REFLECTIONS ON PROBLEM SET 1

Question 1

This simple example illustrates a lot of important concepts:

- Heterogeneous treatment effects: there are four "types" of students, each type differing in the extent to which they benefit from treatment (+10, 40, 20, 15).
- Variation in treatment propensity: the four types of students vary in their likelihood of being treated (20%, 80%, 50%, 40%).
- Selection bias: student types with higher baseline outcomes (Y_0) are more likely to be treated.
- Conditional independence: *conditional on type*, treatment status is randomly assigned. Put another way, within each type, there is no relationship between treatment and potential outcomes (or treatment effects). All students of the same type have the same potential outcomes (and the same treatment effects) and likelihood of treatment.

The example considers several estimators:

- Simple difference in means: fails to recover the ATT because of selection bias. The untreated cases are used to provide the counterfactual for the treated, but we know the untreated have lower mean Y_0 than the treated (selection bias). The simple difference over-estimates the ATT. It also fails to recover the ATE due to selection bias and heterogeneous treatment effect bias. There are two issues here. First, untreated cases are used to provide the counterfactual for the treated, and we know they have lower mean Y_0 (selection bias). Second, *even if there were no selection bias* (in terms of different Y_0) the difference between the treated and untreated group would not reflect the ATU if there are heterogeneous treatment effects. (Recall, the ATE is a weighted average of the ATT and ATU).
- Equally-weighted average of the within-group difference in means: as noted above, conditional on type, treatment status is randomly assigned. So, the within-group difference in means provides the ATE, ATT, and ATU *for that group*. So, this estimator takes us a long way. However, the groups are unequal in size, so equal weights provide a (somewhat) distorted view of the ATE. The bias here generally will be larger the more uneven in size the groups are.
- Group-size-weighted average of the within-group difference in means: this gives us the ATE we want since it uses within-group differences in means and appropriately weights by group size.

Group	N	Y0	Y1	τ	P(treat)
1	100	25	35	10	0.20
2	150	50	90	40	0.80
3	200	40	60	20	0.50
4	150	30	45	15	0.40
All					

- Regression with controls for group: Intuitively, since group is the only confounding factor here one might think that a regression with explicit controls for group would be enough to provide us with the ATE. While it gets close to estimating the ATE, it doesn't estimate this

exactly. This is because OLS provides a *variance-weighted average treatment effect*. Groups that have more variance in treatment get more weight. (Since the variance of a binary variable is *p*1-p* the group closest to p=0.50 gets the most weight). See the problem set do file for a demo of this.

Connection to Lecture 2:

Consider your calculation of the group-size weighted average of the within-group difference in means above. Let *t* represent treated observations and *u* represent untreated observations. The numbers represent groups (1, 2, 3, 4):

$$\frac{n_1}{N}\left(\frac{1}{n_{1t}}\sum Y_{1t} - \frac{1}{n_{1u}}\sum Y_{1u}\right) +$$
$$\frac{n_2}{N}\left(\frac{1}{n_{2t}}\sum Y_{2t} - \frac{1}{n_{2u}}\sum Y_{2u}\right) +$$
$$\frac{n_3}{N}\left(\frac{1}{n_{3t}}\sum Y_{3t} - \frac{1}{n_{3u}}\sum Y_{3u}\right) +$$
$$\frac{n_4}{N}\left(\frac{1}{n_{4t}}\sum Y_{4t} - \frac{1}{n_{4u}}\sum Y_{4u}\right)$$

This can be re-written as:

$$\frac{1}{N}\left(\sum \frac{1}{p_1} Y_{1t} - \sum \frac{1}{(1-p_1)} Y_{1u}\right) + \cdots$$

In other words, a simple average of an *inverse propensity score weighted* Y, where treated observations are multiplied by (1/p) where p is their propensity score and untreated observations are weighted by (-1)*(1/(1-p)).

Question 3

Note 1: This question considers a continuous treatment, years of teacher experience. In the absence of random assignment, one would want to be conscious of selection bias (treatment—in this case, years of teaching experience—is related to potential outcomes) and heterogeneous treatment effects (students may differ in the extent to which they benefit from teaching experience). An additional consideration is that there multiple possible values of teacher experience; treatment effects could be linear in experience or nonlinear. (For this problem, we won't think about nonlinear treatment effects).

Note 2: The regression anatomy formula (part e) often comes in handy in that it allows us to simplify a multivariate regression. In the first step, we regress a variable X1 on *all other explanatory variables* and get the residuals (the "unexplained part" of X1). The regression anatomy formula tells us that the coefficient on X1 in the "long" regression is the same as the coefficient on a "short" regression of Y on the residualized X1. The "unexplained part" of X1 is what is left over after netting out its relationship with the other explanatory variables. The "short" regression formula is simpler and easier to think about than a multivariate one, and captures much of the intuition.