Foundations of Econometrics - Part II
Final Exam Solutions

Exam_2021.pdf

December 14, 2021

Question 1: Randomized Controlled Trials

(20 points)

You are asked to design a randomized experiment for Airbnb to evaluate the effect of taking professional photos on sales (booking rates). The idea is to help Airbnb decide whether it is profitable to offer hosts the free service of having pictures of the accommodation taken by professional photographers.

(a) (10 points)

Describe briefly how you would implement the experiment. Discuss on which level you think the randomization should be (country, city, accommodation)? Are you worried about take-up and why?

Answer: I would randomly offer free professional photography to accommodations across a few cities. Randomly offering the service helps create independence and allows us to treat it like a RCT. I would randomize at the accommodation level because it increases sample size (more accommodations than cities or countries, easily) and avoids problems of covariance we may get from randomizing at the city level (if we give Barcelona hosts free photos but not London, we have to determine if the treatment effect is from the free photos alone or other systemic differences between the two cities).
I would be worried about take up because host characteristics like motivation and engagement as a host may create an issue. Hosts that take the treatment may be those that are more motivated and care more about their listing, thus they may be doing other things (taking better care of flat, marketing etc) to boost sales. Hosts who deny the free photographs may care less about renting it, and their low motivation means less engagement to get sales otherwise. Also if they have a lower quality flat, they may decline professional photos because it actually hurts their listing more (updated HD photos of a bad flat could be worse than older lower quality of the same flat, where they can obscure shortcomings).

(b) (5 points)

Give the formulas for the ATE and ATT, how would you estimate them?

Answer:
The ATE is defined as: E[Yi(1) - Yi(0)] This is the causal relationship of the entire population; difference in expected value of the outcome with and without exposure to treatment.
The ATT is defined as: E[Yi(0) - Yi(1) — Di = 1] This is the causal relationship for the treated only; difference in expected value of the outcome with and without exposure to treatment, for the treated.
The problem is the counterfactual; in real life, you cannot observe what the treated groups' outcome would have been if they were not exposed to treatment.
If we had a truly random assignment/RCT, the ATE=ATT, so we can just calculate ATT and know it gives use the true ATE. Otherwise, we need to use some other method, such as matching, to ensure conditional independence and thus calculate the ATE.

(c) (5 points)

How could you check if treatment and control groups are truly random?

Answer: We need to ensure that there is no systemic differences between the two. We could check the statistical significance in the difference of the covariants, or also run reggressions on the covariants to determine if there is a significant difference between the two groups, and thus a violation of randomness.

Question 2: Matching

(20 points)

You want to study the effect of soft drink consumption on obesity. You have cross-sectional data on high school pupils' soft drink consumption (an indicator), weight and gender.

The table below shows the joint distribution of gender and soft drink consumption (all entries sum up to one.)

$X_i / \boldsymbol{D_i}$	**0**	**1**
F	0.1	0.35
M	0.15	0.4

And you also have mean weight (in kg) for girls and boys that do and do not consume soft drinks.

$X_i / \boldsymbol{D_i}$	**0**	**1**
F	50	53
M	60	65

(a) (12 points)

Calculate the ATE and ATT using direct matching. Simply insert the relevant numbers into the formula, no need to solve for the final number.

Answer:

$$\hat{\alpha}_{\text{ATE}} = \sum_{j=1}^{J} (\bar{Y}_{j1} - \bar{Y}_{j0}) \frac{N_j}{N}$$

Plugging in: (53-50)*(0.45) + (65-60)*(0.55)
ATE

$$\hat{\alpha}_{\text{ATT}} = \sum_{j=1}^{J} (\bar{Y}_{j1} - \bar{Y}_{j0}) \frac{N_j 1}{N_1}$$

Plugging in: (53-50)*(0.35/0.75) + (65-60)*(0.40/0.75)

(b) (8 points)

Which assumption on potential outcomes and treatment is needed for matching? What about assumptions on $P(D = 1|X)$?

