

# Les méthodes économétriques d'évaluation

Denis Fougère\*

## ■ Introduction

Les méthodes d'évaluation sont aujourd'hui très fréquemment utilisées pour estimer les effets des interventions publiques en matière de santé, d'éducation ou d'emploi. Elles ont également donné lieu ces dernières années à de nombreuses avancées méthodologiques<sup>1</sup>. Les techniques disponibles sont certes multiples, mais il est important de rappeler que leurs spécificités et leurs hypothèses conditionnent fortement les résultats<sup>2</sup>. Pour l'essentiel, ces méthodes sont fondées sur une comparaison entre des individus, ménages ou entreprises bénéficiant de la réforme ou de l'intervention publique que l'on souhaite évaluer et des individus, ménages ou entreprises n'en bénéficiant pas. Au cœur de la démarche statistique d'évaluation, est de ce fait souvent posée la question des biais de sélectivité : le recours à une subvention, à un crédit d'impôt ou à un dispositif public d'aide ou d'accompagnement est une décision qui met en œuvre un choix individuel, relevant au moins en partie du domaine de la rationalité. En particulier, cette décision ne peut être indépendante de la façon dont l'agent évalue par lui-même les conséquences de son choix. Si l'on ne tient pas compte de cette évaluation individuelle, qui s'apparente à une autosélection, l'on risque de produire des estimations biaisées des effets de l'intervention publique en comparant directement les situations des deux groupes, bénéficiaires et non bénéficiaires de cette intervention.

Pour limiter les conséquences des biais de sélectivité, les statisticiens et les économètres construisent généralement un groupe de contrôle dont les caractéristiques observables se rapprochent le plus possible de celles des agents bénéficiaires de l'intervention publique. Cette démarche n'est valide que dans le cas où la sélectivité opère seulement sur la base de caractéristiques observables (telles que l'âge des personnes ou l'effectif des entreprises). Elle est toutefois jugée trop restrictive par les statisticiens et économètres,

\* Directeur de recherche au CNRS et membre du Laboratoire de microéconométrie du Centre de recherches en économie et statistique (CREST, Paris).

1. Voir Heckman, Lalonde et Smith (1999) pour une synthèse très complète.

2. Sur la sensibilité des résultats de l'évaluation aux méthodes, le lecteur pourra consulter, entre autres exemples, les contributions de Lalonde (1986), Heckman et Hotz (1989), Heckman et Robb (1985), Heckman, Tobias et Vytlačil (2000).

qui lui préfèrent l'hypothèse d'une sélectivité provenant à la fois des caractéristiques individuelles observables et inobservables <sup>1</sup>.

Il est important à ce stade de bien comprendre la différence entre l'*évaluation ex ante*, reposant sur la simulation des réformes possibles à partir d'un modèle économétrique structurel estimé avant la mise en place des réformes, et l'*évaluation ex post*, proposant une évaluation réalisée à l'aide de données expérimentales ou non, mais collectées après mise en place des réformes. Ces deux approches sont toutefois complémentaires comme l'ont montré notamment Ichimura et Taber (2002), Todd et Wolpin (2005), ou bien encore Lise, Seitz et Smith (2005), en utilisant les résultats de méthodes économétriques d'évaluation *ex post* pour élaborer et calibrer des exercices d'évaluation *ex ante*.

Le cadre statistique général adapté à la démarche évaluative a été à l'origine développé par les statisticiens dans un domaine très éloigné de celui de l'évaluation des politiques économiques, à savoir celui de l'évaluation des traitements dans le domaine biomédical. Mais, dans ce champ, les statisticiens ont proposé des outils, concepts et méthodes, dont le domaine d'application est bien plus large que celui des essais cliniques. En particulier, ces concepts et méthodes ont été fréquemment utilisés en microéconométrie pour l'évaluation des programmes de formation et d'aide sociale. Mais, à n'en pas douter, leur champ d'application potentiel en économétrie est bien plus large. De ce point de vue, notre synthèse ne se propose pas de recenser l'ensemble des applications aujourd'hui réalisées, mais plutôt d'illustrer les méthodes passées en revue à l'aide d'exemples.

Nous présentons tout d'abord le modèle canonique de l'évaluation, introduit par Rubin (1974). Le succès de ce modèle, qui est proche du modèle économétrique d'autosélection de Roy (1951), provient du fait qu'il permet de définir clairement tout à la fois l'effet causal de la politique que l'on souhaite évaluer et la nature du biais de sélectivité. Les deux caractéristiques importantes de l'effet causal tel qu'il est défini dans ce cadre, sont, d'une part, son hétérogénéité dans la population et, d'autre part, son caractère inobservable. Il faut comprendre que c'est cette non-observabilité qui ne permet pas d'identifier librement la distribution de l'effet causal dans la population.

