SDT 279

AN OPTIMIZATION-BASED MATCHING PROCEDURE

Autor: Rodrigo Krell,
Tomás Rau y Jorge Rivera

Santiago, Mayo 2008

Serie Documentos de Trabajo
N 279

An optimization-based matching procedure

Rodrigo Krell Tomás Rau Jorge Rivera

Departamento de Economia
Universidad de Chile

Abstract

This paper presents an alternative matching procedure and analyzes it is performance compared to other popular estimators. These tests show that the estimator performs comparably to the popular propensity score method.

The proposed method involves the advantage of eliminating the inconvenience of an arbitrary common support determination problem, since it solves this problem objectively.

Keywords:

Matching, propensity score, treatment effect, common support, optimization.

JEL codes: C01, C14, C61.

An optimization-based matching procedure[*]

Rodrigo Krell[†]   Tomás Rau[‡]   Jorge Rivera C[§]

January 2008

Abstract

This paper presents an alternative matching procedure and analyzes its performance compared to other popular estimators. These tests show that the estimator performs comparably to the popular propensity score method.

The proposed method involves the advantage of eliminating the inconvenience of an arbitrary common support determination problem, since it solves this problem objectively.

Keywords: Matching, propensity score, treatment effect, common support, optimization.

JEL codes: C01, C14, C61.

[†]Departamento de Economía, Universidad de Chile, e-mail: rkrell@econ.uchile.cl.

[‡]Departamento de Economia, Universidad de Chile, e-mail: trau@econ.uchile.cl.

[§]Departamento de Economia, Universidad de Chile, e-mail: jrivera@econ.uchile.cl.

1 Introduction

Econometricians are often interested on measuring the effect that a treatment has on a group of individuals. In the absence of randomized trials, matching methods are powerful tools for estimating this effect, under the assumption that, conditional on observable variables, the treatment assignment is ignorable. In general, a matching estimator compares the average outcome of the treated group and a group of counterfactuals, which is constructed from the sample of non-treated individuals, meeting some criteria given a priori. This yields an estimate of the treatment effect on the treated individuals.

The most popular matching method is based on the propensity score, introduced by Rosenbaum & Rubin (1983) and posteriorly developed by several authors (see Heckman et al. (1997), Heckman et al. (1998) and others.)

In this paper, we propose an alternative matching estimator. In essence, it pairs individuals whose characteristics are "similar" according to a criteria that depends exclusively on the vector of individual's attributes, not by employing propensity score (or related concepts) as a criteria to match individuals. In this sense, our estimator belongs to the kind of methods which don't reduce the dimensionality of the matching problem, such as Abadie & Imbens (2002).

The counterfactual of a treated individual will arise from the solution of an optimization problem defined on the convex hull of the characteristics of the non-treated agent. This leads us to determine a *virtual* counterfactual of any treated individual whose characteristics are close to him, but where the attributes of the extreme points that define the optimal convex combination are also the nearest possible.

To our mind, this new method has four advantages regarding the standard

procedures. First, the right election of the best counterfactual rests upon an objective criteria that seems to be reasonable in the sense that it takes into account all the individual's characteristics, not based exclusively on a real number that summarizes these attributes.

Second, the proposed method overcomes the decision problem regarding the practical way to calculate propensity score. It does not require an extra criteria to define a set of neighbors that are employed to determine the weights used to calculate the resulting counterfactual's outcome.

Third, the same optimization criteria that chooses the best counterfactual for each treated individual is used to account for the common support determination problem that arises when using other matching methods. Contrasting with propensity score matching, the common support problem is solved endogenously, by an objective and optimization-based criteria.

Finally, alluding to numerical applications (Section 4), we found that the proposed alternative matching method performs as well as the propensity score matching Lalonde (1986), and better than most of non-experimental estimators usually employed in the literature.

As we will show, a weakness in our methodology deals with the right election of a norm that defines the optimization problem previously mentioned. Indeed, as well as in standard methods, the resulting estimations here obtained depends on the norm employed, introducing therefore a non desirable arbitrariness in the procedure.

For the numerical applications yet mentioned, we will estimate the treatment effect using the Mahalanobis norm and also employing the norm induced by the diagonal matrix of the t-test (absolute values), arising from a linear

estimation of outcomes versus covariates. From this experience, the method employing the t-test matrix gets a better performance.