Answer:
For matching, we need two assumptions:
Conditional independence:

$$(Y_i(0), Y_i(1)) \perp D_i | X_i$$

This means that the match is "as good as random", that is, once we control for our observed covariates, we are isolating the treatment effect.
Common support:

$$0 < P(D = 1|X = x) < 1$$

The probability of being treated and not treated is non-zero in both treatment and control groups.

Question 3: Instrumental Variables

(20 points)

You would like to study the effect of treatment D on outcome Y but are worried that there was some self-selection into treatment. You are considering using an instrument Z to address this issue.

(a) (10 points)

Which conditions must Z satisfy?

Answer: We need two conditions: relevance and independence (exogeneity). If we assume heterogeneous effects, we also need monotonicity and exclusion.
Relevance (first stage):

$$\text{Cov}(Z, D) \neq 0$$

Z must have some causal effect on D.
Independence (Exogeneity):

$$\text{Cov}(Z, \varepsilon) = 0$$

There must be no correlation between Z and the error terms, meaning Z must be unrelated to any unobservables that are affecting Y.
For heterogeneous effects, also need:
Exclusion Restriction:

$$\text{Cov}(Z_i, Y_i(d)) = 0 \quad \text{for all } d$$

Means that Z only impacts Y through D.
Monotonicity:

$$D_i(1) \geq D_i(0) \quad \text{for all } i$$

There are no defiers. They would have a value of $D_i(1) = 0$ and $D_i(0) = 1$, which violates this assumption.

(b) (10 points)

Assume that treatment effects are heterogeneous, i.e. not all individuals respond to the treatment in the same way. What type of treatment effect can you compute with your instrument? Are there individuals for whom you cannot compute an effect? State the formula of the treatment effect.

Answer: With heterogenous effects, we are only able to calculate the local average treatment effect (LATE). We can only calculate this for compliers.

$$\text{LATE} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}$$

Question 4: Regression Discontinuity

(10 points)

State the two conditions for Y and D around cut-off z_0 that are necessary for RD. State the regression equation that you would estimate for a sharp design, assuming equal slopes to the right and left of the cut-off.

Answer: For regression discontinuity, we must have both continuity of the potential outcome around the cut off and a discontinuity in the treatment assignment.
Continuity of potential outcome:

$$\lim_{z \to z_0^-} E[Y_i|Z_i = z] = \lim_{z \to z_0^+} E[Y_i|Z_i = z]$$

This states that the only jump is from the treatment effect. Those that are just above and below the cutoff should be similar.
Discontinuous in treatment assignment: Sharp of fuzzy. Sharp means it jumps directly from 0 to 1, fuzzy is more gradual.

$$\lim_{z \to z_0^-} E[D_i|Z_i = z] \neq \lim_{z \to z_0^+} E[D_i|Z_i = z]$$

The probability of treating must jump at the cutoff. Sharp would jump directly from 0 to 1; fuzzy would mean that it still increases at the cutoff, but it may not be all the way from 0 to 1.
The regression equation for equal slopes is:

$$Y_i = \alpha + \tau 1(Z_i \geq z_0) + \gamma(Z_i - z_0) + \varepsilon_i$$

For different slopes would be;

$$Y_i = \alpha + \tau 1(Z_i \geq z_0) + \gamma(Z_i - z_0) + \beta(Z_i - z_0)\mathbb{1}(Z_i \geq z_0) + \varepsilon_i$$

Question 5: Difference-in-Difference

(10 points)

You would like to assess the effect of a treatment on an outcome of interest. There is data for locations A and B over time periods $T = \{0, 1\}$, only location A has been treated in period 1. Graphically illustrate how Diff-in-Diff works on the $Y - T$ plane (replicate the graph from class). In particular, graphically show the effect that is being estimated, making use of the parallel-trends assumption.

Answer: