2

The Evaluation of Policy Interventions with Matching Estimators

2.1 Introduction

The matching approach originated from the statistical literature and shows a close link to the experimental context.[1] The basic idea underlying this approach is to find in a large group of non-participants those individuals who are similar to the participants in all relevant pre-treatment characteristics. That being done, differences in outcomes of this well selected and thus adequate control group and of the participants can then be attributed to the programme. Matching has become a popular method especially when evaluating labour market policies, but examples for usage of the matching method can be found in very diverse fields of study. It applies for all situations where we have a treatment, a group of treated individuals and a group of non-treated individuals. The nature of treatment may be very diverse. For example, Perkins, Tu, Underhill, Zhou, and Murray (2000) discuss the usage of matching in pharmacoepidemiologic research. Hitt and Frei (2002) analyse the effect of online banking on the profitability of customers. Davies and Kim (2003) compare the effect on the percentage bid-ask spread of Canadian firms being interlisted on an US-Exchange, whereas Hujer and Radic (2005) analyse the effects of subsidies on the innovation activities of firms in Germany. Brand and Halaby (2003) analyse the effect of elite college attendance on career outcomes, while Ham, Li, and Reagan (2003) study the effect of a migration decision on the wage growth of young men. Bryson (2002) analyse the effect of union membership on wages of employees and Behrman, Cheng, and Todd (2004) apply matching methods to estimate the impact of preschool programmes on cognitive, psycho-social and anthropometric outcomes of children.

This is only a short listing which could be easily augmented. It should merely point out the popularity of matching in diverse fields of research.

[1] See Rubin (1974), (1977), (1979), Rosenbaum and Rubin (1983), (1985a), (1985b) or Lechner (1998)).

Of course, matching is first of all plagued by the same problem as all non-experimental estimators, which means that the condition in equation (1.5) cannot be expected to hold when treatment assignment is not random. However, following Rubin (1977), treatment assignment may be random given a set of covariates. The construction of a valid control group via matching is based on the identifying assumption that conditional on all relevant pre-training covariates X, potential outcomes are independent of treatment assignment. We have already discussed this assumption, known as unconfoundedness or conditional independence assumption (CIA), in section 1.5. Similar to randomisation in a classical experiment, the role of matching is to balance the distributions of all relevant pre-treatment characteristics X in the treatment and control group, and thus to achieve independence between potential outcomes and assignment into treatment, resulting in an unbiased estimate. We have also shown that matching represents an important extension of standard regression approaches as it does not impose functional forms on the outcome equations, highlights the common support problem, and allows for heterogeneous treatment effects (see our discussion in subsection 1.5.4).

Conditioning on all relevant covariates is, however, limited in case of a high dimensional vector X. For instance, if X contains n covariates which are all dichotomous, the number of possible matches will be 2^n. In this case exact matching on X (cell matching) is in practice not possible, because an increase in the number of variables increases the number of matching cells exponentially. In the dataset underlying our further analyses in chapters 5 and 6 we have approximately 38 discrete and 6 continuous variables. This gives us a possible number of over 278 million cells and makes the use of exact covariate matching impossible.

To deal with this dimensionality problem, Rosenbaum and Rubin (1983) suggest the use of balancing scores $b(X)$, i.e. functions of the relevant observed covariates X such that the conditional distribution of X given $b(X)$ is independent of assignment into treatment. We will discuss their results in the following section. One possible balancing score is the propensity score, i.e. the probability of participating in a programme given observed characteristics X. Matching procedures based on this balancing score are known as propensity score matching (PSM) and will be the focus of this chapter. Whereas exact (or cell) matching is completely non-parametric, propensity score matching is semi-parametric as it combines a parametric model for estimation of the participation probability with a non-parametric comparison of the outcomes.[2]

Even though propensity score matching was developed in the early 1980s (Rosenbaum and Rubin, 1983) and has its roots in a conceptual framework which dates back even further, its use in labour market policy evaluation was established in the late 1990s only. Especially the work of Dehejia and Wahba

[2] Note that Hirano, Imbens, and Ridder (2003) suggest a non-parametric series estimator for the propensity score which makes propensity score matching completely non-parametric.