ the post period
- f_s is 1 if the time period is s and 0 otherwise
- We have group and time effects α_g and α_s
- Heterogeneity is captured by interacting group and time effects
- $\tau_{gs} \equiv ATET(g,t)$
- We use all our data and do not partition them
- If I have covariates, they enter fully interacted
- Extended two-way fixed effects

Wooldridge

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- q is the first treatment time and $q \ldots T$ the post period
- f_s is 1 if the time period is s and 0 otherwise
- We have group and time effects α_g and α_s
- Heterogeneity is captured by interacting group and time effects
- $\tau_{gs} \equiv ATET(g,t)$
- We use all our data and do not partition them
- If I have covariates, they enter fully interacted
- Extended two-way fixed effects

Let's see it work

```
. 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
```

- Out of the 113 contests, the Terrier group has won 47 times

Let's see it work

```
. 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
```

- Out of the 113 contests, the Terrier group has won 47 times

Staggered treatment

```
. tabulate year movie
```

Year	Was a movie protagonist 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

Specification: Panel/Longitudinal

```
xthdidregress estimator (registered best) (movie),
group(breed)
```

- You need to `xtset` with panel ID and time variable
- *estimator* is one of:
  - `ra`
  - `twfe`
  - `ipw`
  - `aipw`

Specification: Panel/Longitudinal

```
xthdidregress estimator (registered best) (movie),
group(breed)
```

- You need to `xtset` with panel ID and time variable
- *estimator* is one of:
  - `ra`
  - `twfe`
  - `ipw`
  - `aipw`

Output I

```
 . 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 omitted*)

Output II

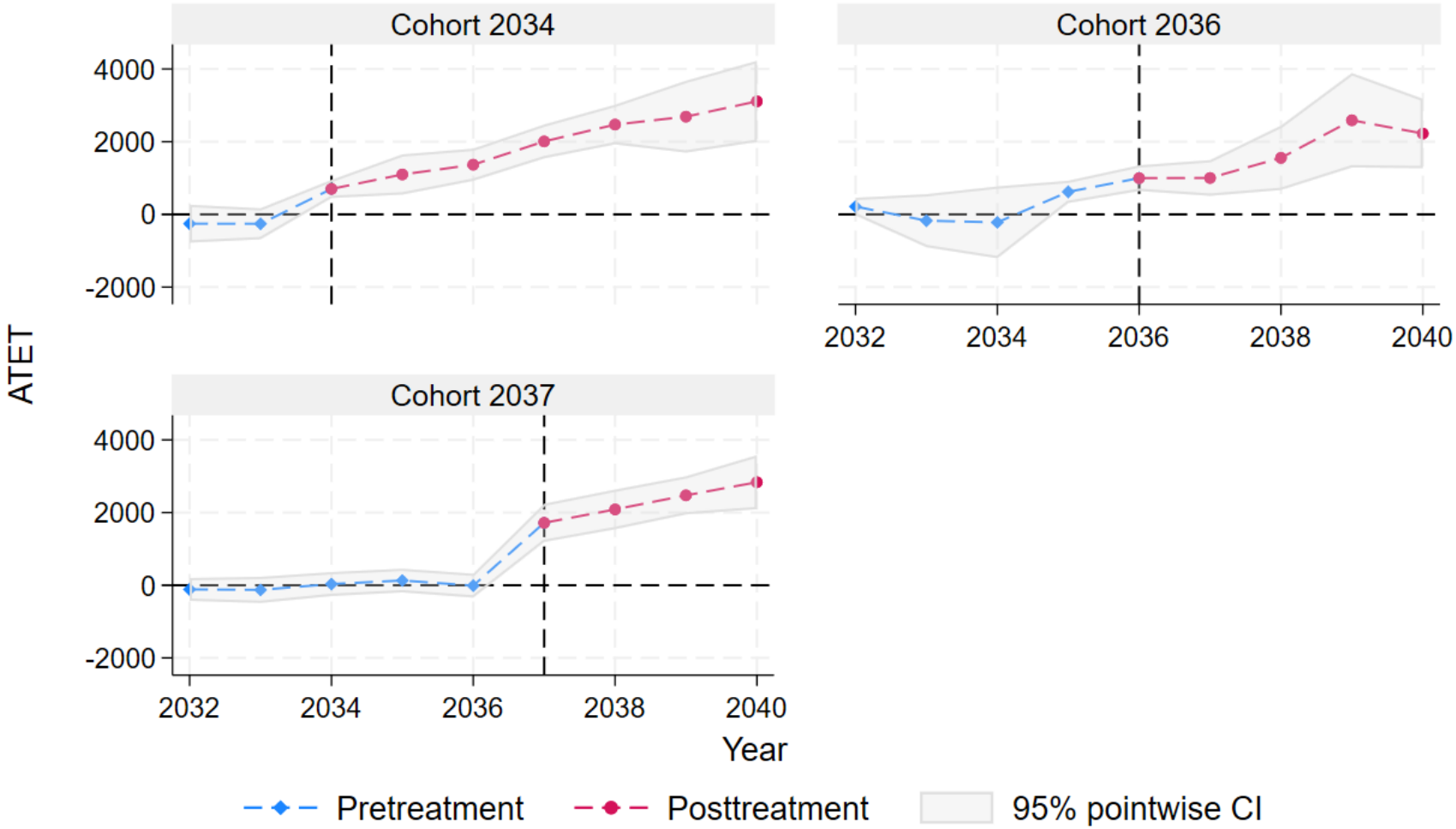
```
. xtset breed year
  (output omitted)
. xthdidregress ra (registered best) (movie), group(breed)
  (output omitted)
Heterogeneous-treatment-effects regression          Number of obs     = 1,410
                                                    Number of panels =   141