Section 2 presents matching methods and briefly discuss the popular propensity score matching. In Section 3 we propose our alternative matching procedure and finally Section 5 is devoted to conclusions.

2 The matching approach at estimating a treatment effect

It has long been documented that when a treatment is not assigned randomly to a particular group, comparing the average outcomes of treated and non-treated individuals yield a poor (biased) estimate of the treatment effect Heckman et al. (1997). Under the assumptions that the treatment assignment is correlated only to observable variables, and that the response of one particular individual does not depend on the treatment assigned to other(s) (Rosenbaum & Rubin 1983), matching methods become an attractive tool for assessing the treatment effect.

Let Y_{1i} be the outcome of individual i if it receives treatment and Y_{0i} if not. A causal effect is a comparison between Y_{1i} and Y_{0i}. Due to the nature of economic data, it's very uncommon for the econometrician to observe both Y_{1i} and Y_{0i}, so a valid estimate of the effect that a treatment has on a group of individuals requires the finding of a counterfactual individual.

In general, matching methods attempt to build a valid counterfactual control group, selecting from the non treated sample individuals which are *similar* to the treated ones. Each treated individual is *matched* to a non-treated indi-

vidual (the latter may not be necessarily in the sample, see Section 3). There are basically two ways by which this can be achieved: One is to select, for each treated individual with characteristics X_{1i}, a non-treated one (or group of non-treated individuals) whose characteristics (X_{0i}) closely resemble X_{1i}. This is a multi-dimensional problem, therefore characteristics must be weighted through a norm.

The other is the popular propensity score matching, which, following Rosenbaum & Rubin (1983) we now proceed to discuss briefly. The propensity score is a pseudo-metric, which summarizes in a real number all the (relevant) observable characteristics of each individual.

We denote $u = 1$ if the treatment is assigned and $u = 0$ otherwise. Let the conditional probability of assignment to the treatment given the covariates (the propensity score) be

$$e(X) = \Pr(u = 1 | X).$$

If a treated individual shares the same propensity score with a non-treated one, then we can think of these individuals as if they were randomly assigned to each group, so the comparison between these individuals' outcome is an unbiased estimate of the treatment effect. In practice, the propensity score is not known but can be estimated, for example using a logistic or probit regression.

Once the propensity score has been estimated for each individual, the econometrician must choose between some alternatives on the next stage, regarding the algorithm by which the matching will be carried out. One way is to match each treated individual with the non-treated one in the sample whose propensity score is the closest one (the so called nearest neighbor matching). Another

4

alternative is to match each treated individual to a weighted average of a subset of non-treated ones, the weight depending on the closeness of the propensity scores (Kernel Matching). This process involves arbitrary choices (the number of neighbors to be included in the matching and/or the specification of the Kernel function).

Apart from the specification problem which arises from the use of propensity score, this method shares with other matching algorithms the common support problem. In a non-experimental study, there may be a group of treated individuals whose characteristics are significantly different from each one of the non-treated. In terms of the propensity score method, some treated individuals would not have a match in the non-treated sample with a sufficiently similar propensity score. This observations must be dropped, since their inclusion would bias the estimates. There is not an objective criteria according to which this process must be conducted, so it involves some degree of arbitrariness.

3   An alternative matching procedure

Let $I = \{1, 2, \ldots, n\}$ be a set of individuals characterized by a vector

$$(Y_i, X_i, u_i) \in \mathbb{R} \times \mathbb{R}^k \times \{0, 1\},$$

where Y_i is the outcome of individual $i \in I$, X_i is a vector of observed pre-treatment characteristics and $u_i = 1$ if individual $i \in I$ is assigned to the experimental treatment ($u_i = 0$ otherwise).

Let $X_1 = \{X_i \mid u_i = 1\}$ and $X_0 = \{X_i \mid u_i = 0\}$ be the sets of characteristics of the treated and not treated individuals respectively (with cardinals n_1 and n_0). Define also $I^1 = \{i \mid u_i = 1\}$ and $I^0 = \{i \mid u_i = 0\}$.

