Applied Econometrics
Prof. Leo Feler
Quiz 3: Propensity-Score Matching

Name: Key

1. What is propensity-score matching? When can we use it? And why do we use it?

A propensity score is an estimated likelihood of treatment calculated based on observable characteristics. Propensity score matching is a method for approximating random assignment to treatment by comparing the outcome of interest for observations with similar propensity scores but different treatment assignment (for example, one smoker and one non-smoker with the same estimated probability of being a smoker).

The propensity score theorem posits that if outcomes are independent of treatment status conditional on the covariates then outcomes are also independent of treatment conditional on the propensity score:

If $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp T_i \mid X_i$

then $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp T_i \mid p(X_i)$

where $Pr(T_i=1 \mid X_i) = E(T_i \mid X_i) = p(X_i)$

See Mostly Harmless pp. 80-81 for proof.

We can use propensity score matching when there is sufficient overlap between the propensity scores of treated and untreated individuals and when the covariates balance between the two groups within the blocks we divide them into for matching.

We use this method because it allows us to approximate the conditions of a random experiment when one is not feasible, and because it is a lot simpler (and more efficient) than controlling for all of the covariates.

2. What makes propensity-score matching *potentially* better than simply using the same controls (used to derive the propensity score) in the structural equation we're trying to estimate? What are the risks to using propensity-score matching?

Propensity score matching is potentially better than controlling for all relevant observable characteristics because it reduces the multi-dimensional X_K into a single index that we can use to match individuals.

The main risks are that if there is insufficient overlap in pscores, or if treatment status and observable traits are not independent, then matching is useless. (unbalanced blocks)

Estimated effects will be sensitive to how you calculate the pscore.

3. What's the "algorithm" used to estimate the propensity score? In other words, what steps do you (or Stata) take to estimate the propensity score?

① Start with parsimonious a logit model for estimating $\hat{p}(x_i)$.

② Stratify the sample into 5 blocks based on $\hat{p}(x_i)$.

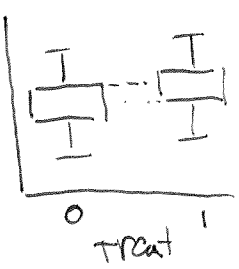
③ Test whether each covariate balances within each block using a t-test to check whether the mean for treated and untreated individuals is not statistically significantly different.

- If the covariates balance, stop.
- If they do not balance, split the blocks that do not balance into two.
  - or, add in polynomials and interaction terms and re-estimate $\hat{p}(x_i)$

"balanced" means that 90% of the covariates in 90% of the blocks fail to reject H_0: $\bar{X}_{1,K} = \bar{X}_{0,K}$, which is a rule of thumb rather than a set cutoff point.

4. Once you've obtained the propensity score, what are four different things you can do with it (be specific: write out equations where appropriate, or show me the commands you would use in Stata)?

① graph a boxplot of $\hat{p}(x_i)$ for treated and untreated individuals to look for overlap.



② Use $\hat{p}(x_i)$ as a control: $Y_i = \alpha + \theta T + \delta_1 \hat{p}_i + \delta T_i (\hat{p}_i - \hat{\mu}_p) + \varepsilon_i$

Can also control for interaction between treatment and deviation from mean pscore.

③ Allow for a different intercept for each block

$$Y_i = \sum_b \alpha_b + \theta T + \varepsilon_i$$

In other words, putting in dummy variables for each of the blocks

④ Allow θ to vary by block as well

$$Y_i = \sum_b \alpha_b D_b + \sum_b \gamma_b T D_b + \varepsilon_i$$

⑤ Calculate the ATE, TOT using the pscores to weight observations in the regression

$E(Y_{1i} - Y_{0i} | T=1) \equiv TOT$, if $T=0$, weight $= \frac{p_i}{1-p_i}$; if $T=1$, weight $= 1$

$E(Y_{1i} - Y_{0i}) \equiv ATE$, if $T=1$, weight $= \frac{1}{p_i}$; if $T=0$, weight $= \frac{1}{1-p_i}$

5. Discuss why we might want to use ATE and TOT weights when estimating a treatment effect; specifically discuss what happens to the kernel density distributions of the estimated propensity scores when we weight by ATE and TOT weights.

We might want to use ATE weights when estimating a treatment effect when we want to put more weight on the average person in order to see what would be the impact on the average person if they changed their treatment status (ex. smokers became non-smokers and vice versa). This method puts more weight on the average individual, which typically comprises a combination of both the treated and the untreated. The weights for ATE are different for the treatment and the control (untreated) groups:

$$\text{Weight for treated: } \frac{1}{p(x_i)}$$

$$\text{Weight for controls (untreated): } \frac{1}{1- p(x_i)}$$

With regards to the kernel density distribution, we would expect both the distributions of the treated and the untreated to shift towards the total mean p-score (towards the middle).

Using TOT weights would be useful in examining the effect of treatment specifically for those who are really likely to select into treatment in order to try to estimate what would be the impact on the average treated person if they did not receive treatment. The weights for TOT only fall on the control group and take the following form:

$$\frac{p(x_i)}{1 - p(x_i)}$$

The higher the $p(x_i)$, the greater the weight. Therefore, this places more weight on control observations that are more likely to select into treatment (but didn't) and less weight on control observations less likely to select into treatment (the treatment group is 'weighted' simply by 1).

With regards to the kernel density distribution, we would expect the distribution of the untreated to move rightward towards the distribution of the treated in that it shifts towards the mean p-score of the treated group.

We would expect to see similar effects from the two kernel density distributions, except when the effects of treatment are heterogeneous (not constant). In the impact of smoking on birth weight example, the kernel density distributions using ATE and TOT weights are similar.