Nous considérons ensuite la situation dans laquelle 1) les observations analysées ne résultent pas d'expérimentations contrôlées, et 2) la sélection des

<sup>1</sup>. Les premiers travaux français d'évaluation reposent sur ce second type d'hypothèse. Ces travaux, recensés dans la synthèse rédigée par Fougère, Kramarz et Magnac (2000), sont fondés sur des modèles de transition en temps continu (Bonnal, Fougère et Sérandon, 1997), ou en temps discret (Magnac, 2000). Ces études ont examiné les effets de deux types de politiques actives, d'une part les stages et dispositifs d'insertion des jeunes sur le marché du travail, et d'autre part, les allègements de charges patronales sur l'emploi des moins qualifiés. Les conclusions de ces études corroborent les résultats des travaux conduits dans d'autres pays : les programmes de formation et de stages plus particulièrement destinés aux jeunes chômeurs peu qualifiés n'ont en général pas d'effets à court terme sur leurs salaires et leur employabilité, à moins que ces stages et formations aient un contenu suffisant en formation. Toutefois, les politiques d'allègement des charges patronales ont des effets substantiels sur l'emploi notamment des travailleurs dont le salaire est proche du salaire minimum.

agents éligibles à l'intervention évaluée dépend uniquement de caractéristiques individuelles observables par l'analyste. Cette situation a été particulièrement examinée par les économètres. Elle les a conduits à proposer de nouveaux estimateurs, en particulier les estimateurs par appariement, introduits à l'origine par Rubin (1977) et développés plus récemment par Heckman, Ichimura et Todd (1998). Ces estimateurs sont présentés dans la partie qui suit. Puis nous détaillons les étapes de la mise en œuvre pratique de l'estimation par appariement et nous présentons plusieurs études utilisant cette procédure. Enfin les deux dernières parties du texte sont respectivement consacrées à des méthodes moins restrictives que l'estimation par appariement puisqu'elles permettent d'envisager des situations dans lesquelles l'accès des individus à l'intervention évaluée peut dépendre tout à la fois de leurs caractéristiques observables et non observables. Ces deux méthodes, relativement faciles à mettre en œuvre (ce qui explique leur fréquente utilisation), sont la méthode des doubles différences (*differences in differences*) et la régression avec discontinuité (*regression discontinuity*)<sup>1</sup>.

Nous concluons cette synthèse en soulignant l'utilité de chacune de ces méthodes non expérimentales, mais aussi en insistant sur leurs limites. Tirant parti des travaux d'Heckman *et al.* (1998), basés sur des échantillons de contrôle expérimentaux et non expérimentaux, nous tentons de dresser une carte des différentes situations possibles et des estimateurs à mettre en œuvre dans chacun des cas.

## ■ Le modèle générique pour l'évaluation

### *Définitions et notations*

Le modèle canonique de l'évaluation a été introduit par Rubin en 1974. Ce modèle, assez général, est adapté à la situation dans laquelle un *traitement* peut être administré ou non à un individu. Le terme de traitement se réfère aux premiers travaux ayant permis de développer ce cadre conceptuel, travaux qui concernaient l'évaluation de l'efficacité des traitements dans le domaine médical. Bien qu'il ne soit pas le plus approprié, il est utilisé en économétrie pour qualifier une intervention publique, une réforme fiscale, une politique de subvention, un programme de formation, ou bien un programme d'aide sociale que l'on cherche à évaluer.

De manière formelle, nous supposons que, pour chacun des individus  $i$  d'un échantillon de taille  $N$ , nous observons l'ensemble suivant de variables aléatoires. L'accès au programme (*i. e.* au traitement) est représenté par une variable aléatoire  $T$ , qui prend la valeur 1 si l'individu accède au programme,

<sup>1</sup>. Ces deux méthodes n'étaient pas exposées dans l'article de synthèse écrit par Brodaty, Crépon et Fougère (2007). De ce fait, l'article présente complète cette synthèse antérieure, à laquelle il emprunte sa première partie.

0 si ce n'est pas le cas. L'efficacité du programme est mesurée au travers d'une variable de résultat, notée  $Y$ . En fait, le modèle de Rubin repose sur l'existence de deux variables latentes de résultat, notées  $Y_1$  et  $Y_0$ , selon que l'individu reçoit le traitement ( $T=1$ ) ou non ( $T=0$ ). Ces variables correspondent aux *résultats potentiels* (*potential outcomes*) du programme. Elles ne sont jamais simultanément observées à la même date pour un même individu. Ainsi, pour un individu traité,  $Y_1$  est observée tandis que  $Y_0$  est inconnue. Dans ce cas, la variable  $Y_0$  correspond au résultat qui aurait été réalisé si l'individu n'avait pas été traité. On dit aussi que la variable  $Y_0$  représente le résultat *contrefactuel*. Pour un individu non traité, on observe au contraire  $Y_0$ , tandis que  $Y_1$  est inconnue. La variable de résultat observée peut donc se déduire des variables potentielles et de la variable de traitement par la relation :

$$Y = T Y_1 + (1-T) Y_0 \quad (1)$$

Seul le couple  $(Y, T)$  est observé pour chaque individu.

### ***Paramètres d'intérêt***

L'*effet causal* du traitement est défini pour chaque individu par l'écart  $\Delta = Y_1 - Y_0$  qui représente la différence entre ce que serait la situation de l'individu s'il était traité et ce qu'elle serait s'il ne l'était pas. L'effet causal a ainsi deux caractéristiques importantes :

- il est *inobservable*, puisque seule une des deux variables potentielles est observée pour chaque individu ;
- il est *individuel*, et de ce fait il existe une distribution de l'effet causal dans la population.

