

Treatment effect estimation in Stata

Eduardo Garcia Echeverri

APSA Annual Meeting and Exhibition, Vancouver, 2025

Outline

1 Treatment effects (TE) estimation – teffects

2 High-dimensional data – telasso

3 Treatment effect heterogeneity – cate

Treatment effect (TE) estimation – Cross-sectional data

We are interested in estimating TE's from observational data.

- Potential outcomes (PO) not observed
- Treatment not randomly assigned

unbalance: Treatment and control groups are not comparable

Estimators:

1. Regression adjustment (RA)
2. Inverse probability weighting (IPW)
3. Augmented IPW (AIPW)
4. IPW regression adjustment (IPWRA)
5. Matching

RA estimator in a nutshell

Assume a model for PO's. For example, a linear model:

- $y_i(t) = \beta_t X_i + \varepsilon_i$

Other models available: probit, logit, hetprobit, poisson ...

ATE estimation (point estimates):

1. Use treated units to estimate the model for $y_i(1)$;
 - Predict $\hat{y}_i(1)$ for all units
2. Use control units to estimate the model for $y_i(0)$;
 - Predict $\hat{y}_i(0)$ for all units
3. Compute $\hat{y}_i(1) - \hat{y}_i(0)$ for all units and take the average.

Remark: SE's come from GMM.

Example – Cattaneo (2010)

Measure effect of smoking during pregnancy (`mbsmoke`) on the weight at birth (`bweight`).

Controls: Age, education, marital status, indicator of a prenatal exam on 1st trimester, whether this baby was the first.

RA estimator – Output

```
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mb smoke)
```

```
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
```

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

IPW estimator

Assume a model for the treatment. For example, a logit model:

- $\Pr[T_i = 1|X_i] = \frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)}$

Other models available: probit and hetprobit.

ATE estimation (point estimates):

1. Estimate the treatment model using all units.
 - Predict the treatment probability for each unit p_i .
2. Compute IPW's.
 - $w_i = 1/p_i$ if unit is treated, $w_i = 1/(1 - p_i)$ otherwise
3. Calculate weighted means of outcome for treated and control units and take the difference.

Remark: SE's come from GMM.

IPW estimator – Output

```
. teffects ipw (bweight) (mbsmoke prenatal1 mmarried mage fbaby)
```

```
Iteration 0: EE criterion = 1.873e-22
```

```
Iteration 1: EE criterion = 3.428e-26
```

```
Treatment-effects estimation      Number of obs      =      4,642
```

```
Estimator      : inverse-probability weights
```

```
Outcome model  : weighted mean
```

```
Treatment model: logit
```

		Robust		z	P> z	[95% conf. interval]	
bweight		Coefficient	std. err.				
ATE							
	mbsmoke						
(Smoker vs Nonsmoker)		-236.1038	23.86187	-9.89	0.000	-282.8722	-189.3354
POmean							
	mbsmoke						
	Nonsmoker	3402.552	9.539555	356.68	0.000	3383.855	3421.249

Model diagnostics – Balanced covariates

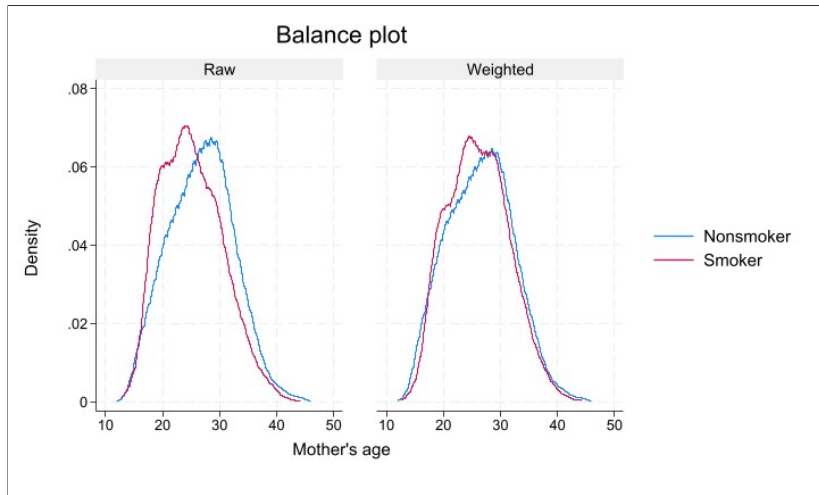
```
. tebalance sum
```

```
Covariate balance summary
```

