



HAL
open science

Appui méthodologique à l'évaluation du développement des zones rurales. Fascicule 3 : note sur l'évaluation des effets propres

Pascale Lofredi, Bertrand B. Schmitt, Francis Aubert

► To cite this version:

Pascale Lofredi, Bertrand B. Schmitt, Francis Aubert. Appui méthodologique à l'évaluation du développement des zones rurales. Fascicule 3 : note sur l'évaluation des effets propres. [0] 2005. hal-02831722

HAL Id: hal-02831722

<https://hal.inrae.fr/hal-02831722v1>

Submitted on 7 Jun 2020

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.



Ministère de l'Agriculture et de la Pêche

Direction des politiques économique et internationale
Service des stratégies agricoles et industrielles
Sous-Direction de l'Évaluation, de la Prospective, des Etudes et de l'Orientation
Bureau de l'évaluation et des programmes d'études

APPUI METHODOLOGIQUE A L'EVALUATION DU DEVELOPPEMENT DES ZONES RURALES

Fascicule 3

Note sur l'évaluation des effets propres

Novembre 2005

Convention du 27 juillet 2005
Ministère de l'agriculture et de la pêche - ENESAD.





Ministère de l'Agriculture et de la Pêche

Direction des politiques économique et internationale
Service des stratégies agricoles et industrielles
Sous-Direction de l'Évaluation, de la Prospective, des Études et de l'Orientation
Bureau de l'évaluation et des programmes d'études

APPUI METHODOLOGIQUE A L'EVALUATION DU DEVELOPPEMENT DES ZONES RURALES

Fascicule 3

Note sur l'évaluation des effets propres

Novembre 2005

Convention du 27 juillet 2005
Ministère de l'agriculture et de la pêche - ENESAD.

Pascale LOFREDI : Doctorante
Bertrand SCHMITT : Directeur de recherche

Francis AUBERT : Professeur, responsable de l'étude



Institut National de la Recherche Agronomique – Etablissement National d'Enseignement Supérieur Agronomique de Dijon
UMR INRA-ENESAD CESAER

Centre d'Economie et Sociologie Rurales appliquées à l'Agriculture et aux Espaces Ruraux
26, Bd Docteur Petitjean – PB 87999 – 21079 DIJON cedex
Tél. : 03 80 77 25 80 (25 70) – Fax : 03 80 77 25 71

SOMMAIRE

Introduction	7
1. Le problème de l'évaluateur	8
1.1. Le paramètre d'intérêt	9
1.2. Estimer le contrefactuel et les biais engendrés par des comparaisons simples	9
2. Les méthodologies disponibles pour corriger les biais	11
2.1. Les méthodes d'appariement ou de <i>matching</i> :.....	11
2.2. L'approche en différences de différences ou doubles différences	14
2.3. Une approche en données de panel : le modèle à taux de croissance aléatoire	14
3. Eléments de conclusion : Quelle méthode pour quel programme ?	16
Eléments bibliographiques	17

Introduction

La question de l'évaluation des politiques publiques est aujourd'hui largement en débat et, dans certains Etats, elle conditionne même la mise en œuvre ou la poursuite de certaines actions publiques. Au cœur des enjeux de l'évaluation figure la notion d'« efficacité », à savoir la confrontation des impacts d'une politique avec les objectifs affichés par ses concepteurs. Dans la plupart des évaluations actuelles et notamment celles qui portent sur les politiques de développement rural, l'analyse des effets d'une politique se limite le plus souvent à l'examen, certes détaillé, des montants engagés et de leur répartition selon les domaines d'intervention envisagés, auquel peut s'ajouter une analyse des caractéristiques des bénéficiaires (individus ou territoires) et de leurs évolutions sous programme. Ainsi, les nombreux guides ou vade-mecum de l'Union Européenne (European Commission, 1999) focalisent souvent les évaluations sur la construction des indicateurs de programme, sans chercher à mesurer les relations de causes à effets des actions publiques. Dans la plupart des cas, l'évaluation ne fournit que des hypothèses sur les facteurs qui influent sur l'efficacité mais sans les confronter aux faits en cherchant à mesurer les effets réels de la politique (Toulemonde, 1995).

