PSC 504 | Prof. Matthew Blackwell | Harkness 307 | m.blackwell@rochester.edu

Homework 3

Hard copy due February 14th, 2013 at the start of class

1 Propensity Scores

Assume that conditional ignorability holds, along with non-degenerate treatment assignment:

$$(Y(1), Y(0)) \perp\!\!\!\perp A | X$$

$$0 < e(X) = \Pr(A = 1|X) < 1$$

Prove that ignorability also holds, conditional on $e(X)$. Hint: use various conditional distributions of A.

2 Stratification

For this problem, you will analyze the data from:

> Christopher Blattman and J Annan. 2010. "The consequences of child soldiering." *Review of Economics and Statistics* 92 (4): 882–898

The data are from a panel survey of male youth in war-afflicted regions of Uganda. The authors want to estimate the impact of forced military service on various outcomes. They focus on Uganda because there were a significant number of abductions of young men into the Lord's Resistance Army.

Blattman and Annan describe the abductions as follows:

> Abduction was large-scale and seemingly indiscriminate; 60,000 to 80,000 youth are estimated to have been abducted and more than a quarter of males currently aged 14 to 30 in our study region were abducted for at least two weeks. Most were abducted after 1996 and from one of the Acholi districts of Gulu, Kitgum, and Pader.
>
> Youth were typically taken by roving groups of 10 to 20 rebels during night raids on rural homes. Adolescent males appear to have been the most pliable, reliable and effective forced recruits, and so were disproportionately targeted by the LRA. Youth under age 11 and over 24 tended to be avoided and had a high probability of immediate release. Lengths of abduction ranged from a day to ten years, averaging 8.9 months in our sample. Youth who failed to escape were trained as fighters and, after a few months, received a gun. Two thirds of abductees were forced to perpetrate a crime or violence. A third eventually became fighters, and a fifth were forced to murder soldiers, civilians, or even family members in order to bind them to the group, to reduce their fear of killing, and to discourage disobedience.

In this problem we will look at the effect of abduction on `educ` (years of education). The `abd` variable is the treatment in this case. Note that `educ`, `distress`, and `logwage` are all outcomes/post-treatment variables.

- `abd`: abducted by the LRA (the treatment)
- `c_ach` – `c_pal`: Location indicators (each abbreviation corresponds to a sub-district; i.e. ach = Acholibur, etc.)
- `age`: age in years
- `fthr_ed`: father's education (years)
- `mthr_ed`: mother's education (years)
- `orphan96`: indicator if parent's died before 1997
- `hh_fthr_frm`: indicator if father is a farmer
- `hh_size96`: household size in 1996
- `educ`: years of education
- `distress`: index of emotional distress (0-15)

- `logwage`: log of average daily wage earned in last 4 weeks

1. Calculate the naive (*prima facie*) ATE of abduction on education and the standard error of this estimate.

2. Create a "Love plot" similar to Figure 7 and Figure 8 that has variables on the y-axis and standardized differences-in-mean for those variables on the x-axis for the overall data. In what ways do abducted and unabducted youths differ?

3. Use a parametric model to calculate the propensity scores for each person in the data. Include whatever covariates or functions of covariates you think maybe be important.

4. Divide the data into three equally-sized strata based on the propensity score and create a "Love plot" to compare the standardized differences-in-mean between the overall data and one of the strata. Does the balance improve after subclassification?

5. Divide the data into 6 equally-sized strata based on the propensity score and repeat the previous exercise. How does increasing the number of strata affect the balance between the treated and control groups?

6. Calculate the average treatment effect within each of the 6 strata and then calculate the overall average treatment effect.

7. Derive an expression for the (population, not sample) variance of the within-strata ATE and the overall ATE. Use these expressions to estimate standard errors for each set of estimates. Hint: within the strata, we are assuming a completely randomized experiment.

8. The last part ignores variation due to part of estimation procedure. Identify that part of the estimation and calculate bootstrapped standard errors to handle these issues. Your bootstrapping should sample (with replacement) from the units 500 times and (1) calculate the propensity score, (2) divide the units into 6 equally-sized strata, and (3) calculate within-strata and overall ATE estimates. How do the bootstrapped standard errors differ from those calculated in the last part?

9. In the Blattman and Annan paper, they rely on non-parametric estimates of the propensity score. What features or properties of using propensity scores might lead them to use non-parametric as opposed to parametric models?