Given a norm $\|\cdot\|$ in $\mathbb{R}^k$, for a treated individual $i \in I^1$ consider the following optimization problem[1]:

$$\min_{\mu} \|X_i - \sum_{j \in I^0} \mu_j X_j\| + \lambda \cdot \sum_{j \in I^0} \mu_j \|X_i - X_j\|$$

$$s.t.\ \mu = (\mu_j) \in \Delta_{n_0},$$ (1)

$$\lambda \geq 0,$$

that is, look for a point in the convex hull of X_0 that approximates X_i as good as possible, but where the corresponding extreme points of the optimal convex combination are the nearest possible to X_i according to the weights given by convex combination factors.

Figure 1: Matching in two variables

---
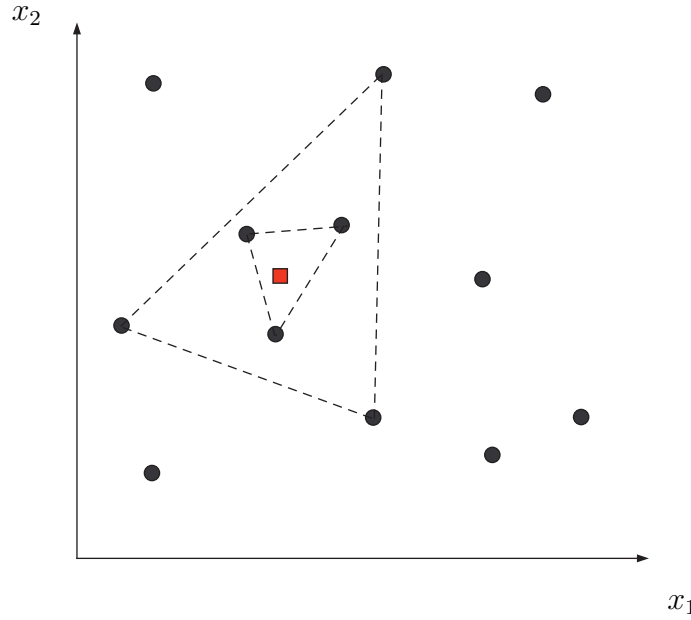
[1]The Simplex in $\mathbb{R}^n$ is denoted by Δ_n, whereas coZ is the convex hull of a set Z, that is, $coZ = \left\{ \sum_{k \in K} \mu_k z_k,\ z_k \in Z,\ \mu_k \geq 0,\ \sum_{k \in K} \mu_k = 1 \right\}.$

Figure 1 shows how this matching works. Notice that the objective function (1) consists of two parts. The first part depends on the difference between the characteristics of the treated individual and those of the convex combination of non-treated individuals $\sum_{j \in I^0} \mu_j X_j$. If the minimization problem consisted only on this expression, it could be the case that a convex combination of non-treated individuals solved the problem, but consisted on observations whose characteristics where far distant from those of the treated one (marked as a square in Figure 1).

The second term in the objective function depends on the distance between the characteristics of the treated individual and those of the non-treated individuals included in the convex combination. The minimization of this term implies a more accurate match. Geometrically, as Figure 1 shows, the inclusion of this term implies the solution will be one with the extreme points as near as possible to the red square (treated individual).

According to our procedure, the best solution between the two *triangles* in the last picture is given by the extreme points that defines the little triangle around the square dot, that is, by those extreme point closest to the the characteristics of the treated individual. Thus, there is a trade-off between the distance of X_i to the convex combination $X \in coX^0$, and the fact that the extreme points we are considering are the nearest as posible to X_i.

We denote by $\mu_i^a = (\mu_{ij}^a) \in \Delta_{n_0}$ a solution of problem (1) and define

$$Y_i^a = \sum_{j \in I^0} \mu_{ij}^a \, Y_j \in \mathbb{R},$$

and

$$v_i = \left\| X_i - \sum_{j \in I^0} \mu_{ij}^a X_j \right\| + \lambda \cdot \sum_{j \in I^0} \mu_{ij}^a \| X_i - X_j \|$$

7

Let $\alpha_i = 1/v_i$. Then, we define

Definition 3.1. *The weighted average treatment effect on the untreated is defined by E^a given by*

$$E^a = \left(\sum_{i=1}^{n_1} \alpha_i\right)^{-1} \cdot \sum_{i=1}^{n_1} \alpha_i(Y_i - Y_i^a)$$

It is easy to check that the solution set of problem (1) is a non-empty and compact set. Since the objective function is only convex (not strictly), we do not have guarantee for the uniqueness of the solution. However, we persist in this formulation for two reasons. First, the uniqueness can be attained if we slightly modify the objective function, for instance introducing a penalizing quadratic term in μ_j. Second, from the numerical experience that we report in next Section, to our mind the results there obtained are satisfactory with the adopted specification.