La distribution de l'effet causal n'est toutefois pas identifiable, tout simplement parce que l'effet causal est inobservable. Néanmoins, grâce à des hypothèses sur la loi jointe du triplet  $(Y_0, Y_1, T)$ , on peut identifier certains paramètres de la distribution de l'effet causal à partir de la densité des variables observables  $(Y, T)$ . Deux paramètres font généralement l'objet d'un examen spécifique. Il s'agit de *l'effet moyen du traitement dans la population* :

$$\Delta^{ATE} = E(Y_1 - Y_0) \quad (2)$$

et de *l'effet moyen du traitement dans la population des individus traités* :

$$\Delta^{TT} = E(Y_1 - Y_0 | T=1) \quad (3)$$

Ces deux paramètres ne sont égaux que sous certaines hypothèses très restrictives. D'une façon générale, les conditions nécessaires à l'identification du premier paramètre sont plus exigeantes que celles nécessaires à l'identification du second. En effet, pour ce dernier, les hypothèses ne portent que sur la loi de  $Y_0$  et  $T$ . En particulier, si les variables de résultat sont indépendantes de la variable d'accès au traitement, c'est-à-dire si  $(Y_0, Y_1) \perp\!\!\!\perp T$ , il est possible d'identifier les deux paramètres d'intérêt  $\Delta^{ATE}$  et  $\Delta^{TT}$  préalablement

définis. En effet, si cette condition (suffisante) est satisfaite, ces deux paramètres d'intérêt deviennent :

$$\begin{aligned}\Delta^{ATE} &= E(Y_1) - E(Y_0) = E(Y_1 | T=1) - E(Y_0 | T=0) \\ &= E(Y | T=1) - E(Y | T=0)\end{aligned}\quad (4)$$

et

$$\begin{aligned}\Delta^{TT} &= E(Y_1 | T=1) - E(Y_0 | T=1) = E(Y_1 | T=1) - E(Y_0 | T=0) \\ &= E(Y | T=1) - E(Y | T=0)\end{aligned}\quad (5)$$

Dans ce cas, les deux paramètres sont égaux et peuvent être estimés simplement comme la différence des moyennes des variables de résultat observées dans le groupe des individus traités et dans le groupe des individus non traités. On peut remarquer que l'identification du second paramètre nécessite une hypothèse moins forte. En ce cas, il suffit en effet que les variables aléatoires  $Y_0$  et  $T$  soient indépendantes.

Dès lors que la propriété d'indépendance précédente n'est plus satisfaite, l'estimateur naturel formé par la différence des moyennes des variables de résultat est affecté d'un *biais de sélection*. En effet,

$$\begin{aligned}E(Y | T=1) - E(Y | T=0) &= E(Y_1 | T=1) - E(Y_0 | T=0) \\ &= E(Y_1 | T=1) - E(Y_0 | T=1) + E(Y_0 | T=1) - E(Y_0 | T=0) \\ &= \Delta^{TT} + B^{TT}\end{aligned}\quad (6)$$

Le biais de sélection est le terme  $B^{TT} = E(Y_0 | T=1) - E(Y_0 | T=0)$  dans l'expression précédente. Ce biais trouve son origine dans le fait que la situation moyenne des individus qui ont reçu le traitement n'aurait pas été la même en l'absence de traitement que celle des individus n'ayant pas reçu le traitement. Il en est ainsi parce que ces deux populations ne sont pas identiques, sauf dans le cas particulier d'une expérience contrôlée.

### ***Le principe des expérimentations contrôlées***

L'évaluation par expérimentation contrôlée essaie de s'affranchir du biais de sélection en proposant d'affecter de manière aléatoire (par tirage au sort) les individus potentiellement intéressés par un dispositif d'aide ou de subvention, un programme de formation, etc., à un groupe de traitement qui bénéficiera de ce dispositif ou programme, ou bien à un groupe de contrôle qui n'en bénéficiera pas. Un bilan des expérimentations relatives à l'évaluation des programmes d'aide à l'emploi introduits durant les vingt dernières années est contenu dans les deux articles de synthèse rédigés par Fougère (2000a, 2000b).

Depuis le début des années 1960, les autorités fédérales et locales des États-Unis d'Amérique ont recouru à des expériences contrôlées pour évaluer des politiques publiques dans des domaines aussi divers que la santé, le logement, l'impôt sur les revenus du travail, l'insertion et la réinsertion sociale, les économies d'énergie ou la tarification de la consommation d'électricité.

Mais c'est dans le domaine des programmes de formation et d'emploi que les expériences menées ont donné lieu aux débats les plus vifs et aux contributions méthodologiques les plus significatives.

Le fait que le tirage au sort ajoute de l'incertitude dans le processus d'accès au programme peut conduire les individus les plus adresses au risque à ne pas se porter candidats. Parallèlement, si les agences chargées de la mise en œuvre des programmes ou interventions offrent un nombre de places limité ou si elles sont évaluées sur la base des résultats de l'expérimentation, elles peuvent être de ce fait incitées à sélectionner les candidats les plus qualifiés ou ceux pour lesquels le programme est potentiellement le plus bénéfique. Les expériences contrôlées peuvent par ailleurs être soumises à des biais de substitution, apparaissant dès lors que les membres du groupe de contrôle ont accès à des substituts proches du traitement ou du programme évalué. Dans le cas du programme américain *Job Training Partnership Act*, les enquêtes de suivi réalisées par l'administration indiquaient que 65 % des membres du groupe de traitement reçurent des aides (aide à la recherche d'emploi, formation, etc.) durant les dix-huit mois suivant la procédure d'affectation par tirage au sort. Toutefois, des enquêtes complémentaires montrèrent que, durant la même période, 32 % des membres du groupe de contrôle et 20 % environ des individus éligibles ne participant pas à l'expérimentation reçurent une formation dispensée par un autre organisme. Au cours de cette expérimentation, les individus du groupe de contrôle eurent donc accès à un volume de formation supérieur à celui dont bénéficia la population éligible ne participant pas au programme.

