



HAL
open science

Matching methods

Pauline Givord

► **To cite this version:**

| Pauline Givord. Matching methods. LIEPP Methods Brief n°18, 2023. hal-04106880

HAL Id: hal-04106880

<https://sciencespo.hal.science/hal-04106880v1>

Submitted on 6 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

LIEPP METHODS BRIEF n°18

2023

This brief is part of a set of methods briefs published by LIEPP. As such, it is supported by the ANR and the State under the "Investissements d'avenir" programme within the framework of the IdEx Université Paris Cité (ANR-18-IDEX-0001).

Matching methods

Pauline GIVORD (INSEE)

pauline.givord@insee.fr



Distributed under a Creative Commons Paternité.
Attribution-NonCommercial-NoDerivatives | 4.0 International License

<https://www.sciencespo.fr/liepp/en.html>

How to cite this publication:
GIVORD, Pauline, **Matching methods**,
LIEPP Methods Brief n°18, 2023-05

Publication initialement rédigée en français :
GIVORD, Pauline, **Méthodes d'appariement**,
LIEPP Fiche méthodologique n°17, 2023-05

IN A NUTSHELL:

Matching is a quantitative method for *ex-post* evaluation in which, in the absence of direct experimentation, a counterfactual situation is reconstructed by comparing the situations of beneficiaries of an intervention with those of non-beneficiaries with very similar characteristics. This method is particularly useful for evaluating the impact of a programme on a whole population, when sufficiently precise data exist to compare beneficiaries and non-beneficiaries.

Keywords: Quantitative method, *ex post* evaluation, causal effect, propensity score, common support

I. What does this method consist of?

Matching methods are among the main quantitative methods for *ex-post* evaluation, aiming to measure the effect of a public policy tool or programme (e.g. a training programme for jobseekers, or localised aids in certain territories) on the situation of the beneficiaries. As with most quantitative evaluation methods, the aim is to estimate the causal effect of the intervention on the situation of the beneficiaries (for example, a return to employment after training, or the economic activity of the targeted territory). The objective of matching methods is to estimate this causal effect by comparing the situation of beneficiaries of the programme with that of people who have not benefited from it, but whose characteristics are so similar that it would have been possible for them to benefit from it. The observation of these non-beneficiaries is supposed to give an idea of the "counterfactual" situation, that which the beneficiaries would have experienced in the absence of the programme.

The challenge here is to reduce the selection effects that can occur when one wishes to estimate the effect of an intervention. In general, the beneficiaries have not been designated by chance, and they have specific characteristics that can explain by themselves a more or less favourable evolution, even in the absence of the programme being evaluated. For example, the evaluation of a training programme aimed at people furthest from employment cannot be done simply by comparing the chances of return to employment of beneficiaries before and after the training, at the risk of underestimating the effect of the programme for the most disadvantaged public. Nor is it possible to compare the return-to-work rates of trainees with those of the non-trained population as a whole: the latter are too different for their employment situation to be a likely reflection of what the trainees would have experienced in the absence of training.

The principle of matching methods is to restrict the comparison of trainees to comparable non-trainees. Specifically, each beneficiary of the programme being evaluated is matched with one or more "twin" non-beneficiaries, in the sense that they have very similar individual characteristics in all dimensions that may influence both benefiting from the programme and their subsequent situation. In the example of the estimation of the training course impact on the chances of returning to employment, we compare for each trainee the chances of having found a job for instance during the year following the entry into training with the same chances of persons identical or at least closest to this trainee at the date of the entry into training in the dimensions considered important for the return to employment. The average effect of training for trainees is obtained by averaging all these comparisons for each beneficiary.

In principle, one wishes to match on as many dimensions as possible, to avoid the risk of missing an important characteristic, whose non-inclusion in the comparisons would lead to incorrect estimates of the causal effect. However, the more dimensions one wishes to match on, the more difficult it will be to find exactly identical non-beneficiaries for each beneficiary in all these dimensions. In the example of the evaluation of a training programme, it may therefore be relevant to match on age, level of education, length of time unemployed and past experience (e.g. number of previous unemployment episodes), past

work experience (e.g. job qualification), type of job sought, possible mobility, which are all variables that may influence both the choice of training and the return to employment (independently of this training). Exact matching on each of these dimensions means that for each vocational trainee one must find a person with exactly the same characteristics in all of these dimensions: the higher the number of variables, the less likely it is to find a perfect "twin", especially if the number of observations is low.