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

```
  (output omitted)
Note: ATET computed using covariates.
```

Graphical representation: `estat atetplot`



How to interpret my results

- I might want an average over all the $ATET(g,t)$
- I might want to know the effect of treatment within each group
- I might want to know the effect of treatment within a particular year
- I might want to see how the effect evolves with the duration of treatment

Overall

```
. 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

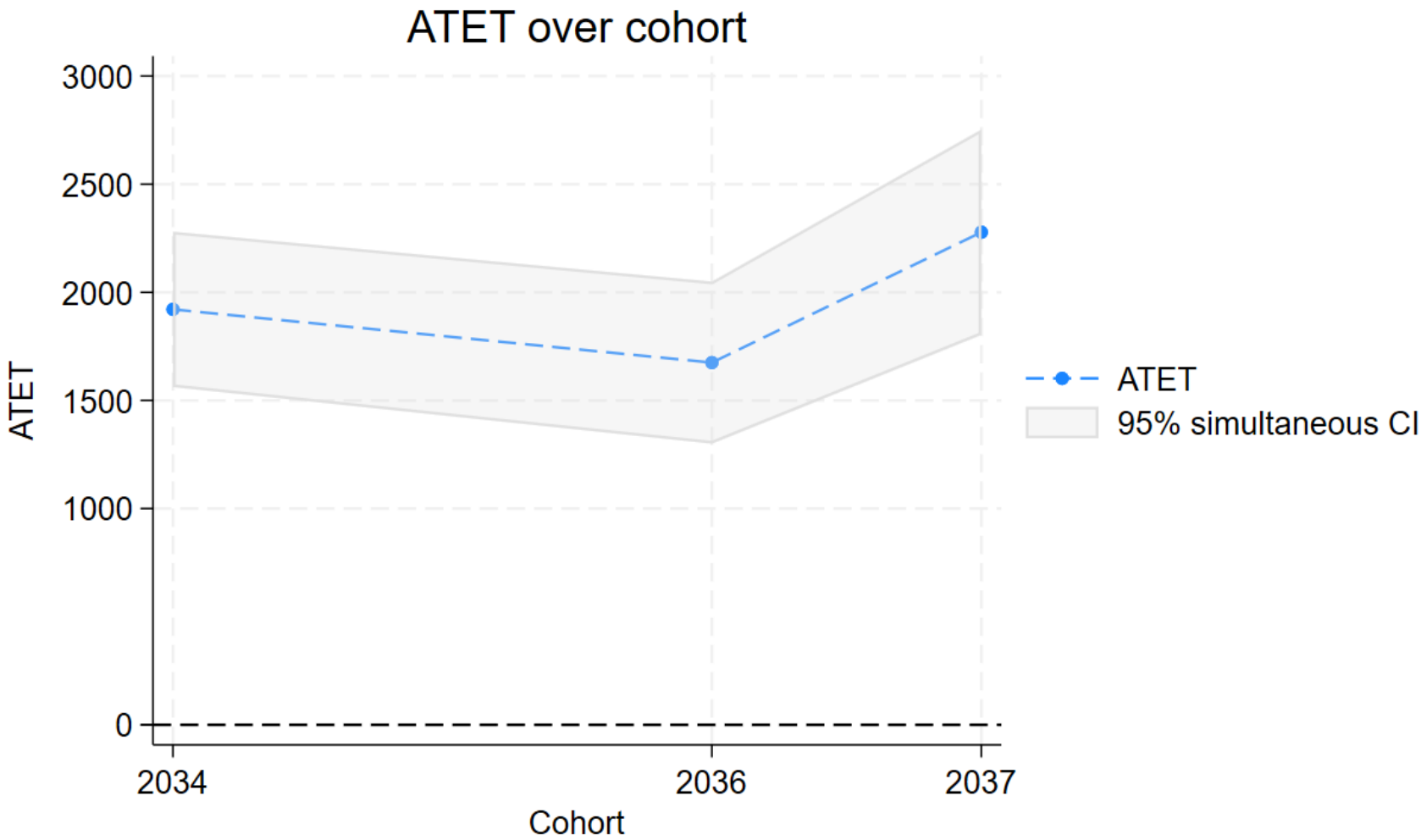
Cohort Tabular