Les conditions qui permettent de recourir à l'expérimentation et déterminent son protocole sont au nombre de trois :

1. les participants à l'expérience doivent être pleinement informés des avantages que leur participation peut leur procurer comme des risques qu'elle peut leur faire courir (*principe de consentement éclairé*);
2. dans les cas où l'on anticipe que le programme peut provoquer des risques ou des torts, il faut prévoir de compenser les individus qui seront affectés au groupe de traitement; dans le cas où l'on anticipe que le programme peut être bénéfique, il faut prévoir de compenser ceux qui seront affectés au groupe de contrôle (*principe de compensation*);
3. enfin, une expérimentation n'est envisageable que dans les cas où l'on ignore *a priori* l'ampleur de l'effet du traitement; si le traitement s'avère préjudiciable, il est préférable d'observer ses effets sur un nombre limité d'individus consentants que sur un très grand nombre d'individus contraints (*principe de précaution*).

Le lecteur intéressé trouvera dans le récent article de Levitt et List (2009) une présentation très complète des champs d'application des expérimentations sociales, et dans la synthèse de Deaton (2009) un point de vue plus critique.

## ■ Identification sous l'hypothèse d'indépendance conditionnelle à des caractéristiques observables

Lorsque l'on souhaite évaluer des interventions publiques, par exemple des programmes de formation ou des politiques de subvention, à l'aide de données d'observation (non expérimentales), on fait face à deux populations, les bénéficiaires et les non-bénéficiaires, qui diffèrent de par la distribution des caractéristiques individuelles observables qui affectent vraisemblablement la participation à ces programmes. En pratique donc, l'indépendance (inconditionnelle) entre les variables latentes de résultat ( $Y_0$ ,  $Y_1$ ) et l'affectation au traitement  $T$  est une hypothèse très improbable. Une condition moins restrictive consiste à considérer qu'il existe un ensemble  $X$  de variables observables conditionnellement auquel la propriété d'indépendance entre les résultats latents et l'affectation au traitement est vérifiée.

Une façon alternative de formuler cette restriction est de considérer que conditionnellement aux variables observables individuelles  $X$ , on se situe dans le cadre d'une expérience contrôlée, c'est-à-dire avec une affectation aléatoire au traitement. La « randomisation » serait maintenant assurée par les covariables  $X$  plutôt que par un processus aléatoire d'affectation. Ainsi, la distribution de chacun des résultats potentiels est identifiable, puisque leur distribution conditionnelle aux variables observables  $X$  l'est. En conséquence, il est également possible d'identifier l'ensemble des paramètres de chacune de ces distributions, et en particulier l'effet moyen du traitement, de même que l'effet moyen du traitement pour le groupe des individus traités.

Remarquons toutefois que, sous l'hypothèse d'indépendance conditionnelle des résultats aux variables observables  $X$ , la distribution des variables latentes de résultat est identifiée, mais leur loi jointe ne l'est pas. Ceci implique en particulier que les moments d'ordre supérieur à 1 de la loi de la différence  $Y_1 - Y_0$  (en particulier sa variance) ne sont pas identifiés.

Sous cette hypothèse d'indépendance conditionnelle, le principe de l'estimation consiste à utiliser les informations dont on dispose sur les individus non traités pour construire pour chaque individu traité un contrefactuel, c'est-à-dire une estimation de ce qu'aurait été sa situation s'il n'avait pas été traité. Considérons par exemple l'effet causal du traitement sur les traités :

$$\begin{aligned}\Delta^{TT} &= E(Y_1 - Y_0 \mid T=1) = E(Y - Y_0 \mid T=1) \\ &= E[Y - E(Y_0 \mid X, T=1) \mid T=1] = E[Y - E(Y_0 \mid X, T=0) \mid T=1] \\ &= E[Y - E(Y \mid X, T=0) \mid T=1]\end{aligned}$$

Le problème est donc d'estimer pour chaque individu traité de caractéristiques  $x_i$  la quantité  $E(Y \mid X = x_i, T=0) = g(x_i)$ . L'estimateur final est alors obtenu comme la moyenne des écarts de la situation des individus traités et du contrefactuel construit.

Sa formule est donc :

$$\hat{\Delta}^{TT} = \frac{1}{N_1} \sum_{i \in I_1} \{y_i - \hat{g}(x_i)\} \quad (8)$$

où  $I_1$  est le sous-échantillon des individus traités, défini par  $I_1 = \{i \mid T_i = 1\}$ ,  $N_1$  est le nombre d'individus traités et  $\hat{g}(x_i)$  est l'estimateur de la fonction  $g(x_i) = E(Y \mid X = x_i, T=0)$ .