	Raw	Weighted
Number of obs =	4,642	4,642.0
Treated obs =	864	2,315.3
Control obs =	3,778	2,326.7

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
prenatal1	-.3242695	-.0156339	1.496155	1.023424
mmarried	-.5953009	-.0105562	1.335944	1.009079
mage	-.300179	-.0672115	.8818025	.8536401
fbaby	-.1663271	.0257705	.9430944	1.005698

Model diagnostics – Balanced covariates



Formally testing for balanced covariates

teffects ipw is an exactly identified GMM model.

However, we can reestimate it with additional moment conditions

- $\mathbb{E}\left[\frac{t_i}{p_i}X_i - \frac{1-t_i}{1-p_i}X_i\right] = 0$

and test for overidentification.

```
Overidentification test for covariate balance
H0: Covariates are balanced

      chi2(5)      = 38.1464
      Prob > chi2   = 0.0000
```

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
- AIPW and IPWRA have this property.

Nearest neighborhood matching

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 do not allow for multivalued treatments

Nearest neighbor matching estimator – Output

```
. teffects nnmatch (bweight prenatal1 mmarried mage fbaby) (mb smoke)
```

```
Treatment-effects estimation      Number of obs      =      4,642
Estimator      : nearest-neighbor matching      Matches: requested =      1
Outcome model  : matching                      min =      1
Distance metric: Mahalanobis                  max =      139
```

	bweight	AI robust		z	P> z	[95% conf. interval]	
		Coefficient	std. err.				
ATE							
	mb smoke						
(Smoker vs Nonsmoker)		-240.3306	28.43006	-8.45	0.000	-296.0525	-184.6087

Propensity score matching

Estimate the treatment probabilities (propensity scores).

- Assign values to unobserved outcomes based on observed ones with similar propensity scores
- Estimate ATE
- These estimators do not allow for multivalued treatments

Propensity score matching estimator – Output

```
. teffects psmatch (bweight) (mb smoke prenatal1 mmarried mage fbaby)
```

```
Treatment-effects estimation      Number of obs      =      4,642
Estimator      : propensity-score matching      Matches: requested =      1
Outcome model  : matching                      min =      1
Treatment model: logit                      max =     139
```

		AI robust		z	P> z	[95% conf. interval]	
bweight		Coefficient	std. err.				
ATE	mb smoke						
	(Smoker vs Nonsmoker)	-235.1714	27.74409	-8.48	0.000	-289.5488	-180.794

Outline

- 1 Treatment effects (TE) estimation – teffects
- 2 High-dimensional data – telasso
- 3 Treatment effect heterogeneity – cate

TE estimation with variable selection

There is conflict between the assumptions in causal inference.
Often, with more conditioning variables:

- Selection on observables assumptions seems more believable,
- but the overlap assumption seems less.

Can we solve this conflict in a data-driven way?

Yes,

Include many potential variables + use lasso to select among them
+ make sure SE's account for model selection

How does lasso select variables?

Lasso solves the following minimization problem

$$\beta_{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{N} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

The penalty term on the right implies corner solutions. Therefore, many coefficients are set to 0.

telasso estimator – Chernozhukov et al. (2018)

6 steps to estimate ATE:

1. Use lasso to select variables in outcome model for each treatment level,
2. Use selected variables in 1 to
 - 2.1 Fit separate regression models of the outcome for each treatment level,
 - 2.2 Obtain treatment-specific outcomes for each unit

telasso estimator – Chernozhukov et al. (2018)

6 steps to estimate ATE:

3. Use lasso to select variables in the treatment model
4. Use selected variables in 3 to
 - 4.1 Estimate parameters of treatment model,
 - 4.2 Predict treatment probabilities and compute IPW's
5. Compute weighted means of treatment-specific predicted outcomes for each treatment level,
6. Take difference of means in 5.

Example – ATE of bilateral lung cancer (BLT)

BLT has a higher short-term death rate than single lung transplant (SLT), but provides a larger improvement on quality of life.

Objective: Measure the effect of BLT (vs. SLT) on forced expiratory volume in one second (FEV1%).

Data:

- Fictional dataset (`lung.dta`) inspired by Koch et al. (2018)
- 31 variables with characteristics of patients and donors.

