



HAL
open science

Méthodes d'appariement

Pauline Givord

► **To cite this version:**

| Pauline Givord. Méthodes d'appariement. LIEPP Fiche méthodologique n°17, 2023. hal-04106562

HAL Id: hal-04106562

<https://sciencespo.hal.science/hal-04106562v1>

Submitted on 25 May 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 FICHE MÉTHODOLOGIQUE n°17

2023

Cette fiche fait partie d'un ensemble de fiches méthodologiques publiées par le LIEPP. A ce titre, elle bénéficie du soutien apporté par l'ANR et l'État au titre du programme d'Investissements d'avenir dans le cadre de l'IdEx Université Paris Cité (ANR-18-IDEX-0001).

Méthodes d'appariement

Pauline GIVORD (INSEE)

pauline.givord@insee.fr



Partage selon les Conditions Initiales 4.0 International License

www.sciencespo.fr/liepp

Comment citer cette publication :
GIVORD, Pauline, **Méthodes d'appariement**,
LIEPP Fiche méthodologique n°17, 2023-05

This publication is also available in English:
GIVORD, Pauline, **Matching methods**,
LIEPP Methods Brief n°18, 2023-05

EN BREF :

L'appariement est une méthode quantitative d'évaluation *ex post* dans laquelle, en l'absence d'expérimentation directe, on reconstitue une situation contrefactuelle en comparant les situations de bénéficiaires d'une intervention à celles de non bénéficiaires aux caractéristiques très proches. Cette méthode est notamment utile pour évaluer l'impact d'un programme sur l'ensemble d'une population, lorsqu'il existe des données suffisamment précises pour comparer les personnes bénéficiaires et non bénéficiaires.

Mots-clés : Méthode quantitative, évaluation *ex post*, effet propre, score de propension, support commun

I. En quoi consiste cette méthode ?

Les méthodes d'appariement font partie des principales méthodes d'évaluation quantitative *ex post*, visant à mesurer l'effet d'un dispositif ou d'un programme (par exemple un programme de formation pour les demandeurs et demandeuses d'emploi, ou des aides localisées sur certains territoire) sur la situation des bénéficiaires. Comme la plupart des méthodes d'évaluation quantitatives, l'enjeu est d'estimer l'effet propre (*causal*) du dispositif sur la situation des bénéficiaires (par exemple, un retour à l'emploi après une formation, ou l'activité économique du territoire visé). L'objectif des méthodes d'appariement (*matching*) est d'estimer cet effet propre du dispositif en comparant la situation des bénéficiaires du dispositif avec celle de personnes qui n'en ont pas bénéficié, mais qui ont des caractéristiques si proches qu'il aurait été possible qu'ils en bénéficient. L'observation de ces non bénéficiaires est supposée donner une idée de la situation « contrefactuelle », celle qu'aurait connue les bénéficiaires en l'absence du dispositif.

L'enjeu est ici de réduire les effets de sélection qui peuvent se produire quand on souhaite estimer l'effet d'un dispositif. En général, les bénéficiaires du dispositif n'ont pas été désigné-e-s par hasard, et ils et elles ont des caractéristiques spécifiques qui peuvent expliquer par elles-mêmes une évolution plus ou moins favorable, même en l'absence du dispositif évalué. Par exemple, l'évaluation d'un programme de formation à destination des personnes les plus éloignées de l'emploi ne peut se faire simplement en comparant les chances de retour à l'emploi des personnes formées avant ou après la formation, au risque de sous-estimer l'effet de la formation qui ciblerait des personnes ayant plus de difficultés. Il n'est pas non plus possible de comparer les taux de retour à l'emploi des personnes formées avec celui de l'ensemble des personnes non formées : ces dernières sont trop différentes pour que leur situation vis-à-vis de l'emploi soit un reflet probable de la situation qu'aurait connue les stagiaires en l'absence de formation.

