Propensity Score Matching and Genetic Matching : Monte Carlo Results

Donzé, Laurent
University of Fribourg (Switzerland), Department of Quantitative Economics
Bd de Pérolles 90
CH – 1700 Fribourg, Switzerland
E-mail: Laurent.Donze@UniFr.ch

Lai, Peychyn
University of Fribourg (Switzerland), Department of Quantitative Economics
Bd de Pérolles 90
CH – 1700 Fribourg, Switzerland
E-mail: Peychyn.Lai@UniFr.ch

The measure of an impact treatment or the economic evaluation of some political programs appears to raise some non trivial problems, especially when the data are coming from observational studies. Bias adjustments are one of them. In this context, matching methods are generally used. As many others, we chose these methods for example in our study on the impact of the Swiss policy on firm innovation performance [1]. Cautiously, a sensitive analysis was undertaken. Nevertheless, the question of the right matching method stays always open.

Propensity scores matching vs genetic matching

Generally, the matching process is based on different covariates. The idea is to match two units with same characteristics in terms of covariates. Or in other words to find the nearest-neighbour. Following Rosenbaum and Rubin [5], the propensity score matching is become a standard. The great advantage is to operate the matching on the propensity scores only, i.e. the probabilities of assignment to treatment, and not on a vector of covariates. The method is fully theoretically justify but unfortunately has severe limitations. One of these is that the propensity score is generally unknown and has to be estimated. In fact, the propensity score matching is valid only if we know the true propensity score model and we have sufficient observations to estimate them.

One important concept is the "*Equal Percent Bias Reducing*" (EPBR) introduced by Rubin [6]. A matching method is said EPBR for a vector of covariates $\boldsymbol{X}$ if the percent reduction of bias is the same for each of the covariates that are matched and for all linear function of them. In particular, one can prove that the propensity scores matching is EPBR only if the covariates have ellipsoidal distributions such as the normal or the student distributions.

We have also to note that if the EPBR condition is not fulfilled then the bias can be increased even if the univariate covariate means are closer after the matching than before. Therefore applying propensity scores matching may conduct to balance worse across key confounding variables.

Sekhon and Grieve [10] advocate the use of Genetic matching method ([3], [8]). The matching criterion is based on a generalization of the Mahalanobis distance. In fact, a weight matrix is included

in the distance function:

$$(1) \qquad d(\boldsymbol{X}^i, \boldsymbol{X}^j) = \left[(\boldsymbol{X}^i - \boldsymbol{X}^j)'(\boldsymbol{\Sigma}^{-1/2})'\boldsymbol{W}(\boldsymbol{\Sigma}^{-1/2})(\boldsymbol{X}^i - \boldsymbol{X}^j)\right]^{1/2},$$

where $\boldsymbol{W}$ is a $(p \times p)$-matrix; p is the number of covariates in $\boldsymbol{X}$; i and j are two distinct units; $\boldsymbol{\Sigma}^{-1/2}$ is a Cholesky decomposition of $\boldsymbol{\Sigma}$, the variance-covariance matrix of $\boldsymbol{X}$. By default, the matrix $\boldsymbol{W}$ is diagonal with p parameters chosen appropriately. The genetic matching allows the incorporation of the propensity scores in the list of covariates. The issue to choose the free p parameters of $\boldsymbol{W}$ is open. In general, one compares the distribution of the covariates for the two groups of units (treatment and control units) before and after matching. Paired t-tests and non-parametric Kolmogorov-Smirnov tests are applied. The p-values of the test statistics may help to choose in the sense that greater p-values signal better balances. The strategy adopted by the genetic matching method is thus to maximise at each step the smallest p-value (across the covariates). Imbalance is accepted to worsen, as long as the maximum discrepancy is reduced.

Empirical study

In two case studies, Sekhon and Grieve [10] show the superiority of genetic matching over a propensity scores matching or a parametric matching. They emphasize that unlike propensity scores matching, genetic matching does not rely on covariates with ellipsoidal distributions. Genetic matching can reduce biases in case of great imbalances and are not sensitive to model specification. Precisely, they pretend that genetic matching can achieve much better balance than a strict propensity scores matching.

Numerous researches are based on propensity scores matching. This is generally done without considering the distribution of the covariates and / or based on an assumed true model for estimating the propensity scores. The results can be very sensitive to the applied method. It is therefore of great importance to verify the above mentioned results. In our attempt to do this, we adopt the following strategy.

We search first to construct true propensity scores based on a vector of covariates $\boldsymbol{X}$. The covariates are generated according to different ellipsoidal or not-ellipsoidal distributions. A sample is drawn with probabilities given by the propensity scores. Four different matching are then used and the balance of the covariates is tested.

The R software [4] is used to perform the simulation. Particularly, we used the library `sampling` [11] to draw the treatment group and the library `Matching` [9] to operate the matching and verify the balance of the covariates.

Simulation set-up

1. We first generate a population of size $N = 1000$ with four variables according to the distributions given in Table 1. These covariates could mean respectively the income, sex, age and a last characteristic — less present in the population — of a person.

