Causal Inference and Treatment Effect Estimation using Stata

Enrique Pinzón

StataCorp LLC

April 16, 2026

Background

- Stata has been adding causal inference tools since Stata 13 (We are in 19)
- [CAUSAL] manual which was before called [TEFFECTS]
- Potential outcome framework
- Applied researchers in social and health sciences

Road map

- Tour
- Selection on observables
- Selection on unobservables
- A bit of theory
- A bit of Stata
- Mostly infinite knowledge of the data
- All material is going to be available

Road map

- Tour
- Selection on observables
- Selection on unobservables
- A bit of theory
- A bit of Stata
- Mostly infinite knowledge of the data
- All material is going to be available

Road map

- Tour
- Selection on observables
- Selection on unobservables
- A bit of theory
- A bit of Stata
- Mostly infinite knowledge of the data
- All material is going to be available

SELECTION ON OBSERVABLES

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

Simulated data

```
. clear all
. version 19.5
.
. ////////////////////// teffects ////////////////////////////////////////
>
. *************** Simulated data for treatment effects ***************
. quietly set obs 30000
. set seed 111
. // Covariates
. generate x  = rchi2(5)-3
. generate a  = int(rbeta(2,3)*4)
. generate z  = rnormal()
. generate e0 = rchi2(5)-5
. generate e1 = rchi2(5)-5
. // Potential outcomes
. generate y0 = 1.5 - x*a + x + a/3 + e0
. generate y1 = 1 + x*a - x - a/3 + e1
. //Treatment
. generate t  = -1 - .5*z*a/5 - z + a/4 - rnormal()>0
. // Treatment effect
. generate te = y1 - y0
. // Observed outcome
. generate y  = t*y1 + (1-t)*y0
. ********************************************************************
```

teffects

`teffects` *`estimator`* `(`*`outcome_model`*`)` `(`*`treatment_model`*`)`

`estimator` corresponds to:

- ra (regression adjustment)
- ipw (inverse-probability weighting)
- ipwra (inverse-probability-weighted regression adjustment)
- aipw (augmented inverse-probability weighting)
- psmatch (propensity-score matching)
- nnmatch (nearest-neighbor matching)

Regression adjustment

```
. teffects ra (y c.x##i.a) (t)
Iteration 0:  EE criterion = 6.276e-28
Iteration 1:  EE criterion = 7.517e-32
Treatment-effects estimation                    Number of obs     =     30,000
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t						
(1 vs 0)	-.8210592	.0535007	-15.35	0.000	-.9259186	-.7161998
POmean						
t						
0	1.658092	.0284536	58.27	0.000	1.602324	1.71386

```
. sum te y0
```

Variable	Obs	Mean	Std. dev.	Min	Max
te	30,000	-.8183005	7.599519	-45.56726	75.86474
y0	30,000	1.670335	4.419516	-37.52642	27.97842

Regression adjustment

```
. teffects ra (y c.x##i.a) (t)
Iteration 0:  EE criterion = 6.276e-28
Iteration 1:  EE criterion = 7.517e-32
Treatment-effects estimation                    Number of obs     =     30,000
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t						
(1 vs 0)	-.8210592	.0535007	-15.35	0.000	-.9259186	-.7161998
POmean						
t						
0	1.658092	.0284536	58.27	0.000	1.602324	1.71386

```
. sum te y0
```

Variable	Obs	Mean	Std. dev.	Min	Max
te	30,000	-.8183005	7.599519	-45.56726	75.86474
y0	30,000	1.670335	4.419516	-37.52642	27.97842

regress I

```
. quietly regress y c.x##i.a if t==1, vce(robust)
. predict double y1hat
(option xb assumed; fitted values)
. quietly regress y c.x##i.a if t==0, vce(robust)
. predict double y0hat
(option xb assumed; fitted values)
. generate double tehat = y1hat - y0hat
. sum tehat
```

Variable	Obs	Mean	Std. dev.	Min	Max
tehat	30,000	-.8210592	6.152497	-45.13368	72.95405

regress II

```
. quietly regress y i.t##(c.x##i.a), vce(robust)
. margins r.t, contrast(nowald) vce(unconditional)
Contrasts of predictive margins                         Number of obs = 30,000
Expression: Linear prediction, predict()
```

	Unconditional Contrast	Unconditional std. err.	[95% conf.	interval]
t (1 vs 0)	-.8210592	.053515	-.9259508	-.7161676

```
. di sqrt(r(V)[1,1]*r(df_r)/r(N))
.05350068
```

Inverse-probability weighting

```
. teffects ipw (y) (t c.z##i.a)
Iteration 0:  EE criterion = 7.667e-19
Iteration 1:  EE criterion = 5.231e-32
Treatment-effects estimation                    Number of obs     =     30,000
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t (1 vs 0)	-.8272432	.0798468	-10.36	0.000	-.98374	-.6707463
POmean						
t 0	1.716408	.0372815	46.04	0.000	1.643337	1.789478

```
. estimates store ipw
```

regress, again, but with some help

```
. quietly logit t c.z##i.a
. predict double p
(option pr assumed; Pr(t))
. generate double ps = 1/p
. quietly replace ps = 1/(1-p) if t==0
. regress y t [pw=ps], vce(robust)
(sum of wgt is 57,760.9102587447)
Linear regression                               Number of obs     =     30,000
                                                F(1, 29998)       =     102.36
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0088
                                                Root MSE          =     4.3829
```

y	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
t	-.8272432	.0817666	-10.12	0.000	-.9875091	-.6669772
_cons	1.716408	.0381247	45.02	0.000	1.641681	1.791134

Double robustness

```
. generate double b = runiformint(1,20)
. teffects aipw (y c.x##i.a) (t i.b#c.x)
Iteration 0:  EE criterion = 2.566e-17
Iteration 1:  EE criterion = 4.804e-32
Treatment-effects estimation                    Number of obs     =     30,000
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t (1 vs 0)	-.8205445	.0534699	-15.35	0.000	-.9253436	-.7157453
POmean						
t 0	1.657873	.0284506	58.27	0.000	1.602111	1.713635

telasso – A step further but not too much

```
. telasso (y c.x##c.z##(i.a##i.b)) (t c.x##c.z##(i.a##i.b))
Estimating lasso for outcome y if t = 0 using plugin method ...
Estimating lasso for outcome y if t = 1 using plugin method ...
Estimating lasso for treatment t using plugin method ...
Estimating ATE ...
Treatment-effects lasso estimation    Number of observations      =     30,000
Outcome model:   linear               Number of controls          =        419
Treatment model: logit                Number of selected controls =          8
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t						
(1 vs 0)	-.8494596	.0719932	-11.80	0.000	-.9905637	-.7083555
POmean						
t						
0	1.679623	.0329581	50.96	0.000	1.615027	1.74422

```
. di  e(tmvars_sel)
0bn.a 2bn.a 2bn.a#c.z 3bn.a z
. di  e(omvars_sel)
0bn.a#c.x 2bn.a#c.x 3bn.a#c.x
```

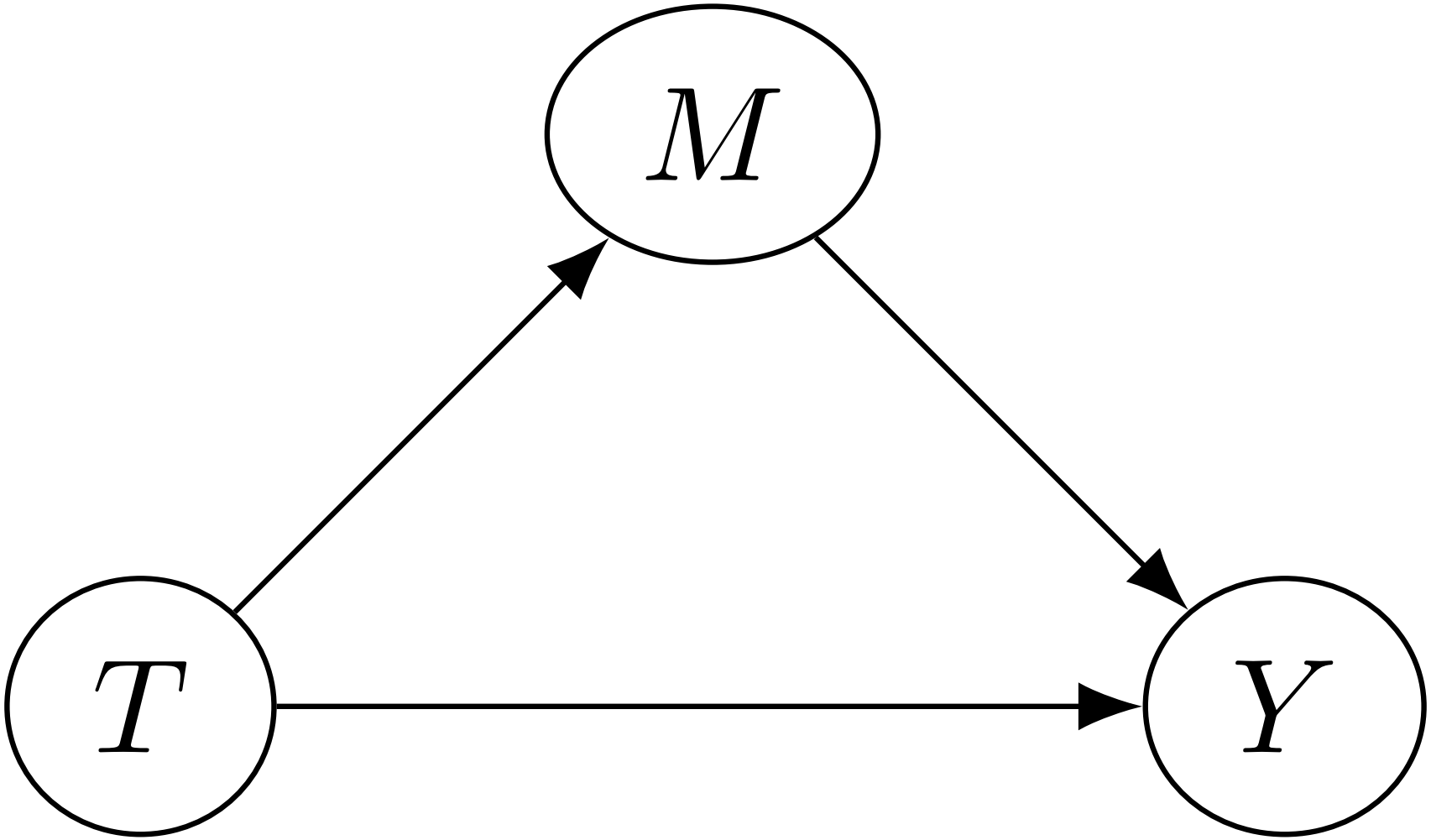
telasso – A step further but not too much

```
. telasso (y c.x##c.z##(i.a##i.b)) (t c.x##c.z##(i.a##i.b))
Estimating lasso for outcome y if t = 0 using plugin method ...
Estimating lasso for outcome y if t = 1 using plugin method ...
Estimating lasso for treatment t using plugin method ...
Estimating ATE ...
Treatment-effects lasso estimation    Number of observations      =     30,000
Outcome model:   linear               Number of controls          =        419
Treatment model: logit                Number of selected controls =          8
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t						
(1 vs 0)	-.8494596	.0719932	-11.80	0.000	-.9905637	-.7083555
POmean						
t						
0	1.679623	.0329581	50.96	0.000	1.615027	1.74422

```
. di  e(tmvars_sel)
0bn.a 2bn.a 2bn.a#c.z 3bn.a z
. di  e(omvars_sel)
0bn.a#c.x 2bn.a#c.x 3bn.a#c.x
```

