Commentaire

“Expérimentation contrôlée contre appariement”

Sylvain Chabé-Ferret

Dans Économie & prévision 2014/1-2 n° 204-205 , pages 183 à 191
Éditions Direction générale du Trésor

ISSN 0249-4744

DOI 10.3917/ecop.204.0183

Date de mise en ligne : 14/01/2016

Article disponible en ligne à l’adresse

https://shs.cairn.info/revue-economie-et-prevision-2014-1-page-183?lang=fr

CAIRN.INFO

Découvrir le sommaire de ce numéro, suivre la revue par email, s’abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour Direction générale du Trésor.
Vous avez l’autorisation de reproduire cet article dans les limites des conditions d’utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.
Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l’autorisation de l’Éditeur ou, le cas échéant, de l’organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l’organisme agréé en la matière.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Commentaire

"Expérimentation contrôlée contre appariement"

Sylvain Chabé-Ferret(*)

Je suis enchanté d'avoir été invité à faire un commentaire sur l'article de Marianne Bléhaut et Roland Rathelot (B&R dans la suite). Je pense qu'il s'agit d'une contribution extrêmement importante au débat méthodologique sur l'évaluation des politiques publiques en France. Je suis convaincu que ces résultats peuvent aussi avoir une répercussion internationale. Ils viennent en effet se positionner dans une littérature où les résultats les plus récents sont controversés et où la grande majorité des travaux s'appuie sur les données issues d'un seul et même programme.

Dans cet article, B&R montrent que les méthodes d'appariement (ou de *matching*) sont incapables de reproduire les résultats d'une expérimentation randomisée. C'est un résultat extrêmement important car l'appariement est une méthode d'évaluation simple et quasiment toujours applicable. Dans ce commentaire, je discute la portée des résultats présentés par B&R et les conséquences que l'on peut en tirer pour l'utilisation de l'appariement. Dans une première partie, je décris la question fondamentale à aborder : les variables observées suffisent-elles à éliminer le biais de sélection ? Je détaille ensuite pourquoi, en toute rigueur, les résultats de B&R ne permettent pas d'y répondre. En effet, la présence d'attrition et d'effets de déplacement peut aussi expliquer ces résultats. Dans une troisième partie, je précise que les résultats présentés par B&R démontrent néanmoins de manière indiscutable que les variables observées sont insuffisantes pour neutraliser simultanément sélection, attrition et effets de déplacement. Ainsi, même s'ils n'en apportent pas la preuve définitive, les résultats de B&R jettent un doute sérieux sur la capacité de l'appariement à résoudre le problème de biais de sélection. Il est en effet difficile d'imaginer que le biais de sélection serait le seul à pouvoir être capté entièrement par les variables observées. Ce doute est encore renforcé à la lumière des résultats de travaux similaires, dont je résume les principaux enseignements dans la quatrième partie. Certains de ces travaux ne sont en effet pas affectés par des problèmes d'attrition et/ou d'effets de déplacement, et parviennent généralement à des résultats extrêmement proches de ceux de B&R. Dans la dernière partie de ce commentaire, je tire les conséquences de ces résultats pour l'utilisation de l'appariement pour l'évaluation des politiques publiques, notamment de l'emploi. La première de ces conclusions est qu'il est primordial de promouvoir et de développer des évaluations randomisées dès lors que c'est possible.

(*) *Toulouse School of Economics* et Inra (Lerna).
E-mail : sylvain.chabe-ferret@tse-fr.eu

Je remercie la rédactrice en chef, Hélène Erkel-Rousse, de m'avoir donné l'opportunité d'écrire ce commentaire. Je remercie en outre les auteurs pour les échanges passionnants que nous avons eus autour de leur travail.

Ce commentaire n'engage que son auteur et non les institutions auxquelles il appartient. Il n'engage *a fortiori* ni la Direction générale du Trésor, ni le ministère des Finances et des Comptes publics, ni le ministère de l'Économie, de l'Industrie et du Numérique.

La question centrale : les variables observées suffisent-elles à éliminer le biais de sélection ?