Le principe des méthodes d'appariement est de restreindre la comparaison des bénéficiaires aux non bénéficiaires comparables. Précisément, on apparie chaque bénéficiaire du programme ou dispositif à évaluer avec un-e ou des non bénéficiaires « jumeaux » ou « jumelles », au sens où ils et elles ont des caractéristiques individuelles très proches dans toutes les dimensions pouvant influencer à la fois sur le fait de bénéficier du dispositif et sur la situation ultérieure. Dans l'exemple de l'estimation de l'impact du stage de formation sur les chances de retour à l'emploi, on compare pour chaque stagiaire par exemple le fait d'avoir retrouvé un emploi dans l'année suivant l'entrée en formation avec les chances de retour à l'emploi de personnes identiques ou tout au moins les plus proches à ce ou cette stagiaire à la date de l'entrée en formation dans les dimensions considérées comme importantes pour le retour à l'emploi. L'effet moyen de la formation pour les stagiaires est obtenu en faisant la moyenne de l'ensemble de ces comparaisons menées pour chaque bénéficiaire.

En principe, on souhaite appairer sur le plus grand nombre de dimensions possibles, pour se prémunir du risque de manquer une caractéristique importante, et dont la non prise en compte dans les comparaisons conduirait à des estimations incorrectes de l'effet propre du dispositif. Cependant, plus les dimensions sur lesquelles on souhaite appairer sont nombreuses et plus il sera difficile de trouver pour chaque bénéficiaire des non bénéficiaires exactement identiques dans toutes ces dimensions. Dans l'exemple de l'évaluation d'un programme de formation, il pourra ainsi être pertinent d'appairer sur l'âge, le niveau de diplôme, l'ancienneté au chômage et l'expérience passée (par exemple le nombre d'épisodes de chômage antérieur), l'expérience

professionnelle passée (par exemple la qualification de l'emploi), le type d'emploi recherché, la mobilité possible, qui sont autant de variables qui peuvent influencer tant sur le choix de se former que sur le retour à l'emploi (indépendamment de cette formation). Faire un appariement exact sur chacune de ces dimensions signifie qu'on doit trouver pour chaque stagiaire de la formation professionnelle une personne ayant des caractéristiques exactement identiques dans l'ensemble de ces dimensions : plus le nombre de variables est élevé, et moins il est probable, surtout si le nombre d'observations est faible, de trouver un « jumeau » parfait ou une « jumelle » parfaite.

Une réponse fréquemment utilisée est d'apparier non pas sur l'ensemble de ces caractéristiques, mais sur un résumé de celles-ci fourni par le « score de propension ». Celui-ci correspond à la probabilité d'être bénéficiaire, conditionnellement aux dimensions retenues comme importantes pour l'appariement. Cela signifie que l'estimation se fait en deux étapes. Dans un premier temps, on estime ce score de propension, c'est-à-dire comment les différentes dimensions prédisent l'entrée en formation, ce qui permet pour chaque observation de définir une probabilité *a priori* d'être bénéficiaire, en fonction de ses caractéristiques. Dans notre exemple, on estimera la probabilité d'entrer en formation comme une fonction de l'âge, du diplôme, *etc...* On utilisera cette estimation pour calculer pour chaque personne, stagiaire de la formation ou non, sa « propension » à entrer en formation, c'est-à-dire la probabilité prédite en fonction de ces caractéristiques individuelles. Les valeurs du score de propension se situent en général strictement entre zéro et un (sauf condition d'exclusion particulière, il est rare qu'une personne n'ait aucune chance d'entrée en formation, et inversement il est peu vraisemblable qu'une des caractéristiques utilisées se traduise automatiquement par une entrée en formation). Leurs répartitions se recouvrent entre les bénéficiaires et les non bénéficiaires. Si les personnes qui ont *a priori* une probabilité élevée d'entrer en formation sont plus nombreuses parmi les personnes effectivement entrées en formation, certaines ne le font pas et peuvent servir de comparaison. Inversement, certaines personnes ayant *a priori* une propension faible à entrer en formation peuvent néanmoins choisir de se former – et il sera aussi possible de les comparer avec des personnes qui ne se sont pas formées, ayant également une faible propension à le faire. On peut montrer que lorsque l'on utilise un appariement sur le score de propension, les caractéristiques importantes sont en moyenne identiques entre le groupe des bénéficiaires et les non bénéficiaires.