Causal mediation



Linear outcome and linear mediator model

$$
\begin{aligned}
y &= \beta_0 + \beta_1 t + \beta_2 m + \beta_3 mt + \varepsilon_y \\
m &= \alpha_0 + \alpha_1 t + \varepsilon_m \\
y &= \beta_0 + \beta_1 t + \beta_2 \left(\alpha_0 + \alpha_1 t' + \varepsilon_m\right) \\
&+ \beta_3 t \left(\alpha_0 + \alpha_1 t' + \varepsilon_m\right) + \varepsilon_y \\
E\left(y | t, m \equiv m(t')\right) &= \beta_0 + \beta_1 t + \left(\beta_2 + \beta_3 t\right)\left(\alpha_0 + \alpha_1 t'\right)
\end{aligned}
$$

Linear outcome and linear mediator model

$$
\begin{aligned}
y &= \beta_0 + \beta_1 t + \beta_2 m + \beta_3 mt + \varepsilon_y \\
m &= \alpha_0 + \alpha_1 t + \varepsilon_m \\
y &= \beta_0 + \beta_1 t + \beta_2 \left(\alpha_0 + \alpha_1 t' + \varepsilon_m\right) \\
&+ \beta_3 t \left(\alpha_0 + \alpha_1 t' + \varepsilon_m\right) + \varepsilon_y \\
E\left(y|t, m \equiv m(t')\right) &= \beta_0 + \beta_1 t + \left(\beta_2 + \beta_3 t\right)\left(\alpha_0 + \alpha_1 t'\right)
\end{aligned}
$$

Counterfactual intuition

$$E\left(y|0, m \equiv m(0)\right)$$
$$E\left(y|0, m \equiv m(1)\right)$$
$$E\left(y|1, m \equiv m(0)\right)$$
$$E\left(y|1, m \equiv m(1)\right)$$

- Natural indirect effects $E\left(y|t, m \equiv m(1)\right) - E\left(y|t, m \equiv m(0)\right)$
  - We report $E\left(y|1, m \equiv m(1)\right) - E\left(y|1, m \equiv m(0)\right)$
- Natural direct effect $E\left(y|1, m \equiv m(t)\right) - E\left(y|0, m \equiv m(t)\right)$
  - We report $E\left(y|1, m \equiv m(0)\right) - E\left(y|0, m \equiv m(0)\right)$
- Total effect $E\left(y|1, m \equiv m(1)\right) - E\left(y|0, m \equiv m(0)\right)$

What can we do here

Table: Outcome and Mediator Combinations

Outcome Model	Mediation Model
Linear	Linear
Linear	Probit
Linear	Logit
Linear	Poisson
Linear	Exponential mean
Probit	Linear
Probit	Probit
Probit	Logit
Probit	Poisson
Probit	Exponential mean
Logit	Probit
Logit	Logit
Logit	Poisson
Poisson	Linear
Poisson	Probit
Poisson	Logit
Poisson	Poisson
Poisson	Exponential mean

Simulated data II

```
. clear all
. quietly set obs 10000
. set seed 111
. generate t   = .1 + rnormal()>0
. generate m0  = 2 + rchi2(5)-5
. generate m1  = 1 + rchi2(5)-5
. generate m   = t*m1 + (1-t)*m0
. generate y0  = 2 - m + rchi2(5)-5
. generate y1  = 1 - m + rchi2(5)-5
. generate y   = t*y1 - (1-t)*y0
```

mediate

```
. mediate (y) (m) (t)
Iteration 0:  EE criterion = 2.620e-29
Iteration 1:  EE criterion = 1.743e-29
Causal mediation analysis                       Number of obs = 10,000
Outcome model:     Linear
Mediator model:    Linear
Mediator variable: m
Treatment type:    Binary
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
t (1 vs 0)	1.039025	.0649129	16.01	0.000	.9117978	1.166252
NDE						
t (1 vs 0)	-1.136193	.1132047	-10.04	0.000	-1.35807	-.9143157
TE						
t (1 vs 0)	-.097168	.089638	-1.08	0.278	-.2728553	.0785192

Note: Outcome equation includes treatmentmediator interaction.

Total effects

```
. teffects ra (y) (t)
Iteration 0:  EE criterion = 2.953e-33
Iteration 1:  EE criterion = 1.234e-33
Treatment-effects estimation                    Number of obs     =     10,000
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t (1 vs 0)	-.097168	.0897509	-1.08	0.279	-.2730764	.0787405
POmean						
t 0	.0000264	.0663783	0.00	1.000	-.1300727	.1301256

SELECTION ON 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

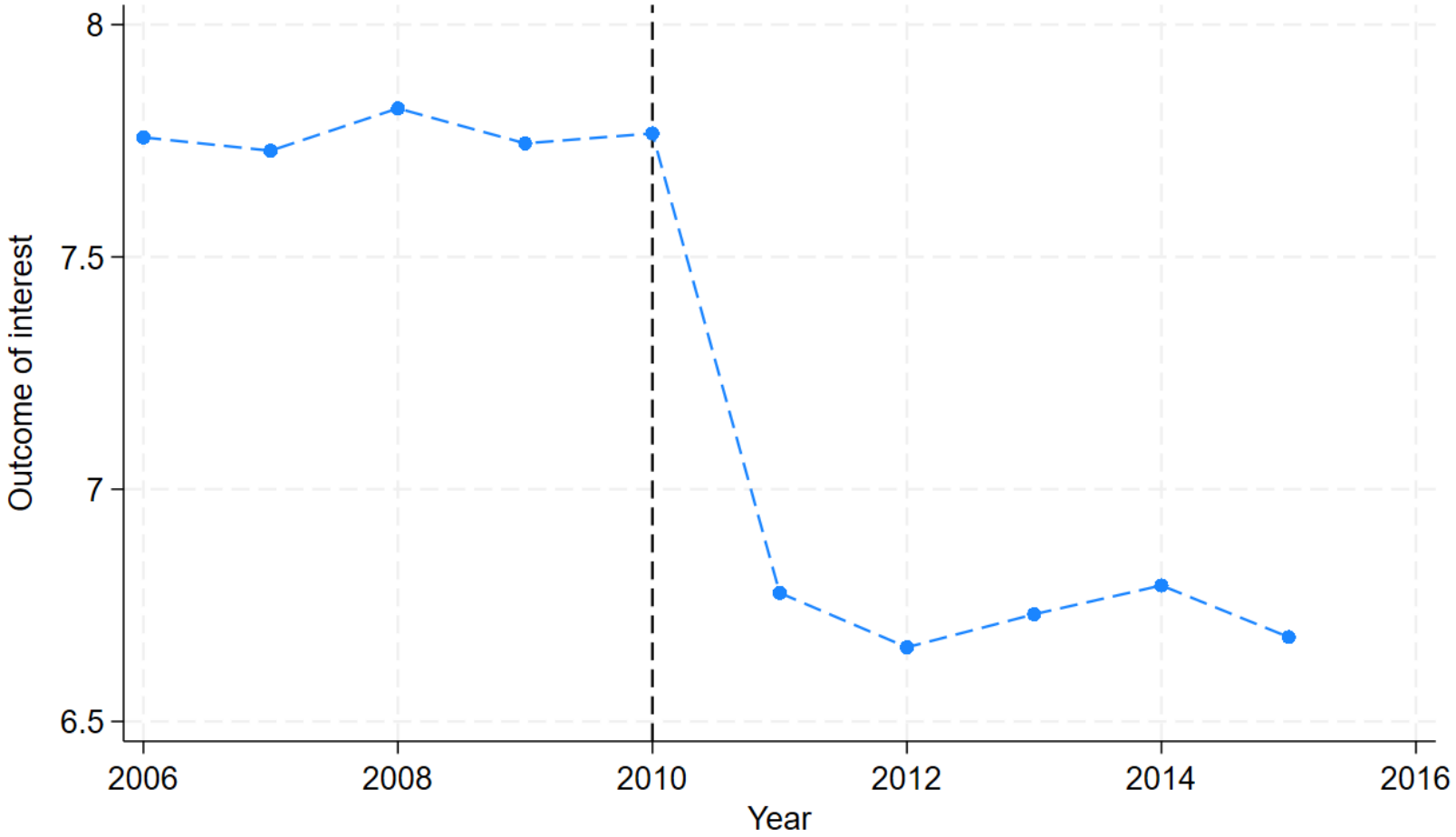
Difference-in-differences (DID)

