

Université Paris-Saclay  
Master of Economics – 1st year  
Advanced Econometrics – Tutorial #5  
Matching Estimators \*

March 28th 2022

## Tutorial context

**Randomized experiments cannot always be implemented.** Thus, estimating effects from microeconomic policy, such as the effect of job training program on earnings, must deal with **observational studies and data**. Several techniques have been developed, among which the **matching method** using propensity score. The matching technique is a **systematic method for creating comparison groups with similar covariate values than the treated individuals**, which can then be used to average the differences with the treated individuals and identify a causal effect.

One way to assess the ability of such an econometric method to solve the evaluation problem is to try the method in a situation where an experimental estimate is available. This is the strategy in LaLondes classic 1986 paper and in the papers that reanalyze his data: Dehejia and Wahba (1999), Becker and Ichino (2002), Smith and Todd(2005).

LaLonde (1986) is the first to examine a randomized experiment (the National Supported Work Demonstration, NSW) in order to assess the performance of several evaluation methods. He obtains **an unbiased estimate of the training effect from the randomized experiment**, and then compares the experimental result to those obtained from a range of parametric selection models applied to the NSW observations that received training and a set of comparison observations constructed from population survey data sets (CPS and PSID). Following papers have reanalyzed Lalondes results, using more recent estimation methods, including matching.

The goal of this tutorial is to analyze **the performance of regression and matching estimators** using observational study, replicating some of the results of these studies.

## 1 Data description

We here use a subset of the data used in LaLonde (1986), constructed by Dehejia and Wahba (1999). These data are available online: <http://www.nber.org/~rdehejia/nswdata2.html>

There are two types of samples:

- One sample of experimental data, which is a subset of Lalonde's data, and they are obtained from an experimental setting (see below for details on the experiment)

---

\*Contact: thibault.richard(at)ens-paris-saclay.fr

- Two samples of observational data, which form the comparison groups. These two samples are constructed from two surveys: the Population Survey of Income Dynamics (PSID), and the Current Population Survey (CPS).

The experimental data come from a randomized experiment from the National Supported Work (NSW) Demonstration. This program was a transitional, subsidized work experience program that provided trainees with work in a sheltered training environment and then assisted them in finding regular jobs. To participate in NSW, potential participants had to satisfy a set of eligibility criteria that were intended to identify individuals with significant barriers to employment. The main criteria were: (1) the person must have been currently unemployed (defined as having worked no more than 40 hours in the 4 weeks preceding the time of selection into the program), and (2) the person must have spent no more than 3 months on one regular job of at least 20 hours per week during the preceding 6 months. Actual treatment among participants was randomized. However, as a result of these criteria as well as of self-selection into the program, persons who participated in NSW are likely to differ in many ways from the general U.S. population.

Candidates eligible for the NSW program were randomized into treatment between March 1975 and July 1977. Treated individuals in our dataset concerned individuals assigned to treatment in 1976, so that retrospective earnings information from the experiment include calendar 1974 and 1975 earnings (variables RE74 and RE75). Thus, the pretreatment variables are earnings in 1974 and 1975. Postintervention data includes calendar 1978 earnings.

Variables, for both types of data, are (from left to right) treatment indicator (1 if treated, 0 if not treated), age, education, Black (1 if black, 0 otherwise), Hispanic (1 if Hispanic, 0 otherwise), married (1 if married, 0 otherwise), nodegree (1 if no degree, 0 otherwise), RE74 (earnings in 1974), RE75 (earnings in 1975), and RE78 (earnings in 1978), and dummy variables to identify data origin with exp (1 if experimental data, 0 otherwise) and cps (1 if CPS observational data and 0 otherwise).

1. We want to verify that the treatment and control groups have the same distribution of variables **before the program intervention**. Calculate the descriptive statistics. What can you conclude from this table?

## 2 Econometric analysis and replication

### 2.1 Standard estimation of treatment effect

2. We are now interested in replicating the comparison of earnings and estimated training effects of the NSW program in Table 2 of Dehejia and Wahba (1999). We are first interested in the estimation of the effect of training on earnings using the experimental observations, which correspond to the first row called “NSW” for the sets of columns B. and C. in their table.
  - (a) From the experimental observations, what is the most simple unbiased estimator of the average treatment effect? Write down the linear model which allows to recover

this estimator. Then, estimate it with the experimental subset of data. How do you interpret your result?