## ■ Estimation des paramètres d'intérêt

### *Estimateur par appariement sur les caractéristiques observables*

La méthode d'estimation par appariement a été initialement proposée par Rubin (1977). Selon lui, c'est la méthode qui se rapproche le plus des expérimentations contrôlées parce que la mise en œuvre de ces deux méthodes (expérimentation et appariement) ne dépend pas directement de la valeur prise par les variables de résultat (par exemple, la qualité de l'emploi ou le niveau de salaire atteint) permettant d'évaluer l'efficacité du dispositif. L'expérimentation repose sur le tirage au sort des individus traités et non traités, indépendamment de ce que sera leur résultat à l'expérience, et la méthode de l'appariement repose sur l'élaboration préalable d'un outil, le score de propension, qui n'en dépend également pas. Ce score mesure la probabilité d'accéder au dispositif pour chaque individu, qu'il en soit bénéficiaire ou non, indépendamment des résultats du passage par le dispositif. Si ce score est bien construit, il joue le rôle du tirage au sort dans les expériences contrôlées : plus précisément, il rend les caractéristiques observables (âge, sexe, qualification, etc.) des individus, tout au moins celles qui sont censées agir sur l'accès au dispositif, indépendantes du résultat du passage par ce dispositif. Ce score permet donc avant tout d'*équilibrer* la distribution de ces variables dans les groupes de traitement et de contrôle<sup>1</sup>, c'est-à-dire de rendre ces deux groupes semblables du point de vue de la distribution des variables agissant sur la probabilité d'accès au dispositif. Les autres méthodes non expérimentales, telles que la régression avec discontinuité ou la méthode des doubles différences, que nous évoquerons plus loin, ne permettent pas d'atteindre cet objectif, car elles sont directement et exclusivement fondées sur l'analyse des variables de résultat<sup>2</sup>.

1. Rosenbaum et Rubin (1983) insistent fortement sur cet aspect crucial de la méthode, malheureusement ignoré par de nombreux praticiens : un score de propension adapté est avant tout un outil d'équilibrage (un *balancing score*) des distributions des variables individuelles, autres que les variables de résultat, dans les deux groupes de traitement et de contrôle.

2. Pour autant, il n'est pas possible d'affirmer que la méthode de l'appariement avec score de propension élimine systématiquement le biais de sélection. Ainsi, Heckman, Ichimura et Todd (1998) ont montré qu'en certains cas au moins, les biais résiduels obtenus après application de la méthode d'appariement peuvent être de l'ordre de grandeur de l'effet du programme lui-même.



Cette méthode associe à chaque individu  $i$  traité un individu non traité, noté  $\tilde{i}(i)$ , dont les caractéristiques sont identiques à celles de l'individu  $i$ . Si l'on note  $X$  le vecteur des caractéristiques individuelles (supposées observables), cette méthode suppose que l'on observe, pour chaque individu traité, un individu non traité comparable, au sens des caractéristiques  $X$ , c'est-à-dire tel que  $X_{\tilde{i}(i)} = X_i$ . La quantité  $Y_{\tilde{i}(i)}$  est un estimateur de l'espérance du résultat potentiel  $Y_{0i}$  conditionnellement au score de l'individu  $i$  :

$$Y_{\tilde{i}(i)} = \hat{E}(Y_{0i} | T_i = 0, x_i) = \hat{E}(Y_{0i} | T_i = 1, x_i) \quad (9)$$

L'estimateur de Rubin pour l'effet moyen du traitement sur les traités est alors égal à :

$$\hat{\Delta}_R^{TT} = \frac{1}{N_1} \sum_{i \in I_1} (y_i - y_{\tilde{i}(i)}) \quad (10)$$

La propriété d'indépendance conditionnelle nécessite en général la prise en compte d'un nombre important de variables de conditionnement. Pour cette raison, l'appariement peut être difficile à réaliser en pratique. En effet, pour certains individus traités, on peut n'observer aucun individu non traité ayant exactement les mêmes caractéristiques. L'estimateur proposé par Rubin consiste en fait à choisir un individu non traité aussi proche que possible de l'individu traité.

### ***Estimation par appariement sur le score de propension***

Le problème de dimension du vecteur  $X$  est fortement réduit par une propriété mise en évidence par Rosenbaum et Rubin (1983). La propriété d'indépendance conditionnellement à des variables observables implique celle d'indépendance conditionnellement à un résumé de dimension un, qui est la probabilité de traitement, ou score de propension à être traité, noté  $P(X) = \Pr(T = 1 | X)$ .

**Proposition** (Rosenbaum, Rubin, 1983) : Si la variable de résultat  $Y_0$  est indépendante de l'accès au traitement  $T$  conditionnellement aux observables  $X$ , alors elle est également indépendante de  $T$  conditionnellement au score de propension  $P(X)$  :

$$Y_0 \perp\!\!\!\perp T | X \implies Y_0 \perp\!\!\!\perp T | P(X)$$

En raison de cette propriété, il n'est pas nécessaire d'apparier les individus sur l'ensemble des variables de conditionnement. Il suffit de les apparier sur leur score de propension, lequel constitue un résumé unidimensionnel de l'ensemble de ces variables. L'individu non traité noté  $\tilde{i}$ , qui est apparié avec l'individu traité  $i$ , est alors défini par  $P(x_{\tilde{i}}) = P(x_i)$ .

