Computer Exercice 1 – Matching

Bart Cockx

2014

For this exercise use the sample data (`dataBlun_noc.dta`) extracted from the Blundell and Dias (2009) paper and used in the lectures for purposes of illustration.

1) Use the procedure `pscore` written by Becker and Ichino (2002) for Stata[1] to check whether the balancing property is satisfied for the model in which `y = $d*ln(y1)+(1-$d)*ln(y0)`as outcome variable and
   a) *x* and *theta* are taken as matching variables without including any interactions between these variables or higher order terms
   b) *x* and *theta* are taken as matching variables including *theta* squared, and the interaction between *x* and *theta.*

   Use the option "details" and try to understand how this procedure exactly works. If the balancing property is not satisfied, report in which block and for which variable it is not satisfied. If it's satisfied, report how many blocks are required to satisfy the balancing property. Discuss your findings.

2) Use psmatch2 (of Leuven and Sianesi) to
   a) Estimate the ATT, ATNT (=atu) and ATE using of Epanechnikov kernel matching with bandwidth = .05 using the outcome and matching variables defined in Question 1)b), **without** imposing a common support. Estimate the standard error by bootstrapping with 50 replications. Set the seed equal to "321" so that I can check the results. Report the statistics of pstest before and after matching.
   b) As in a), but imposing a common support.
   c) As in a) , but estimate the nearest neighbor matching estimator using the analytical error proposed by Abadie and Imbens (2006) using one neighbor to calculate the conditional variance (i.e. option "ai(1)").

   Use "atts" of the procedure of Becker and Ichino to

   d) As in a), but estimate the blocking estimator for ATT and ATNT, using the specification of the propensity score described in b) and determining the number of blocks according to the "pscore" procedure to balance the propensity within blocks. Estimate the analytical standard error: you do **not** need to bootstrap. Estimate ATE

[1] Define the variables such as needed to estimate *ATT* and *not ATNT.* This does not fundamentally affect the results, but if the balancing property is not satisfied, the order of the blocks in which it's not satisfied is reversed.

using the fact that ATE=P*ATT+(1-P)*ATNT, where we use the fraction of treated units in the sample as an estimate for P.

e) According to the statistics of pstest in a), the propensity score is not completely balanced. Provide the statistics which are a basis for concern and explain why. This contrasts to the findings in 1.b). How do you explain this contrast?

f) Imposing a common support in 2.b) hardly changes the results. This seems to suggest that common support is not really an issue. However, if you analyze the density functions of the propensity score by treatment group (which you can do by the command "psgraph") there are nevertheless reasons for concern. Explain this. Also mention for which estimator is the concern more important: ATT or ATNT? Why?

g) Compare the ATT, ATNT and ATE of a), b) and d). Discuss the potential sources of the observed differences. Use in this discussion the fact that the true ATT=0.471 and the ATNT=0.315 (see Blundell and Costa Dias (2009), Table 3).

3)

a) Estimate ATE=ATT=ATNT by standard OLS using x and theta (without second order terms) as controls and imposing *homogenous* returns to education.

b) Estimate ATNT, ATT and ATE using linear regression according to the procedure explained in the lecture (cf. slides entitled "Linear Regression for Average Treatment Effects"). Use the same controls as in a). No need to estimate the standard errors.

c) Estimate ATE, ATT and ATNT using the re-weighting estimation procedure. Use the same specification of the propensity score as in Question 1.b).

Compare the findings and discuss briefly their differences using the theoretical arguments from the course.