- (b) In the Table 2 of Dehejia and Wahba (1999), we note that in addition to the simple treatment estimation (columns called “Unadjusted”), the authors also compute estimated treatment effect while controlling for age, age squared, years of schooling, high school dropout status and race (columns called “Adjusted”). What is the purpose of controlling for other covariates in our experimental setting? Reproduce these two results.
  - (c) Finally, the authors add other variables which are pretreatment earnings (earnings in 1975 for the columns “Unrestricted differences in differences: quasi difference in earnings growth 1975–1978”, and earnings in 1974 for all columns of panel C.). Replicate their results using these variables.
  - (d) What do you conclude on your results when you add pretreatment observables relative to your first simple estimate in question a?
3. We are now interested in assessing whether the regression method applied on observational data (no random assignment of treatment) can replicate the treatment impact. We separately use the PSID and CPS samples as comparison groups of the treated individuals (and thus exclude the non-treated experimental data).
- (a) Using the observational data as a comparison group, what can we expect on the simple estimator of the average treatment effect while knowing the results of the data analysis in question 1? What about when accounting for pretreatment covariates?
  - (b) Perform the same regressions as in question 2 while using first the CPS data as a control group, and second the PSID data as a control group. What do you conclude on the estimation of the treatment impact while using the observational samples as comparison groups?

## 2.2 Estimating the treatment effect using the propensity score

The propensity score can be an intermediary tool to compare the average differences by matching experimental treated individuals with similar non-experimental individuals. The intuition of the propensity score is that whereas in the standard regressions we are trying to condition on  $X_i$  (intuitively, to find observations with similar covariates), we are here trying to condition just on the propensity score because observations (experimental and non-experimental) with the same propensity score have the same distribution of the full vector of covariates  $X_i$  (see class Session materials on matching and article of Dehejia and Wahba (1999) for further demonstration). Therefore, if we condition our regressions on the propensity score (that is by comparing individuals with similar propensity score), the distribution of pretreatment regressors will be the same across the treatment and comparison group. Each individual has the same probability of assignment to treatment, as in a randomized experiment.

The estimation is done in two steps. In the first step, the propensity score is estimated **on the experimental treated units and on the PSID non-experimental sample.**

Second, given the estimated propensity score, **we estimate a regression of the effect of job training on earnings for individuals with similar estimated propensity score**. Finding the individuals with similar estimated propensity score, that is, *the matching method*, can be done with varying techniques. The two techniques developed in Dehejia and Wahba (1999) are the stratification strategy and the nearest-neighbor matching strategy.

4. First, we estimate the propensity score using the sample of the **experimental treated observations** and the **PSID observations** (thus discard CPS observations and experimental control observations). One issue is what functional form of the preintervention variables to include in the propensity score model. We thus want to test several functional forms for the propensity score and select the specification which leads to distributions of covariates for observations with similar propensity score that are the same across the treatment and the comparison groups.
  - (a) Generate the following variables: a dummy variable denominated **U74** which is equal to 1 if the individual is unemployed in 1974 (identified by 1974 earnings equal to 0) and 0 otherwise; a dummy variable denominated **U75** which is equal to 1 if the individual is unemployed in 1975 (identified by 1975 earnings equal to 0) and 0 otherwise; a dummy variable denominated **blackU74** which is equal to 1 if the individual is black *and* unemployed in 1974 and 0 otherwise; the squared education denominated **ed2**; and the 1974 and 1975 squared earnings denominated **re742** and **re752**.
  - (b) We now try a first specification of propensity score. Estimate a **logit** specification of the propensity score (the probability of being treated) as a function of **age age2 ed ed2 black hisp married re75 re74 re742 re752 blackU74**. Then produce an histogram to compare the predicted propensity score between treated and non-treated observations. What do you conclude on your propensity score specification?
  - (c) There are STATA packages which estimate the propensity score, and then can perform a matching strategy. We here produce a matching strategy which stratifies the data by propensity score, which can be opposed to a nearest-neighbor matching strategy. Observations with estimated propensity score which are in the same range are then grouped into blocks (strata) and we check whether we succeed in having similar distributions of covariates within each blocks for treated and comparison observations. Use the **findit** command to look for the **st0026\_2** package developed in Becker and Ichino (2002) which test the balancing property for each block and each covariate and install it. Then, use the **pscore** command using the same covariates as for the preceding **logit** command, and with the **pscore(mypscore) blockid(myblock) logit numblo(5) level(0.005) and detail** in the options of the command. Explain how the algorithm processes and discuss the results obtained on your STATA output.
5. We now perform the OLS regression after restricting the NSW-treated / PSID-comparison sample using propensity score.
  - (a) First estimate the propensity score using a Logit specification and where you add

control for `re75`, `age`, `age2`, `ed`, `black`, `hisp`, `nodeg` and `married`. Keep the observations of predicted values of propensity score which verify  $0.1 < \hat{p}(X) < 0.9$  — such as we can exclude effects caused by outliers.

- (b) Finally, run the same previous regressions as those in questions 2 but on these matched observations. Compare your results, with the one obtained in the previous questions.