Or, ce sont bien les effets propres d'un programme qu'il est indispensable de mesurer si l'on veut évaluer complètement les impacts d'une intervention publique sur les trajectoires des individus ou des territoires qu'elle vise. L'évaluation des effets propres d'un programme se heurte alors à deux problèmes classiques. Le premier relève du choix et de la disponibilité d'indicateurs de résultats ou de performances à partir desquels il s'avère pertinent d'évaluer les résultats de l'action publique ; étant envisagé par ailleurs dans ce rapport, il ne sera pas traité ici. On focalisera dans cette partie sur le problème dit de « l'évaluateur », *i.e.* l'impossibilité dans laquelle se trouve celui-ci de comparer directement les performances réelles d'un individu bénéficiaire de l'intervention (*policy on*) avec ce qu'auraient été ses performances en absence de cette intervention (*policy off*). L'obtention de la « vraie » mesure de l'impact du programme sur ses bénéficiaires est de ce fait délicate et la comparaison entre individus aidés et individus non aidés, qui peut paraître une alternative commode pour pallier cette difficulté, n'est pas sans poser de problèmes graves sur la valeur de l'évaluation réalisée¹ : celle-ci risque en effet d'être entachée de certains biais que nous exposerons plus loin.

Une abondante littérature économique et économétrique s'est développée au cours des dernières décennies (et notamment depuis les travaux fondateurs de Rubin, 1974) en vue de clarifier le problème de l'évaluateur et de permettre la correction des biais auquel il doit faire face. Pour se faire une idée du caractère très actif de ces questions sur le plan scientifique, on pourra consulter le récent numéro spécial que le *Journal of Econometrics* a consacré à cette question (Ham et LaLonde, eds, 2005), et notamment l'article de Smith et Todd (2005), ainsi que, par exemple, Ravaillon (2001), Heckman (1979 et 2000), Heckman, Lalonde et Smith (1999), Rosenbaum et Rubin (1985). La mise en œuvre des méthodologies élaborées dans ce cadre a surtout concerné l'évaluation des politiques d'emploi

¹ Notons que ce problème se pose en termes moins cruciaux s'il est possible de se placer dans un cadre expérimental. En effet, dans ce cas, le caractère aléatoire de la sélection des bénéficiaires et des non bénéficiaires du programme permet une estimation non biaisée des effets du programme en comparant les performances des bénéficiaires à celles des non bénéficiaires. Ce cadre d'analyse délicat à mettre en œuvre dans le cas de politiques publiques n'est *a fortiori* pas envisageable dans le cas des politiques de développement rural.

(Dehejia et Wahba, 2002 ; Heckman, Ichimura et Todd, 1997 ; Heckman et Hotz, 1989) ou d'éducation (voir Bénabou, Kramarz et Prost, 2004, pour un bel exemple d'évaluation des effets d'une politique française de discrimination positive, celle des Zones d'éducation prioritaire, les ZEP). Les autres politiques économiques, notamment celles visant à inverser des situations de crise de développement localisée et dont l'action, positivement discriminante, repose sur un zonage, sont rarement évaluées (Bondonio, 2000). Notons cependant l'existence d'une littérature empirique récente évaluant les effets des politiques d'incitations fiscales à la localisation des firmes dans le cadre des « *empowerment zones* » (Papke, 1994 ; Boarnet et Bogart, 1996 ; Bondonio et Engberg, 2000), l'impact régional des infrastructures de transport (Rephmann, 1993) ou les effets du « *credit pooling for infrastructure financing* » sur la croissance économique locale (Ray et Hite, 2000) ou encore les effets des politiques structurelles européennes (Bondonio, 2002 ; Schmitt *et al.*, 2002).

Après avoir exposé les problèmes auxquels est confronté l'évaluateur lorsqu'il cherche à mesurer les effets propres d'une politique, on évoquera dans la suite de cette note les méthodes actuellement disponibles pour remédier aux problèmes soulevés, tout en examinant les conditions de leur application dans le cas de l'évaluation du futur RDR.

1. Le problème de l'évaluateur

Une modélisation du problème de l'évaluation a été proposée par Rubin (1974) dans un contexte biomédical marquant ainsi toute la terminologie qui y est associée. Les notations utilisées ici sont celles des travaux d'Heckman, Lalonde et Smith (1999) appliquées non plus dans un contexte d'évaluation de politiques d'emploi en direction d'individus, mais de programmes de développement territorialisés.

Soit un espace géographique donné avec, en son sein, des subdivisions de cet espace, ici des cantons (mais ce pourrait des individus comme des entreprises, des exploitations agricoles ou des ménages ruraux), reçoivent des fonds publics pour stimuler leur développement tandis que d'autres n'en reçoivent pas. Cet état d'appartenance au programme est noté $D=1$, et son événement contraire, $D=0$. L'indicateur de performances ou un vecteur d'indicateurs, c'est-à-dire les caractéristiques de la zone ou de l'individu sur lesquelles le programme cherche explicitement à avoir un effet favorable ou dont on peut penser a priori qu'elles sont affectées par le programme (évolution démographique, de l'emploi, valeur ajoutée locale, ...), est noté Y_0 en l'absence de programme et Y_1 avec le programme.

Au total, on peut donc disposer de deux populations statistiques : les cantons ou individus tels que $D_i=1$ que l'on appelle les « bénéficiaires »², et les autres « non bénéficiaires », tels que $D_i=0$ et pour chaque canton ou individu, de deux niveaux de performance possibles Y_1 et Y_0 , respectivement avec et sans programme. La distribution de l'impact du programme ou vecteur de gain est alors donné par $\Delta_i = Y_{1i} - Y_{0i}$.

Poser de cette façon, la teneur du problème à résoudre apparaît immédiatement : on pourra réellement observer (i) pour les bénéficiaires, que leurs seules performances avec programme et (ii) pour les non

² Terminologie qui évite celle issue des approches biomédicales qui distingue individus « traités » et « non traités ».

bénéficiaires que leurs performances sans programme. Or, ce que l'on cherche, c'est à évaluer l'écart entre les performances des bénéficiaires après intervention du programme et ce qu'auraient été les performances de ces mêmes bénéficiaires si le programme n'avait pas eu lieu. Ainsi, les deux « contrefactuels », à savoir les performances sans programme des bénéficiaires mais aussi les performances avec programme des non bénéficiaires sont par nature inobservables. Les distributions Y_1 et Y_0 sont donc incomplètes (respectivement disponibles seulement pour les i tels que $D_i=1$ et pour les i tels que $D_i=0$) ; et, la distribution de l'impact du programme totalement inconnue.

1.1. Le paramètre d'intérêt

Une manière de surmonter ce problème, et donc d'obtenir une mesure de l'impact du programme, est de chercher à estimer une caractéristique de la distribution de D_i , encore appelé « paramètre d'intérêt ». Typiquement, il s'agit d'une moyenne. Le choix du paramètre d'intérêt à retenir se fait en lien avec la question particulière soulevée par la démarche évaluative elle-même et est donc liée à la nature du programme. Dans le cas des programmes territorialisés (comme les fonds structurels européens ou les Zones de revitalisation rurale), le législateur vise à homogénéiser le développement économique sur l'ensemble de l'espace géographique d'intérêt (un pays, par exemple) *via* des subsides destinés à stimuler les zones économiquement les plus fragiles ou les entreprises qui sont ou pourraient s'y localiser. Il s'agit donc d'une politique de discrimination positive territorialisées ; le programme n'étant pas destiné à être généralisé à l'ensemble du pays, il cible au contraire certaines zones au sein desquelles les agents économiques sont susceptibles d'être aidés. Ainsi, le programme sera considéré comme ayant atteint son but si les zones bénéficiaires comblent au moins une partie de leur retard de développement ou si les tendances d'évolution vont dans le sens d'un rattrapage.

L'évaluation tentera donc de répondre à la question suivante : les territoires ciblés par le programme en ont-ils tiré profit ? L'impact que l'on souhaite estimer ne concerne alors que la sous-population constituée des individus statistiques tels que $D_i=1$, et l'analyse se centre sur l'impact moyen du programme sur les bénéficiaires (ou dit autrement, l'impact moyen du traitement sur les traités).³ Un tel impact s'écrit :

$$\Delta_{TT} = E(Y_1 - Y_0 | D = 1)$$

Pour ce paramètre d'intérêt, la valeur de $E(Y_1 | D = 1)$ est directement observable tandis que $E(Y_0 | D = 1)$ ne l'est pas et doit donc être estimée.

1.2. Estimer le contrefactuel et les biais engendrés par des comparaisons simples

De manière générale, deux grandes stratégies d'estimation de $E(Y_0 | D = 1)$ sont possibles : soit on utilise un groupe de comparaison ou de référence qui correspond au groupe des zones ou individus non bénéficiaires et on compare les performances obtenues après programme par les individus bénéficiaires ; soit on mobilise des données antérieures au programme concernant les seuls bénéficiaires et on compare leurs performances après programme avec celles qu'ils obtenaient avant

³ D'autres paramètres d'intérêt peuvent également être envisagés. Le plus classique est l'« effet moyen du traitement » (ou τ_{ATE}) qui cherche à estimer l'impact d'un programme s'il était appliqué à l'ensemble des individus. Un tel paramètre s'écrit alors : $\tau_{ATE} = E(Y_1 - Y_0)$

programme. La première stratégie correspond à un *estimateur en cross-section*, le second à un *estimateur avant/après* (cf. tableau 1). Le choix de l'une ou l'autre de ces deux voies d'estimation dépend des données disponibles, sachant que l'une et l'autre introduisent d'importants biais dans l'évaluation de l'impact d'un programme.

Tableau 1 - Les trois approches classiques de l'évaluation et les biais qu'elles engendrent

Estimateurs	Hypothèse fondatrice	Données nécessaires	Violation d'hypothèse(s)
Avant/Après	$E(Y_{0t} - Y_{0t'} D = 1) = 0$	2 coupes temporelles en t' et t	(i) Changement global de situation économique : Biais conjoncturel (ii) Changement de position dans un cycle
Cross-section	$E(Y_{0t} D = 1) = E(Y_{0t} D = 0)$	Observation en t d'individus non bénéficiaires	Participation au programme liée aux caractéristiques avant programme : Biais de sélection
Doubles différences	$E(Y_{0t} - Y_{0t'} D = 1) = E(Y_{0t} - Y_{0t'} D = 0)$	2 coupes temporelles pour bénéficiaires et non bénéficiaires	Changement de position dans un cycle

où t = date de l'évaluation, t' = date antérieure au programme.

La qualité de l'estimateur avant/après dépend principalement des évolutions conjoncturelles qui ont eu lieu au cours de la période étudiée et une valeur positive (resp., négative) de cet estimateur peut simplement relever d'une amélioration (resp., dégradation) de la situation économique générale. On a donc à faire face à un biais conjoncturel. Le second estimateur, le plus couramment mobilisé car le moins difficile à construire, souffre, quant à lui, d'un biais de sélection dû à la confusion dans la mesure entre effet du programme et effet de la différence de caractéristiques initiales entre bénéficiaires et non bénéficiaires. Il correspond à l'écart entre ce que l'on veut mesurer : $c_{propre} = E(Y_1 | D = 1) - E(Y_0 | D = 1)$ et ce que l'on peut mesurer : $c_{cross} = E(Y_1 | D = 1) - E(Y_0 | D = 0)$, c'est-à-dire à la quantité suivante :

$$b = E(Y_0 | D = 1) - E(Y_0 | D = 0)$$

Cette quantité correspond à la différence entre les moyennes de la variable de performance Y pour les zones ou individus aidés et non aidés si aucun n'avait bénéficié de la politique mise en œuvre. Elle biaise la mesure de l'effet de la politique. Les différences de performances entre les zones ou individus bénéficiaires et non bénéficiaires ont ainsi deux sources : l'effet réel de la politique menée et la différence initiale entre les deux groupes d'individus.

La combinaison de ces deux approches, qui nécessite l'accès à des données tant antérieures aux programmes qu'externes aux individus bénéficiaires, débouche sur un troisième estimateur beaucoup plus performant puisque basé sur une hypothèse plus réaliste, l'estimateur en doubles différences, sur lequel nous reviendrons par la suite.

2. Les méthodologies disponibles pour corriger les biais

Les méthodes statistiques visant à corriger ces biais, et notamment le biais de sélection, ont été développées à partir de problématiques d'évaluations économiques de politiques pour lesquelles il est beaucoup moins aisé de mettre en place des procédures expérimentales que dans un cadre biomédicales. On trouve, dans cette littérature d'économétrie appliquée, deux grands types de méthodes visant à corriger le biais de sélection dans les évaluations sur données non expérimentales, les méthodes dites de *matching* et l'approche en doubles différences. Nous penchant sur la mesure des effets propres de politiques à ciblage géographique, il convient d'ajouter à ces deux approches le modèle dit «à taux de croissance aléatoire» dont Bondonio (2002) suggère qu'il est le plus adapté à ce type de problématiques.

2.1. Les méthodes d'appariement ou de *matching* :

Les méthodes dites d'appariement ou de *matching* s'appuient sur l'idée originale de Rosebaum et Rubin (1985). Elle consiste à dépasser l'absence d'un processus aléatoire de sélection au programme et permettre la comparaison du groupe des bénéficiaires et d'un groupe de non bénéficiaires en contrôlant les caractéristiques initiales observables conjointement liées au processus de sélection et aux indicateurs de performance. Il s'agit donc de partitionner un échantillon de zones (ou d'individus) bénéficiaires et non bénéficiaires, pour lequel est disponible un jeu de données en coupe, en sous-populations au sein desquelles les individus sont « proches » sur un nombre important de caractéristiques initiales x . Dans chaque sous-population, l'effet propre pourra alors être approximé par le calcul de la différence entre les performances moyennes des individus aidés et celles des individus non aidés. Il reste alors à prendre la moyenne de ces effets par sous-groupe, pondérés par leur taille, pour aboutir à l'effet global sur l'ensemble de la population des bénéficiaires.

Comme on s'en doute, le nombre de sous-populations à définir s'accroît rapidement à mesure qu'augmente le nombre de caractéristiques initiales x à prendre en compte. Il peut en outre y avoir des sous-groupes pour lesquels aucun non bénéficiaire suffisamment proche n'est disponible. Pour surmonter ces problèmes et faciliter la procédure d'appariement, il est d'usage d'estimer (au moyen d'un modèle logit ou probit) pour chaque individu statistique, une probabilité d'être sélectionné au programme en fonction de ses caractéristiques initiales observables x . Sous certaines conditions, on pourra utiliser cette probabilité conditionnelle estimée, $P(D=1|x)$ souvent appelée *propensity score*, pour effectuer les appariements.

Les méthodes de *matching* permettent donc en fait de constituer un groupe de comparaison acceptable. Si l'on note P_i le *propensity*, l'appariement pour chaque participant est construit comme une moyenne pondérée des résultats des non participants où les coefficients de pondération ou poids $W(i,j)$ dépendent de la distance entre P_i et P_j . Différents estimateurs de *matching* peuvent alors être construits

selon deux aspects : la façon dont le voisinage $C(P_i)$ est défini et/ou le mode de construction des poids $W(i,j)$.

Ainsi, un des estimateurs les plus traditionnels de *matching* utilise la règle *du seul plus proche voisin* sans possibilité de remise. Selon cette définition du voisinage, le groupe de comparaison est construit en choisissant un et un seul non bénéficiaire j pour être apparié à un bénéficiaire i de telle sorte que la distance en probabilité entre P_i et P_j soit minimale. De même, et pour peu que l'effectif du groupe de comparaison soit suffisamment important, plusieurs « plus proches voisins » peuvent être utilisés pour construire une moyenne (pondérée par les P_j) qui servira de point de comparaison. Les applications traditionnelles de cet estimateur n'impose pas de condition de « support commun » et apparie « sans remise », de telle sorte que chaque non bénéficiaire ($D=0$) peut servir de voisin pour au plus un bénéficiaire ($D=1$). Cette règle de non remise lors d'un *matching* au plus proche voisin engendre un inconvénient important : les estimations et leurs résultats dépendent de l'ordre dans lequel les observations sont appariées. Pour pallier cette difficulté et notamment dans les cas où la qualité des appariement est médiocre⁴, on peut utiliser l'appariement avec remise des non bénéficiaires, c'est-à-dire qu'un non participant peut être apparié à plus d'un participant. L'appariement avec remise induit cependant un arbitrage entre biais et variance dans la mesure où l'amélioration de la qualité moyenne du *matching* induit une utilisation moindre d'information distincte, ce qui augmente la variance.

Il existe de nombreuses variantes de l'appariement au plus proche voisin, dont par exemple le *caliper matching* (Cochran et Rubin, 1973) qui tente d'éviter les mauvais appariements (ceux pour lesquels P_j est loin de P_i) en imposant une tolérance sur la distance maximale autorisée. Les individus bénéficiaires pour lesquels le *matching* ne peut respecté cette condition sont exclus de l'échantillon. Le *caliper matching* impose donc un support commun, c'est-à-dire une zone où les probabilités des bénéficiaires et des non bénéficiaires soient suffisamment proches pour permettre l'appariement⁵.

Quel que soit le mode d'appariement retenu, l'approche par *matching* débouche sur la construction de divers estimateurs des effets propres dont nous ne présentons ici que quelques exemples (Crépon et Iung, 1999 ; Ravallion, 2000).

Le premier d'entre eux, l'estimateur pondéré, correspond à une « simple » comparaison de moyennes entre les performances des individus aidés et non aidés, ceux-ci étant convenablement pondérés.

L'estimateur de l'effet propre s'écrit dans ce cas :

⁴ C'est-à-dire dans les cas où il n'existe que peu de non bénéficiaires ayant des caractéristiques proches des bénéficiaires, ce qui est le cas quand les individus éligibles sont quasiment mécaniquement bénéficiaires d'aide, comme dans le cas d'évaluation d'impacts territoriaux de politique zonée.

⁵ D'autres types de *matching* existent. Citons le *radius matching*, le *kernel matching* ou l'*interval matching*. Voir Smith et Todd (2004) pour une revue détaillée.

$$c_{pond} = E\left(\frac{(Y_1|D=1)}{P(D=1|x)}\right) - E\left(\frac{(Y_0|D=0)}{1-P(D=1|x)}\right)$$

où $P(D=1|x)$ correspond à la probabilité qu'a une zone ou un individu d'être bénéficiaire du programme évalué au vu des valeurs prises par ses caractéristiques x , probabilité estimée, à l'aide d'un modèle logit ou probit dont la variable dépendante est une indicatrice prenant la valeur 1 quand la zone ou l'individu est bénéficiaire et la valeur 0 sinon.⁶

Une autre approche plus classique de l'effet propre par *matching* passe par l'estimation d'une régression de la variable de performance sur une fonction des caractéristiques x et de la variable dichotomique d'appartenance au groupe bénéficiaire :

$$y = a_0 + a_2 D + x a_3 + e$$

L'estimateur de l'effet causal correspond simplement à \hat{a}_2 .⁷ La mise en œuvre de cet estimateur par régression peut être simplifiée en substituant au vecteur des caractéristiques x la probabilité estimée de l'appartenance à la catégorie des bénéficiaires au vu des caractéristiques x , $P(D=1|x)$. Cette façon de procéder présente l'avantage de pouvoir estimer des fonctions plus flexibles en introduisant, par exemple, cette probabilité estimée et son carré, ce qui débouche sur des variantes de l'estimateur par régression du type :

$$y = d_0 + d_2 D + d_3 P(D=1|x) + h$$

$$y = I_0 + I_2 D + I_3 P(D=1|x) + I_4 [P(D=1|x)]^2 + m$$

\hat{d}_2 et \hat{I}_2 sont alors les estimateurs de l'effet propre de la politique.

La qualité de ces estimateurs dépend, on s'en doutera, de la qualité des appariements entre individus ou zones bénéficiaires et non bénéficiaires et donc du mode de sélection du groupe de comparaison. Plus les individus de ce groupe seront proches en termes de caractéristiques initiales (et donc, on suppose, en termes d'évolution potentielle sous programme), plus l'estimateur utilisé révélera l'effet propre du programme. Il y a donc intérêt à resserrer au plus près des caractéristiques des bénéficiaires le mode de sélection des individus choisis pour appartenir au groupe de comparaison. Il va de soi que ce rapprochement entre individus aidés et non aidés a ces limites car, si les individus non bénéficiaires de la politique sont tous très semblables aux bénéficiaires, on est en droit de se demander pourquoi ces individus n'ont pas été aidés. De multiples raisons peuvent être évoquées pour expliquer un tel phénomène, toutes n'étant pas exemptes de biais. Il n'en demeure pas moins que, si les critères d'éligibilité d'un territoire à une politique ont été convenablement mis en œuvre et/ou si une mesure a attiré une large part de la population visée, la constitution du groupe de comparaison peut s'avérer très délicate dans la mesure où les non bénéficiaires aux caractéristiques initiales proches des bénéficiaires peuvent être peu nombreux. L'appariement sera alors de mauvaise qualité et la mesure de l'effet ne sera qu'imparfaitement corrigée du biais de sélection. Smith et Todd (2005) proposent, dans le cas où

⁶ Il faut noter que le calcul de l'écart-type de cet estimateur nécessite de tenir compte du fait que $P(D=1|x)$ est elle-même estimée.

⁷ Signalons que la formulation retenue suppose que l'impact du programme sur les performances ne varie pas avec les caractéristiques x des zones ou individus aidés. Si tel n'est pas le cas, il est nécessaire d'introduire dans la régression un terme croisé supplémentaire, $x.D$ (Ravaillon, 2000).

l'appariement est de piètre qualité, d'avoir recours à un *matching* avec remise. La sélection des non bénéficiaires membres du groupe de comparaison s'effectue alors sur la base du *propensity score*. Mais, au lieu de limiter le *matching* à des paires d'individus aidés et non aidés en fonction de la proximité de leur *propensity score*, on autorise un même non bénéficiaire à être apparié avec tous les individus bénéficiaires ayant une valeur proche de son *propensity score*. Chaque individu non bénéficiaire du groupe de comparaison est alors pondéré en fonction du nombre de bénéficiaires qu'il est censé représenter. Cette façon de procéder possède le double avantage de réduire le groupe de comparaison aux seuls non bénéficiaires les plus proches des bénéficiaires et d'en pondérer les individus en fonction de leur proximité aux bénéficiaires.

2.2. L'approche en différences de différences ou doubles différences

Comme on l'a relevé précédemment, la combinaison d'une approche en *cross-section* (ou en coupe) et d'une approche avant/après correspond à une approche dite en doubles différences. Celle-ci permet de se parer des éventuels biais liés aux caractéristiques individuelles inobservables affectant à la fois la variable de performance et la variable d'aide publique. Ceux-ci peuvent en effet compenser le biais de sélection (Heckman *et al.*, 1998). Un tel estimateur mesure l'impact du programme évalué par la différence de performances observables avant et après programme entre les individus ou zones bénéficiaires et non bénéficiaires. Il nécessite donc l'accès à des données antérieures et postérieures au programme (en t' et t) pour un échantillon d'individus ou de zones aidés ($D=1$) et non aidés ($D=0$). L'estimateur de l'effet propre, \hat{q} , correspond à la valeur du paramètre q obtenu à l'issue de la régression en moindres carrés ordinaires de :

$$Y_t - Y_{t'} = \mathbf{j}(X_t) - \mathbf{j}(X_{t'}) + D \cdot \mathbf{q} + \mathbf{x}$$

où X_t et $X_{t'}$ correspondent aux caractéristiques en t et en t' des individus aidés et non aidés.