- 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

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

Treated group



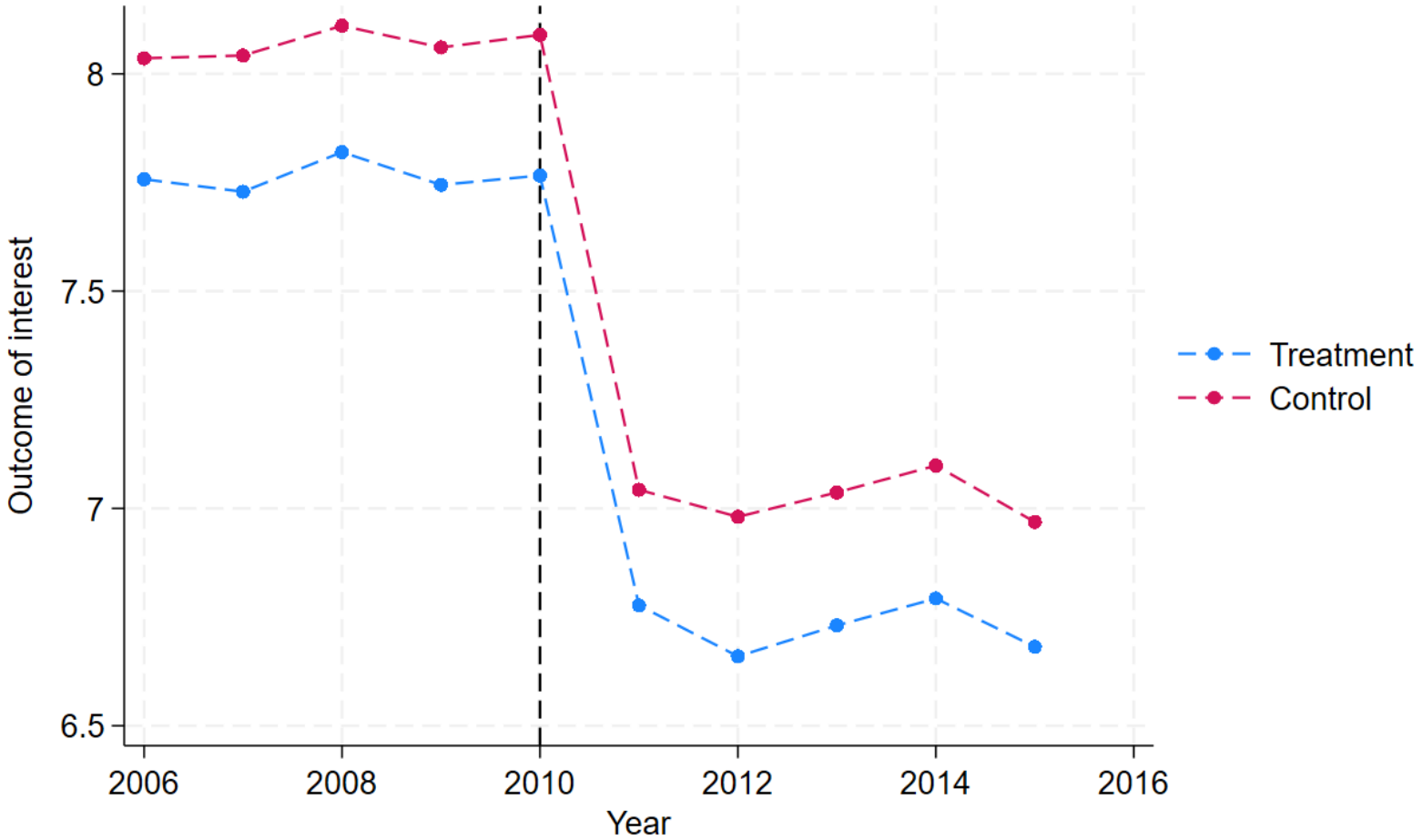
What have we learned

- 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

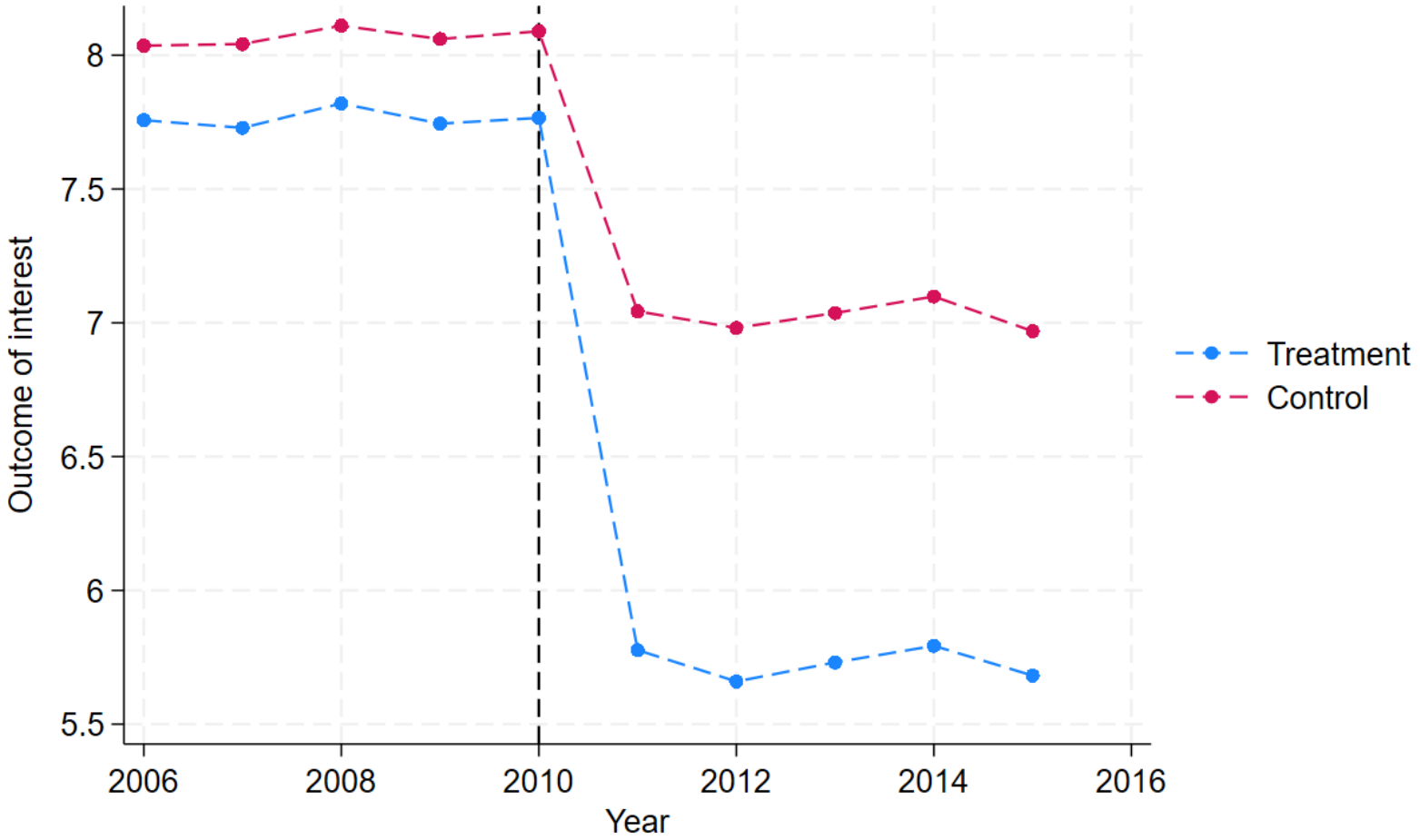
What have we learned

- 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

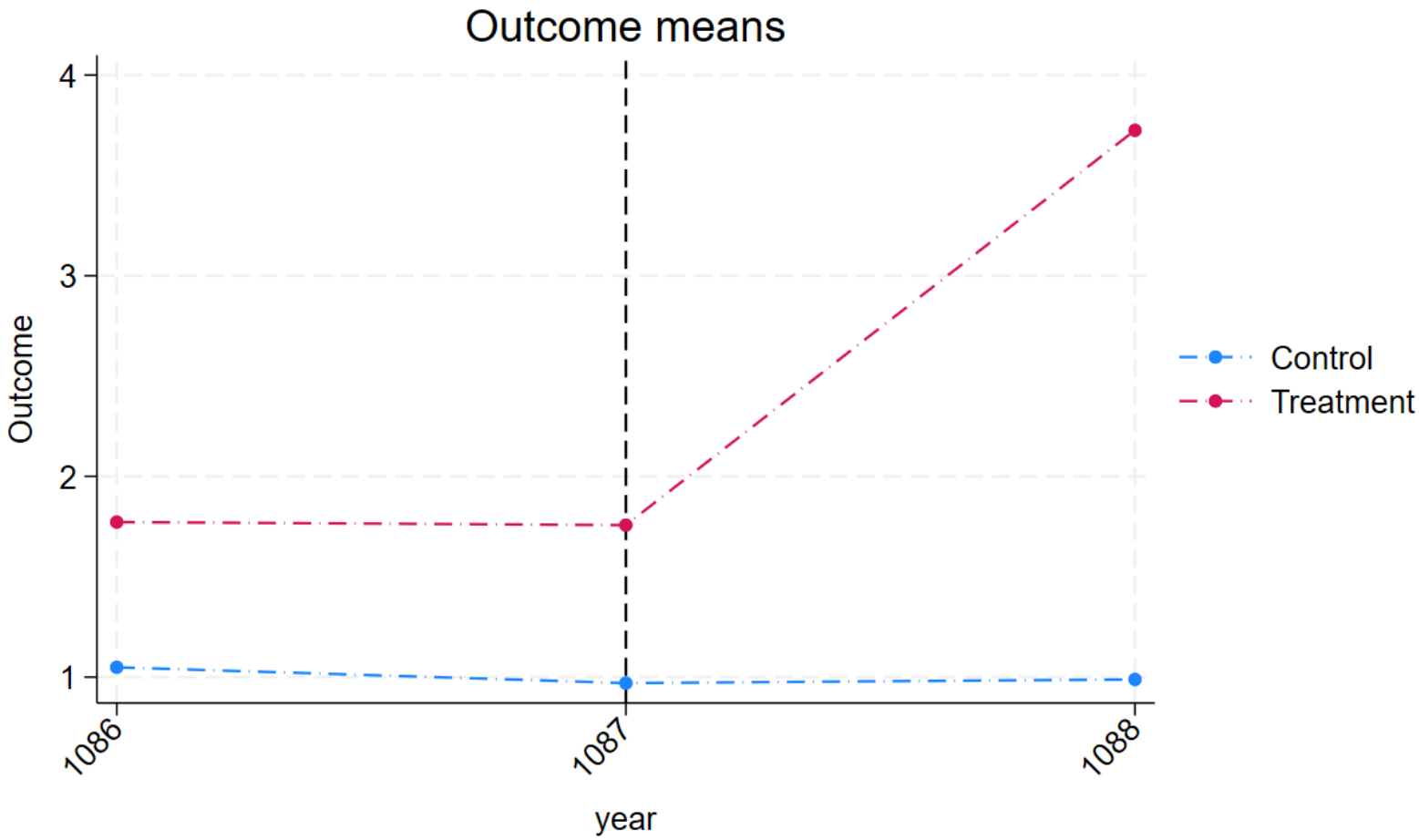
Graphical representation I



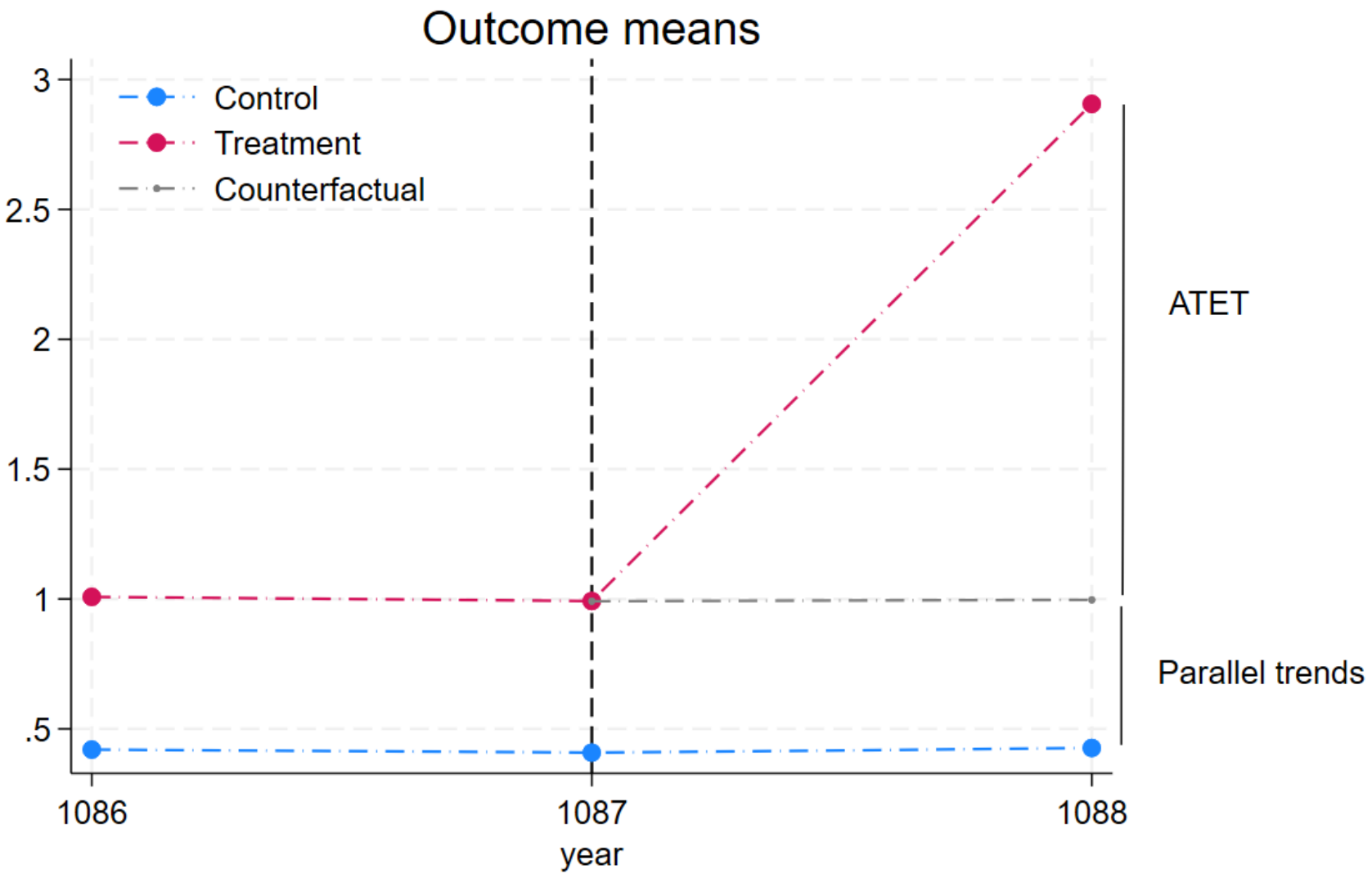
Graphical representation II



2 by 2 framework



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



Heterogeneous ATET and DID

