Treatment effects

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

November 2024

Introduction

- These slides give an introductory summary and data application of treatment effects estimation
  - for a binary treatment.
- The methods give a causal estimate **after suitable use of control variables**
  - this may be done to balance an unbalanced RCT
  - or to control for self-selection into treatment
    - ★ then we additionally need to make the nontestable assumption of selection-on-observables-only (or unconfoundedness).
- The methods allow different (heterogeneous) treatment effects for different individuals
  - so compute an average treatment effect
  - using a framework called the potential outcomes model.

- Separately the Stata file `treat.do` implements these methods
  - using dataset `mus224ohiesmallrecode.dta`
- The data are from chapter 24.8 of A. Colin Cameron and Pravin K. Trivedi (2022), *Microeconometrics using Stata*, Second edition, Volume 2
  - chapter 24 details the methods, the data and the application.
- The data source is NBER: The Oregon health insurance experiment - Data. Public use data archive https://www.nber.org/research/data/oregon-health-insurance-experiment-data.
- For analysis see Katherine Baicker, S. L. Taubman, H. L. Allen, M. Bernstein, J. H. Gruber, J. P. Newhouse, E. C. Schneider, B. J. Wright, A. M. Zaslavsky, and A. N. Finkelstein (2013), "The Oregon experiment - Effects of Medicaid on clinical outcomes," *New England Journal of Medicine*, 368, pages 1713-1722.

Outline

Potential Outcomes Model

- Potential outcomes model or Rubin causal model
  - standard framework that is used.
- Consider a binary treatment D
  - $D_i = 1$ for individual i if treated
  - $D_i = 0$ if individual i is not treated (a control).
- There are two potential outcomes for Y_i
  - Y_{1i} if $D_i = 1$ and Y_{0i} if $D_i = 0$.
- Interest lies in estimating the treatment affect $\gamma_i \equiv Y_{1i} - Y_{0i}$
  - note that γ_i can vary across individuals (heterogeneous effect)
  - we cannot estimate γ_i as we only observe one of Y_{1i} and Y_{0i}
    - ★ so restrict attention to more aggregated measures.
- The average treatment effect (ATE) in the population is
  - ATE$= E[\gamma_i] = E[Y_{1i} - Y_{0i}]$.
- The average treatment effect on the treated (ATET) is
  - ATET$= E[\gamma_i | D_i = 1] = E[(Y_{1i} - Y_{0i}) | D_i = 1]$.

Randomized control trial

- A random controlled trial (RCT) is an experiment where randomly assign people to treatment and control.
  - then estimate ATE (and ATET) by the difference in means
  - $\widehat{\text{ATE}} = \bar{y}_{1i} - \bar{y}_{0i}$.
- Note that this can be mechanically computed using OLS
  - $\widehat{\text{ATE}} = \widehat{\gamma}$ from OLS in the model $y_i = \alpha + \gamma d_i + u_i$
  - but it is still the case that γ_i varies across individuals
  - $\widehat{\gamma}$ is the average over the individuals.
- An RCT may not perfectly randomize and may be unbalanced
  - covariates (that determine in part the outcome) may systematically differ between control and treated individuals
  - so use methods (detailed below) to control for the difference in covariates.

Observational data

- RCT's are difficult to run in economics due to high cost and/or ethical reasons.
  - so we rely on observational data where individuals select into treatment
- With observational data individuals may additionally differ on unobservables (model errors) that determine in part self-selection into treatment or nontreatment.
- A stereotypical example is returns to training where self-select into training
  - $y_i = \beta_1 + \gamma d_i + u_i$ where d_i is a binary indicator for training
- People choose to get training, so we expect that those with higher (unobserved) expected benefits to training will select training
  - then $\mathrm{E}[u_i|d_i = 1] > \mathrm{E}[u_i|d_i = 0]$ so $\mathrm{E}[u_i|d_i] \neq 0$ and OLS is inconsistent.

Observational data (continued)

- Again use methods that control for the difference in covariates
  - but now need to assume that these also control for unobservables that determine in part selection into treatment
    - ★ nontestable crucial assumption called selection-on-observables only, or unconfoundedness, or ignorable selection.
- In summary, the same adjustment methods (detailed below) may be used for RCTs and for observational data.
- But in the latter case we need to assume that the adjustments also control for unobservables that determine self-selection into treatment.
- When this assumption is not reasonable to make we need to use other methods (if possible) including instrumental variables, differences in differences, regression discontinuity design and synthetic control.