Néanmoins, les propriétés asymptotiques de cet estimateur sont insuffisamment connues : l'estimateur est-il convergent ? Est-il asymptotiquement normal ? Quelle est sa vitesse de convergence ? Aucun résultat ne permet

de décrire le comportement de l'estimateur lorsque le nombre d'individus traité devient grand.

### ***Estimateur par appariement avec fonction noyau (kernel matching)***

Les travaux d'Heckman, Ichimura et Todd (1997, 1998) ont permis de répondre à ces questions. Heckman et ses coauteurs proposent en particulier d'utiliser des estimateurs à noyau, qu'ils montrent être convergents (en racine carrée de  $N$ ) et asymptotiquement distribués selon des lois normales. Le plus simple de ces estimateurs s'écrit :

$$\hat{E}(Y_0 | P(x) = P(x_i)) = \sum_{j \in I_0} \frac{K((P(x_j) - P(x_i))/h)}{\sum_{j \in I_0} K((P(x_j) - P(x_i))/h)} y_j \quad (13)$$

où  $I_0$  est l'ensemble des individus non traités, défini par  $I_0 = \{i | T_i = 0\}$ ,  $N_0$  est le nombre d'individus non traités,  $K$  est une fonction noyau (*kernel function*), continûment différentiable, symétrique par rapport à 0, et telle que  $\int_{-\infty}^{+\infty} K(u) du = 1$ , et  $h$  la fenêtre d'estimation (*bandwith function*)<sup>1</sup>.

Chaque individu non traité participe ainsi à la construction du contrefactuel de l'individu  $i$ , avec une importance qui varie selon la distance entre son score et celui de l'individu considéré. L'estimateur final de l'effet du traitement conditionnellement au fait d'être traité est ainsi :

$$\hat{\Delta}_{KM}^{IT} = \frac{1}{N_1} \sum_{i \in I_1} \left\{ y_i - \sum_{j \in I_0} \frac{K[(P(x_j) - P(x_i))/h]}{\sum_{j \in I_0} K[(P(x_j) - P(x_i))/h]} y_j \right\} \quad (14)$$

Heckman, Ichimura et Todd (1998) montrent que, sous certaines hypothèses de régularité, cet estimateur est convergent, asymptotiquement normal, avec une vitesse de convergence en racine carrée de  $N$ .

1. La fonction « noyau » est un opérateur qui pondère symétriquement les valeurs des observations  $y_j$ , en accordant un poids plus élevé à celles de ces observations (d'individus non traités) dont la probabilité d'être traité est « proche » de celle de l'individu traité  $i$ . Cette distance est inversement proportionnelle à la largeur de la « fenêtre »  $h$ , qui peut être choisie de façon optimale. Les fonctions « noyau » les plus usuelles sont le noyau gaussien, le noyau quartique et le noyau d'Epanechnikov. Le lecteur trouvera dans Härdle et Linton (1994) une très bonne présentation des méthodes d'estimation non paramétrique.

## ■ Mise en œuvre des procédures d'estimation par appariement d'échantillons

### *Le support de la distribution du score de propension*

La question du support des distributions du score conditionnel au traitement est essentielle dans ce type d'analyse<sup>1</sup>. Son importance a été soulignée par Heckman, Ichimura et Todd (1998) qui ont montré qu'elle constitue une source prépondérante de biais dans l'estimation de l'effet causal du traitement. Dans les méthodes d'estimation par appariement ou par régression, il est nécessaire de pouvoir construire pour chaque individu traité de score  $p_0$  un contrefactuel à partir des individus non traités, c'est-à-dire de pouvoir estimer  $E(Y|P(X) = p_0, T = 0)$  pour déterminer l'effet causal du traitement dans la population des individus traités.

Une estimation non paramétrique de cette quantité impose que l'on dispose pour chaque individu traité d'individus non traités dont les scores ont des valeurs proches du score de l'individu traité. En d'autres termes, la densité du score des individus non traités ne doit pas être nulle au voisinage des valeurs du score de l'individu traité que l'on considère.

On ne peut donc construire de contrefactuel que pour les individus dont le score appartient à l'intersection des supports des distributions des scores des individus traités et des individus non traités. Ceci conduit à la conclusion que même sous l'hypothèse d'indépendance conditionnelle à des observables, on ne peut pas systématiquement estimer  $\Delta^{ATE}$ , puisque  $E(\Delta|P(X))$  ne peut être estimé que pour les individus dont le score appartient au support commun de la distribution des scores. *L'estimateur obtenu in fine est alors un estimateur local* : on est seulement en mesure d'estimer  $E(\Delta|P(X) \in S \cap T = 1)$  ou  $E(\Delta|P(X) \in S \cap T = 0)$ ,  $S$  étant le support commun défini par  $S = S_T \cap S_{NT}$ ,  $S_T$  étant le support de la distribution du score des individus traités et  $S_{NT}$  celui des individus non traités.

Cette condition de support a une autre implication : le modèle servant à la construction du score, c'est-à-dire expliquant le traitement à partir des variables de conditionnement, ne doit pas être trop « bon ». Dans le cas extrême où l'on expliquerait parfaitement le traitement, les densités du score dégénéreraient chacune en un point (de probabilité 1), le score pour les individus non traités prenant toujours la valeur 0, celui pour les individus traités toujours la valeur 1. Les supports seraient alors disjoints et aucun appariement ne serait possible.