L'appariement est une méthode économétrique non-expérimentale relativement simple à mettre en œuvre. Il consiste à comparer chaque bénéficiaire d'une mesure (ou traitement) à un ou plusieurs non bénéficiaires "jumeaux" ayant des caractéristiques observées similaires (Givord, 2014). Cette méthode permet d'obtenir des résultats dans la majorité des cas où elle est employée[1]. Contrairement à d'autres méthodes non expérimentales, elle ne requiert l'existence, ni d'une période d'observation avant la mise en place de la politique, ni d'une source d'exogénéité dans la mise en œuvre de la politique publique[2]. Elle a donc un fort attrait dans la pratique. En conséquence, la question de sa validité est centrale.

Pour examiner la validité de l'appariement, il faut *déterminer si les variables observées sont suffisantes pour éliminer le biais de sélection*. Le biais de sélection est le problème majeur qui se pose lorsque l'on cherche à estimer l'effet d'une politique publique en comparant le résultat des bénéficiaires à celui des non bénéficiaires. Dans l'exemple étudié par B&R, le programme d'accompagnement semble avoir été choisi par des individus plus éloignés de l'emploi que la moyenne. Les jeunes les plus proches de l'emploi ont pu considérer qu'ils n'avaient pas besoin de l'aide apportée par le programme pour parvenir à retrouver un emploi, voire que participer au programme pourrait les retarder dans leur recherche d'emploi. En conséquence, la comparaison du taux d'emploi des bénéficiaires avec celui des non bénéficiaires sous-estime l'effet réel du programme d'accompagnement. Comme c'est le processus d'auto-sélection des bénéficiaires qui crée une corrélation entre la participation au programme et la probabilité initiale de retrouver un emploi, on parle de *biais de sélection*.

La question méthodologique majeure abordée par B&R est la suivante : est-il suffisant de comparer bénéficiaires et non bénéficiaires ayant les mêmes caractéristiques observées pour éliminer le biais de sélection ? Autrement dit, l'éloignement initial de l'emploi est-il entièrement capté par le niveau de diplôme, la durée de la dernière période de chômage, le type d'emploi recherché, l'âge, le sexe, etc. ? Si c'est le cas, on dit que la sélection est "*sur les observées*". Dans le cas contraire où des variables inobservées déterminent simultanément la participation au programme et la probabilité initiale de retrouver un emploi, on parle de "*sélection sur les inobservées*".

B&R ne peuvent rigoureusement répondre à cette question

Pour évaluer les performances de l'appariement, B&R testent sa capacité à reproduire les résultats d'une expérience randomisée. Ils importent, à ma connaissance pour la première fois en France, la méthode initiée par LaLonde (1986) consistant à confronter les méthodes non expérimentales à la référence que constitue une expérience randomisée. Cette approche permet-elle de répondre à la question posée ? En théorie, la réponse est oui. En effet, la randomisation permet de répartir le bénéfice de la politique de manière indépendante des caractéristiques initiales des individus et donc, par exemple, de leur proximité initiale de l'emploi. L'effet estimé par la méthode expérimentale n'est affecté d'aucun biais de sélection et offre donc une excellente référence à laquelle comparer la méthode non expérimentale[3].

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Est-ce à dire que les résultats extrêmement négatifs de B&R (l'appariement est incapable de détecter l'effet positif du programme d'accompagnement sur le taux d'emploi des jeunes) permettent de conclure que l'appariement ne parvient pas à éliminer le biais de sélection ? En toute rigueur, ce n'est malheureusement pas le cas. En effet, l'expérience qu'utilisent B&R souffre de deux problèmes qui viennent compliquer la comparaison entre expérimentation randomisée et appariement : l'attrition et les effets de déplacement. On parle d'attrition lorsque certains individus disparaissent de l'échantillon après la randomisation. Si ce phénomène affecte différemment groupe de traitement et groupe de contrôle, les deux groupes ne sont plus parfaitement comparables. On parle d'effets de déplacement lorsque les individus du groupe de contrôle sont affectés indirectement par le traitement reçu par les individus du groupe traité. Par exemple, les individus ayant bénéficié du programme de suivi peuvent occuper des emplois qui auraient été occupés par des non bénéficiaires en l'absence du programme. On souhaiterait pouvoir interpréter la différence entre l'estimateur expérimental et l'appariement comme due uniquement au phénomène de sélection sur les inobservées. Cependant, elle peut tout aussi bien être due au fait que l'attrition et les effets de déplacement affectent différemment l'appariement et l'estimateur expérimental.