- 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

Heterogeneous ATET and DID

- 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

Heterogeneous ATET and DID

- 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

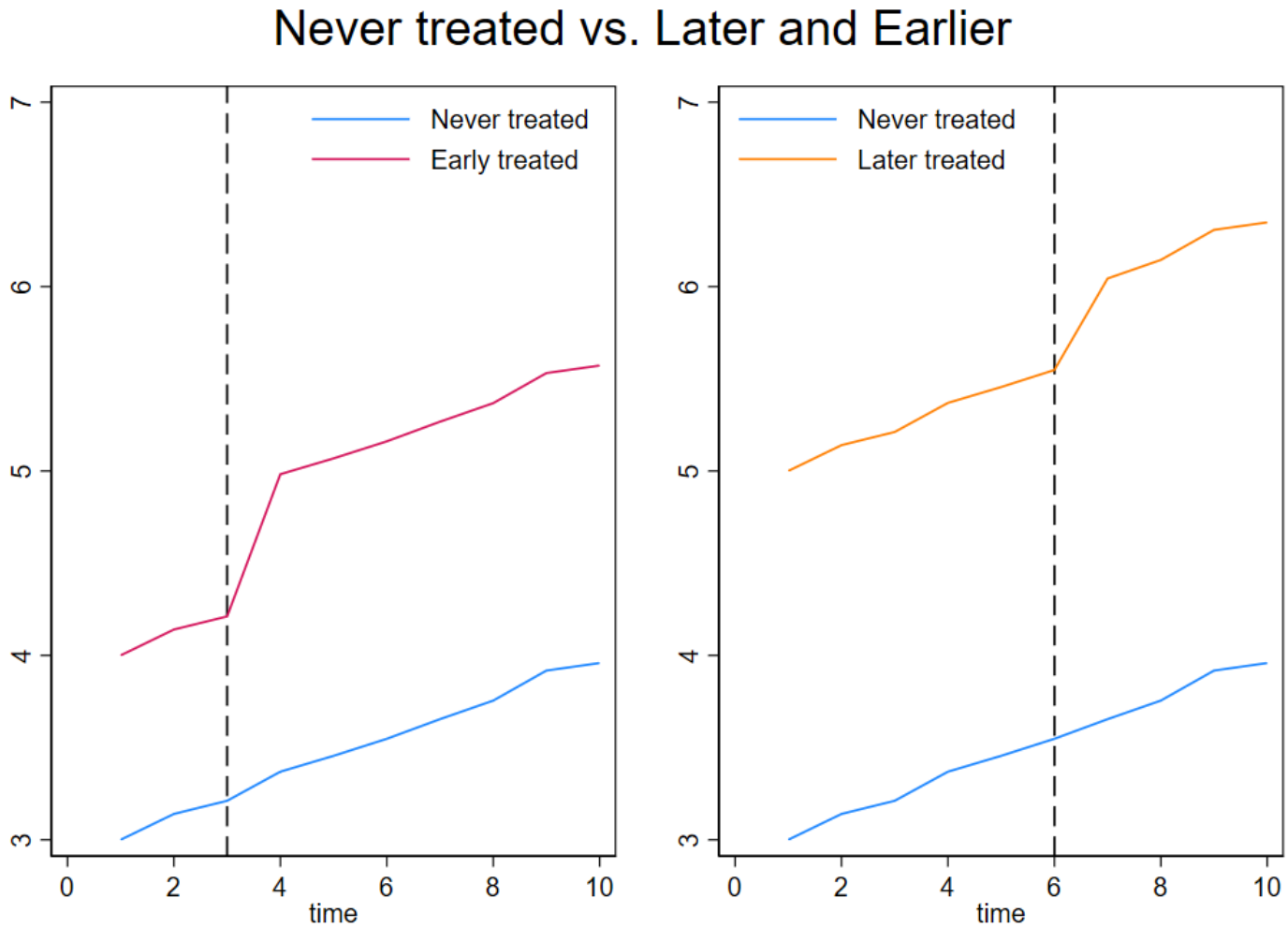
Homogeneous treatment with multiple treatment times I

- 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

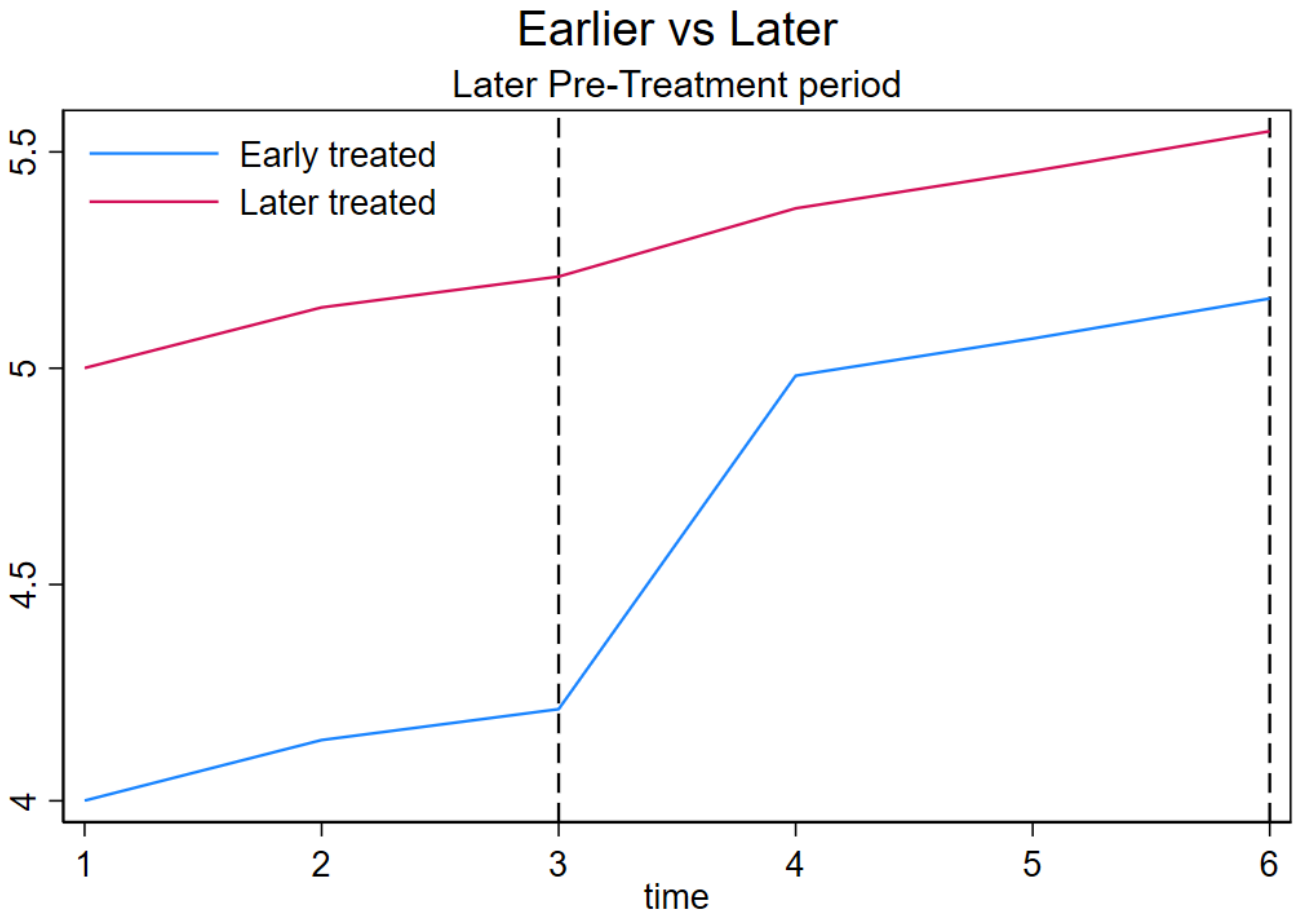
Homogeneous treatment with multiple treatment times I

- 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

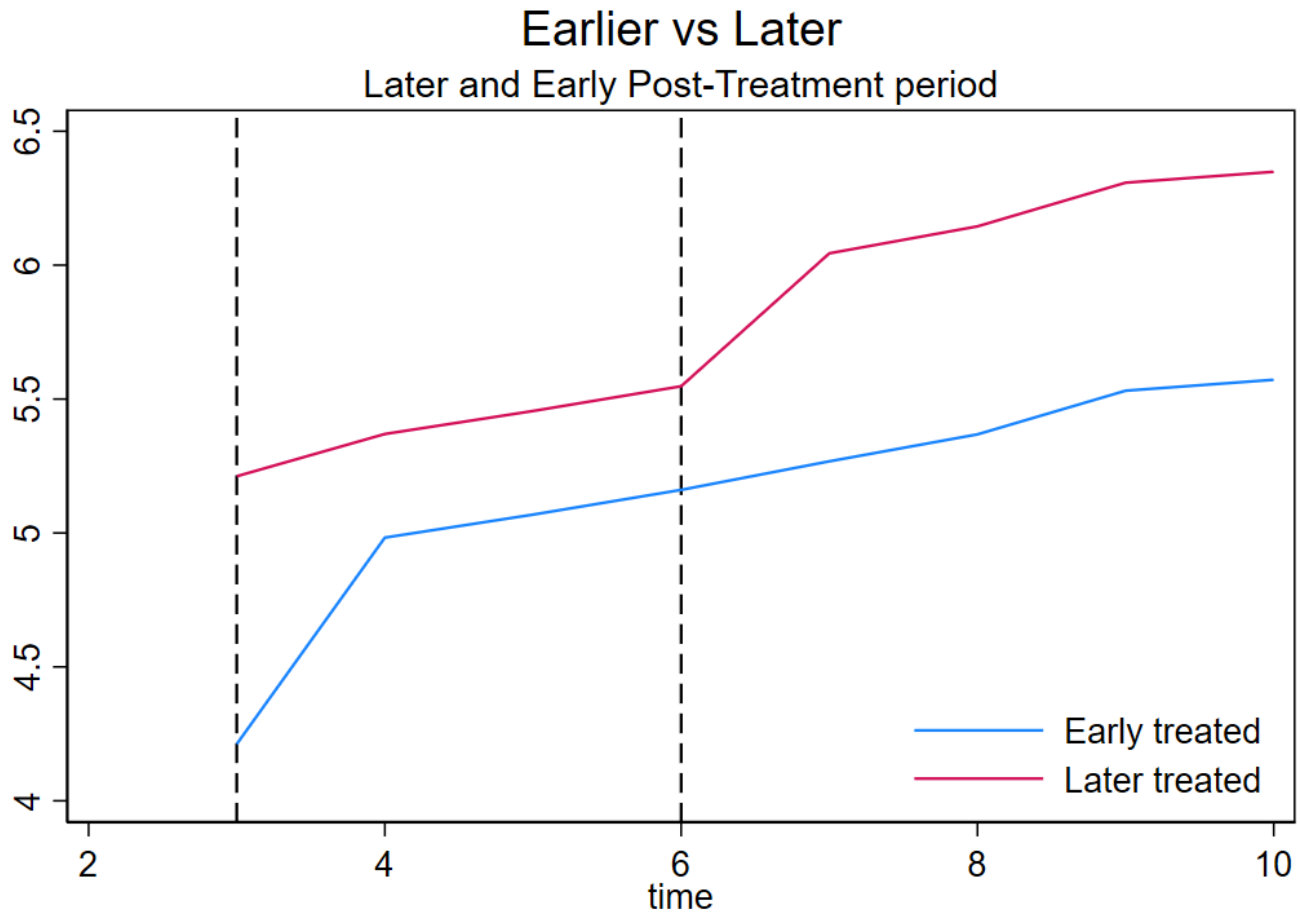
Homogeneous treatment with multiple treatment times I