Que l'appariement se fasse sur une seule dimension (le score de propension), ou sur plusieurs d'entre elles, il est difficile d'avoir des valeurs exactement identiques pour l'appariement : celui-ci se fait donc en utilisant les « plus proches voisin-e-s » des bénéficiaires, c'est-à-dire les non bénéficiaires qui se rapprochent le plus du ou de la bénéficiaire selon les dimensions retenues (ou selon le score de propension). Il existe ensuite plusieurs variantes notamment sur le nombre de voisin-e-s qu'on retient (on peut préférer en retenir plusieurs pour éviter de comparer par malchance avec un-e non bénéficiaire dont le comportement serait atypique) et la distance maximale qu'on autorise entre le ou la bénéficiaire et les comparaisons (des voisin-e-s trop éloigné-e-s étant par définition moins adapté-e-s pour la comparaison).

Quelle que soit la méthode d'appariement utilisée, il est nécessaire pour l'appliquer de disposer de données individuelles permettant de décrire finement la situation et les caractéristiques individuelles, et de très nombreuses observations pour avoir plus d'assurance de pouvoir trouver des voisin-e-s proches.

II. En quoi cette méthode est-elle utile pour l'évaluation des politiques publiques ?

Les méthodes d'appariement permettent d'estimer *ex post* l'effet d'un programme sur les bénéficiaires, sur un ensemble de dimensions objectivables et mesurables. Elles permettent par exemple de répondre à des questions du type : des demandeurs d'emploi qui ont choisi de se former (au risque d'interrompre une recherche d'emploi) ont-ils une probabilité de retour à l'emploi durable *in fine* plus élevée que des demandeurs d'emploi qui ne se forment pas ? Cette formation leur permet-elle d'espérer un niveau de rémunération plus élevée ? Quels demandeurs d'emploi bénéficient le plus de la formation ?

Il s'agit donc de mesurer des écarts entre la situation qui a été effectivement vécue par les bénéficiaires d'un programme, et une situation « contrefactuelle », qui aurait prévalu en l'absence de ce programme. En général, ces méthodes sont adaptées pour évaluer l'impact de la mise en œuvre d'un programme (par rapport à une

situation où ce programme n'existerait pas), mais le sont moins pour mesurer l'effet des différentes modalités de cette mise en œuvre (dans notre exemple, plusieurs dispositifs plus ou moins intensifs de formation des demandeurs d'emploi).

III. Deux exemples d'application : politiques actives de l'emploi et exonérations fiscales territoriales

Les méthodes d'appariement sont très classiquement utilisées pour évaluer les effets des mesures dites « actives » de l'emploi (formation, aides à la recherche d'emploi, etc.), notamment depuis l'étude méthodologique d'Heckman, Ichimura et Todd (1997). Cette méthode a été utilisée par exemple pour étudier une politique d'emploi active en Suède (Sianesi, 2004), des programmes de formation en Allemagne ou plus récemment les formations à destination des personnes en recherche d'emploi en France (Chaubaud et al., 2022).

Un autre exemple est l'évaluation des effets des Zones franches urbaines (ZFU), dispositifs d'exonérations fiscales et sociales prévues pour favoriser l'implantation des entreprises dans des zones urbaines défavorisées, à l'image des Entreprises Zones mises en place aux Etats-Unis à partir des années 1980. Givord, Rathelot et Sillard (2013) s'intéressent aux effets de ces exonérations sur l'installation des entreprises et l'évolution de l'emploi dans les quartiers ciblés, en comparaison avec d'autres quartiers initialement très proches (voir aussi Malgouyres et Py, 2016). Ces études suggèrent un effet positif des zones sur l'emploi et l'activité économique, mais au détriment des zones immédiatement voisines. Une autre étude suggérait par ailleurs des effets non persistants au-delà de la durée des exemptions (Givord et al., 2022).