Par exemple, on peut tout-à-fait imaginer que les seuls individus subissant les effets de déplacement sont ceux qui auraient souhaité bénéficier de l'accompagnement mais qui, n'ayant pas eu la chance d'être tirés au sort, ne se le sont pas vu proposer. Il est aussi possible que les individus qui ne souhaitent pas participer au programme ne subissent aucun effet de déplacement. Supposons maintenant que les variables observées permettent d'éliminer le biais de sélection et ignorons l'attrition, par souci de simplicité. Sous ces hypothèses, l'appariement, en comparant les bénéficiaires à ceux qui ont refusé de participer au programme, estime sans biais l'effet du programme sur les bénéficiaires. Par contre, l'estimateur expérimental surestime cet effet, puisqu'il lui ajoute l'effet de déplacement. Les deux approches sont en désaccord alors que le biais de sélection a été éliminé par l'appariement, qui est donc valide dans ce cas.

La question à laquelle répondent B&R : est-ce que les variables observées permettent de neutraliser biais de sélection et d'attrition et effets de déplacement ?

Que peut-on alors conclure des résultats de B&R ? À mon sens, leur contribution majeure est de répondre à la question suivante : est-ce que les variables observées sont suffisantes pour neutraliser ces trois phénomènes (sélection, attrition, effets de déplacement) ? Les résultats de B&R permettent de répondre à cette question car, si la réponse est positive, appariement et approche expérimentale devraient aboutir au même résultat[4]. C'est par ailleurs une question centrale. Si la réponse est positive, le travail de l'évaluateur et du chercheur est grandement simplifié. Sinon, il est important de penser à des solutions pour corriger les biais induits par des variables qui sont pour le moment inobservées.

Les résultats de B&R sont sans appel : il est impossible de résoudre ces trois problèmes en utilisant uniquement les variables observées dont ils disposent. C'est une conclusion frappante. Ce résultat doit inciter tous les chercheurs en sciences sociales à la modestie et nous rappeler combien mal nous comprenons les déterminants des comportements individuels. Jim Heckman avait déjà rappelé dans son discours Nobel que les R^2 des régressions sur données microéconomiques sont très faibles. Des variables inobservées expliquent en conséquence une large portion de la variance de variables comme l'emploi ou le salaire (Heckman, 2001). En d'autres termes, notre ignorance est grande et l'hétérogénéité des comportements individuels extrêmement forte. De plus, cette hétérogénéité interagit avec le choix de bénéficier des politiques évaluées et est responsable de biais.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Finalement, les résultats de B&R enterrent-ils définitivement l'appariement ? Pour le savoir, il faudrait pouvoir déterminer quelle est la cause prépondérante de l'échec de l'appariement parmi les trois possibles. En effet, si le biais de sélection est éliminé par les variables observées et que seule l'attrition et/ou les effets de déplacement sont responsables du biais, l'appariement reste toujours extrêmement intéressant. Il suffirait en effet de l'appliquer au cas où l'attrition est inexistante et où il n'y a aucun effet de déplacement. Est-il possible de progresser sur ces points avec les données utilisées par B&R ? Examinons tout d'abord le rôle de l'attrition. B&R montrent que l'attrition n'est pas corrélée avec les variables observées. Toutefois, cela n'implique pas que l'attrition n'est pas corrélée avec les variables inobservées. Ainsi, les différences significatives de durée de la dernière période de chômage entre les deux échantillons expérimentaux semblent suggérer un possible problème de biais d'attrition. Une manière de résoudre ce problème serait de remplacer l'estimateur expérimental par les bornes robustes à l'attrition suggérées par Horowitz et Manski (2000) et Lee (2009). Si l'effet estimé par l'appariement se situe en dehors de ces bornes, il ne resterait comme explication concurrente au biais de sélection que les effets de déplacements. Le rôle des effets de déplacement est plus délicat à étudier, car la recherche est extrêmement peu développée sur ce point[5].