Homogeneous treatment with multiple treatment times II



Homogeneous treatment with multiple treatment times III



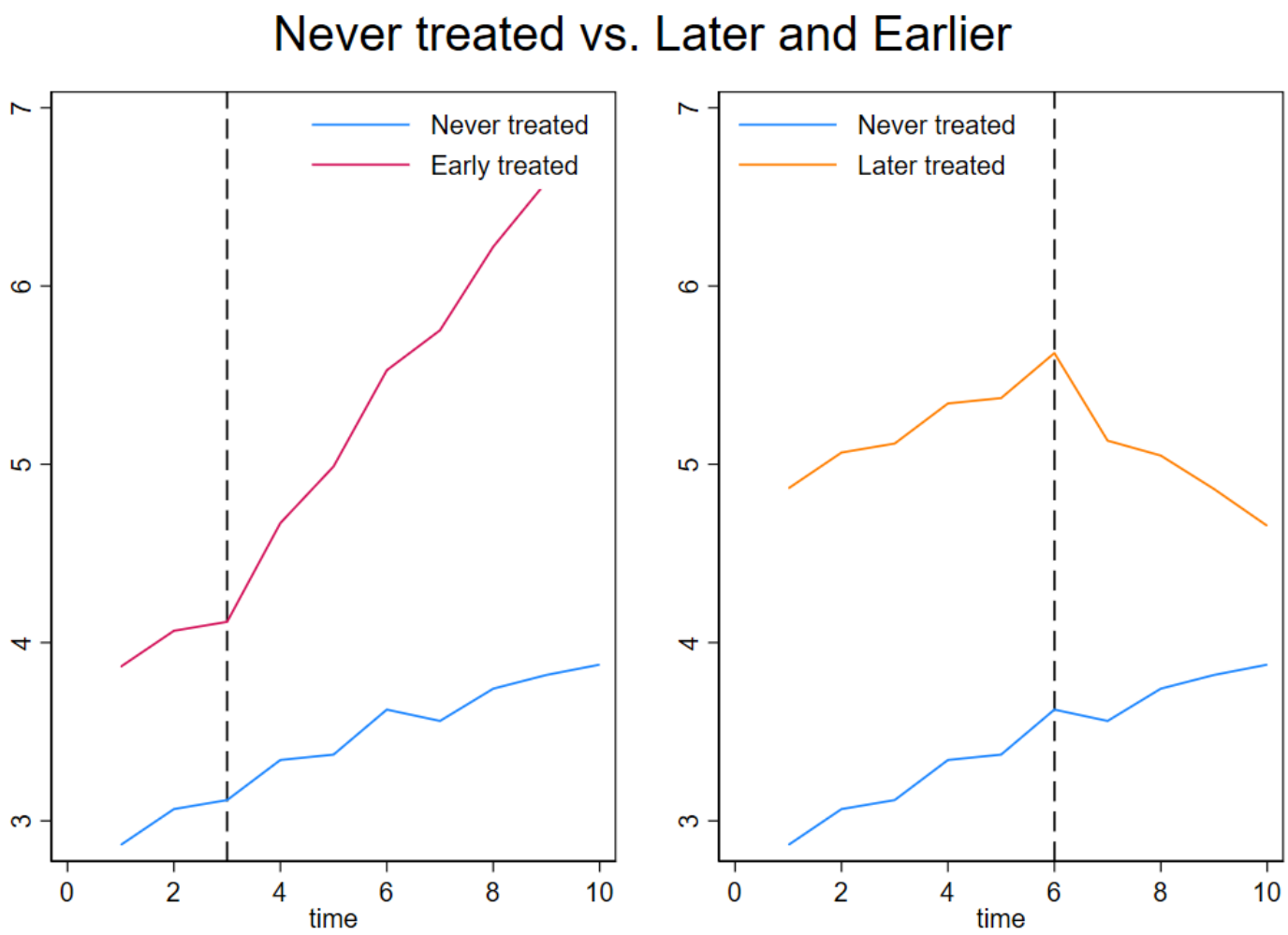
Homogeneous treatment: `estat bdecomp`

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

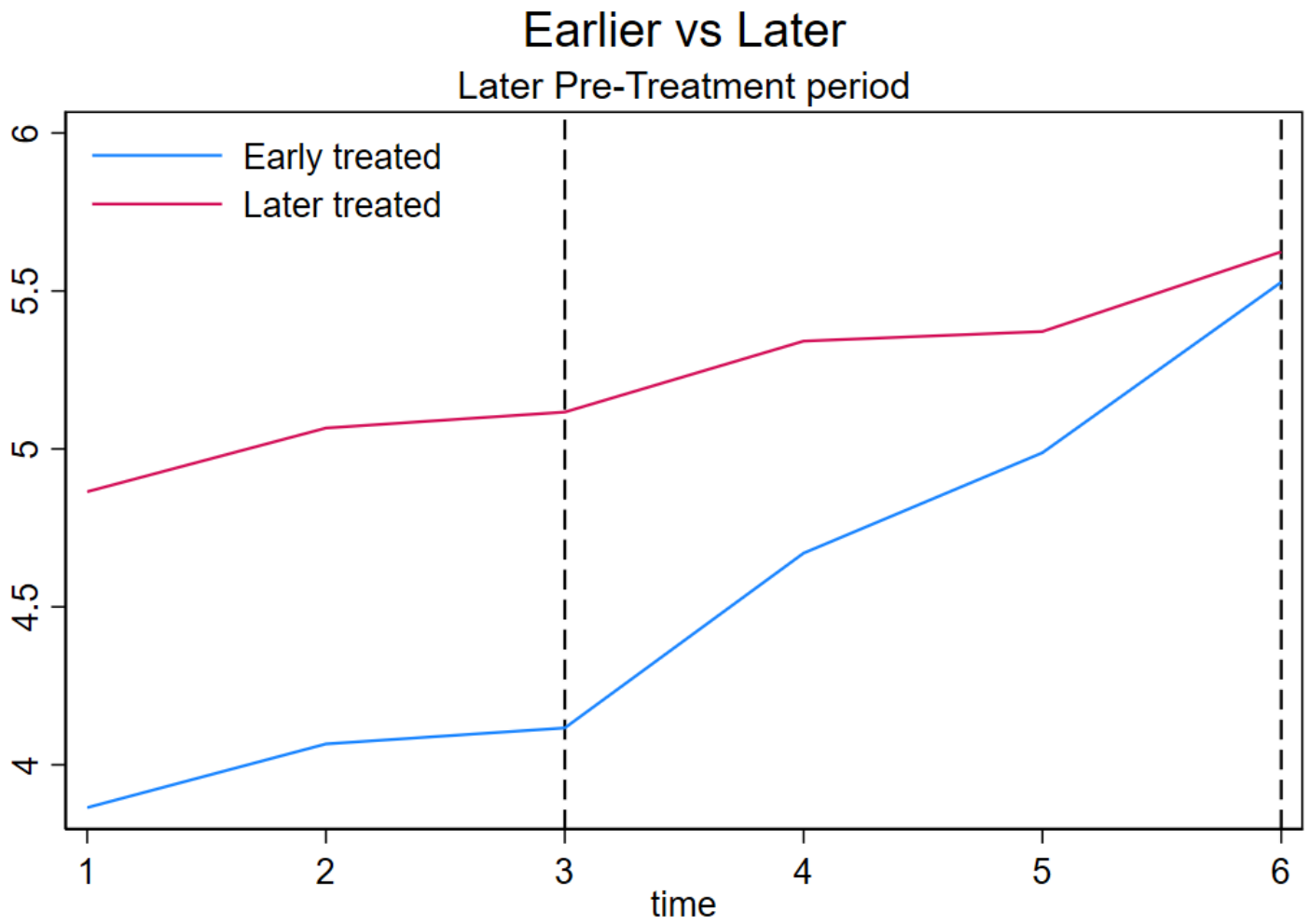
ATET decomposition summary	ATET component	Weight
Treated vs never treated	4.0435917	0.865605
Treated earlier vs later	4.0245287	0.057598
Treated later vs earlier	4.0331799	0.076797

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**.

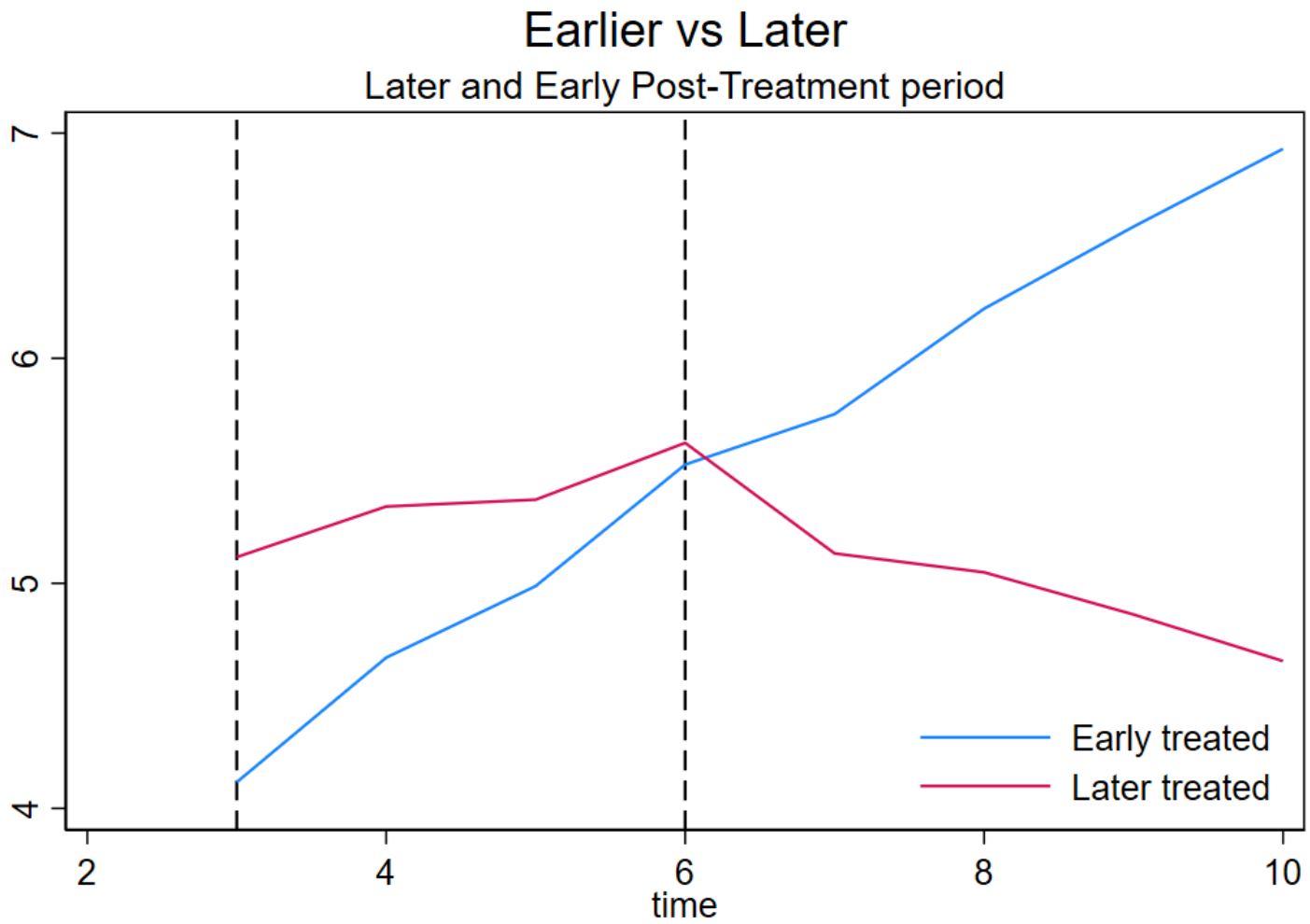
Heterogeneous treatment effect I



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 (2024)

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 (equivalent).

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

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

- 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

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

- 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

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

- 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

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

- 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/r19/akc.dta
 Observations:         1,410                  Fictional dog breed and AKC
                                                registration data
    Variables:             5                  1 Feb 2253 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/r19/akc.dta
 Observations:         1,410                  Fictional dog breed and AKC
                                                registration data
    Variables:             5                  1 Feb 2253 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 using
      the estimation sample.
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 II

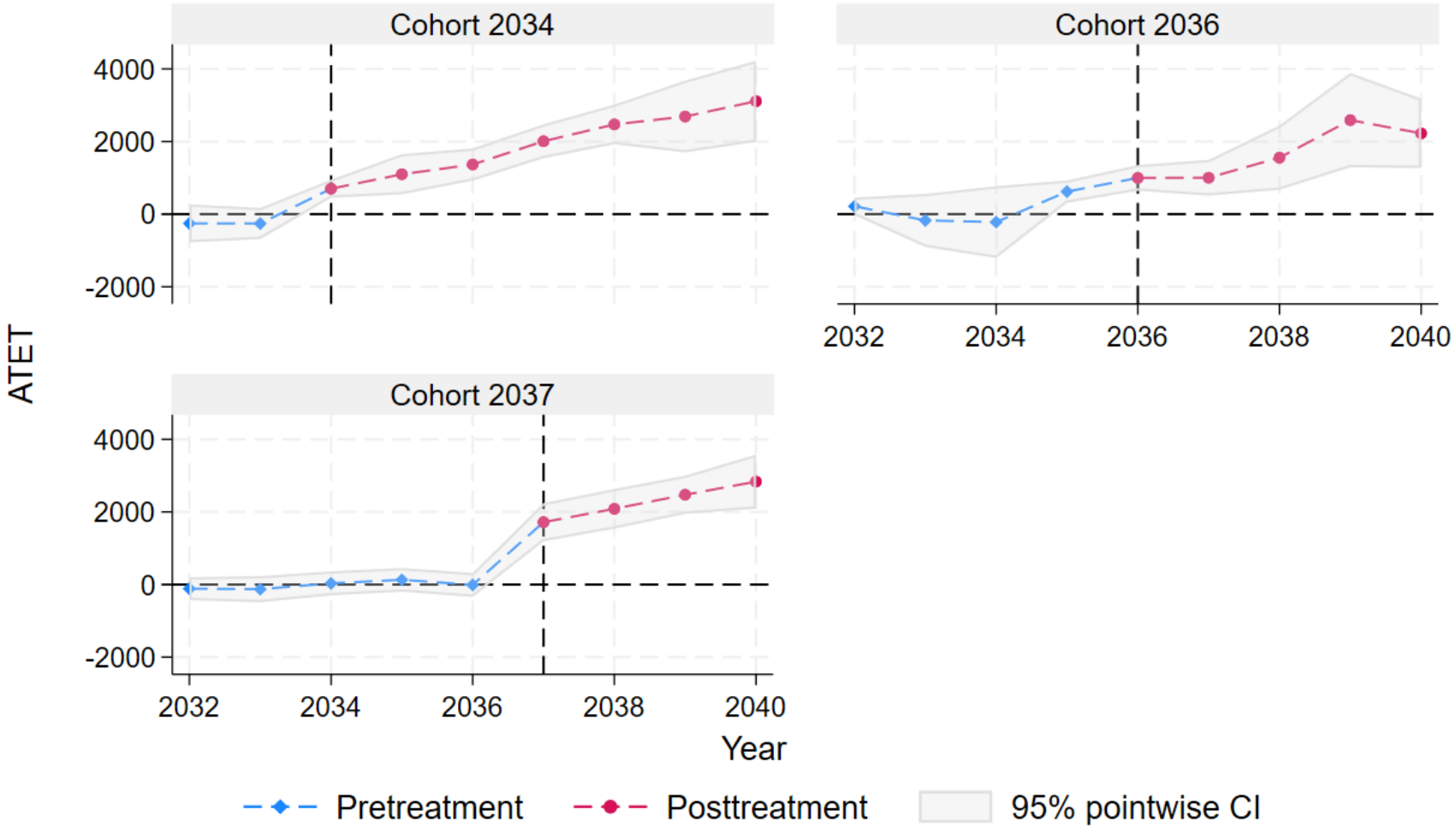
```
Heterogeneous-treatment-effects regression          Number of obs     = 1,410
                                                    Number of panels  =   141

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.
Note: Base time for pretreatment ATETs is the previous period.
```

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

SELECTION ON OBSERVABLES, AGAIN

ATE versus CATE

- ATE is a popular way to measure the treatment effects.

$$ATE = \mathbf{E}[y(1) - y(0)] \tag{1}$$

When each individual or group has different (heterogeneous) treatment effects, ATE might be inadequate.

- Conditional Average Treatment Effects (CATE) measure the treatment effects conditional on a set of variables.

$$CATE = \mathbf{E}[y(1) - y(0)|\mathbf{x}] \tag{2}$$

CATE measures the treatment effects as a function of $\mathbf{x}$.

cate

- `cate` is for cross-sectional data
- Estimate heterogeneous treatment effects
- Test and explore the behavior of heterogeneity
- Think about policy evaluation

Different versions of CATE

$$CATE = \mathbf{E}[y(1) - y(0)|\mathbf{x}]$$

Depending on the definition of **x**, CATE helps us to understand the heterogeneous treatment effects at different levels.

- IATE: Individualized average treatment effects when **x** is individual characteristics (finest level of treatment effects).
- GATE: Group average treatment effects when **x** is a group (prespecified group analysis).
- GATES: Sorted group average treatment effects when **x** ranks IATEs (data-driven group hypothesis testing).

Methodological building blocks

- Generalized random forest: estimates the IATE function $\tau(\mathbf{x}) = \mathbf{E}[y(1) - y(0)|\mathbf{x}]$
- Debiased/double machine learning: partialling-out and AIPW estimators + cross-fitting my

Advantages:

1. Flexible IATE estimation without assuming parametric assumptions
2. High-dimensional controls in both the outcome and the treatment models
3. Guard against machine learning bias

CATE estimation aipw

$$y(1) = g_1(\mathbf{x}, \mathbf{w}) + \epsilon_1$$
$$y(0) = g_0(\mathbf{x}, \mathbf{w}) + \epsilon_0$$
$$d = f(\mathbf{x}, \mathbf{w}_2) + u$$

The AIPW version of the potential outcomes are

$$y(1)_{AIPW} = g_1(\mathbf{x}, \mathbf{w}) + \frac{I(d = 1)(y - g_1(\mathbf{x}, \mathbf{w}))}{f(\mathbf{x}, \mathbf{w}_2)}$$
$$y(0)_{AIPW} = g_0(\mathbf{x}, \mathbf{w}) + \frac{I(d = 0)(y - g_0(\mathbf{x}, \mathbf{w}))}{1 - f(\mathbf{x}, \mathbf{w}_2)}$$

We can estimate the function $g_1(\mathbf{x}, \mathbf{w})$, $g_0(\mathbf{x}, \mathbf{w})$, and $f(\mathbf{x}, \mathbf{w}_2)$, so we can also estimate $y(1)_{AIPW}$ and $y(0)_{AIPW}$.
Let

$$\widehat{\Gamma} = \widehat{y(1)}_{AIWP} - \widehat{y(0)}_{AIWP}$$

AIPW scores

- For IATE, solve $\tau(\mathbf{x})$ in

$$\sum_{i=1}^{N}[\alpha(\mathbf{x}_i)(\widehat{\Gamma}_i - \tau(\mathbf{x}))] = 0$$

  We use $\widehat{\Gamma}$ as the dependent variable in a machine learning prediction problem.
- For GATE or GATES, we

  `regress` $\widehat{\Gamma}$ on `i.groupvar`

  Mean of $\widehat{\Gamma}$ within each group.
- For linear or series projection, we do linear or series projection of $\widehat{\Gamma}$ on the specific variables.
- For policy evaluation, we need to evaluate the weighted mean of the AIPW potential outcomes.

$$\Pi(\pi) = \mathbf{E}[\pi_i y(1)_{AIPW} + (1 - \pi_i) y(0)_{AIWP}]$$

CATE partialling-out estimator

$$y = d\tau(\mathbf{x}) + g(\mathbf{x}, \mathbf{w}) + \epsilon$$
$$d = f(\mathbf{x}, \mathbf{w}) + u$$

Taking conditional expecation

$$\mathbf{E}[y|\mathbf{x}, \mathbf{w}] = f(\mathbf{x}, \mathbf{w}) * \tau(\mathbf{x}) + g(\mathbf{x}, \mathbf{w})$$

Taking differences we abstract from $g(\mathbf{x}, \mathbf{w})$

$$\overbrace{y - \mathbf{E}[y|\mathbf{x}, \mathbf{w}]}^{\widetilde{y}} = \overbrace{(d - f(\mathbf{x}, \mathbf{w}))}^{\widetilde{d}} * \tau(\mathbf{x}) + \epsilon$$