```
. 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	135.6432	1567.112	2275.548
2036	1675.093	136.8109	1317.825	2032.361
2037	2278.136	168.6133	1837.819	2718.452

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

Cohort Graphical



Cohort list

```
. estat aggregation, cohort(2036 2034)
ATET over cohort                                Number of obs = 1,410
                                (Std. err. adjusted for 141 clusters in breed)
```

Cohort	ATET	Robust std. err.	z	P>|z|	[95% conf.	interval]
2034	1921.33	187.2787	10.26	0.000	1554.271	2288.389
2036	1675.093	130.4929	12.84	0.000	1419.332	1930.855

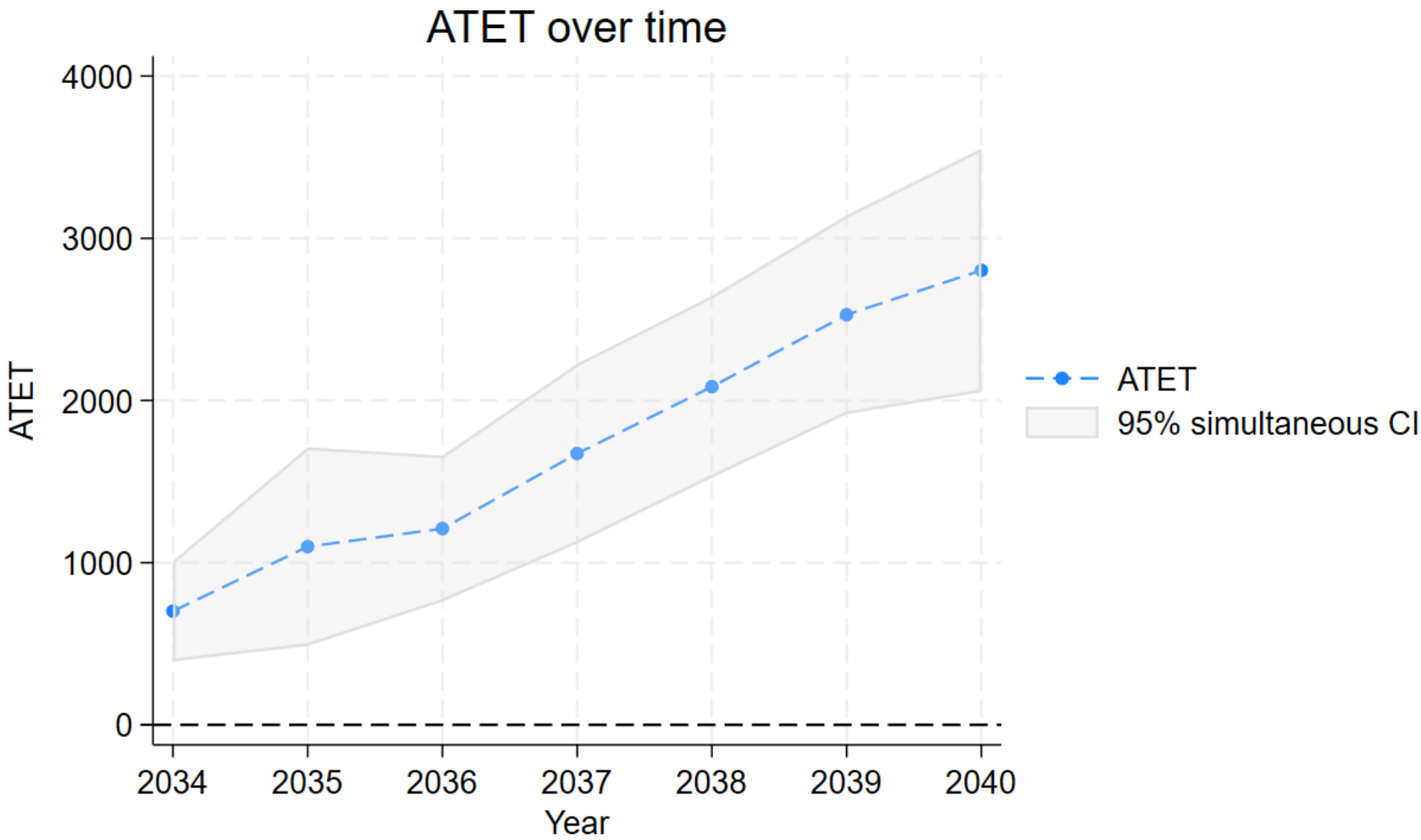
Time Tabular

```
. 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	125.496	384.0382	1018.225
2035	1099.044	273.6595	407.5822	1790.506
2036	1209.68	182.688	748.0782	1671.282
2037	1672.655	196.3049	1176.647	2168.664
2038	2084.658	226.3332	1512.777	2656.539
2039	2528.847	231.9688	1942.727	3114.968
2040	2802.171	300.7844	2042.172	3562.17

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

Time Graphical



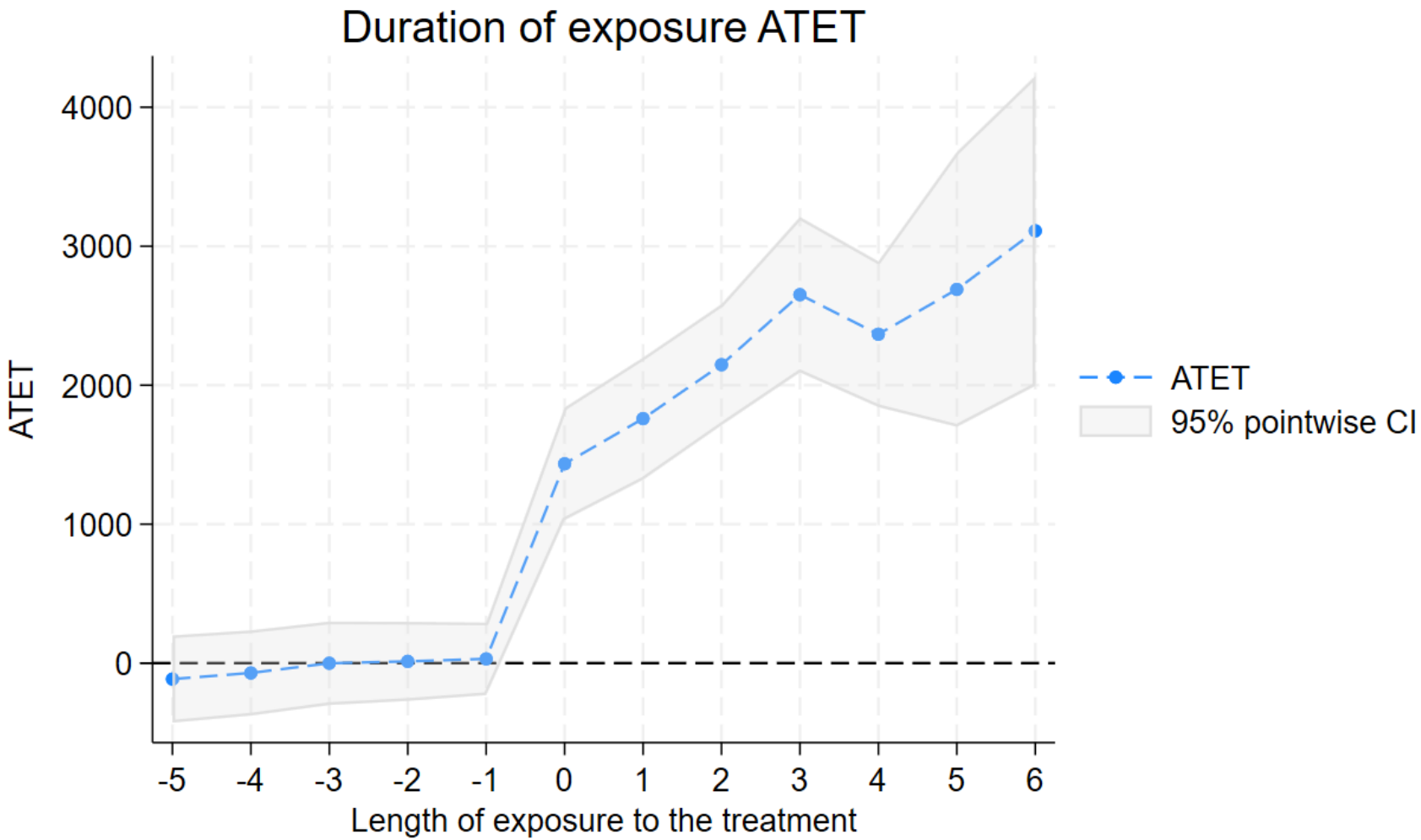
Duration Tabular