Au bilan, les résultats de B&R permettent de rejeter avec certitude le fait que les variables observées dont ils disposent suffisent à rendre compte des trois phénomènes (sélection, attrition, déplacement). Il semble, mais c'est ici une déduction sans preuve formelle, qu'il soit difficile d'imaginer que le biais de sélection serait le seul à pouvoir être capté entièrement par des variables observées. Donc, même s'ils n'en apportent pas la preuve définitive, les résultats des auteurs jettent un doute sérieux sur la capacité de l'appariement à résoudre le problème de biais de sélection. Lorsqu'on les lit à la lumière des résultats de travaux similaires dans la littérature, ce doute est encore renforcé.

La position des résultats de B&R dans le reste de la littérature

B&R ne sont pas les premiers à tester la capacité de l'appariement à reproduire les résultats d'une expérience randomisée. J'ai connaissance de sept études similaires dans le contexte de l'évaluation de politiques de l'emploi[6]. Cinq de ces études utilisent la même expérimentation, la *National Supported Work demonstration* (NSW) et peu ou prou le même échantillon, de taille modeste (de 300 à 600 traités). B&R apportent donc une nouvelle base de données à l'étude de ce problème, et elle est la plus fournie en nombre de traités concernés. Hormis deux études utilisant des données de salaire issues des registres de la sécurité sociale[7], les autres études font face à des niveaux d'attrition comparables à B&R. L'attrition semble toutefois être un problème mineur dans ces études. En effet, les résultats obtenus avec des données administratives sont qualitativement similaires à ceux obtenus avec des données d'enquête souffrant d'attrition. Par ailleurs, dans le cas de l'expérimentation NSW, les effets de déplacements sont, semble-t-il, inexistants.

Le bilan de ces travaux est à mon sens extrêmement similaire aux résultats présentés par B&R. L'appariement ou sa version paramétrique utilisant l'estimateur des moindres carrés ordinaires (MCO) ne parvient généralement pas à reproduire les résultats expérimentaux, notamment dans le cas de programmes concernant des jeunes travailleurs[8]. Ces résultats ont néanmoins fait récemment l'objet d'un débat important sur les performances de l'appariement utilisant le score de propension. Dans une série d'articles extrêmement influents, Dehejia et Wahba (1999, 2002) montrent que l'appariement sur le score de propension permet de reproduire les résultats expérimentaux sur l'échantillon des jeunes du NSW. Smith et Todd (2005a), en réexaminant ces résultats, concluent que c'est surtout le choix de l'échantillon qui permet de supprimer tout biais de sélection sur les inobservées. Ce débat s'est poursuivi[9] sans qu'aucune conclusion définitive n'en émerge. À ma connaissance, B&R présentent la première contribution à cette littérature depuis cette controverse. Leurs résultats, à mon sens, confirment la majorité des résultats

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

précédents : l'appariement ne parvient généralement pas à reproduire les résultats expérimentaux, notamment dans le cas de programmes visant les jeunes. Ils confirment aussi le résultat généralement obtenu, qui est que l'appariement sous-estime l'effet réel du traitement[10, 11].

Que faire en pratique ?

Au bilan, les résultats des auteurs et ceux de la littérature similaire, malgré leurs imperfections, sont extrêmement négatifs pour l'appariement. Je détaille dans cette partie les enseignements principaux que je tire des résultats de B&R et de la littérature similaire pour la pratique de l'évaluation et pour la recherche sur les méthodes d'évaluation.