\$allvars: Global with all controls plus interactions (454 variables)

Control variables

Variable name	Storage type	Display format	Value label	Variable label
agep	byte	%10.0g		Patient age (years)
bmip	double	%10.0g		Patient body mass index
diabetesp	byte	%12.0g	lbdiab	Patient diabetes status
heightp	double	%10.0g		Patient height (cm)
o2amt	double	%10.0g		Oxygen delivered
karn	byte	%8.0g	lbyes	Karnofsky score > 60
lungals	double	%10.0g		Lung allocation score
racep	byte	%8.0g	lbrace	Patient race
sexp	byte	%8.0g	lbsex	Patient gender
lifesvent	byte	%8.0g	lbyes	Life support ventilator needed
assisvent	byte	%8.0g	lbyes	Assisted ventilation needed
centervol	double	%10.0g		Center volume
walkdist	double	%10.0g		Walking distance in 6 minutes
o2rest	byte	%8.0g	lbyes	Oxygen needed at rest
aged	byte	%10.0g		Donor age (years)
raced	byte	%8.0g	lbrace	Donor race
bmld	double	%10.0g		Donor body mass index
smoked	byte	%8.0g	lbyes	Donor if has history of smoking
cmv	byte	%8.0g	lbyes	Positive cytomegalovirus test
deathcause	byte	%8.0g	lbyes	Cause of death - traumatic brain injury
diabetesd	byte	%12.0g	lbdiab	Donor diabetes status
expandd	byte	%8.0g	lbyes	Expanded donor needed
heightd	double	%10.0g		Donor height (cm)

telasso – Output

```
. telasso (fev1p $allvars) (transtype $allvars)

Estimating lasso for outcome fev1p if transtype = 0 using plugin method ...
Estimating lasso for outcome fev1p if transtype = 1 using plugin method ...
Estimating lasso for treatment transtype using plugin method ...
Estimating ATE ...

Treatment-effects lasso estimation      Number of observations      =      937
Outcome model:  linear                  Number of controls         =      454
Treatment model: logit                 Number of selected controls =       8
```

fev1p	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE						
transtype						
(BLT vs SLT)	37.51841	.1606703	233.51	0.000	37.20351	37.83332
POMean						
transtype						
SLT	46.4938	.2021582	229.99	0.000	46.09757	46.89002

Outline

- 1 Treatment effects (TE) estimation – `teffects`
- 2 High-dimensional data – `telasso`
- 3 Treatment effect heterogeneity – `cate`

Conditional average treatment effect

The conditional average treatment effect (CATE) is defined as:

$$CATE = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] \equiv \tau(x)$$

It helps to answer questions such as:

1. How do TE's change with covariates?
2. Do TE's change vary across prespecified groups?
3. Are there unknown groups in the data for which TE's differ?
4. Are TE's homogeneous or heterogeneous?
5. Among different treatment-assignment rules, which is best?

The cate command

Different versions of CATE:

IATE: Individualized average treatment effects,

x : characteristics of a specific observation

$\tau(x)$: expected TE for units with same characteristics as observation

GATE: Group average treatment effects,

x : pre-specified group indicators, denoted as G_i

$\tau(g)$: expected TE for units in group g

GATES: Sorted group average treatment effects,

x : data-driven group indicators based on quantiles on IATE's,

$\tau(q)$: expected TE for units in IATE's q -th quantile

Partialling-out (PO) estimator – Nie and Wager (2021)

We can estimate IATE with the partial linear model:

outcome: $y = d * \tau(x) + g(x, w) + \varepsilon$

treat.: $d = f(x, w) + u$

where,

d : Binary treatment

$g(\cdot)$: nuisance function for outcome model

$f(\cdot)$: nuisance function for treatment model

$\tau(x)$: IATE function

PO estimator

The partial linear model and unconfoundedness imply

$$y - \mathbb{E}[y|x, w] = (d - f(x, w))\tau(x) + \varepsilon$$

PO estimator (heuristics):

1. Estimate $\mathbb{E}[y|x, w]$ and $f(x, w)$ via lasso, random forest, or parametric regression.
2. Partial-out the residuals:
 - 2.1 $\tilde{y} = y - \hat{\mathbb{E}}[y|x, w]$
 - 2.2 $\tilde{d} = d - \hat{f}(x, w)$
3. Estimate $\tau(x)$ from \tilde{y} and \tilde{d} via generalized random forest or linear regression.

