Treatment Effects

Qingqing Zong
School of Public Economics and Administration
SHUFE

主讲教师：宗庆庆

- 上海财经大学数量经济学博士、北京大学光华管理学院博士后
- 美国宾夕法尼亚大学访问学者
- 上海财大公共经济与管理学院副院长，副教授
- 北京大学经济政策研究所兼职研究员
- 研究方向：劳动经济、产业经济、城市经济
- 在*Research Policy*，《经济研究》，《金融研究》等中英文权威学术期刊上发表论文多篇
- 研究成果获得“全国财政理论研究成果奖”等奖项
- 主持国家自然科学基金项目一项，参与国家和省部级课题多项，主持或参与财政部，卫计委等国家部委委托的横向课题多项

Course Information

- E-mail: zong.qingqing@mail.shufe.edu.cn
- Office： Room 518
- Office Hour： Tuesday 2-4pm or by appointment

Overview

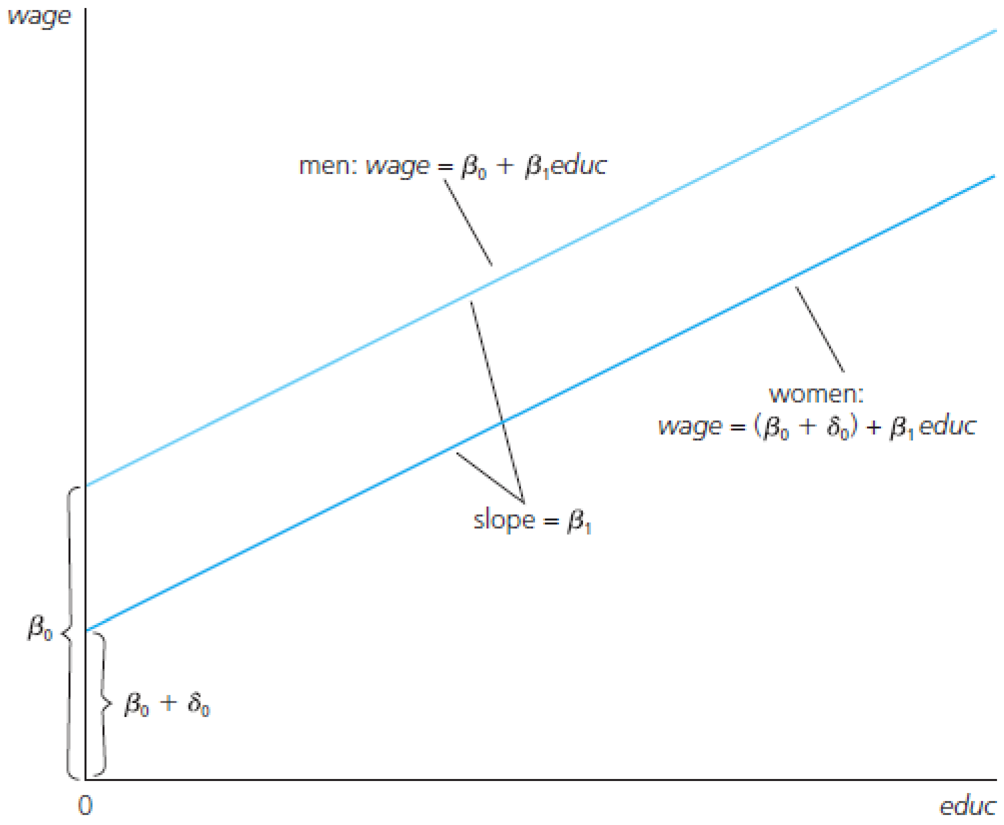
- Preliminaries
  - Dummy Variables
- Introduction to Treatment Model
  - Basic Setup
  - Counterfactual Framework
  - Randomization

Preliminaries: Dummy Variables

- In econometrics, qualitative information is usually captured by defining a zero-one variable. It's called a dummy variable or a binary variable.
- For example, $female_i = \begin{cases} 1 & \text{if individual } i \text{ is female} \\ 0 & \text{if individual } i \text{ is male} \end{cases}$.
- In regression, parameters of dummy variables have natural interpretations: differences between groups divided by dummy variables.
- For more examples, dummy variables can be races, industries, regions, etc.
- Thus, dummy variables are commonly used in policy analysis or program evaluations when the policy or program takes place with only one level, which is exactly the case we'll discuss.