Promouvoir et développer des évaluations randomisées dès lors que c'est possible

C'est exactement la conclusion qu'ont tirée les pouvoirs publics aux États-Unis après les travaux de LaLonde (1986) et Fraker et Maynard (1987). Le programme national *Job Training Partnership Act* (JTPA) a notamment été évalué de manière expérimentale au milieu des années 1980. Il semble que les évaluations randomisées sont encore sous-utilisées en France, même si l'on assiste depuis quelques années à une augmentation bienvenue de leur nombre. Bien entendu, l'approche randomisée n'est pas exempte de problèmes (Heckman, 1992 ; Heckman et Smith, 1995). Le traitement de l'attrition est ainsi fondamental. L'utilisation des données administratives est une option permettant de réduire grandement ce phénomène. Il est aussi toujours possible d'estimer des bornes robustes à l'attrition (Horowitz et Manski, 2000 ; Lee, 2009). La prise en compte des effets de déplacement est de même une frontière importante de la recherche dans ce domaine, qui est explorée notamment par Crépon, Duflo, Gurgand, Rathelot et Zamora (2013).

Est-ce qu'il faut complètement abandonner l'appariement ?

Je ne le pense pas. Toutefois, il faut l'utiliser en toute conscience de ses limites. D'un point de vue théorique, tout d'abord, l'appariement requiert qu'un processus aléatoire ait alloué le traitement entre individus ayant les mêmes caractéristiques observées. C'est dans la réalité extrêmement difficile à justifier. L'appariement est donc généralement biaisé. En revanche, même si ce n'est théoriquement pas garanti (Heckman et Navarro-Lozano, 2004), l'appariement réduit généralement le biais par rapport à des comparaisons sans aucune variable de contrôle. Enfin, dans le cas de programmes de formation et/ou d'emploi, l'appariement sous-estime généralement l'effet réel du traitement et le biais peut être du même ordre de grandeur que l'effet du traitement (Chabé-Ferret, 2015). À terme, on peut espérer qu'une méta-analyse des résultats des travaux comparant estimateurs expérimentaux et non expérimentaux permettra de dégager l'approche la plus satisfaisante.

Est-il possible de collecter des variables supplémentaires pour améliorer les résultats de l'appariement ?

Il est en effet courant de voir l'adéquation de l'appariement justifiée en arguant du nombre important de variables de contrôle à la disposition de l'analyste. C'est généralement insuffisant, puisqu'il faut pouvoir expliquer en quoi le bénéfice du traitement est réparti aléatoirement entre individus identiques. En pratique, on peut imaginer collecter plus de variables sur le passé des individus. Les salaires précédents et les transitions récentes sur le marché du travail sont notamment des déterminants importants de la sélection dans ce type de programme (Ashenfelter, 1978) et leur prise en compte semble rapprocher grandement les résultats de l'appariement des résultats expérimentaux (Heckman, Ichimura, Smith et Todd, 1998). Des résultats récents de Sianesi (2014) suggèrent pourtant que l'inclusion de

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

ces variables ne suffit pas à reproduire des résultats expérimentaux. Andersson, Holzer, Lane, Rosenblum et Smith (2013) montrent que l'inclusion des trajectoires longues sur le marché du travail ne modifie pas l'estimation par l'appariement de l'effet causal d'un programme de formation.

En outre, ces variables sont indisponibles dans le cas de jeunes ayant peu d'expérience professionnelle et, donc, n'ayant complètement révélé ni leurs aptitudes ni leurs préférences sur le marché du travail. À ce titre, il serait intéressant d'explorer la prise en compte des capacités cognitives et non cognitives des individus en complément des variables habituellement recueillies. Ces variables paraissent en effet être de bonnes prédictrices des résultats observés sur le marché de l'emploi (Heckman, Stixrud et Urzua, 2006 ; Almlund, Duckworth, Heckman et Kautz, 2011). Des résultats récents de Caliendo, Mahlstedt et Mitnik (2014) suggèrent néanmoins que la prise en compte de ces variables ne modifie pas l'estimation de l'effet de programmes de formation. Enfin, il serait aussi déterminant de capter la proximité de l'emploi, c'est-à-dire les opportunités auxquelles les individus font face. Ces variables semblent déterminantes mais extrêmement délicates à capter.

