Policy Evaluation
Supervision 1

Samuel Lee

Question 1

(a)

The population expectation of the observed average difference between treated individuals and non-treated individuals is

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\text{Expected treatment effect on the treated}} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection bias}}$$

where Y_{1i} is the (possibly hypothetical) outcome variable after treatment for individual i, regardless of whether they are treated or not, and Y_{0i} is the (possibly hypothetical) outcome variable without treatment for individual i, regardless of whether they are treated or not. We are able to equate $E[Y_i|D_i = 1] = E[Y_{1i}|D_i = 1]$ because $\Pr[Y_i = Y_{0i}|D_i = 1] = 0$, and likewise when we equate $E[Y_i|D_i = 0] = E[Y_{0i}|D_i = 0]$.

If selection into the treatment is not random, the terms referred to as the "selection bias" can be non-zero. We can see why those terms are referred to as a selection bias with a concrete example. Y_{0i} is the outcome variable had individual i not received the treatment; the outcome variable could be all-cause mortality (ignoring for now how that would be measured) and the treatment could be any positive amount of running. We might observe that people who run exhibit lower mortality than people who do not run, but it is plausible that the sorts of people who would make the effort to run any non-zero amount are also the sorts of people who would have other healthy habits that reduce their mortality even if they don't run. That is, $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] > 0$, which is how we get articles in the British Journal of Sports Medicine saying that "any amount of running" lowers the risk of premature death, or even that "higher doses of running may not necessarily be associated with greater mortality benefits".

(b)

We can decompose Y_i into

$$Y_i = \underbrace{E[Y_{0i}]}_{\alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\rho_i} D_i + \underbrace{Y_{0i} - E[Y_{0i}]}_{\eta_i}$$

and it is easy to see that this is analogous to a linear regression function in D_i; ρ_i represents the unbiased treatment effect given the zero conditional mean assumption $E[\eta_i|D_i] = 0$.

If we run a regression of Y_i on D_i, the estimated coefficient on D_i is an estimator for the observed treatment effect, but we might want to go further if we think the selection bias is substantial. If we could be persuaded to make an additional assumption that D_i is independent of Y_i conditional on a vector of observed covariates X_i, that is, $Y_i \perp\!\!\!\perp D_i | X_i$, then we have

$$E[Y_i|X_i, D_i = 1] - E[Y_i|X_i, D_i = 0] = E[Y_{1i}|X_i, D_i = 1] - E[Y_{0i}|X_i, D_i = 0]$$
$$= E[Y_{1i}|X_i, D_i = 1] - E[Y_{0i}|X_i, D_i = 1] = E[Y_{1i} - Y_{0i}|X_i, D_i = 1]$$

where we used $Y_i \perp\!\!\!\perp D_i | X_i \implies E[Y_{0i}|X_i, D_i = 0] = E[Y_{0i}|X_i, D_i = 1]$. The above expression is the expected treatment effect on the treated if we iterate expectations over X_i (keeping it conditioned on $D_i = 1$), which means we could recover the unbiased treatment effect by conditioning on X_i. The population regression function is the best linear approximation of $E[Y_i|X_i, D_i]$, and we could estimate it by running a regression of Y_i on D_i and X_i. The coefficient on D_i would now be the (hopefully) unbiased estimator for the treatment effect on the treated, conditional on X_i.

(c)

1. We have the classic example of earnings and education: it may be that people with higher earnings capabilities select into more years of education, making the observed average difference an overstatement of the returns to education. (Duflo (2001) examines arguably exogenous increases in the supply of education in Indonesia to recover less biased estimates of the returns to education.)

2. Looking at periods where Trump tweets about Muslims, we might conclude that tweets about Muslims lead to more anti-Muslim hate crimes, but it could be that Trump chooses to tweet about Muslims in periods when anti-Muslim sentiments are relatively higher. (Müller and Schwarz (2019) find that the association remains when using Trump's golf days as an instrument variable.)

3. Observed differences in the health outcomes of vaccinated and non-vaccinated children may be biased downwards given that anti-vaccine parents tend to be rich and white (McNutt et al. 2016; Tomeny, Vargo, and El-Toukhy 2017; Yang et al. 2016).

Question 2

(a)

Assignment to the treatment group is guaranteed to be independent of any unobserved covariates, so the estimated intent-to-treat effect will be unbiased. However, attendance for the treatment group is not guaranteed to be independent of unobserved covariates that also affect employment, so comparing the employment rates of participants and non-participants could yield biased estimates of the average treatment effect. What we do get is consistent estimates of the local average treatment effect, or the treatment effect on the treated. How substantial the difference is depends on what story we have for why some people don't attend. We see that those who did not attend have a lower employment rate than the control group; one possible reason is that these

include people who have difficulty making childcare arrangements. They may not have been able to attend the program because their children needed taking care of, and they may have a lower employment rate as they require jobs with more flexible hours.