Très performante car, par cette double comparaison, elle contrôle le biais de sélection tant en niveau qu'en trend sans encourir de biais conjoncturel, cette approche s'appuie cependant sur une hypothèse de similarité d'évolution des performances sans programme des bénéficiaires et des non bénéficiaires. Elle est alors sensible à des changements de position dans le «cycle de vie» des individus ou des zones. En outre, procédant comme les méthodes de *matching* par recours à un groupe de comparaison, sa mise en œuvre peut être améliorée, dans les cas de *matching* difficile (comme pour les politiques zonées), en recourant à un *matching* avec remise et pondération lors de la constitution du groupe de comparaison.

2.3. Une approche en données de panel : le modèle à taux de croissance aléatoire

Lorsque sont disponibles des données de panel, c'est-à-dire un suivi dans le temps d'un ensemble d'individus aidés et non aidés, il est possible et potentiellement efficace de recourir aux méthodes économétriques correspondant à ce type de données. En effet, la modélisation d'effets spécifiques aux individus permet de contrôler les différences initiales liées à l'indicateur de performances. Si l'hypothèse de trends spécifiques aux zones ou aux individus liés à la variable de résultat est plausible, il faut alors utiliser le modèle dit à *taux de croissance aléatoire* proposé par Heckman et Hotz (1989).

Une première correction du biais de sélection peut être obtenue par un contrôle des caractéristiques initiales des zones (ou individus) susceptibles d'influer sur la variable de performance en estimant un modèle à effets fixes individuels dans lequel les influences externes variables au cours du temps sont contrôlées *via* l'introduction d'un trend linéaire. Le modèle formel s'écrit alors :

$$\text{Log}Y_{it} = \mathbf{b}_1 t + \mathbf{d}_1 D_{it} + \mathbf{a}_{1i} + u_{1it}$$

où Y_{it} est la variable de performances, a_i correspond aux effets fixes individuels, β est un trend linéaire et D_{it} une variable dichotomique d'appartenance au programme à l'instant t .

Le modèle à taux de croissance aléatoire tel que défini par Heckman et Hotz (1989) et utilisé notamment par Papke (1994) lors de son évaluation de programmes d'entreprises en Indiana, prend en compte le fait que la croissance économique affecte les zones ou les individus à un rythme différent et que celui-ci peut dépendre de l'appartenance ou non au programme d'aides. Il s'écrit alors :

$$\text{Log}Y_{it} = \mathbf{a}_{2t} + \mathbf{a}_{2i} + \mathbf{b}_{2i} t + \mathbf{d}_2 D_{it} + u_{2it}$$

où β_{2i} correspond à la croissance spécifique de la zone i . Une estimation de cette équation en différences premières conduit à éliminer les effets fixes, ce qui donne l'équation suivante :

$$\Delta \text{Log}Y_{it} = \mathbf{b}_{2i} + \mathbf{d}_2 \Delta D_{it} + \Delta \mathbf{a}_{2t} + \Delta u_{2it}$$

que l'on estime par un traitement standard des effets fixes en écart à la moyenne ou en calculant une différence seconde.

3. Eléments de conclusion : Quelle méthode pour quel programme ?

La ou les méthodologies mobilisées pour estimer l'effet propre d'un programme sont intimement liées à la disponibilité de données concernant non seulement le groupe des bénéficiaires mais aussi un groupe de comparaison. Ces données doivent comprendre des caractéristiques descriptives des individus notamment celles qui ont initialement déterminé la sélection au programme, des indicateurs de performances dont le choix doit être conditionné par les objectifs attendus du programme ainsi que ses effets potentiels et, dans la mesure du possible un ensemble de variables explicatives des indicateurs de performance retenus. De façon évidente, plus l'information disponible est abondante en coupes temporelles et plus la qualité des estimations sera améliorée en termes de correction de biais potentiels et de précision des estimateurs. *A minima*, trois dates de recueil des données avant la mise en place du programme et trois dates postérieures sont requises pour obtenir un niveau de fiabilité acceptable. Disposer d'un suivi temporel régulier des différentes variables pour l'ensemble des individus et qui débute au moins trois ans avant la mise en place du programme correspond à la situation « idéale ».