A frequently used response to this limitation is to match not on all these characteristics, but on a summary of them provided by the "propensity score". This corresponds to the probability of being a beneficiary, conditional on the dimensions selected as important for the matching. This means that the estimation is done in two steps. First, the propensity score is estimated, i.e. how the different dimensions predict entry into training, which makes it possible to define an *a priori* probability of being a beneficiary for each observation, depending on its characteristics. In our example, the probability of entering training will be estimated as a function of age, diploma, etc.... This estimate will be used to calculate for each person, whether or not a trainee, his/her "propensity" to enter training, i.e. the probability predicted as a function of these individual characteristics. The values of the propensity score are generally strictly between zero and one (unless a particular exclusion condition is met, it is rare that a person has no chance of entering training, and conversely it is unlikely that any of the characteristics will automatically result in entry into training). Their distributions overlap between beneficiaries and non-beneficiaries. While those who have *a priori* a high probability of entering training are more numerous among those who actually enter training, some do not and can be used for comparison. Conversely, some people with an *a priori* low propensity to enter training may nevertheless choose to train - and it will also be possible to compare them with people who did not train, also having a low propensity to do so. It can be shown that when matching on propensity scores, the important characteristics are on average identical between the beneficiary and non-beneficiary groups.

Whether the matching is done on a single dimension (the propensity score), or on several of them, it is difficult to have exactly identical values for the matching: it is therefore done by using the "closest neighbours" of the beneficiaries, i.e. the non-beneficiaries who are closest to the beneficiary according to the dimensions retained (or according to the propensity score). There are then several variants, notably on the number of neighbours retained (it may be preferable to retain several to avoid comparing by misfortune with a non-beneficiary whose behaviour would be atypical) and on the maximum distance allowed between the beneficiary and the comparisons (neighbours who are too far away being by definition less suitable for comparison).

Whichever matching method is used, it is necessary to have individual data to describe the situation and individual characteristics in detail, and a large number of observations to be more confident of finding close neighbours.

II. How is this method useful for policy evaluation?

Matching methods make it possible to estimate *ex post* the effect of a programme on beneficiaries, on a set of objectively measurable dimensions. For example, they make it possible to answer questions such as: do jobseekers who have chosen to train (at the risk of interrupting a job search) have a higher probability of returning to sustainable employment than jobseekers who do not train? Does this training allow them to expect a higher level of pay? Which jobseekers benefit most from training?

The goal, therefore, is to measure the differences between the situation that was actually experienced by the beneficiaries of a programme and a "counterfactual" situation that would have prevailed in the absence of this programme. In general, these methods are suitable for evaluating the general impact of a programme (compared to a situation where this programme would not exist), but are less suitable for measuring the effect of the different modalities of this programme (in our example, several more or less intensive programmes for training jobseekers).

III. Two examples of application: active employment policies and territorial tax exemptions

Matching methods are very commonly used to evaluate the effects of so-called "active" employment measures (training, job search assistance, etc.), particularly since the methodological study by Heckman, Ichimura and Todd (1997). This method has been used, for example, to study an active employment policy in Sweden (Sianesi, 2004), training programmes in Germany or, more recently, training for job seekers in France (Chabaud et al., 2022).

Another example is the evaluation of the effects of the Zones franches urbaines (ZFU), a public policy tool designed to encourage the establishment of companies in disadvantaged urban areas, similar to the Enterprises Zones set up in the United States in the 1980s. Givord, Rathelot and Sillard (2013) look at the effects of these exemptions on the establishment of businesses and the evolution of employment in the targeted neighbourhoods, compared with other neighbourhoods that were initially very close (see also Malgouyres and Py, 2016). These studies suggest a positive effect of the zones on employment and economic activity, but at the expense of the immediately neighbouring zones. Another study also suggested that the effects were not persistent beyond the duration of the exemptions (Givord et al., 2022).