Existe-t-il d'autres méthodes non expérimentales plus performantes ?

Les méthodes utilisant la régression par la discontinuité (RDD) ont généralement une bonne capacité à reproduire les résultats expérimentaux (Cook et Wong, 2008). Les méthodes de double ou triple différence semblent avoir de meilleures performances, notamment lorsqu'associées à l'appariement, pour estimer l'effet de programmes d'emploi (Heckman et Hotz, 1989 ; Heckman, Ichimura, Smith et Todd, 1998 ; Smith et Todd, 2005a). Il existe des raisons théoriques pour cela (Chabé-Ferret, 2015), même si certaines des conditions de validité (stationnarité) paraissent peu compatibles avec les échantillons de jeunes travailleurs. D'autres méthodes plus récentes n'ont pas encore été confrontées à des tests similaires (Athey et Imbens, 2006 ; Abadie, Diamond et Hainmueller, 2010).

Peut-on utiliser des tests de spécification pour détecter les applications où l'appariement (ou toute autre méthode non expérimentale) est invalide ?

Heckman et Hotz (1989) et Friedlander et Robins (1995) montrent que cette approche permet d'éliminer les estimateurs les plus éloignés des résultats expérimentaux. Il serait intéressant de voir si cette approche permettrait de détecter que l'appariement échoue dans l'application de B&R.

Mieux comprendre le processus de sélection dans les programmes

En conclusion, le biais qui affecte chaque estimateur non expérimental dépend de la nature de la règle de sélection (Chabé-Ferret, 2015). Nous comprenons encore assez mal les déterminants de la participation à des programmes de placement ou de formation. Les études existantes sont généralement plutôt descriptives (Heckman et Smith, 2004). Les modèles théoriques restent assez simplistes, oscillant entre modèle d'acquisition de capital humain et modèle de recherche d'emploi (Heckman, LaLonde et Smith, 1999). Mieux comprendre les processus de sélection, par exemple en utilisant des méthodes expérimentales pour tester des modèles théoriques, semble une voie de recherche prometteuse.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Notes

(1) Il arrive parfois qu'il soit impossible de remplir la condition de support commun : il n'existe alors que très peu d'individus non bénéficiaires similaires aux individus bénéficiaires. C'est le cas pour certaines des mesures évaluées par Chabé-Ferret et Subervie (2009, 2013). C'est aussi le problème que semblent rencontrer B&R avec leur spécification la plus riche (5).

(2) Cette dernière affirmation n'est en réalité vraie qu'en apparence. En toute rigueur, la validité de l'appariement requiert que le bénéfice de la politique publique soit déterminé de manière aléatoire parmi les individus ayant les mêmes caractéristiques observées. Cette nécessité d'une variable instrumentale inobservée est généralement oubliée en pratique.

(3) Encore faut-il que l'appariement soit parfaitement mis en œuvre. B&R utilisent les développements techniques les plus récents. On aurait pu simplement souhaiter un test plus probant du bon alignement des variables observées après appariement (Rosenbaum et Rubin, 1985 ; Lee, 2011). Il serait aussi intéressant de vérifier la robustesse des résultats des auteurs lorsqu'ils restreignent la comparaison au sein des quintuplets d'agences locales d'emploi (ALE). En effet, les différences de conditions entre marchés du travail sont une source importante de biais de l'appariement identifiée par Heckman, Ichimura, Smith et Todd (1998). Les deux autres sources identifiées par ces auteurs (différences de définition des variables entre groupes et manque de support commun, c'est-à-dire difficulté à trouver des individus traités et non traités qui puissent être comparés) n'affectent pas les estimations présentées par B&R.

(4) C'est le cas parce que B&R conditionnent sur les mêmes variables observées lorsqu'ils analysent les résultats de leur expérimentation et lorsqu'ils estiment l'effet du programme d'accompagnement par l'appariement.

(5) À ce titre, il pourrait être intéressant d'exploiter la randomisation de premier niveau (entre ALE), moins affectée par les effets de déplacement.