Remark 3.1. The condition of strong ignorable treatment assignment (and weaker assumptions) introduced by Rosenbaum & Rubin (1983) should imply that

$$coX^0 \cap coX^1 \neq \emptyset, \tag{2}$$

since in the contrary case would exist some characteristic that could *separate* these sets[2]

We also remark that under condition (2), it may happen that for some individuals $i \in I^1$ could have several points in coX^0 such that the distance with X_i is zero. In such case, the optimal solution of problem (1) will not

---

[2]That is, would exists a non-zero vector $p \in \mathbb{R}^k$ such that $p \cdot X_i \leq p \cdot X_j$, $\forall i \in I^0$, $j \in I^1$.

necessarily select one of them due to the component in the objective function involving the distance among characteristics.

Note that the treatment effect estimate we propose is a weighted average of the differences between the treated individuals outcomes and their counterfactuals. The weights are the inverse of the minimized objective function. Thus, this matching method provides an optimization-based procedure to solve the common support determination problem. Observations involving a poor matching are automatically given little weight compared to those in which matching has been accurate, instead of dropping them.

Remark 3.2. The objective function in problem (1) involves a specific norm, which must be chosen according to a criteria given ex ante, introducing therefore certain arbitrariness in the procedure. In the following section, we test the performance of the estimator E^a using two different norms induced by matrices: the usual Mahalanobis norm and the diagonal matrix with the absolute values of the resulting t-tests from a linear regression between covariates and the outcome of individuals. We propose this last matrix since, intuitively, those characteristics which are more tightly linked to the outcome should be given a higher weight than those that are not.

Remark 3.3. In problem (1), λ determines the relative importance of the closeness between the extreme points of the convex combination and the treated individual. If $\lambda \to 0$, then problem (1) has multiple solutions. In terms of Figure 1, there are many possible convex combinations of non-treated individuals that equal a single red square.

On the other hand, if $\lambda \to \infty$, then the solution to the problem implies the choice of the nearest neighbor.

In the following section, we present a test of the estimator's performance assuming $\lambda = 1$.

4   Testing the matching estimator

4.1   Lalonde's evaluation

Following Lalonde (1986), we use non experimental data to evaluate the impact of training on earnings of individuals participating in the National Supported Work Demonstration project. Instead of the experimental controls included in the NSW[3], we use the same sub-samples developed in Lalonde's study. These groups consist on sub-samples drawn from the *Panel Study of Income Dynamics (PSID)* and the *Current Population Survey - Social Security Administration File (CPS_SSA)*.

The unbiased estimate of the treatment effect is found comparing the mean earnings of both treatment and control experimental groups. The need for homogeneity in the data implies a reduction in the sample, amounting to 297 treated and 425 non-treated individuals. This estimate amounts to $866 with a standard error of $476 for male participants and $851 with a standard error of $317 for females.

Lalonde also estimates the treatment effect for a further subsample of data, which contains one more year of pre-intervention data (earnings in 1974). This data includes 185 treated and 260 non-treated observations, all male partici-

---

[3]The NSW program assigned individuals to training positions randomly. One group received all the benefits of this program, while those assigned to the control group received nothing.

pants. The unbiased estimated treatment effect for this observations is $1794, with a standard error of $633. As in Dehejia & Wahba (1998), we restrict our analysis to this subset of data.

Other matching estimators have been tested using the NSW program. For instance, Dehejia & Wahba (1998) evaluate propensity score methods (PSM) with a subset of Lalonde's data, and show that PSM result in accurate estimates of the treatment impact; Abadie & Imbens (2002) propose a bias-corrected matching estimator and test it against the experimental results. They find that their bias-correction improves the estimates of the training program once it's applied to the non-experimental control groups.