Dummy Variables

- This is an example in Wooldridge's textbook when the gender difference in wage in constant as: $wage = \beta_0 + \delta_0 female + \beta_1 educ + u$.



Dummy Variables

- If there're male and female both married or single in study of wage, we can use three dummy variables in the regression.
- $\ln(wage) = \alpha_0 + \alpha_1 marrmale + \alpha_2 marrfem + \alpha_3 singfem + \cdots$
- Or, we can use interactions of dummy variables.
- $\ln(wage) = \beta_0 + \beta_1 female + \beta_2 marr + \beta_3 marr \cdot female + \cdots$
- In both examples, we regard the single male as a baseline and allow the "married effect" to differ between male and female.
- α_1,α_2,α_3 capture the differences between other groups and the baseline.
- Note that we should introduce m-1 dummy variables to avoid dummy variable trap when the qualitative variable in the model contains m categories.

Dummy Variables

- When the dependent variable is binary, we have a binary choice model.

- Linear Probability Model: $E(Y|X) = X\beta$

- Probit Model: $E(Y|X) = \Phi(X\beta)$ where $\Phi(\cdot)$ is the CDF of Standard Normal Distribution.

- Logit Model: $E(Y|X) = e^{X\beta}/(1 + e^{X\beta})$

Setup of Treatment Model

- Rubin (1974), Holland (1986), Pearl (2000), Rosenbaum (2002):
  - One has two potential outcomes, Y_1 and Y_0.
  - Y_1 describes what would happen if one is treated.
  - Y_0 describes what would happen if one is not treated.
  - A binary variable D represents for the treatment.
  - So only one of Y_d is observed depending on D:
  - $Y = D \cdot Y_1 + (1 - D) \cdot Y_0$

- D can be many interested variables, like medicine, education, job-training.

- We're interested in the causal effect of D on Y.
  - In other words, the difference between Y_1 and Y_0, so-called treatment effects.

Three No-Effect Concepts

- Exact Same: $Y_1 = Y_0$

- Exchangeability:
  $$P(Y_0 \le t_0, Y_1 \le t_1) = P(Y_1 \le t_0, Y_0 \le t_1) \text{ for } \forall t_0, t_1$$

- Zero mean(or median): $E(Y_1) = E(Y_0)$ or $Med(Y_1) = Med(Y_0)$

- Their relation:
  Exact Same ⇒ Exchangeability ⇒ Zero mean/median

Definition of Treatment Effects

- These treatment effects are defined by statistics:
  - Mean treatment effects: $E(Y_1 - Y_0)$ or $E(Y_1 - Y_0|X)$
  - Quantile treatment effects: $Q_{Y_1}(\tau|X) - Q_{Y_0}(\tau|X)$
  - Distributional treatment effects: $F_{Y_1|X}(y) - F_{Y_0|X}(y)$
- These mean treatment effects are defined by populations:
  - Average treatment effects (ATE): $E(Y_1 - Y_0)$
  - Average treatment effects on treated (ATT): $E(Y_1 - Y_0|D = 1)$
  - Average treatment effects on untreated (ATUT): $E(Y_1 - Y_0|D = 0)$
- ATE is useful when treatment has broad applicability.
- ATT is mostly for those "focused" program. We don't care the college return for one who doesn't finish senior school.