All details: [cate documentation](#)

AIPW estimator

We can relax the partial linear assumption:

PO 0: $y = g_0(x, w) + \varepsilon_0$

PO 1: $y = g_1(x, w) + \varepsilon_1$

treat.: $d = f(x, w) + u$

$$\tau(x) = \mathbb{E}[g_1(X, W) - g_0(X, W) | X = x]$$

where,

d : Binary treatment

g_0, g_1 : nuisance functions for outcome model

$f(\cdot)$: nuisance function for treatment model

Example – IATE estimation ($w = \emptyset$)

Objectives:

1. Measure effect of 401k eligibility (e401k) on net financial assets (asset)
2. Study treatment effect heterogeneity.

$$\tau(x) = \mathbb{E}[\text{asset}(1) - \text{asset}(0)|x]$$

Data:

- 1990 Survey of Income and Program Participation (excerpt)
- Data on demographics; household income and assets; and pension contributions.

Loading the data

```
. use https://www.stata-press.com/data/r19/assets3, clear
(Excerpt from Chernozhukov and Hansen (2004))

. global catecovars age educ i.(incomecat pension married twoearn ira ownhome)

. describe

Contains data from https://www.stata-press.com/data/r19/assets3.dta
Observations:      9,913      Excerpt from Chernozhukov and Hansen (2004)
Variables:         11        27 Feb 2025 19:19
                        (_dta has notes)
```

Variable name	Storage type	Display format	Value label	Variable label
assets	float	%9.0g		Net total financial assets
age	byte	%9.0g		Age
income	float	%9.0g		Household income
educ	byte	%9.0g		Years of education
pension	byte	%16.0g	lbpen	Pension benefits
married	byte	%11.0g	lbmar	Marital status
twoearn	byte	%9.0g	lbyes	Two-earner household
e401k	byte	%12.0g	lbe401	401(k) eligibility
ira	byte	%9.0g	lbyes	IRA participation
ownhome	byte	%9.0g	lbyes	Homeowner
incomecat	byte	%9.0g		Income category

PO estimator – estimation log

```
. cate po (assets $catecovars) (e401k), rseed(12345671)
```

```
Cross-fit fold 1 of 10 ...
```

```
Performing lasso for outcome assets ...
```

```
Performing lasso for treatment e401k ...
```

```
Cross-fit fold 2 of 10 ...
```

```
Performing lasso for outcome assets ...
```

```
Performing lasso for treatment e401k ...
```

```
Cross-fit fold 3 of 10 ...
```

```
Performing lasso for outcome assets ...
```

PO estimator – Results

Performing random forest for IATE ...

Estimating AIPW scores ...

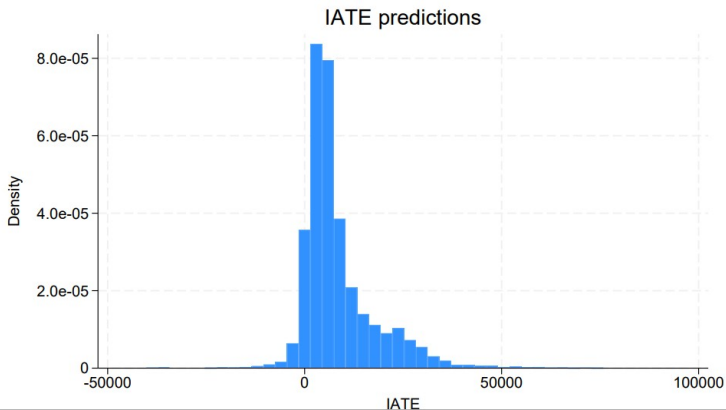
Estimating ATE ...