(6) LaLonde (1986), Fraker et Maynard (1987), Heckman et Hotz (1989), Friedlander et Robins (1995), Heckman, Ichimura, Smith et Todd (1998), Dehejia et Wahba (1999, 2002), Smith et Todd (2005a).

(7) Il s'agit malheureusement de moyennes pour des groupes de sept à dix personnes environ.

(8) Ces méthodes obtiennent de meilleurs résultats, dans le cas de l'expérimentation NSW, sur l'échantillon des femmes adultes.

(9) Dehejia (2005) ; Smith et Todd (2005b).

(10) Seuls Friedlander et Robins (1995) observent des surestimations. Ces auteurs utilisent les groupes de contrôle de chaque expérimentation comme groupes de comparaison pour les autres expérimentations, ce qui rend délicate l'interprétation de leurs résultats.

(11) C'est un résultat attendu au vu d'un modèle théorique de choix d'entrée dans un tel programme (Chabé-Ferret, 2015) : pour des individus possédant de meilleurs chances d'obtenir un emploi, le programme a un coût d'opportunité plus élevé en retardant leur accès à l'emploi. Cette information est souvent inobservée, ce qui conduit l'appariement à sous-estimer l'effet réel du programme.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Bibliographie

Abadie A., Diamond A. et Hainmueller J. (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program", *Journal of the American Statistical Association*, vol. 105, n° 490, pp. 493-505.

Almlund M., Duckworth A.L., Heckman J. et Kautz T. (2011). "Personality, Psychology and Economics", *in* Eric S.M., Hanushek A. et Woessmann L. (éds), *Handbook of The Economics of Education*, vol. 4, chap. 1, Elsevier, pp. 1-181.

Andersson F., Holzer H.J., Lane J.I., Rosenblum D. et Smith J. (2013). "Does Federally-Funded Job Training Work? Nonexperimental Estimates of WIA Training Impacts Using Longitudinal Data on Workers and Firms", National Bureau of Economic Research, Working Paper n° 19446, septembre, 88 pages.

Ashenfelter O. (1978). "Estimating the Effect of Training Programs on Earnings", *The Review of Economics and Statistics*, vol. 60, n° 1, pp. 47-57.

Athey S. et Imbens G.W. (2006). "Identification and Inference in Nonlinear Difference-in-Differences Models", *Econometrica*, vol. 74, n° 2, pp. 431-497.

Bléhaut M. et Rathelot R. (2014). "Expérimentation contrôlée contre appariement : le cas d'un dispositif d'accompagnement de jeunes diplômés demandeurs d'emploi", *Économie et Prévision*, n° 204-205, pp. 163-181, ce numéro.

Caliendo M., Mahlstedt R. et Mitnik O.A. (2014). "Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies", Institute for the Study of Labor (IZA) Discussion Papers n° 8337, juillet, 41 pages.

Chabé-Ferret S. (2015). "Analysis of the Bias of Matching and Difference-In-Difference under Alternative Earnings and Selection Processes", *Journal of Econometrics*, vol. 185, n° 1, pp. 110-123.

Chabé-Ferret S. et Subervie J. (2009). "Évaluation de l'effet propre des mesures agro-environnementales du Plan de développement rural national 2000-2006 sur les pratiques des agriculteurs," Rapport d'évaluation, Centre national du machinisme agricole du génie rural, des eaux et des forêts (Cemagref), juillet, 222 pages
(http://agriculture.gouv.fr/sections/publications/evaluation- politiques / evaluations / estimation - effets).

Chabé-Ferret S. et Subervie J. (2013). "How Much Green for the Buck? Estimating Additional and Windfall Effects of French Agro-Environmental Schemes by Difference- In-Difference Matching", *Journal of Environmental Economics and Management*, vol. 65, n° 1, pp. 12-27.

Cook T.D. et Wong V.C. (2008). "Empirical Tests of the Validity of the Regression Discontinuity Design", *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, n° 91/92, pp. 127-150.