1. Rappelons qu'en topologie, le support d'une fonction continue numérique (à valeurs réelles ou complexes) définie sur un espace topologique  $X$  est l'adhérence de l'ensemble des points en lesquels la fonction ne s'annule pas. En particulier, c'est une partie fermée de  $X$ . La notion de support d'une distribution généralise la notion de support d'une fonction continue. Puisque le score de propension est une probabilité, donc un nombre réel compris entre 0 et 1, le support de la distribution des scores estimés est le sous-ensemble de l'intervalle ]0, 1[ sur lequel la densité des scores estimés n'est pas nulle.

Pour bien comprendre cette condition de support, il faut garder présente à l'esprit l'idée initiale de Rubin : conditionnellement à un ensemble de variables explicatives  $X$  (ou conditionnellement au score de propension), on se trouve dans le cas d'une expérience contrôlée. La persistance de cette composante aléatoire de l'affectation au traitement conditionnellement à des observables est ainsi essentielle dans la procédure d'appariement.

### *Les étapes de l'estimation*

Les différentes méthodes d'évaluation précédemment exposées se font toutes en plusieurs étapes. La première consiste à «expliquer» la variable  $T$  d'affectation au traitement par les caractéristiques observables  $X$ . Pour cela, une solution commode est d'estimer un modèle de type Logit (ou un modèle Logit multinomial lorsqu'il existe plusieurs catégories de traitement). Cette étape est informative dans la mesure où elle propose une description de l'affectation au traitement. Le choix des variables de conditionnement est essentiel dans cette étape, et il faut conserver à l'esprit que ce qui importe n'est pas une description aussi fidèle que possible de la probabilité du traitement, mais simplement la détermination des variables nécessaires à l'obtention de la propriété d'indépendance. Introduire un trop grand nombre de variables peut avoir des conséquences néfastes sur l'estimation à plusieurs titres. D'abord, la description de la variable de traitement étant meilleure, les supports des distributions des scores des individus traités et des individus non traités risquent de se dissocier davantage ; les possibilités d'appariement seront alors plus restreintes. Mais surtout, introduire trop de variables de conditionnement peut conduire à biaiser les estimations. Il est possible en effet que la propriété d'indépendance soit satisfaite pour un ensemble de variables de conditionnement, mais qu'elle ne le soit plus lorsque l'on ajoute d'autres variables de conditionnement, quand bien même celles-ci seraient significatives dans l'estimation du score.

La deuxième étape consiste à déterminer le support commun des densités des scores des deux groupes d'individus. Pour cela, plusieurs techniques peuvent être employées :

1. on peut par exemple exclure les observations dont le score de propension estimé est proche de 1 ou de 0 ;
2. Dehejia et Wabba (1999) suggèrent de supprimer toutes les observations du groupe de contrôle pour lesquelles le score de propension estimé est inférieur au minimum des scores de propension estimés dans le groupe de traitement, et appliquer la même règle pour le maximum ;
3. Heckman, Ichimura et Todd (1998) préfèrent supprimer les observations du groupe de contrôle dont les covariables ont une densité inférieure à un certain seuil.

Cette deuxième étape est importante, et l'oublier pourrait, là encore, biaiser les estimations. C'est en outre une importante source d'imprécision des estimations, l'estimation du contrefactuel pour les individus n'appartenant pas au support commun étant biaisée et très imprécise.

Mais il faut en cet endroit, vérifier une autre conséquence de l'hypothèse. L'appariement réalisé sur la base du score de propension doit en principe équilibrer la distribution des variables  $X$  dans les deux groupes, traitement et contrôle. Cette propriété du score de propension, ainsi défini comme un score d'équilibrage (*balancing score*), est un corollaire du théorème de Rosenbaum et Rubin (1983). Elle signifie que :

$$X \perp\!\!\!\perp T \mid P(X)$$

En d'autres termes, si après conditionnement par le score  $P(X)$ , la variable  $T$  dépend encore de  $X$ , le modèle utilisé pour estimer  $P(X)$  est mal spécifié, ou bien l'hypothèse CIA ( $Y_0 \perp\!\!\!\perp T \mid X$ ) n'est pas valide. Pour tester la validité de la propriété d'équilibrage des variables  $X$  dans les deux groupes, plusieurs procédures ont été proposées : parmi elles, citons le calcul du biais standardisé (Rosenbaum, Rubin, 1985), le  $t$ -test (Rosenbaum, Rubin, 1985), le pseudo- $R^2$  (Sianesi, 2004) et le test de stratification (Dehejia, Wabba, 1999).

