Political Science 239

Problem Set 5

Due date: Wednesday, October 11th, 2006

In this problem set, we will use the experimental data used by Lalonde to estimate the impact of the NSW employment program on earnings. The dataset is available from the "Matching" library (after loading the library, use the command `data(lalonde)`). The dependent variable is earnings in 1978 (`re78`), the treatment is whether the person was assigned to the training program (`treat`), and the rest are covariates. You can read the details of the experiment in Lalonde (1986) and Dehejia and Wahba (1999). Remember that these are data from a randomized experiment, since the training program was randomly assigned.

Exercise 1 *Compute difference in means (ttests) and variance ratios on the covariates between treatment and control observations. Are they balanced? Why should they be?*

Exercise 2 *Estimate the average treatment effect of training on earnings by subtracting the mean of the 1978 income of controls from the mean of the 1978 income of treatments. Is this an unbiased estimator of the average treatment effect of training on earnings? Is this an unbiased estimator of the average treatment effect on the treated of training on earnings? Why/Why not?*

Exercise 3 *Run an OLS model to estimate the impact of training on earnings. Run two models: one with no covariates and one including all covariates. Remember exercise 3 in Problem Set 2. Does your estimated effect change? What about your standard errors? What changes should you expect and what changes do you see? What do these changes tell you about the design?*

Exercise 4 *Now we will do matching using the propensity score, orthogonalized covariates and Mahalanobis distance. Estimate the propensity score using all covariates. Match treatment and controls using the* `Match()` *function. Match on the estimated linear predictor and the orthogonalized covariates (i.e., you should transform every covariate such that it is orthogonal to the linear predictor). Use the Mahalanobis distance metric as your weighting scheme (this is, set* `Weight=2`*). Now analyze the balance of your matched sample on all the covariates using the* `MatchBalance()`*function. Has your balanced improved? Using these matched sample, report the estimated average treatment effect on the treated (ATT). How does it compare to the average treatment effect estimated in Exercise 2?*

Exercise 5 *Now re-run the matching that you did in Exercise 4, including higher order terms and interactions in the estimation of the propensity score. Analyze the balance of this new matched sample using* `MatchBalance()`*. How does it compare to the balance you had under the first matching procedure? What is the ATT? Is it similar to the one you estimated in Exercise 4? What about the ATT that you estimated in Exercise 1? Should they be similar? Why/Why not?*