Lalonde constructs different subsamples from PSID and CPS_SSA databases, in order to resemble the experimental control group from NSW. The subsamples, denoted respectively by PSID-2, PSID-3, CPS_SSA-2 and CPS_SSA-3. All of them are built drawing from the original sample individuals that meet some eligibility criteria for the NSW program.

Table 1: Characteristics of different groups (standard deviations)

Variable	NSW T	NWS C	PSID-1	PSID-2	PSID-3	CPS-1	CPS-2	CPS-3
Age	25,681	25,05	34,85	36,09	38,26	33,23	28,25	28,03
	(6,69)	(6,59)	(10,44)	(12,08)	(12,89)	(11,05)	(11,7)	(10,79)
Years of School	10,35	10,09	12,12	10,77	10,3	12,03	11,24	10,24
	(1,82)	(1,62)	(3,08)	(3,18)	(3,18)	(2,87)	(2,58)	(2,86)
Black	0,84	0,83	0,25	0,39	0,45	0,07	0,11	0,2
	(0,4)	(0,4)	(0,43)	(0,49)	(0,5)	(0,26)	(0,32)	(0,4)
Hispanic	0,059	0,1	0,03	0,07	0,12	0,07	0,08	0,14
	(0,29)	(0,32)	(0,18)	(0,25)	(0,32)	(0,26)	(0,28)	(0,35)
Married	0,19	0,15	0,87	0,74	0,7	0,71	0,46	0,51
	(0,37)	(0,36)	(0,34)	(0,44)	(0,46)	(0,45)	(0,5)	(0,5)
Dropout	0,71	0,83	0,31	0,49	0,51	0,3	0,45	0,6
	(0,44)	(0,39)	(0,46)	(0,5)	(0,5)	(0,46)	(0,5)	(0,49)
Real Earnings 1974	2096	2107	19429	11027	5566	14016	8728	5619
	(4964)	(5011)	(14263)	(8921)	(5353)	(9176)	(8032)	(3341)
Real Earnings 1975	3066	3027	19063	7569	2610	13651	7397	2467
	(4875)	(5201)	(13597)	(9041)	(5573)	(9270)	(8112)	(3292)
No of obs	185	260	2940	253	128	15992	2369	429

Table 1 shows the averages and standard deviations of each characteristic across different groups. We see that the subsamples of different control groups are quite different from the experimental control group. However, the subsample that most closely resembles a valid comparison group is PSID-3. Lalonde (1986) shows that the results from other non-experimental estimates, (not including any matching method, see Smith & Todd (2003)) yield very poor estimates, even when the PSID-3 subsample is used for comparison, although the accuracy of these estimates improves relative to the other subsamples.

4.2 Estimation with non-experimental data

We calculate two estimates of the treatment effect for each comparison group identified in Lalonde (1986) and compare them to both the mean difference between groups and a propensity score matching estimator.

Our first matching estimator requires a weighting matrix when calculating the norm inside the objective function to be minimized. We calculate our estimate using two possible weighting matrices, the Mahalanobis distance and a rather simple but intuitive one: diagonal matrix which contains the absolute value of simple t-tests resulting from an OLS regression of the outcome on pre-treatment variables.

Our point estimates of the treatment effect vary between 1355 and 2064. The performance of our estimator is comparable to the propensity score matching one. Results from both methods are very similar in terms of standard errors. Remarkably, the propensity score estimate for the CPS-3 subsample is far below the unbiased estimate, being significantly improved by outçr proposed method. Table 2 shows the estimated treatment effect for each one of the comparison groups proposed by Lalonde (1986). The last two columns show the estimates obtained by our method, using both the Mahalanobis weighting matrix and the diagonal with the t-tests. [4]

---

[4]Propensity score matching replicates the results by Dehejia & Wahba (1998). Standard errors calculated by bootstrap method.

Table 2: Matching treatment effect estimate. (Numbers in parentheses are the standard errors)

Control group	NWS less comparison group	P. Score	Mahalanobis	t-test adjusted
NSW	1794			
	(633)			
PSID-1	-15205	1691	1803	1752
	(1154)	(2209)	(1933)	(2003)
PSID-2	-3647	1455	1487	1676
	(959)	(2303)	(2366)	(2152)
PSID-3	1069	2120	2064	1997
	(899)	(2335)	(1431)	(1254)
CPS-1	-8498	1582	1321	1455
	(712)	(1069)	(1154)	(996)
CPS-2	-3822	1788	1966	2012
	(670)	(1205)	(1421)	(1246)
CPS-3	-635	587	1355	1416
	(657)	(1496)	(1677)	(1596)

Table 3: Sample means of (Mahalanobis) matched comparison groups characteristics.