Estimate $\tau(\mathbf{x})$ by solving a local moment condition via generalized random forest.

$$\mathbf{E}\left[\alpha(\mathbf{x}) * \widetilde{d} * \left(\widetilde{y} - \widetilde{d} * \tau(\mathbf{x})\right)\right] = 0$$

Simulating Heterogeneous Treatments

```
. clear all
. program define mkdata,
  1.          syntax [anything], [                                    ///
>                                 n(integer 1000)                   ///
>                                 k(integer 1)                      ///
>                                 seed(integer 1234)                ///
>                                ]
  2.          clear
  3.          set obs `n´
  4.          set seed `seed´
  5.          generate x1 = rnormal()
  6.          generate x2 = (rchi2(5)-5)/10
  7.          generate a  = runiformint(1,3)
  8.          generate y0 = .5     - x1 + x2 + x1*a  + rnormal()
  9.          generate y1 = (-1)^(`k´) - x1 - x2 - x1*a  + rnormal()
 10.          generate t  = .5*(1 - x1) + rnormal()>0
 11.          generate ig = `k´
 12.          label var ig "income group"
 13.          label define ig 1 "low" 2 "medium-low" 3 "medium" 4 "high"
 14.          label values ig ig
 15.          generate y  = y1*t + (1-t)*y0
 16. end
```

Generating data

```
. mkdata
Number of observations (_N) was 0, now 1,000.
. save te1, replace
file te1.dta saved
. mkdata, k(2)
Number of observations (_N) was 0, now 1,000.
. save te2, replace
file te2.dta saved
. mkdata, k(3)
Number of observations (_N) was 0, now 1,000.
. save te3, replace
file te3.dta saved
. mkdata, k(4)
Number of observations (_N) was 0, now 1,000.
. append using te1 te2 te3
(label ig already defined)
(label ig already defined)
(label ig already defined)
```

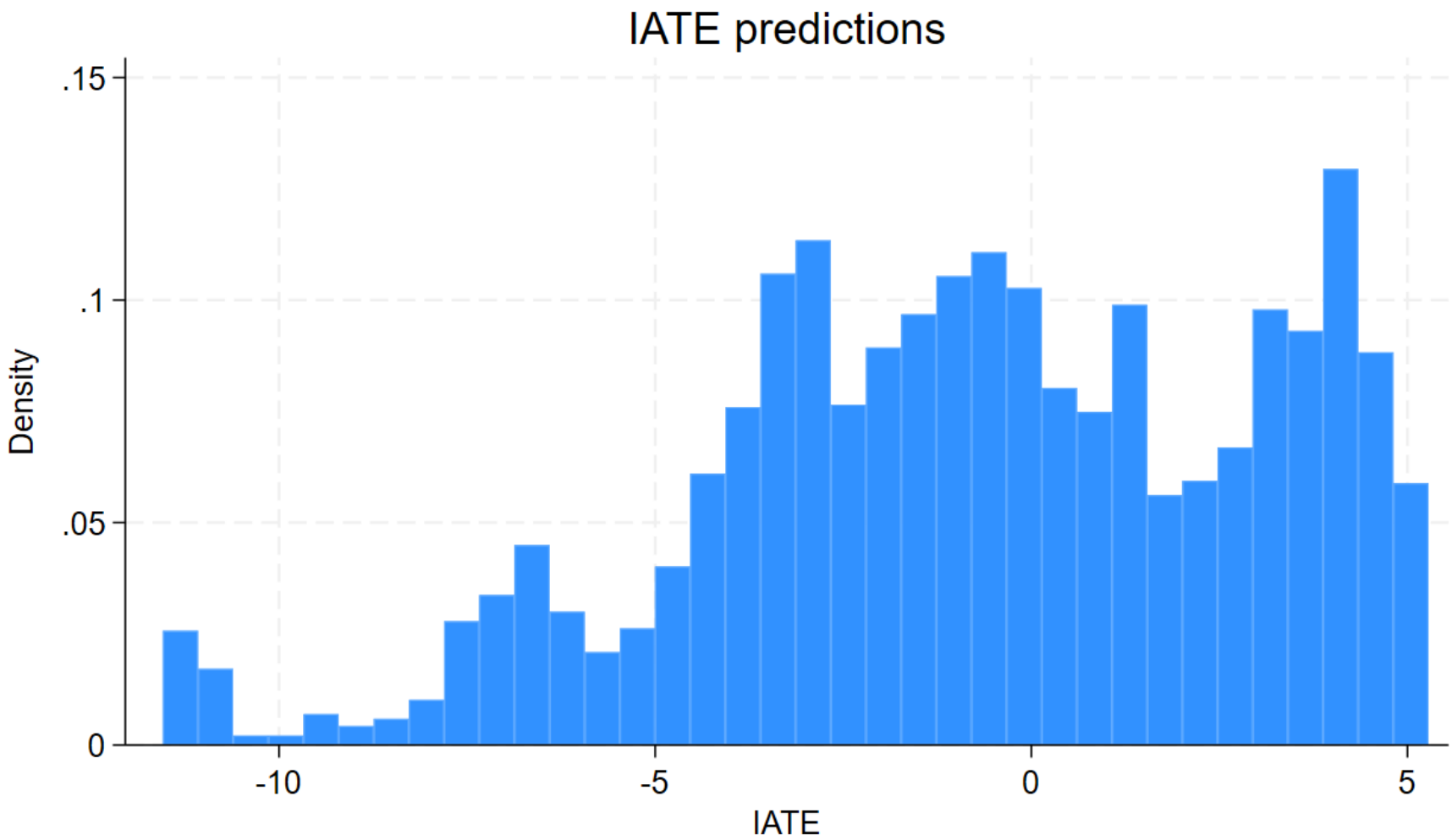
Estimation

```
. cate po (y (c.x1 c.x2 i.a)##(c.x1 c.x2 i.a)##(c.x1 c.x2 i.a)##i.ig) ///
>                 (t), nolog rseed(111)
Conditional average treatment effects     Number of observations       = 4,000
Estimator:       Partialing out           Number of folds in cross-fit =    10
Outcome model:   Linear lasso             Number of outcome controls   =   139
Treatment model: Logit lasso              Number of treatment controls =   139
CATE model:      Random forest            Number of CATE variables     =   139
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
t						
(1 vs 0)	-.578933	.0873457	-6.63	0.000	-.7501274	-.4077386
POmean						
0.t	.5057817	.0527184	9.59	0.000	.4024556	.6091078

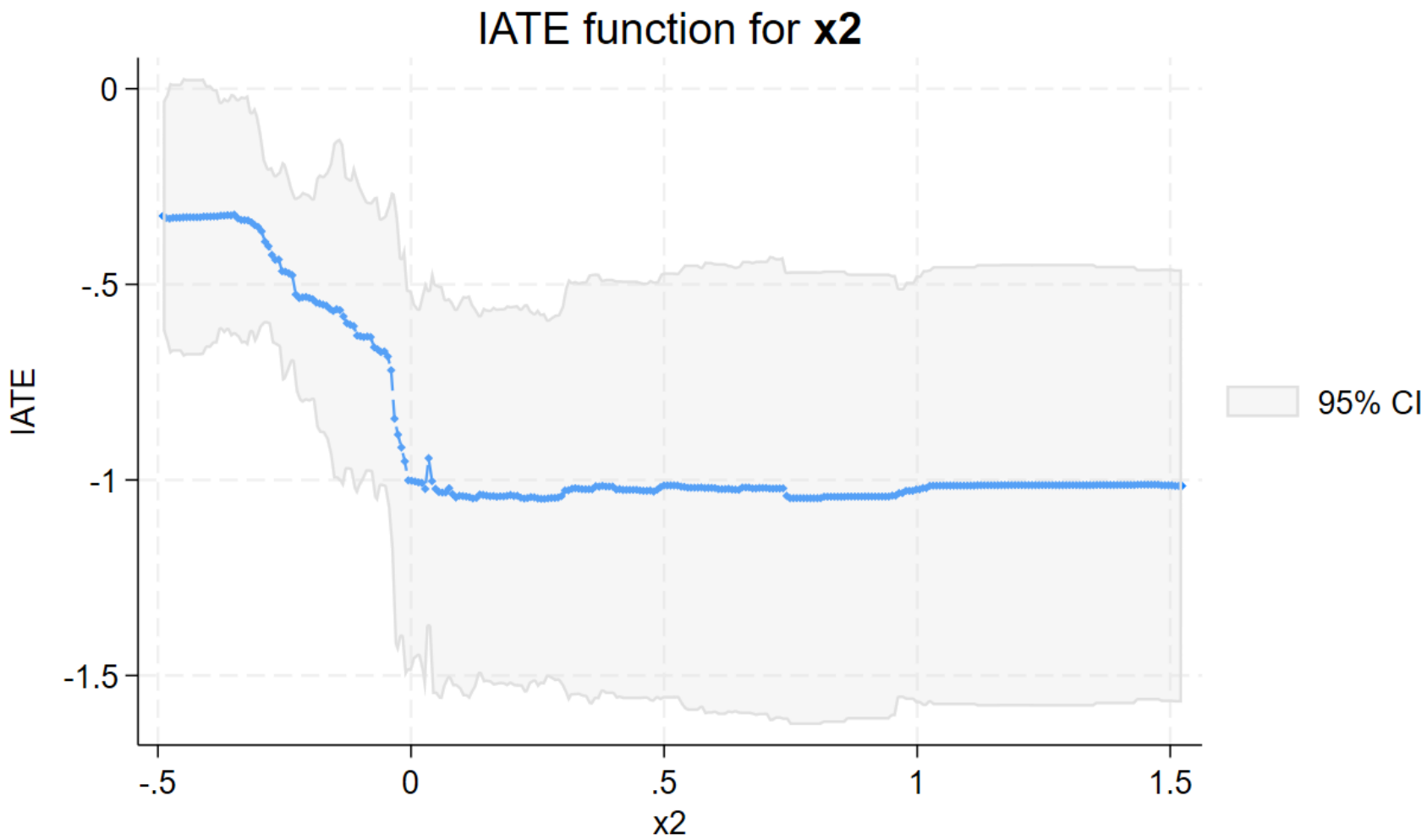
Exploring heterogeneity I