Table 1: Distributions of the variables

Variables	Distribution
X_1	$X_1 \sim \mathrm{N}\,(6000, 1500)$
X_2	$X_2 \sim \mathrm{Bernoulli}\,(0.6)$
X_3	$X_3 \sim \mathrm{N}\,(35, 5)$
X_3	$X_4 \sim \mathrm{Bernoulli}\,(0.1)$

2. Then, we generate for each unit in the population a propensity score according to the following model:

$$ps(\boldsymbol{X}) := \boldsymbol{\Phi}(2.606 - 0.0004032X_1 - 0.5198X_2 + 0.008032X_3 - 0.7896X_4), \tag{2}$$

   where $\boldsymbol{\Phi}$ is the cumulative normal distribution.

3. The next step is the simulation. We fix the number of each simulation to 1000.

4. At each iteration, a treatment group is drawn. Each unit is sampled with a Brewer algorithm with inclusion probabilities given by the propensity scores. The size of the group is equal to the sum of the propensity scores. In our case, the size is of 574 units and therefore the control group has 426 units.

5. The units of the treatment group are then matched. We use four matching methods:

   (a) traditional matching without propensity scores;

   (b) traditional matching with propensity scores;

   (c) genetic matching without propensity scores;

   (d) genetic matching with propensity scores.

6. For each matching method used, we control the balance of the covariates. As in Sekhon and Grieve [10], a paired t-test — difference in the mean of the covariate for the treatment and the control groups — is done for each covariate and the p-value of the test is stored. Remember that high p-values suggest better balance.

7. Finally, at the end of the simulation we compute for each covariate the mean of the p-values. The results are given in table 2.

Results and conclusion

The results are given in table 2. The first covariate, X_1, is suppose issued from a normal distribution. Before matching, the mean of the p-values is close to zero and suggests that there is strong imbalance between the two groups. Traditional matching without propensity scores and genetic

matching without propensity scores don't really improve the balance. On the contrary, the inclusion of the propensity scores in the vector of covariates improve greatly the balance. This is the case for the traditional matching and for the genetic matching. With 0.3178, genetic matching with propensity scores appears the better method.

The third covariate, X_3, comes also from a normal distribution. We see that before matching there is already a great balance (0.3585). Surprisingly, only the traditional matching without propensity scores is able to improve the balance. We note again that genetic matching without performs the worst.

For the second covariate, X_2, a Bernoulli's type, it is the traditional matching without propensity scores which improves the best the balance (0.8467). Again, genetic matching without propensity scores seems to be the worst method.

Finally, for the last covariate, X_4, a Bernoulli's type, traditional matching without propensity scores appears again the best method with a perfect balance (1.0). In this case, the second best is matching without propensity scores. But, we have to notice that the balance isn't better than before matching (0.4880 vs. 0.4907).

In conclusion, our results are very contrasted and don't show the superiority of genetic matching, particularly without propensity scores.

Table 2: p-values means

	X_1	X_2	X_3	X_4
Before Matching	0.0000	0.0001	0.3585	0.4907
Traditional Matching without PS	0.0000	0.8467	0.3731	1.0000
Traditional Matching with PS	0.2847	0.3255	0.3092	0.3320
Genetic Matching without PS	0.0110	0.3105	0.2866	0.4880
Genetic Matching with PS	0.3178	0.3517	0.3201	0.3461

REFERENCES

[1] S. Arvanitis, L. Donzé, and N. Sydow, *Impact of Swiss technology policy on firm innovation performance : an evaluation based on a matching approach*, Science and Public Policy, 37 (2010), pp. 63–78.

[2] P. Lai, *Genetic matching : Eine bessere Alternative zu den bisherigen Matching-Methoden? Eine empirische Studie über die Auswirkung der Projektförderung der Kommission für Technologie und Innovation*, Master's thesis, Department of Quantitative Economics, University of Fribourg, 23. Juni 2010.

[3] W. R. Mebane, Jr. and J. S. Sekhon, *Genetic Optimization Using Derivatives: The rgenoud Package for R*, Journal of Statistical Software, (2009).

[4] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[5] P. R. Rosenbaum and D. B. Rubin, *The central role of the propensity score in observational studies for causal effects*, Biometrika, 70 (1983), pp. 41–55.

[6] D. B. Rubin, *Multivariate Matching Methods That Are Equal Percent Bias Reducing, I: Some Examples*, Biometrics, 32 (1976), pp. 109–120.

[7] J. S. Sekhon, *Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R.* unpublished, 2007.

[8] J. S. Sekhon, *Matching: Multivariate and Propensity Score Matching with Balance Optimization*, 4th May 2011. R package version 4.7-13.

[9] ——, *Multivariate and Propensity Score Matching. Software with Automated Balance Optimization: The Matching package for R*, Journal of Statistical Software, (2011). forthcoming.

[10] J. S. Sekhon and R. Grieve, *A new non-parametric matching method for bias adjustment with applications to economic evaluations.*, Working Paper, (2008), p. 46.

[11] Y. Tillé and A. Matei, *sampling: Functions for drawing and calibrating samples*, 21st January 2011. R package version 2.4.

ABSTRACT

The matching methods have been extensively used to evaluate the impact of a specific treatment in the case as instance of non-experimental studies. In this context, the matching methods based on propensity scores have become a well-known standard. However it appears in many cases that the propensity score matching methods are not really a panacea. Following the work of Sekhon and Grieve [10] we explore by a Monte Carlo study, the genetic matching method in comparison to the traditional methods. Our results are contrasted and don't proof the superiority of the genetic matching method.