(b)

Among the treatment group, $80\% \times \frac{1500}{2000} + 40\% \times \frac{500}{2000} = 70\%$ were employed after one year. Therefore the estimated intent-to-treat effect is 20 percentage points (where we assume the treatment effect is multiplicative rather than additive, or alternatively, additive in the employment rate). This estimate is unbiased given the assignment mechanism, and can reflect the efficacy of implementing the program while taking into account non-compliance. The limitation is that this isn't an estimate of the efficacy of the training sessions, interviews, and job search plans. That would be nice to know, but is arguably less important than the intent-to-treat estimate; a government deciding whether or not to implement a program should want an estimate of its benefits that accounts for attrition. It (in theory) doesn't set new taxes without first thinking about substitution effects, and it probably doesn't want to implement a program that has a 99% non-compliance rate even if the treatment effect is substantial for the 1% who comply. This would be different if it wanted to, say, evaluate the efficacy of a new drug.

(c)

We can estimate the treatment effect on the treated by scaling the estimated intent-to-treat effect by the inverse proportion of compliers: $\hat{\beta}_{TOT} = \hat{\beta}_{ITT} \div \hat{\Pr}(D_i = 1 | Z_i = 1) = 20\% \div 0.75 = 26\frac{2}{3}\%$. Despite what was said before, this figure has its uses. If the treatment effect on the treated is large, but compliance is low, then there is a good chance that the marginal returns to enforcing or otherwise encouraging compliance is high, and a government might like to know if doing that is a good idea. The only caveat is that this assumes the treatment effect on the non-compliers will be of similar magnitude; there is no guarantee that this is the case. Returning to the example of people with childcare needs, the effect of forcing these people to go for training sessions and interviews is likely smaller since their reasons for unemployment are probably less related to skills compared to the others in the treatment group. It might be better to subsidise childcare for this subset of people.

(d)

We could compare summary statistics on compliers and non-compliers to have an idea of how substantial selection into non-compliance is. Again, if we are willing to assume that non-compliance is more or less independent conditional on these characteristics, then generalising the treatment effect on the treated to the whole sample might be more forgivable. But in any case, I would argue that this is not something we should lose sleep over if we can at least estimate the intent-to-treat effect.

Question 3

Answering each suggested point briefly,

1. *What is the main conclusion in the existing literature on the effect of class size?*

   The extent of my literature review is the one paper by Krueger (1999) mentioned in the lectures. With the usual caveats of external validity, covariate balance between the treatment and control, and selection into the voluntary study, he found that smaller class sizes led to a 5% improvement in percentile scores among the sample studied.

2. *Can we directly apply their conclusion when studying the size of supervision groups? Discuss the external validity issue.*

   It is probably dangerous to import the above conclusions wholesale into the Cambridge context. Student characteristics are likely very different between the two contexts. For one, the STAR kids were in kindergarten. We can come up with many other stories for why the results wouldn't carry over (differences in socio-economic diversity, aptitude, independence, etc.)

3. *If we want to design our own empirical study, what is the fundamental Trade-off? When should the effect expect to be positive? When should the effect expect to be negative?*

   Money is (I think) the fundamental trade-off if we're talking about doing an RCT versus an observational study, and it might be unethical or at least very unpopular to randomise the treatment in cases where the treatment (or withholding of the treatment) might do harm.

4. *How might the effect of supervision size be tested? what strategy should we use? RCT? OLS?*

   - *RCT: Who should be the control group and who should be the treatment group? group?*
   - *OLS: what is the specification of OLS regression? What data variation should we use?*

   We could randomly assign the treatment group, ideally at the individual level, but it would probably be more feasible to assign the treatment at the college level. One sacrifice is that we'd probably get more imprecise estimates with clustering at the college level. OLS is more of an estimation method than an empirical strategy or research design; if the question is whether we want to work with observational data instead, we might first run a naïve OLS regression on Tripos results and supervision group size, to at least see what the observed association is. From that (very low) baseline there are at least two things to deal with in getting closer to a causal interpretation:

   (a) Selection into group sizes (perhaps weaker students ask for, or are put into, smaller supervision groups)

   (b) Specification of regressor: group sizes likely vary by supervisors even within the same subject, and can vary even with the same supervisor if there are mid-term swaps and whatnot

The selection problem is probably more fundamental, and a researcher would probably deal with it using something more exotic like a regression discontinuity design or instrumental variables: perhaps colleges with odd numbers of student intake tend to have smaller or larger group sizes, depending on whether they round up or down from the typical group size. (Either examples can be estimated using OLS, but perhaps the question is more targeted at the typical linear regression model.)

5. *Could you describe a situation when we need to combine the RCT approach and the OLS approach? What is the advantage of this mixed approach compared with pure RCT? what is the advantage of this mixed approach compared with pure OLS?*

   Again, OLS is an estimation method while randomised controlled trials are more of a research design, and it's not entirely clear what a "pure RCT" is (it's a bit easier to tell what a "pure OLS" is supposed to refer to). Assuming that a "pure RCT" means comparing simple averages observed from a randomised controlled trial, then the advantage of a regression approach to estimation is that we can condition the estimated outcome on observed covariates; useful if we are willing to make the conditional independence assumption. The advantage of running an OLS with data from an RCT is that we are at least guaranteed that the first-stage regression yields unbiased estimates.