Crépon B., Duflo E., Gurgand M., Rathelot R. et Zamora Ph. (2013). "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment", *The Quarterly Journal of Economics*, vol. 128, n° 2, pp. 531-580.

Dehejia R.H. (2005). "Practical Propensity Score Matching: A Reply to Smith and Todd", *Journal of Econometrics*, vol. 125, n° 1-2, pp. 355-364.

Dehejia R.H. et Wahba S. (1999). "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, vol. 94, n° 448, pp. 1053-1062.

Dehejia R.H. et Wahba S. (2002). "Propensity Score-Matching Methods For Nonexperimental Causal Studies", *The Review of Economics and Statistics*, vol. 84, n° 1, pp. 151-161.

Fraker T. et Maynard R. (1987). "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs", *The Journal of Human Resources*, vol. 22, n° 2, pp. 194-227.

Friedlander D. et Robins P.K. (1995). "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods", *The American Economic Review*, vol. 85, n° 4, pp. 923-937.

Givord P. (2014). "Méthodes économétriques pour l'évaluation de politiques publiques", *Économie et Prévision*, n° 204-205, pp. 1-28, ce numéro.

Heckman J.J. (1992). "Randomization and Social Policy Evaluation", *in* Manski C.F. et Garfinkel I. (éds), *Evaluating Welfare and Training Programs*, Harvard University Press, pp. 201-230.

Heckman J.J. (2001). "Micro data, Heterogeneity and the Evaluation of Public Policy: Nobel Lecture", *Journal of Political Economy*, vol. 109, n° 4, pp. 673-748.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)

Heckman J.J. et Hotz V.J. (1989). "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, vol. 84, n° 408, pp. 862-874.

Heckman J.J., Ichimura H., Smith J.A. et Todd P.E. (1998). "Characterizing Selection Bias Using Experimental Data", *Econometrica*, vol. 66, n° 5, pp. 1017-1099.

Heckman J.J., LaLonde R.J. et Smith J.A. (1999). "The Economics and Econometrics of Active Labor Market Programs", *in* Ashenfelter O.C. et Card D. (éds), *Handbook of Labor Economics*, vol. 3, chap. 31, Elsevier, North Holland, pp. 1865-2097.

Heckman J.J. et Navarro-Lozano S. (2004). "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models", *The Review of Economics and Statistics*, vol. 86, n° 1, pp. 30-57.

Heckman J.J. et Smith J.A. (1995). "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, vol. 9, n° 2, pp. 85-10.

Heckman J.J. et Smith J.A. (2004). "The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program", *Journal of Labor Economics*, vol. 22, n° 2, pp. 243-298.

Heckman J.J., Stixrud J. et Urzua S. (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", *Journal of Labor Economics*, vol. 24, n° 3, pp. 411-482.

Horowitz J.L. et Manski C.F. (2000). "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data", *Journal of the American Statistical Association*, vol. 95, n° 449, pp. 77-84.

LaLonde R.J. (1986). "Evaluating the Econometric Evaluation of Training Programs with Experimental Data", *American Economic Review*, vol. 76, n° 4, pp. 604-620.

Lee D.S. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies*, vol. 76, n° 3, pp. 1071-1102.

Lee W.-S. (2011). "Propensity Score Matching and Variations on the Balancing Test", *Empirical Economics*, vol. 44, n° 1, pp. 47-80.

Rosenbaum P.R. et Rubin D.B. (1985). "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, vol. 39, n° 1, pp. 33-38.

Sianesi B. (2014). "Dealing with Randomisation Bias in a Social Experiment: The Case of ERA", Institute for Fiscal Studies (IFS) Working Paper n° W14/10, 52 pages.

Smith J.A. et Todd P.E. (2005a). "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?", *Journal of Econometrics*, vol. 125, n° 1-2, pp. 305-353.

Smith J.A. et Todd P.E. (2005b). "Rejoinder", *Journal of Econometrics*, vol. 125, n° 1-2, pp. 365-375.

Direction générale du Trésor | Téléchargé le 27/10/2026 sur https://shs.cairn.info par Courtier Théophile (IP: 81.1.6.66)