Variable	NWS T	PSID-1	PSID-2	PSID-3	CPS-1	CPS-2	CPS-3
Age	25,81	26,05	27,12	27,26	25,23	28,25	25,03
Years of School	10,35	11,12	10,77	10,3	11,03	11,24	10,24
Black	0,84	0,75	0,89	0,75	0,74	0,81	0,8
Hispanic	0,06	0,03	0,04	0,02	0,07	0,08	0,04
Married	0,19	0,17	0,24	0,15	0,21	0,26	0,2
Dropout	0,71	0,61	0,49	0,51	0,49	0,65	0,61
Real Earnings 1974	2096	2458	1972	2586	2020	1745	2633
Real Earnings 1975	1532	1340	1938	1773	1291	1392	1661

Table 4: Sample means of (t-test adjusted) matched comparison groups characteristics.

Variable	NWS T	PSID-1	PSID-2	PSID-3	CPS-1	CPS-2	CPS-3
Age	25,81	26,46	27,91	26,36	24,67	28,95	26,21
Years of School	10,35	10,55	11,1	10,15	10,96	11,74	10,19
Black	0,84	0,81	0,79	0,71	0,8	0,89	0,77
Hispanic	0,06	0,04	0,04	0,03	0,06	0,08	0,05
Married	0,19	0,2	0,21	0,13	0,24	0,26	0,22
Dropout	0,71	0,63	0,55	0,52	0,42	0,7	0,64
Real Earnings 1974	2096	1167	1521	1782	1813	1694	1993
Real Earnings 1975	1532	1340	1938	1773	1291	1392	1661

Tables 3 and 4 show that the characteristics of the matched comparison groups are very close to those of the treatment group. This suggests that the method is indeed succeeding in "creating" a valid non-experimental control group, despite the big differences between the comparison and treatment groups, as Table 1 shows. This indicates that our matching method successfully replicates a valid counterfactual comparison group.

Table 5: Square roots of mean squared errors of estimates.

Control group	NWS less comparison group	P. Score	Mahalanobis	t-test adjusted
PSID1	17038	2211	1933	2003
PSID2	5525	2328	2386	2155
PSID3	1155	2358	1456	1270
CPS1	10317	1090	1247	1052
CPS2	5656	1205	1431	1265
CPS3	2516	1922	1734	1640

Table 5, presents the square roots of the mean squared errors of the different estimates, assuming the experimental estimate as the true treatment

effect. The table shows that our estimator has lower MSE than propensity score matching with both norm specifications for 3 out of 6 subsamples, and only for the CPS-2 subsample, displays a higher MSE in both norm specifications.

5 Conclusions

We propose an intuitive alternative matching method for estimating the effect of a treatment, which relies on the minimization of the distance between the characteristics of the treatment and control groups. This method is appealing since it avoids some of the arbitrariness of propensity score matching method, such as the choice of a Kernel specification, the nearest neighbor criterion, the specification issues involving the estimation of the propensity in practice and the procedure according to which a common support region is determined. In turn, our method needs the choice of a norm according to which the characteristics are weighted, introducing thus certain non desirable arbitrariness.

The propensity score matching method requires that the estimated probability of treatment assignment between both the treatment and comparison groups overlap sufficiently. Likewise, the proposed estimator relies on the existence of a common support between the comparison and treatment groups, which is precisely a condition we assume to our method. However, propensity score matching relies on the choice of a common support region, and this choice involves an arbitrary decision by the econometrician. Our method solves this problem with an objective and optimization-based criterion.

Our results show that the proposed alternative matching method succeeds

in estimating the treatment effect for the benchmark database of Lalonde's (1986) study. We choose to test this method with two alternative weighting matrices: the Mahalanobis distance and a diagonal matrix containing the t-tests of a simple regression between outcomes and characteristics. The estimator performs comparably to the popular Propensity Score Matching. In terms of mean squared error, our method appears to behave better than Propensity Score

References

Abadie, A. & Imbens, G. W. (2002), Simple and bias-corrected matching estimators for average treatment effects, NBER Technical Working Papers 0283, National Bureau of Economic Research, Inc. available at http://ideas.repec.org/p/nbr/nberte/0283.html.