IV. Quels sont les critères permettant de juger de la qualité de la mobilisation de cette méthode ?

La validité des méthodes d'appariement dépend de manière cruciale de la manière dont elles peuvent corriger des effets de sélection, et donc des informations disponibles pour comparer les bénéficiaires et les non bénéficiaires. Il faut avoir une assurance que le processus de sélection dans le dispositif ne se fasse pas en fonction de variables qui ne sont pas disponibles dans les données (par exemple, les résultats d'un entretien de motivation utilisé pour entrer dans un programme de formation, qui viserait à mesurer des dimensions peu objectivables et donc non disponibles pour un regard extérieur). Le fait de disposer d'informations individuelles sur la variable d'intérêt sur le passé (par exemple, la trajectoire professionnelle antérieure à l'entrée en formation) est en général considéré comme indispensable pour éviter de capter des effets de sélection : les méthodes d'appariement sont dans ce cas combinées avec des « différences de différences » (voir fiche séparée).

Ensuite, il est nécessaire qu'il soit possible d'apparier l'ensemble des bénéficiaires avec des non bénéficiaires (on parle de « support commun »). Cette dernière condition signifie notamment qu'il existe un certain aléa dans le fait de bénéficier du programme : si celui est totalement déterministe en fonction des caractéristiques observables (par exemple un programme systématiquement proposé aux jeunes non diplômés, qui excluraient en revanche les personnes au-delà d'un seuil d'âge ou de revenu), il ne sera pas possible d'apparier les bénéficiaires.

Enfin, les méthodes d'appariement fournissent une estimation statistique, et donc comme telles ne permettent pas de mesurer avec une totale certitude la « vraie » valeur de l'effet mais seulement une approximation dont on peut quantifier la précision, c'est-à-dire le degré de confiance avec laquelle on peut utiliser cette estimation. Cette précision se mesure via l'écart-type (plus celui-ci est faible, et plus il est possible d'avoir confiance dans le fait que le « vrai » effet est proche de la valeur estimée) ou encore un intervalle de confiance, qui correspond à l'intervalle de valeurs au sein duquel le vrai effet se trouve avec une probabilité déterminée : par exemple, l'intervalle de valeur où la vraie valeur de l'effet se trouve avec une probabilité de 95% (plus l'intervalle de confiance est réduit, et plus la valeur estimée est connue avec précision). Cette mesure de précision est notamment utilisée pour vérifier que l'effet du dispositif évalué est « significatif » ou « significativement différent

de zéro », c'est-à-dire qu'on peut dire avec une certaine assurance que le programme a effectivement un effet strictement positif, ou strictement négatif.

V. Quels sont les atouts et les limites de cette méthode par rapport à d'autres ?

L'un des atouts des méthodes d'appariement est de pouvoir estimer les effets en « population générale », c'est-à-dire sur l'ensemble de la population (à condition de disposer de suffisamment d'observations pour pouvoir trouver des comparaisons et que l'assignation au dispositif comporte suffisamment d'aléas pour que l'on puisse disposer de bénéficiaires sur l'ensemble). Cela peut constituer un atout par rapport à la plupart des méthodes d'évaluation quantitative *ex post* ne permettent d'estimer sans biais un effet causal que sur des populations « marginales » : par exemple, les personnes autour d'un seuil d'éligibilité pour les régressions sur discontinuités (voir fiche séparée), ou les personnes qui sont sensibles au signal donné par un instrument.

En revanche, les méthodes d'appariement peuvent ne pas être suffisantes pour corriger des biais de sélection. Les estimations sont très sensibles au choix des variables utilisées pour l'appariement, et il est en général difficile de faire confiance à des estimateurs en l'absence de mesures individuelles passées de la variables d'intérêt.

Quelques références bibliographiques pour aller plus loin

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

Givord, Pauline. et Rathelot, Roland. et 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. et Quantin, Simon. et 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. et Hidehiko, Ichimura. et 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. et 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.