Problem Set 3 - Treatment Effects

Profs. Charlie Murry and Richard L. Sweeney

Matching

This problem replicates an excercise by Dehejia and Wahbia (<u>JASA 1999</u>, <u>ReStat 2002</u>) exploring non-experimental treatment effects from the Lalonde (<u>AER 1986</u>) data.

All of the refrenced datasets can be downloaded at <u>Rajeev Dehejia's website</u>. See the descriptions for variable names if you use the text files.

Start with the experimental data. Use the DW sample (nswre74_[group].txt or nsw_dw.dta).

- 1. Test for balance across the treatment and control group.
- 2. Estimate the main treatment effet of the training program on 1978 earnings (re78). Estimate the model with and without controls. Do the covariates matter here?

Now estimate the treatment effect using a different control group. First, drop the NSW control group. Then append the PSID sample (psid_controls.txt or .dta), none of which recieved the treatment.

- 3. Test for balance between the treatment group and this new control group.
- 4. Esimate the effect of the training with and without controls.
- 5. Construct a propensity score for the treatment. Plot the propensity score across the treatment and control groups.
- 6. Re-estimate the treatment effect after matching on the estimated propensity score. Use nearest-neighbor matching (with replacement).
- 7. Finally, estimate the treatment effect using coarsened exact matching. Which matching method performs "better" here?

Regression Discontinuity Design

Use the RDD checklist from Lee and Lemieux (<u>JEL 2010</u>) and the data <code>yelp.Rdata</code> to estimate the causal effect of an additional star (not half-star) on the yelp platform. Your data contain the following variables:

- logrev monthly store revenue in (log) dollars.
- stars the number of displayed stars on the Yelp site
- score this is the true Yelp score that is rounded to produce stars
- rest_id this is the restaurant identifier (1 to 1500)
- time this is the time identifier (1 to 10)