Dehejia, R. H. & Wahba, S. (1998), Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs, NBER Working Papers 6586, National Bureau of Economic Research, Inc. available at http://ideas.repec.org/p/nbr/nberwo/6586.html.

Heckman, J., Ichimura, H. & Todd, P. (1997), 'Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme', *Review of Economic Studies* **64**, 605–654.

Heckman, J., Ichimura, H. & Todd, P. (1998), 'Matching as an econometric evaluation estimator', *Review of Economic Studies* **65**, 261–294.

Lalonde, R. J. (1986), 'Evaluating econometric evaluations of training programs with experimental data', *The American Economic Review* **76**(4), 604–620.

Rosenbaum, P. & Rubin, D. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* (70), 41–55.

Smith, J. & Todd, P. (2003), Does matching overcome lalonde's critique of nonexperimental estimators?, Technical report.

A Matlab codes

```
    function match=matching(Xt,Xc,Yc,Yt)

global Xtim Xcm I0 I1 testt;

I0=length(Yc); % number of non-treated individuales
I1=length(Yt); % number of treated individuals

Y=[Yt;Yc];
X=[Xt;Xc];

I0=length(Yc);
I1=length(Yt);
Xc=Xc(1:I0,:);
Xt=Xt(1:I1,:);
Yt=(Yt(1:I1));
Yc=(Yc(1:I0));

% Scaling of variables.
Xtm=escal(Xt);
Xcm=escal(Xc);
Ym=escal(Y);
Ycm=escal(Yc);
Ytm=escal(Yt);
Xm=escal(X);



% Xtm=Xt*inv(diag(mean(Xt)));
% Xcm=Xc*inv(diag(mean(Xc)));
% Ym=Y/(mean(Y));
% Ycm=Yc/(mean(Yc));
% Ytm=Yt/(mean(Yt));
% Xm=X*inv(diag(mean(X)));

% OLS regression: t tests will be used in weighting matrix
R=ols(X,Y);
testt=R.t;

% Restrictions at optimization
 Aeq=ones(1,I0);
```

```
 beq=1;
 lb=zeros(I0,1);
 ub=ones(I0,1);
 u0=ones(I0,1)/I0;


options=optimset('MaxFunEvals',100000);
u=[];
time=0;
for i=1:I1;
    tic;
    Xtim=Xtm(i,:);
    [mui,wi]=fmincon(@gener1,u0,[],[],Aeq,beq,lb, ub,[],options);
    u=[u mui];
    toc;
    t=toc;
    fprintf(1,'Completing matching for i=%g .\n',i)
    time=time+toc;
    falta=(I1/i-1)*time;
    if falta>120*2*60
        fprintf(1,'Time left: %g hours.\n',round(falta/360)/10)
    elseif falta>=120 & falta<=120*3600*2
        fprintf(1,'Time left: %g minutes.\n',round(falta/60))
    elseif falta<120
        fprintf(1,'Time left: %g seconds.\n',falta)
    end
end

% COUNTERFACTUALS

counterf=u'*Yc; %

effect=round(sum(Yt-counterf)/I1); % Estimated treatment effect

dif=round(mean(Yt)-mean(Yc)); % Difference estimator

match.testt=testt;
match.u=u;
match.counterf=counterf;
match.Yt=Yt;
match.Yc=Yc;
```

```
match.effect=effect;
match.dif=dif;
match.I0=I0;
match.I1=I1;
match.time=time


now=datestr(fix(clock),30);
save(now,'match','match');


function norma1=norma1(z)

global testt

norma1=sqrt(z'*abs(diag(testt(3:end,:)))*z);


function gener1=gener1(ui)

global Y Yc Yt X Xt Xti Xc Xtm Xtim Xcm I0 I1;

p1=norma1((Xtim(:,2:end)-ui'*Xcm(:,2,end))');
p2=0;
    for j=1:I0;
        p2j=ui(j)*norma1((Xtim(:,2:end)-Xcm(j,2:end))');
        p2=p2j+p2;
    end

gener1=p1+p2;
```