Counterfactual Framework

- For example, we want to know the return of education.
- For college graduates who have D=1, we observe their wage Y_1.
- And for those not entering college, we observe their wage Y_0.
- So, we only observe one of Y_1 and Y_0 for an individual.

$$E(Y_1 - Y_0|D = 1) = E(Y|D = 1) - E(Y_0|D = 1)$$
$$E(Y_1 - Y_0|D = 0) = E(Y_1|D = 0) - E(Y|D = 0)$$

- The key is how to estimate the **counterfactual** results.
- For a treated unit, Y(0) is called the counterfactual. For a untreated unit, Y(1) is the counterfactual.

Random Assignment/Experiment

- Random assignment guarantees the treatment is independent of everything (including all observed X and unobserved ε which may affect Y_d). In technical notation:

$$[Y_1, Y_0] \perp D$$

- Under such assumption, all the characteristics of the individuals are equally distributed between treated and untreated groups (i.e., the proportions are the same).

$$E(Y_0|D=1)=E(Y_0|D=0)=E(Y_0)$$
$$E(Y_1|D=1)=E(Y_1|D=0)=E(Y_1)$$

- Thus

$$E(Y|D=1)-E(Y|D=0)=E(Y_1|D=1)-E(Y_0|D=0)$$
$$=E(Y_1-Y_0)=ATE=ATT=ATUT$$

Estimation

- Under randomization assumption: treatment effects are estimated as,

$$\hat{\beta} = \frac{\sum D_i Y_i}{\sum D_i} - \frac{\sum(1 - D_i)Y_i}{\sum(1 - D_i)}$$

- It's same to the difference between sample averages of two groups.

$$\hat{\beta} = \bar{Y}_{treated} - \bar{Y}_{control}$$

- It's same to run OLS for the following equation:

$$Y_i = \alpha + \beta D_i + U_i$$

Relation between Randomization and Regression

- Consider the following linear regression:

$$Y_i = \alpha + \beta D_i + U_i$$

- The strictly exogenous assumption is

$$E(U_i|D_i = 1) = E(U_i|D_i = 0) = 0$$

- Thus

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$
$$= \big(\alpha + \beta + E(U_i|D_i = 1)\big) - \big(\alpha + E(U_i|D_i = 0)\big) = \beta$$

Three Types of Experiments

- Controlled Experiment
  - ➢ Impossible in Social Science

- Randomized Experiment
  - ➢ Rarely exists in Social Science
  - ➢ Hard to be accomplished
  - ➢ Bears very expensive cost or may be blamed ethically

- Natural Experiment
  - ➢ Can be viewed as quasi-randomized experiment

Internal and External Validity

- Internal Validity: our conclusion truly represents the sample.
- External Validity: our conclusion can be applied to other populations for prediction and so on.

- Threats to Internal Validity:

➢ Psychological effect: If one knows being treated, he may feel better. That's why control group will take placebo in medicine experiments. But placebo isn't available for many other treatments, like job-training.

➢ Substitution effect: If one is told to sleep less or do less sport, he may make up for it in other ways.

- Threats to External Validity:

➢ Non-participation effect: The sample is different from the whole population. e.g., if data is collected on Internet, those not using Internet are excluded.

Example 1

- Hongbin Cai, Yuyu Chen, Hanming Fang, Li-An Zhou, The Effect of Microinsurance on Economic Activities: Evidence from a Randomized Field Experiment, *Review of Economics and Statistics*,2015,97(2).