Unconfoundedness assumption

- Also called selection-on-observables only, or unconfoundedness.
- After controlling for other variables, selection in to treatment can be viewed as if treatment was randomly assigned.
- Define
  - y_0 and y_1 are potential outcomes
  - d is treatment
  - $\mathbf{x}$ are control variables
- The unconfoundedness assumption is that conditional on $\mathbf{x}$, the treatment assignment d and the potential outcomes y_0 and y_1 are independent
  - this rules out, for example that people with high $y_1 - y_0$ are more likely to receive treatment compared to those with low $y_1 - y_0$
  - formally this is denoted $(y_0, y_1 \perp \mathbf{x})|d$.

Method 0: Control function

- A simple control function approach adds controls
  - $\widehat{\text{ATE}} = \widehat{\gamma}$ from OLS of $y_i = \beta_1 + \gamma d_i + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$.
- For consistent estimation of γ we need to assume d_i and u_i uncorrelated once the controls are added.
- The subsequent methods are preferred richer methods.

Method 1: Regression Adjustment

- The regression adjustment estimator runs separate regressions for the treated and untreated
  - and estimates ATE by the difference in predicted means.
- 1. Regress y_i on intercept and $x_{2i}, ..., x_{ki}$ for $d_i = 1$ only observations Compute $\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{1i}$ where $\widehat{y}_{1i} = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_1$ is resulting prediction for observations with $d_i = 0$ and with $d_i = 1$.
- 2. Regress y_i on intercept and $x_{2i}, ..., x_{ki}$ for $d_i = 0$ only observations Compute $\frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{0i}$ where $\widehat{y}_{0i} = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_0$ is resulting prediction for observations with $d_i = 0$ and with $d_i = 1$.

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{1i} - \frac{1}{n}\sum_{i=1}^{n}\widehat{y}_{0i}.$$

Method 2: Inverse-probability weighting (IPW)

- Inverse-probability weighting uses weighted averages of the outcome.
- We use a model such as a logit model to estimate the propensity score, the predicted probability of treatment
  - $\widehat{p}_i = \widehat{\Pr}[d_i = 1|(x_{2i}, ..., x_{ki})]$.
- The higher is $\widehat{p}_i$ the larger the treatment effect is likely to be
  - so if person is treated $(d_i = 1)$ we should downweight their outcome y_i
  - inverse-probability weighting uses $y_i/\widehat{p}_i$ for those with $d_i = 1$
  - and for similar reasons use $1/(1 - \widehat{p}_i)$ if untreated $(d_i = 0)$.

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\frac{d_i y_i}{\widehat{p}_i} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-d_i)y_i}{1-\widehat{p}_i}.$$

- Propensity score overlap requires that $0 < \widehat{p}_i < 1$ so that for each value of $\widehat{p}_i$ there are both treated and untreated observations.

Method 3: Doubly-robust estimator

- The regression adjustment estimator that gives $\widehat{y}_{1i}$ and $\widehat{y}_{0i}$ requires correct specification of the regression model.
- The IPW regression adjustment model that gives $\widehat{p}_i$ requires correct specification of the propensity score model.
- The doubly-robust estimator or augmented IPW estimator
  - combines regression adjustment and inverse-propensity score estimation
  - for consistency requires correct specification of just one of the regression model and the propensity score model.

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i(y_i-\widehat{y}_{1i})}{\widehat{p}_i} - \widehat{y}_{1i}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-d_i)(y_i-\widehat{y}_{0i})}{1-\widehat{p}_i} - \widehat{y}_{0i}\right).$$

Method 4: Matching

- Matching compares each treated individual to a similar (on x's) untreated individual
  - ideally we match on exactly the same $x's$ but with many x's and/or continuous x's this is not possible.
- Nearest neighbor matching compares y_i for each treated individual to the average y_i of the k untreated individuals (e.g. $k = 10$) whose values of $x_2, ..., x_k$ are closest to those for the treated observation
  - several different measures of closeness have been proposed.
- Propensity score matching instead compares outcomes with those with similar probability of treatment
  - where $\widehat{p}_i = \widehat{\Pr}[d_i = 1|(x_{2i}, ..., x_{ki})]$ is a prediction from e.g. a logit model.

Regressor balance check

- In the simplest RCT, regressors are independent of the outcome.
- With observational data this is no longer the case, so we ideally rebalance so that the covariates for treatment and control are similar.
- For a single variable let $\bar{z}_1$ and $\bar{z}_0$ (and s^2_{z1} and s^2_{z0}) denote means (and variances) for treated and untreated individuals.
- Two measures of the difference across treatment groups are
  - standardized difference: $(\bar{z}_1 - \bar{z}_0)/\sqrt{(s^2_{z1} + s^2_{z0})/2}$
  - variance ratio: s^2_{z1}/s^2_{z0}.
- For methods that use weights that are the inverse of the propensity scores we want
  - weighted means to be similar for treated and untreated
  - weighted variances to be similar for treated and untreated
  - weighted standardized difference to be close to 0
  - weighted variance ratio to be close to 1.
- Qualitatively similar balance measures can be constructed for nearest-neighbors matching.

