Problem Set 1
Quantitative and Statistical Methods II

General Instructions

- The problem set is due by January 29 at noon;
- You should send it by email *bruno.conte@barcelonagse.eu*; it must include your Stata code, *a unique* log file and your answer sheet. Tidiness is appreciated – e.g. material correcly labeled and organized in zip files;
- Working in teams is allowed and strongly recommended (keeping the same groups as in the presentations is encouraged);
- The datasets needed to solve the computational questions are uploaded in the Classroom material.

Part 1: "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?", Angelucci and De Giorgi, American Economic Review, 2009

1. Using the data in `angeluccidegiorgi.dta`, replicate the results of the rows *Control*, *Treatment* and *ITE (No Controls)* of Table 1. Try to use both difference in sample averages and regressions; [Hint: *ITE* here does not necessarily mean *Intentions to Treat Effects*. Read the paper, Section III]
2. Interpret the outcome of the row *ITE (No Controls)*. Is this estimate causal? Motivate your answer;
3. Show analytically how to test the significance of outcome of the row *ITE*;
4. Show that:
$$var(\bar{Y_1} - \bar{Y_0}) = var(\bar{Y_1}) + var(\bar{Y_0}) = \frac{var(Y|D=1)}{n_1} + \frac{var(Y|D=0)}{n_0},$$
where $\bar{Y_d} = \frac{1}{n_d}\sum_{i=1}^{n_d} Y_i$ and $n_d = \sum_{i=1}^{n} \mathbb{1}\{D_i = d\}$. State and discuss any assumptions you make in the derivation.

Part 2: "Causal Effects in Nonesperimental Studies", Dehejia and Wahba, Journal of the American Statistical Association, 1999

1. Consider Table 2, rows 2-7. Why do these results fail to replicate the result reported in Table 2, row 1?
2. Describe a method to estimate the average treatment effect on the treated under the conditional independence assumption;
3. Discuss the credibility of the conditional independence assumption in the context of the paper;
4. Read Section 3 carefully and replicate Table 3, column 7 using the datasets `nsw_dw.dta` (experimental data), `psid_controls.dta`, `cps_controls.dta`,...(6 control groups)[1]. Implement nearest neighbor matching based on the propensity score. You will likely not replicate the exact numbers but you should obtain something close to what is reported in the paper. [Hint: use `STATA` commands `pscore` and `attnd` as shown in `class 01.do`. Note also that the specification used by the authors in the first stage (pscore calculation) is reported in the table's footnote. Please ignore variable `nodegree`. Finally, variables `u74` and `u75` mean "unemployed status" in years 1974 and 1975, respectively. An individual is unemployed if her/his reported wage is zero.]

[1] These datasets are available here.