Variables	Whole Sample Mean	Whole Sample SD	By Group: Control Mean	Control SD	LIG Mean	LIG SD	HIG Mean	HIG SD	p-value
Pre-Experiment Variables:									
No. of Sows in Dec. 2006	16.3	21.4	17.9	26.5	13.2	14.2	16.9	21.5	0.17
No. of Sows in Sept. 2007	29.1	31.8	28.8	43.1	28.1	20.5	29.8	29.4	0.90
No. of Sows in Dec. 2007	31.2	34.5	32.5	46.5	26.4	23.7	32.5	31.6	0.43
No. of Pigs in Dec. 2006	356.2	228.4	363.9	248.3	338.3	228.1	361.3	218.4	0.61
Village Population	1029.1	677.8	1048.7	654.2	1017.9	672.0	1025.0	694.5	0.93
No. of Villagers as Migrant Workers	196.0	116.4	188.8	127.1	193.7	103.1	200.8	117.5	0.64
Ave. Villager Age	33.2	2.1	33.1	2.1	33.4	2.3	33.2	1.9	0.45
Ave. Villager Education (Years)	5.95	0.75	6.00	0.78	6.00	0.77	5.90	0.73	0.40
Fraction Male in Village	0.54	0.03	0.54	0.03	0.55	0.03	0.54	0.02	0.31
Land per Household (Mu)	4.31	1.97	4.09	1.91	4.28	1.95	4.43	2.00	0.31
Log House Value	9.83	.63	9.87	.61	9.75	.63	9.84	.63	0.30
No. of Surnames in the Village	5.36	2.68	5.41	3.08	5.19	2.47	5.42	2.57	0.72
No. of Villagers in New Medical Coop. Scheme	551.5	300.7	560.3	322.3	532.1	299.2	556.9	291.2	0.72
No. of Households Receiving Gov. Subsidy	182.2	92.2	183.9	90.4	178.8	99.6	183.0	89.5	0.90

Example 2

- 宗庆庆、张熠、陈玉宇，2020：《老年健康与照料需求：理论和来自随机实验的证据》，《经济研究》，第2期。

表　平衡性检验

变量	实验组	控制组	差异	p-value
年龄	67.717	68.897	1.18	0.06*
是否男性	0.57	0.62	0.05	0.07*
是否已婚	0.847	0.832	-0.01	0.47
是否城镇户口	0.495	0.605	0.11	0.03**
受教育年限	6.576	6.51	-0.07	0.75
子女数量	2.316	2.525	0.21	0.16
与子女同住	0.308	0.269	-0.04	0.11
是否退休	0.952	0.975	0.02	0.14
月收入（元）	1878.554	1883.363	4.81	0.97

Violations to Randomization

- In economics, our data is rarely from experiments.

- Consider the case:

➢ Half are Females with $Y_1 = 70 \ and \ Y_0 = 75$. And 80% take the treatment.

➢ Others are Males with $Y_1 = 50 \ and \ Y_0 = 55$. And 20% take the treatment.

➢ So $E(Y|D = 1) - E(Y|D = 0) = 66 - 59 = 7$. Positive.

- Because of unbalance of observables (here is gender), we have overt bias. To solve it, we should control for covariates.

Violations to Randomization

- Consider the case of college:
  - Half are of high ability with $Y_1 = 70\ and\ Y_0 = 50$.
  - Others are of low ability with $Y_1 = 40\ and\ Y_0 = 30$.
  - If high ability goes to college while low ability not, then $E(Y|D = 1) - E(Y|D = 0) = 70 - 30 = 40$.
  - Much higher than those of high ability (20) or low ability (10).

- Because of unbalance of unobservables (here is ability), we have hidden bias. To solve it, we may need some instruments.

Selection on Observables and Unobservables

- Selection on observables (like gender):

➢ (1) D is independent of Y_d, conditional on X.

$$f(Y_d|D,X) = f(Y_d|X)$$

➢ (2) D is mean-independent of Y_d, conditional on X.

$$E(Y_d|D,X) = E(Y_d|X)$$

➢ Only overt bias, no hidden bias.

- Selection on unobservables (like ability):

$$f(Y_d|D,X) \neq f(Y_d|X) \text{ and } f(Y_d|D,X,\varepsilon) = f(Y_d|X,\varepsilon)$$

$$\text{or } E(Y_d|D,X) \neq E(Y_d|X) \text{ and } E(Y_d|D,X,\varepsilon) = E(Y_d|X,\varepsilon)$$

➢ Have hidden bias.