Example: The Oregon Health Insurance Experiment

- Oregon expanded access to Medicaid, its health insurance program for low income people.
- It did so through a lottery.
  - 90,000 people registered for the lottery
  - the lottery occurred in eight waves over six months
  - 35,000 people were selected.
- We compare the out-of-pocket (OOP) expenditures of those who won the lottery with those who entered the lottery and did not win
  - for those who filled out a subsequent mail questionnaire.
- Note that not all lottery winners subsequently entered Medicaid
  - formally we are estimating intention to treat.

Data summary

- Although the state randomly sampled from individuals on the list, the entire household of any selected individual was considered selected and eligible to apply.
  - so selected (treatment) individuals are disproportionately drawn from larger households.
- Additionally for the sample at hand, winning the lottery varied with the time of the lottery and the survey.
- So even though this is an RCT we should control for household size and time of lottery.
- Additionally, to possibly improve estimator efficiency, we control for some individual characteristics
  - smoker, household income, education and employment.

Variables: Outcome, treatment and individual controls

Variable name	Storage type	Display format	Value label	Variable label
oop	float	%9.0g		Out of pocket costs (cost_tot_oop_mod_12m)
lottery	byte	%12.0g	lottery	Selected in the lottery
household_id	float	%9.0g		Scrambled household identifier
dsmoke	byte	%10.0g	smk_lbl	Currently smoke cigs (smk_curr_12m)
hhinc	float	%3.0f		Household income as % of federal poverty line (hhinc_pctfpl_12m)
deduc2	byte	%8.0g		HS diploma or GED (edu_12m_2)
deduc3	byte	%8.0g		Voc or 2yr degree (edu_12m_3)
deduc4	byte	%8.0g		Four year degree (edu_12m_4)
demploy2	byte	%8.0g		Work < 20 hrs/wk (employ_hrs_12m_2)
demploy3	byte	%8.0g		Work 20-29 hrs/wk (employ_hrs_12m_3)
demploy4	byte	%8.0g		Work 30+ hrs/wk (employ_hrs_12m_4)

Summary statistics: Outcome, treatment and individual controls

```
. global xlist dsmoke hhinc deduc2-deduc4 demploy2-demploy4

. summarize oop lottery $xlist
```

Variable	Obs	Mean	Std. dev.	Min	Max
oop	22,679	269.0062	733.0821	0	9400
lottery	22,679	.4972001	.5000032	0	1
dsmoke	22,154	2.262661	.9171565	1	3
hhinc	20,478	76.97273	69.16905	0	461.6898
deduc2	21,986	.4982716	.5000084	0	1
deduc3	21,986	.2204585	.4145653	0	1
deduc4	21,986	.1137997	.317575	0	1
demploy2	22,411	.0912052	.2879071	0	1
demploy3	22,411	.0999509	.2999412	0	1
demploy4	22,411	.2638436	.4407254	0	1

Variables: Lottery controls

Variable name	Storage type	Display format	Value label	Variable label
dhhsize2	byte	%8.0g		2 in hh (dddnumhh_li_2)
dhhsize3	byte	%8.0g		3 in hh (dddnumhh_li_3)
dlotdraw2	byte	%8.0g		draw_lottery==2 (llldraw_lot_2)
dlotdraw3	byte	%8.0g		draw_lottery==3 (llldraw_lot_3)
dlotdraw4	byte	%8.0g		draw_lottery==4 (llldraw_lot_4)
dlotdraw5	byte	%8.0g		draw_lottery==5 (llldraw_lot_5)
dlotdraw6	byte	%8.0g		draw_lottery==6 (llldraw_lot_6)
dlotdraw7	byte	%8.0g		draw_lottery==7 (llldraw_lot_7)
dlotdraw8	byte	%8.0g		draw_lottery==8 (llldraw_lot_8)
dsurvdraw2	byte	%8.0g		draw_survey==2 (ddddraw_sur_2)
dsurvdraw3	byte	%8.0g		draw_survey==3 (ddddraw_sur_3)
dsurvdraw4	byte	%8.0g		draw_survey==4 (ddddraw_sur_4)
dsurvdraw5	byte	%8.0g		draw_survey==5 (ddddraw_sur_5)
dsurvdraw6	byte	%8.0g		draw_survey==6 (ddddraw_sur_6)
dsurvdraw7	byte	%8.0g		draw_survey==7 (ddddraw_sur_7)