Conditional average treatment effects	Number of observations	= 9,913
Estimator: Partialing out	Number of folds in cross-fit	= 10
Outcome model: Linear lasso	Number of outcome controls	= 17
Treatment model: Logit lasso	Number of treatment controls	= 17
CATE model: Random forest	Number of CATE variables	= 17

	assets	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE	e401k						
(Eligible vs Not eligible)		7937.182	1153.017	6.88	0.000	5677.309	10197.05
POmean	e401k						
Not eligible		14016.38	833.4423	16.82	0.000	12382.87	15649.9

Visualizing the IATE function $\tau(x)$

```
. categraph histogram  
(bin=39, start=-40204.13, width=2975.4332)
```



Testing for heterogeneous treatment effects

```
. estat heterogeneity  
  
Treatment-effects heterogeneity test  
H0: Treatment effects are homogeneous  
  
      chi2(1) =    4.11  
Prob > chi2 = 0.0427
```

Analyzing $\tau(x)$ – Linear projection

```
. estat projection age educ i.incomecat
```

Treatment-effects linear projection

```
Number of obs =    9,913
F(6, 9906)     =     3.56
Prob > F       =    0.0016
R-squared      =    0.0040
Adj R-squared  =    0.0034
Root MSE      = 1.146e+05
```

	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
age	243.7552	127.6935	1.91	0.056	-6.550091	494.0606
educ	-284.8309	472.381	-0.60	0.547	-1210.794	641.1321
incomecat						
1	-2440.883	1952.137	-1.25	0.211	-6267.468	1385.702
2	1425.83	1948.687	0.73	0.464	-2393.993	5245.653
3	4773.172	2442.928	1.95	0.051	-15.46371	9561.807
4	16675.44	4613.107	3.61	0.000	7632.813	25718.07
_cons	-2396.73	8114.533	-0.30	0.768	-18302.87	13509.41

Analyzing $\tau(x)$ – Nonparametric regression

```
. estat series age

Computing approximating function

Minimizing cross-validation criterion

Iteration 0: Cross-validation criterion = 1.32e+10

Computing average derivatives

Nonparametric series regression for IATE
Cubic B-spline estimation      Number of obs      =      9,913
Criterion: cross-validation    Number of knots     =       1
```

	Effect	Robust std. err.	z	P> z	[95% conf. interval]
age	326.5787	156.9877	2.08	0.037	18.88848 634.2689

Note: Effect estimates are averages of derivatives.

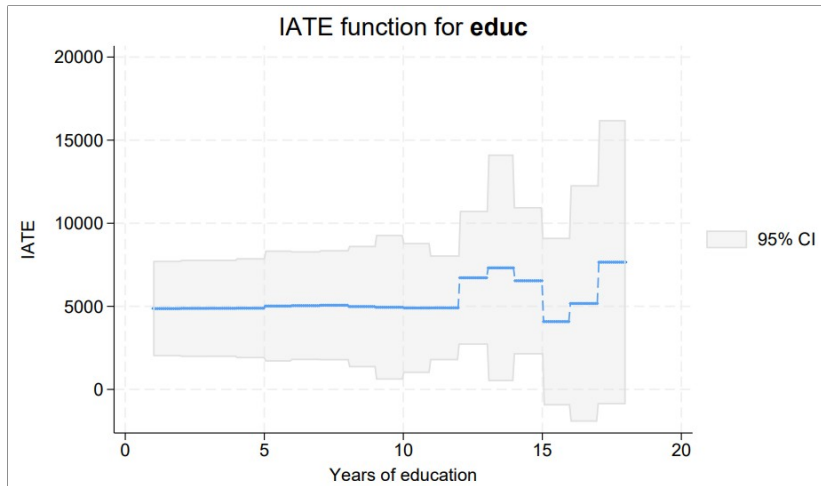
Plotting $\tau(x)$

```
. categraph iateplot educ
```

Note: IATE estimated at fixed values of covariates other than **educ**

Variable	Statistic	Value	Type
age	mean	41.05891	continuous
incomecat	base	0	factor
ira	base	0	factor
married	base	0	factor
ownhome	base	0	factor
pension	base	0	factor
twoearn	base	0	factor

Plotting $\tau(x)$



Options – High-dimensional controls ($w \neq \emptyset$)

```
. global fvars incomecat pension married twoearn ira ownhome
. global controls c.(educ age)#i.($fvars)
. cate po (assets $catecovars) (e401k), rseed(12345671) controls($controls) nolog
```

```
Conditional average treatment effects      Number of observations      = 9,913
Estimator:      Partialing out              Number of folds in cross-fit =   10
Outcome model:   Linear lasso                Number of outcome controls  =   47
Treatment model: Logit lasso                 Number of treatment controls =   47
CATE model:      Random forest               Number of CATE variables    =   17
```

	assets	Robust				
		Coefficient	std. err.	z	P> z	[95% conf. interval]
ATE	e401k					
	(Eligible vs Not eligible)	8107.563	1144.817	7.08	0.000	5863.763 10351.36
P0mean	e401k					
	Not eligible	13902.88	838.5924	16.58	0.000	12259.27 15546.49