IV. What are the criteria for judging the quality of the mobilisation of this method?

The validity of matching methods depends crucially on how well they can be corrected for selection effects, and therefore on the information available to compare beneficiaries and non-beneficiaries. There must be some assurance that the selection process in the intervention is not based on variables that are not available in the data (e.g. the results of a motivational interview used to enter a training programme, which would aim to measure dimensions that are not very objective and therefore not available to an outside eye). Having individual information on the variable of interest in the past (e.g. the professional trajectory prior to entering the training programme) is generally considered indispensable to avoid capturing selection effects: matching methods are in this case combined with "difference-in-differences" (see separate brief on difference-in-differences).

Secondly, the method requires the possibility of matching all beneficiaries with non-beneficiaries (this is called "common support"). This last condition means in particular that there is a certain amount of randomness in the fact of benefiting from the programme: if the programme is totally deterministic in terms of observable characteristics (for example, a programme systematically offered to young people without diplomas, which would exclude people above a certain age or income threshold), it will not be possible to match the beneficiaries to non-beneficiaries on these dimensions.

Finally, matching methods provide a statistical estimate, and therefore as such do not allow the "true" effect value to be measured with complete certainty, but only an approximation whose precision, i.e. the degree of confidence with which this estimate can be used, can be quantified. This precision can be measured by means of the standard deviation (the smaller the standard deviation, the greater the confidence that the "true" effect is close to the estimated value) or by means of a confidence interval, which corresponds to the interval of values within which the true effect is found with a given probability: for example, the interval of values within which the true value of the effect is found with a probability of 95% (the smaller the confidence interval, the greater the precision of the estimated value). This measure of precision is used, for example, to check that the effect of the intervention being evaluated is "significant" or "significantly different from zero", i.e. it can be said with some confidence that the programme does indeed have a strictly positive or strictly negative effect.

V. What are the strengths and limitations of this method compared to others?

One of the strengths of matching methods is that they can estimate effects in the "general population", i.e. on the whole population (provided that there are enough observations to be able to find comparisons and that the assignment to the programme is sufficiently random to allow for the availability of beneficiaries on the whole). This can be an advantage over most ex-post quantitative evaluation methods, which only allow an unbiased estimate of a causal effect on 'marginal' populations: for example, people around an eligibility threshold for discontinuity regressions (see separate brief on discontinuity regressions), or people who are sensitive to the signal given by an instrument.

On the other hand, matching methods may not be sufficient to correct for selection bias. Estimates are very sensitive to the choice of variables used for matching, and it is generally difficult to trust estimators in the absence of past individual measurements of the variable of interest.

Some bibliographical references to go further

Biewen, Martin. and Fitzenberger, Bernd. and Osikominu, Aderonke. and Paul, Marie. 2014. « The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices ». *Journal of Labor Economics*, 32 : 837-897.

Fitzenberger, Bernd. and Völter, Robert. 2007. « Long-run effects of training programs for the unemployed in East Germany ». *Labour Economics*. 14 (4) : 730-755.

Givord, Pauline. and Rathelot, Roland. and Sillard, Patrick. 2013. « Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines program ». *Regional Science and Urban Economics*, 43(1) : 151-163.

Givord, Pauline. and Quantin, Simon. and Trevien, Corentin, 2018. "A long-term evaluation of the first generation of French urban enterprise zones," *Journal of Urban Economics*, n° 105(C) : 149-161.

Heckman, James. and Hidehiko, Ichimura. and Petra, Todd. 1997. « Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme ». *Review of Economic Studies*, 64(4): 605-654.

Lechner, Martin 2002. « Program Heterogeneity And Propensity Score Matching: An Application To The Evaluation of Active Labor Market Policies ». *The Review of Economics and Statistics*, vol. 84, n° 2 : 205-220.

Malgouyres, Clément. and Py, Loriane. 2016. « Les dispositifs d'exonérations géographiquement ciblées bénéficient-ils aux résidents de ces zones? État des lieux de la littérature américaine et française ». *Revue économique*. 67 : 581-614.

Sianesi, Barbara. 2004. « An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s ». *Review of Economics and Statistics*, 86 : 133-155.