Summary statistics: : Lottery controls

```
. global zlist dhhsize2 dhhsize3 dlotdraw* dsurvdraw*

. summarize $zlist   // Household size and lottery and survey draws
```

Variable	Obs	Mean	Std. dev.	Min	Max
dhhsize2	22,679	.2963094	.4566392	0	1
dhhsize3	22,679	.0025133	.0500713	0	1
dlotdraw2	22,679	.1050311	.3066002	0	1
dlotdraw3	22,679	.1028705	.3037964	0	1
dlotdraw4	22,679	.1014595	.3019429	0	1
dlotdraw5	22,679	.0970501	.2960325	0	1
dlotdraw6	22,679	.1974955	.3981181	0	1
dlotdraw7	22,679	.1934389	.3950027	0	1
dlotdraw8	22,679	.0995194	.2993647	0	1
dsurvdraw2	22,679	.1124829	.3159666	0	1
dsurvdraw3	22,679	.1127034	.3162368	0	1
dsurvdraw4	22,679	.1404383	.3474488	0	1
dsurvdraw5	22,679	.1414084	.34845	0	1
dsurvdraw6	22,679	.2043741	.4032524	0	1
dsurvdraw7	22,679	.173244	.3784664	0	1

Results: Difference in means

- Throughout use standard errors that cluster on household.
- Simple differences in means is OLS with no controls.
- The ATE for out-of-pocket spending is -$45
  - this is 17% of the sample mean of $269
  - and is highly statistically significant ($p = 0.000$).

```
. regress oop lottery, vce(cluster household_id) noheader
                     (Std. err. adjusted for 20,148 clusters in household_id)
```

oop	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
lottery	-44.66267	9.921594	-4.50	0.000	-64.1098	-25.21553
_cons	291.2125	7.120921	40.90	0.000	277.2549	305.1701

Results: Control function

- Results change little with controls
  - Diff_clu: no controls
  - zlist: Lottery controls
  - xlist: Individual controls
  - Both: Lottery controls and individual controls.

Variable	Diff_clu	zlist	xlist	Both
lottery	-44.6627	-40.9243	-45.7420	-40.1693
	9.9216	10.1302	10.6484	10.8169
N	22679	22679	19393	19393
r2	0.0009	0.0018	0.0147	0.0154

Legend: b/se

Results: Regression adjustment

- ATE is −$40 (compared to −$45 difference in means)

```
. * Regression-adjusted ATE using $zlist and $xlist
. teffects ra ($y $xlist $zlist ) (lottery), nolog vce(cluster household_id)

Treatment-effects estimation                    Number of obs     =     19,393
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
                        (Std. err. adjusted for 17,348 clusters in household_id)
```

oop	Coefficient	Robust std. err.	z	P>|z|	[95% conf. interval]	
ATE						
lottery (Selected vs Not selected)	-40.44588	10.94579	-3.70	0.000	-61.89924	-18.99252
POmean						
lottery Not selected	292.1394	7.693716	37.97	0.000	277.06	307.2188

Results: Inverse-probability weighting

- Separate analysis shows that the `$xlist` are well balanced but the `$zlist` are not
  - so use just the `$zlist` variables as regressors for the propensity score.
- ATE is $-\$40$ (compared to $-\$45$ difference in means)

```
. * IPW ATE using $zlist
. teffects ipw ($y) (lottery $zlist), nolog vce(cluster household_id)

Treatment-effects estimation                    Number of obs     =     22,679
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
                        (Std. err. adjusted for 20,148 clusters in household_id)
------------------------------------------------------------------------------
             |               Robust
         oop | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
     lottery |
  (Selected  |
          vs |
Not selected) |  -39.56951   10.12338    -3.91   0.000    -59.41098   -19.72804
------------------------------------------------------------------------------
```

Results: Doubly-robust augmented IPW

- ATE is −$39 (compared to −$45 difference in means)

```
. * Doubly-robust augmented IPW
. teffects aipw ($y $xlist) (lottery $zlist), aequations nolog ///
>     vce(cluster household_id)

Treatment-effects estimation                    Number of obs     =     19,393
Estimator      : augmented IPW
Outcome model  : linear by ML
Treatment model: logit
                        (Std. err. adjusted for 17,348 clusters in household_id)
```

	oop	Robust Coefficient	std. err.	z	P>|z|	[95% conf.	interval]
ATE							
	lottery (Selected vs Not selected)	-38.66964	10.79616	-3.58	0.000	-59.82973	-17.50956