```
. categraph histogram
```



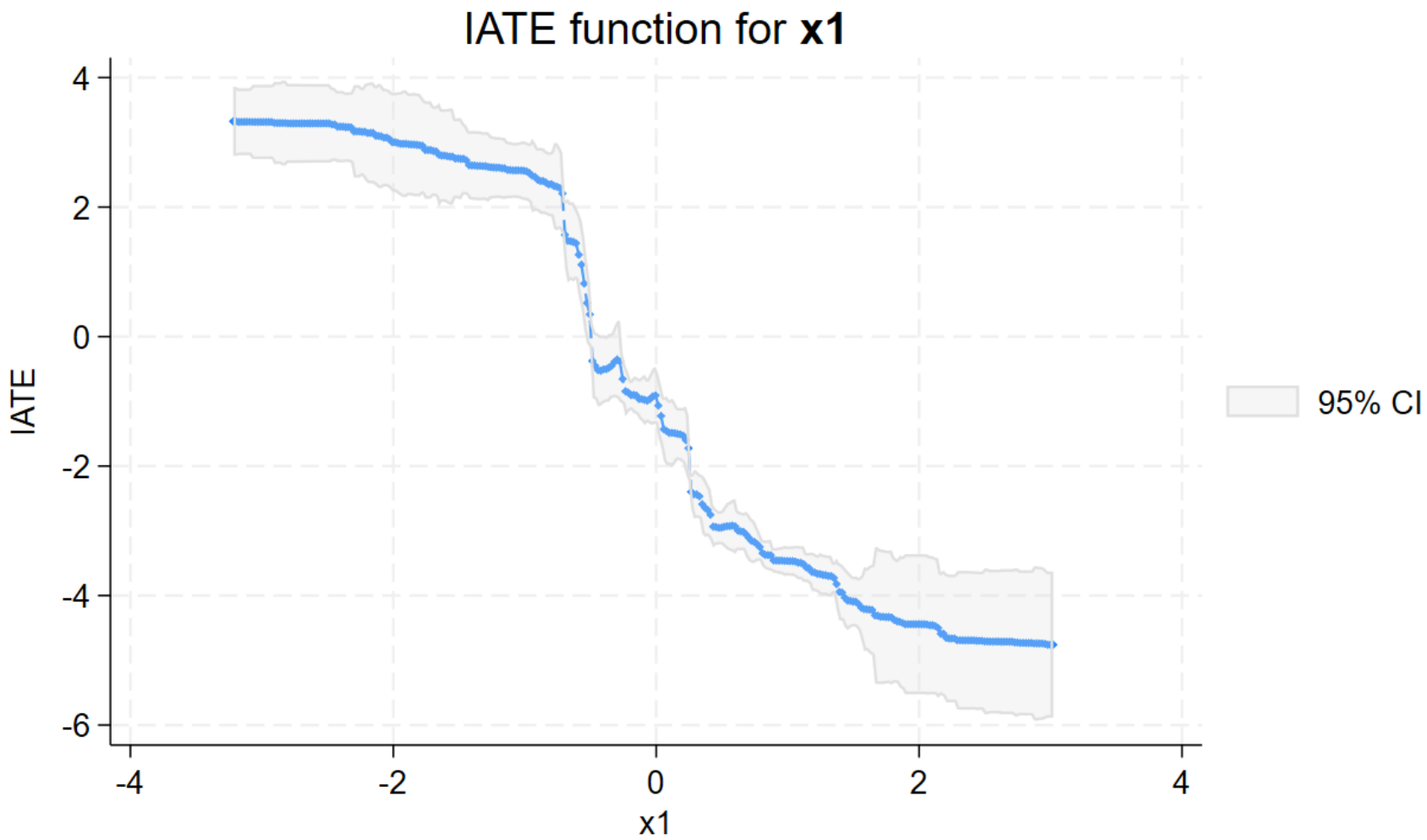
Exploring the IATE function `x2`

```
. categraph iateplot x2
```



Exploring the IATE function `x1`

```
. categraph iateplot x1
```



Exploring heterogeneity II

```
. estat heterogeneity

Treatment-effects heterogeneity test
H0: Treatment effects are homogeneous

    chi2(1) = 6799.84
Prob > chi2 = 0.0000
```

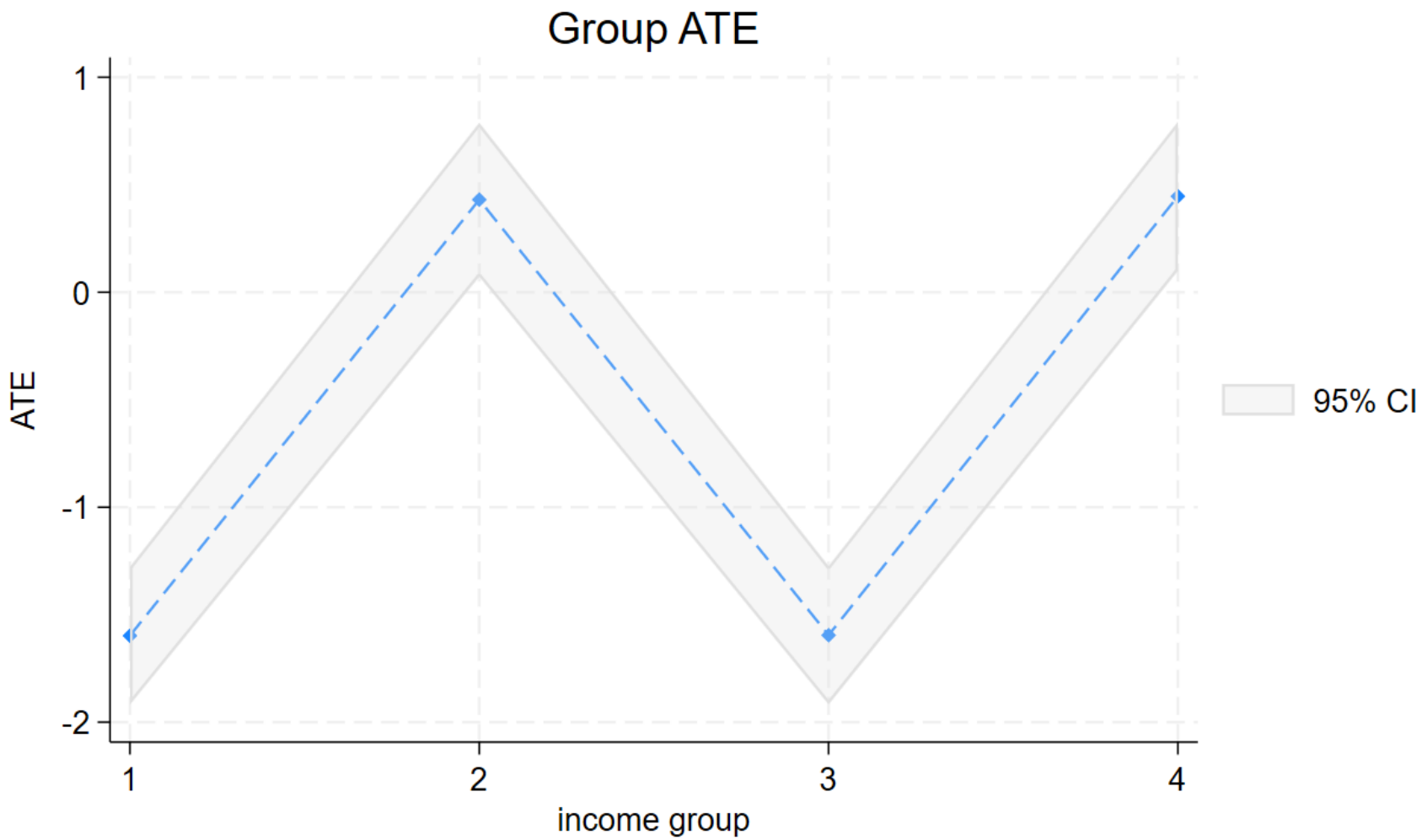
Exploring group specific heterogeneity

```
. cate, reestimate group(ig)
Estimating GATE ...
Conditional average treatment effects      Number of observations       = 4,000
Estimator:       Partialing out            Number of folds in cross-fit =    10
Outcome model:   Linear lasso              Number of outcome controls   =   139
Treatment model: Logit lasso               Number of treatment controls =   139
CATE model:      Random forest             Number of CATE variables     =   139
```

y	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
GATE						
ig						
low	-1.597546	.1631923	-9.79	0.000	-1.917397	-1.277695
medium-low	.4307375	.1822659	2.36	0.018	.0735028	.7879721
medium	-1.595453	.1638551	-9.74	0.000	-1.916603	-1.274303
high	.446529	.176707	2.53	0.012	.1001896	.7928683
ATE						
t						
(1 vs 0)	-.578933	.0873457	-6.63	0.000	-.7501274	-.4077386
POmean						
0.t	.5057817	.0527184	9.59	0.000	.4024556	.6091078

Visualizing heterogeneity

```
. categraph gateplot
```



Testing

```
. estat gatetest
Group treatment-effects heterogeneity test
H0: Group average treatment effects are homogeneous
 ( 1)  [GATE]1bn.ig - [GATE]2.ig = 0
 ( 2)  [GATE]1bn.ig - [GATE]3.ig = 0
 ( 3)  [GATE]1bn.ig - [GATE]4.ig = 0
    chi2(3) = 140.60
Prob > chi2 = 0.0000
```

Treatment-assignment policy

- Policy value:

$$\Pi(\pi) = \mathbf{E}\left[\pi_i y_i(1) + (1 - \pi_i) y_i(0)\right]$$

  where $\pi_i \in [0, 1]$ is a prespecified treatment-assignment probability for the ith observation. π_i is also refered to as the policy.

- Notice that, from IATE estimates, we can already estimate $y(1)$ and $y(0)$. Thus, policy evaluation is closely related to CATE.
- Compare two policies:

$$\Pi\left(\pi_A\right) - \Pi\left(\pi_B\right)$$

ATE is a special case of policy comparison

Let $\pi_A = 1$ and $\pi_B = 0$. Then

$$\begin{aligned} ATE &= \mathbf{E}[y(1)] - \mathbf{E}[y(0)] \\ &= \mathbf{E}[1 * y(1) + 0 * y(0)] - \mathbf{E}[0 * y(1) + 1 * y(0)] \\ &= \Pi(\pi_A) - \Pi(\pi_B) \end{aligned}$$

Thus, ATE is the contrast of the two special policy values. π_A means treat all the units, while π_B means treat none.

Smoking moms

```
. quietly webuse cattaneo2, clear
. cate aipw (bweight c.medu##(mmarried mage fbaby foreign)) (mbsmoke),    ///
>         nolog rseed(111)
Conditional average treatment effects     Number of observations       = 4,642
Estimator:       Augmented IPW            Number of folds in cross-fit =    10
Outcome model:   Linear lasso             Number of outcome controls   =    79
Treatment model: Logit lasso              Number of treatment controls =    79
CATE model:      Random forest            Number of CATE variables     =    79
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-239.1968	23.42006	-10.21	0.000	-285.0993	-193.2943
POmean						
mbsmoke Nonsmoker	3402.832	9.590317	354.82	0.000	3384.035	3421.629

ATE

```
. generate alltreated   = 1
. generate alluntreated = 0
. estat policyeval alltreated alluntreated
Treatment-assignment policy evaluation                  Number of obs = 4,642
```

	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
Value						
policy						
alltreated	3163.635	21.40817	147.78	0.000	3121.676	3205.594
alluntreated	3402.832	9.590317	354.82	0.000	3384.035	3421.629
Contrast						
policy						
(alltreated vs alluntrea~d)	-239.1968	23.42006	-10.21	0.000	-285.0993	-193.2943

Let them smoke

```
. generate onlysmoke    = (mbsmoke==1)*(alcohol==0)*(prenatal1==1)
. estat policyeval onlysmoke alluntreated if onlysmoke==1
Treatment-assignment policy evaluation                  Number of obs = 555
```

	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
Value						
policy						
onlysmoke	3415.753	162.7556	20.99	0.000	3096.758	3734.748
alluntreated	3372.81	4.40451	765.76	0.000	3364.177	3381.442
Contrast						
policy						
(onlysmoke vs alluntrea~d)	42.9435	162.4739	0.26	0.792	-275.4994	361.3864