```
. 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.

Duration Graphical



Costs

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- Number of parameters increases with t, g, and covariates
- This is true for Callaway and Sant'Anna for each 2 by 2 subset
- `twfe` gives us a chance to address this at estimation with `hettype()`

Costs

$$Y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\alpha_s + \sum_{g=q}^{T}\sum_{s=g}^{T} D_{it}G_{ig}f_s\tau_{gs}$$

- Number of parameters increases with t, g, and covariates
- This is true for Callaway and Sant'Anna for each 2 by 2 subset
- `twfe` gives us a chance to address this at estimation with `hettype()`

Cohort heterogeneity

```
. xthdidregress twfe (registered best) (movie), group(breed) hettype(cohort)
  (output omitted)
Computing ATETs using margins ...
  (output omitted)
Heterogeneous-treatment-effects regression          Number of obs     = 1,410
                                                    Number of panels =   141

Estimator:         Two-way fixed effects
Panel variable:    breed
Treatment level:   breed
Control group:     Never treated
Heterogeneity:     Cohort
                                   (Std. err. adjusted for 141 clusters in breed)
```

Cohort	ATET	Robust std. err.	t	P>|t|	[95% conf.	interval]
2034	1662.492	108.002	15.39	0.000	1448.966	1876.017
2036	1978.645	54.21043	36.50	0.000	1871.468	2085.822
2037	2276.223	70.63244	32.23	0.000	2136.579	2415.867

```
Note: ATET computed using covariates.
```

Heterogeneous DID in Stata

- Continued interest and development in causal inference/treatment effects
- Heterogeneous DID
- Estimation and postestimation tools displayed in tabular and graphical forms
- Different ways of thinking about heterogeneity