Selon les données finalement disponibles, les méthodologies d'estimation d'effet propre utilisables pourront être :

- Des estimations sur données de panel d'un modèle à taux de croissance aléatoire, si le programme est zoné et que l'on souhaite mesurer de performances sur données agrégées ;
- des méthodes de *matching* avec et/ou sans remise ainsi que des estimations en doubles différences éventuellement combinés pour un programme zoné sans possibilité d'accès à des données de panel mais un accès à des informations avant et après programme par zones et/ou par individus ciblés ;
- enfin, si les données disponibles se « limitent » à des caractéristiques initiales et à des performances finales des individus et/ou zones ciblées (bénéficiaires et non bénéficiaires), on devra se « contenter » d'utiliser de *matching* avec ou sans remise.

Eléments bibliographiques

- Bénabou R., Kramarz F. et Prost C., 2004. Zones d'éducation prioritaire : quels moyens pour quels résultats ? Une évaluation sur la période 1982-1992. *Economie et Statistique*, 380 : 3-34.
- Boarnet M.G. et Bogart W.T., 1996. Enterprises Zones and local employment: Evidence from New Jersey. *Journal of Urban Economics*, 40: 198-215.
- Bondonio D. et Engberg J., 2000. Enterprises Zones and Local Employment: Evidence from the State's Programs. *Regional Science and Urban Economics*, 20: 510-549.
- Bondonio D., 2000. Statistical Methods to Evaluate Geographically Targeted Economic Development Programs. *Statistica Applicata*, 12(2): 177-204.
- Bondonio D., 2002. Evaluating Decentralized Policies: A Method to Compare the Performance of Economic Development Program across Different Regions or States. *Evaluation*, 8(1): 101-124.
- Bondonio D., Greenbaum R., 2005. Do Business Investment Incentives Promote Employment in Declining Areas? Evidence from EU Objective 2 Regions. *European Urban and Regional Studies*, à paraître.
- Cochran W. et Rubin D., 1973. Controlling Bias in Observational Studies. *Sankhya*, 35, 417-446.
- Crépon B. et Iung N., 1999. *Innovation, emploi et performances*. Document de travail, Direction des Etudes et Synthèses Economiques, G9904, Paris : INSEE.
- Dehejia R.H. et Wahba S., 2002. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94: 1053-1061.
- European Commission (1999). *Evaluating Socio-structural Programmes. Vol 2: Selection and Use of Indicators for Monitoring and Evaluation. Vol 3: Principal Evaluation Techniques and Tools*. MEANS Collection. Brussels: Office for Official Publications of the European Communities.
- Heckman J., Lalonde R. et Smith J., 1999. The Economics and Econometrics of Active Labor Market Programs. In O.Ashenfelter and D.Card (eds), *Handbook of Labor Economics*, vol.III, North Holland, pp.1865-2097.
- Heckman J.J. et Hotz V.J., 1989. Choosing among Alternative Non experimental methods for Estimating the Impact of Social Programs: the Case of Manpower Training. *Journal of the American Statistical association*, 84(408): 862-890.
- Heckman J.J., 1979. Sample Selection as a Specification Error. *Econometrica*, 47(1): 153-161.
- Ham J. C. et LaLonde R.J. (eds), 2005. Experimental and non-experimental evaluation of economic policy and models. Numéro spécial du *Journal of Econometrics*, 125(1-2): 1-377.
- Papke L.E., 1994. Tax Policy and Urban Development. Evidence from the Indiana Enterprise Zone Program. *Journal of Public Economics*, 54: 37-49.
- Ravallion M., 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. *World Bank Economic Review*, 15(1): 115-140.
- Ray M. et Hite J., 2000. A Quasi-Experimental Analysis of the Effectiveness of Credit Pooling for Infrastructures Financing on Economic Growth in Rural Southern Counties. *Contractor Paper 00-07*, Lexington: University of Kentucky, TVA Rural Studies Program.
- Rephann T.J. et Isserman A., 1994. New Highways as Economic Development Tools: An Evaluation using Quasi-experimental Matching Methods. *Regional Science and Urban Economics* 24: 723-751.
- Rosebaum P. et Rubin D., 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods. *American Statistician*, 39: 35-39.
- Rubin D., 1974. Estimating causal effect of treatment in randomized and non randomized studies. *Journal of Educational Methods*, 66: 688-701.

Smith J. A. et Todd P. E., 2005. Does matching overcome LaLonde's critique of non experimental estimators?, *Journal of Econometrics*, 125: 305-353.

Schmitt, Lofredi, Berriet-Sollic et Lépicier, 2003. *Impact Evaluation of the EU Programme for Rural Development in Burgundy: Correcting for Selection Bias*. CESAER Working Papers: 2003/6.

Smith J.A. et Todd P.E., 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, 125: 305-353.

Toulemonde J., 1995. Should Evaluation Be Freed From Its Causal Links? *Evaluation and Program Planning*, 18(2): 79-90.