Results: Propensity score matching

- ATE is $-\$49$ (compared to $-\$45$ difference in means)

```
. * Propensity score matching usign $zlist
. teffects psmatch ($y) (lottery $zlist)

Treatment-effects estimation                    Number of obs      =     22,679
Estimator      : propensity-score matching      Matches: requested =          1
Outcome model  : matching                                      min =          1
Treatment model: logit                                         max =       1018
```

oop	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
lottery (Selected vs Not selected)	-48.935	14.08062	-3.48	0.001	-76.53251	-21.33749

Results: Nearest-neighbor matching

- ATE is −$49 (compared to −$45 difference in means)

```
. * Nearest neighbor matching
. * Exact match on household size either 1 or > 1
. * Match on mahalanobis distance for the remaining variables
. generate dhhbig = dhhsize2 + dhhsize3

. teffects nnmatch ($y $wave) (lottery), ematch(dhhbig) metric(mahalanobis)

Treatment-effects estimation                    Number of obs      =     22,679
Estimator      : nearest-neighbor matching      Matches: requested =          1
Outcome model  : matching                                   min =          5
Distance metric: Mahalanobis                                max =       1018
```

oop	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
lottery (Selected vs Not selected)	-48.68898	13.83561	-3.52	0.000	-75.80629	-21.57168

Check: Regressor balance

- The number of weighted and unweighted observations are similar.

```
. qui teffects ipw ($y) (lottery $zlist), nolog vce(cluster household_id)

. tebalance summarize

Covariate balance summary

                          Raw      Weighted
-------------------------------------------
Number of obs =        22,679      22,679.0
Treated obs   =        11,276      11,151.4
Control obs   =        11,403      11,527.6
-------------------------------------------
```

- More statistics are given on the next slide
- Aside: if we add `$xlist` it shows that the raw data are already well balanced, so no need to include them in the propensity score model.

Regressor balance (continued)

- The weighted data are well balanced
  - whereas the unweighted data were not.

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
dhhsize2	.1939117	.0060577	1.19006	1.005543
dhhsize3	.0797051	.0021402	8.56137	1.04358
dlotdraw2	.0331758	.0015502	1.089294	1.004479
dlotdraw3	.0365961	.0046157	1.100489	1.013669
dlotdraw4	-.0257401	-.0069954	.9342648	.9817981
dlotdraw5	.0003948	-.0069064	1.001076	.9816374
dlotdraw6	-.0083996	-.0028377	.9873158	.9960122
dlotdraw7	-.0192988	.0054252	.970481	1.007623
dlotdraw8	-.0254451	.0079989	.9341439	1.020275
dsurvdraw2	.2260162	-.0055742	1.772475	.9868448
dsurvdraw3	.2397501	-.0094081	1.838261	.9780227
dsurvdraw4	-.0404115	.0106734	.9196795	1.022566
dsurvdraw5	-.0458372	.0091592	.9098577	1.019404
dsurvdraw6	-.1878871	.0052308	.7566851	1.007905
dsurvdraw7	-.290524	-.0054474	.5950349	.9903239

Further details

- More complete analysis of the preceding methods is given in chapter 25 of A. Colin Cameron and Pravin K. Trivedi (2022), *Microeconometrics using Stata*: Volume 2, Second Edition.
  - This includes checks for covariate balance and propensity score overlap.

References for treatment effects

- These econometrics books are given in approximate order of increasing difficulty.
- A. Colin Cameron (2022), Analysis of Economics Data: An Introduction to Econometrics, chapter 17.5.
- Cunningham, Scott (2021), Causal Inference: The MixTape, Yale University Press, chapters 5-6.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata: Volume 2, Second Edition, Stata Press, chapter 24.
- A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press, chapters 25.1-25.4.
- Wooldridge, Jeffrey M. (2010), Econometric Analysis of Cross Section and Panel Data, Second Edition, MIT Press, especially chapters 21.1-21.3.
- Guido W. Imbens and Donald B. Rubin (2015), "Causal Inference in Statistics, Social, and Biomedical Sciences," Cambridge University Press, chapters 1-3, 12-24.

References by non-economists

- Stephen L. Morgan and Christopher Winship (2015), Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second edition, Cambridge University Press, chapters 1-7.
- Richard J. Murnane and John B. Willett (2010), Methods Matter: Improving Causal Inference in Educational and Social Science Research, Oxford University Press, chapters 4, 12.
- Andrew Gelman, Jennifer Hill and Aki Vehtari (2022), Regression and Other Stories, Cambridge University Press, especially chapters 19–20.