Options – Changing estimation methods, oob

```
. cate po (assets $catecovars) (e401k), rseed(12345671) nolog omethod(rforest) tmethod(rforest) oob
```

Conditional average treatment effects	Number of observations	= 9,913
Estimator: Partialing out	Number of folds in cross-fit	= 1
Outcome model: Random forest	Number of outcome controls	= 17
Treatment model: Random forest	Number of treatment controls	= 17
CATE model: Random forest	Number of CATE variables	= 17

assets	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE e401k (Eligible vs Not eligible)	8225.258	1173.862	7.01	0.000	5924.53	10525.99
POmean e401k Not eligible	14016.34	850.1257	16.49	0.000	12350.13	15682.56

Options – Changing estimation methods

```
. cate po (assets $catecovars) (e401k), rseed(12345671) nolog omethod(regress) tmethod(logit) cmethod(regress)
```

```
Conditional average treatment effects      Number of observations      = 9,913
Estimator:      Partialing out              Number of folds in cross-fit = 10
Outcome model:  Linear regression            Number of outcome controls  = 17
Treatment model: Logit                      Number of treatment controls = 17
CATE model:     Linear regression            Number of CATE variables    = 17
```

	assets	Robust				
		Coefficient	std. err.	z	P> z	[95% conf. interval]
ATE	e401k					
	(Eligible vs Not eligible)	7904.218	1155.565	6.84	0.000	5639.351 10169.08
P0mean	e401k					
	Not eligible	13977.45	831.0932	16.82	0.000	12348.54 15606.37

Example – GATE estimation

```
. cate, group(incomecat) reestimate
```

```
Estimating GATE ...
```

```
Conditional average treatment effects   Number of observations   = 9,913
Estimator: Partialing out                Number of folds in cross-fit = 10
Outcome model: Linear lasso              Number of outcome controls = 17
Treatment model: Logit lasso             Number of treatment controls = 17
CATE model: Random forest                 Number of CATE variables   = 17
```

	assets	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
GATE	incomecat						
	0	4087.014	987.7124	4.14	0.000	2151.133	6022.895
	1	1399.398	1663.193	0.84	0.400	-1860.4	4659.196
	2	5154.329	1349.842	3.82	0.000	2508.688	7799.97
	3	8532.238	2287.664	3.73	0.000	4048.499	13015.98
	4	20510.94	4723.741	4.34	0.000	11252.58	29769.3
ATE	e401k (Eligible vs Not eligible)	7937.182	1153.017	6.88	0.000	5677.309	10197.05
P0mean	e401k Not eligible	14016.38	833.4423	16.82	0.000	12382.87	15649.9

Example – GATES estimation

Data generating process:

- We generate 4 groups with TE's of 0, 4, 8, and 16;
- Groups are unknown during estimation;
- They must be recovered from the data.

Details on the DGP are in the do-file.

Example – GATES 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) group(4)
```

```
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
```

		Robust				
	y	Coefficient	std. err.	z	P> z	[95% conf. interval]
GATES						
	rank					
	1	15.56204	.4066135	38.27	0.000	14.76509 16.35899
	2	8.742403	.4030428	21.69	0.000	7.952454 9.532353
	3	4.415533	.3665444	12.05	0.000	3.697119 5.133946
	4	-1.096897	.3451045	-3.18	0.001	-1.773289 -.4205041
ATE						
	t					
	(1 vs 0)	6.917373	.2134812	32.40	0.000	6.498958 7.335789
P0mean						
	0.t	.2118568	.1561817	1.36	0.175	-.0942538 .5179673

GATES testing

```
. estat gatetest
```

Sorted group treatment-effects heterogeneity test

H0: Sorted group average treatment effects are homogeneous

```
( 1)  [GATES]1bn.rank - [GATES]2.rank = 0
```

```
( 2)  [GATES]1bn.rank - [GATES]3.rank = 0
```

```
( 3)  [GATES]1bn.rank - [GATES]4.rank = 0
```

```
      chi2(3) = 1040.50
```

```
Prob > chi2 = 0.0000
```

Questions?

15 minute break