La dernière étape est celle de l'estimation à proprement parler. Il suffit pour cela d'appliquer les formules, par exemple celle proposée par Heckman, Ichimura et Todd (1998) pour l'estimateur par appariement fondé sur les noyaux. Les résultats d'Heckman *et al.* (1998) permettent en outre de garantir la convergence de l'estimateur et sa normalité asymptotique, avec une vitesse de convergence en  $\sqrt{N}$ . L'écart-type de l'estimateur est obtenu en appliquant les méthodes du bootstrap, ce qui consiste à répliquer l'ensemble de la procédure d'estimation sur un échantillon tiré aléatoirement avec remise dans l'échantillon initial, et à déterminer l'écart-type de la distribution de l'ensemble des estimateurs ainsi obtenus. Notons que l'écart-type doit aussi prendre en compte le fait que le score n'est pas connu et est de ce fait estimé. Chaque étape du bootstrap doit ainsi comprendre non seulement l'appariement sur l'échantillon tiré, mais aussi l'estimation du score. L'estimation de l'écart-type peut être de ce fait coûteuse en temps de calcul.

### ***Cas avec plusieurs traitements***

Supposons maintenant qu'il existe non plus un seul, mais  $K$  traitements, et que l'accès à l'un de ces traitements soit représenté par la variable aléatoire  $T \in \{1, \dots, K\}$ ,  $K > 1$ . À ces  $K$  traitements correspondent  $K$  variables de résultats  $Y_1, \dots, Y_K$ . On ajoute à ces  $K$  variables, la variable de résultat associée à l'absence de traitement  $Y_0$ . Seule l'une de ces  $(K+1)$  variables de résultats potentiels est observée.

Comme dans le cas avec un seul traitement, on suppose que les variables de résultat sont conditionnellement indépendantes du processus d'affectation aux traitements, étant donné les valeurs prises par l'ensemble de covariables observables  $X$  :

$$(Y_0, Y_1, \dots, Y_K) \perp\!\!\!\perp T \mid X$$

Imbens (2000) et Lechner (2001) ont montré que sous cette hypothèse :

$$(Y_k, Y_{k'}) \perp\!\!\!\perp T \mid \Pi^{k/k'}(X), \forall \{k, k'\}$$

où :

$$\Pi^{k/k'}(X) = \frac{\Pi^k(X)}{\Pi^k(X) + \Pi^{k'}(X)}$$

et

$$\Pi^k(X) = \Pr(T = k \mid X)$$

Cette généralisation produit plusieurs paramètres d'intérêt qui ne sont pas symétriques :

$$E(Y_k - Y_{k'} \mid T = k, X) \neq E(Y_{k'} - Y_k \mid T = k', X)$$

L'estimation de ces paramètres est réalisée en appariant le résultat pour chaque bénéficiaire du traitement  $k$  à la moyenne pondérée des résultats des bénéficiaires du traitement  $k'$  sur la base du score conditionnel  $\Pi^{k/k'}(X)$ .

### ***Quelques exemples d'application***

De nombreuses études ont été consacrées à l'évaluation des programmes de formation offerts aux chômeurs. Deux études françaises récentes ont utilisé les méthodes d'estimation par appariement pour évaluer les effets de politiques actives de l'emploi. Ces deux études reposent sur l'hypothèse d'une sélection sur caractéristiques observables, et sur celle d'indépendance entre la variable de résultat et la variable d'accès au dispositif conditionnellement aux variables observables.

L'étude de Brodaty, Crépon et Fougère (2002) examine les relations entre la durée de chômage avant l'entrée dans un dispositif d'aide à l'emploi et l'effet de ce dispositif sur l'employabilité ultérieure des bénéficiaires. Elle met en évidence l'existence d'effets différenciés sur l'employabilité des chômeurs de courte et de longue durée. À cette fin, Brodaty *et al.* (2002) développent un outil général d'évaluation qui permet de tenir compte à la fois de la multiplicité des dispositifs et de celle des dates d'entrée dans ces dispositifs. Un modèle de durée de chômage à risques concurrents sert à obtenir les scores de propension qui sont ensuite utilisés pour mettre en œuvre les estimations par appariement de l'effet des dispositifs. L'application concerne certaines mesures d'aide à l'emploi mises en place durant les quinze dernières années. L'analyse empirique exploite deux ensembles de données individuelles, qui ont été collectées au cours de deux périodes distinctes mais relativement similaires du point de vue de la conjoncture macroéconomique. Il s'agit de l'enquête « Suivi des chômeurs » conduite par l'INSEE entre 1986 et 1988, et de l'enquête « Trajectoires des demandeurs d'emploi » réalisée par la DARES entre 1995 et 1998. Les résultats montrent que, pour ce qui est de l'employabilité des jeunes chômeurs, les dispositifs d'aide à l'emploi

ont été généralement moins efficaces à la fin des années 1990. Toutefois, certains dispositifs semblent avoir bénéficié plus aux jeunes chômeurs de longue durée. Cela a été particulièrement le cas des stages de formation à la fin des années 1980, et des emplois aidés dans le secteur marchand dix ans plus tard. Plus récemment, toujours à l'aide des méthodes d'appariement, Cavaco, Fougère et Pouget (2008) ont évalué les effets des passages par des conventions de conversion, dispositif aujourd'hui supprimé, sur le retour à l'emploi stable.

## ■ L'estimateur des doubles différences

### *Le cadre général*

La méthode des doubles différences est très souvent utilisée dans les applications économétriques. Son principe est très simple. Il se déduit directement du cadre des modèles à effets fixes individuels et temporels, utilisés en économétrie des données de panel. La forme générale du modèle est la suivante :

$$Y_{it} = \gamma T_{it} + \alpha_i + \lambda_t + \varepsilon_{it} \quad \text{avec } i = 1, \dots, N \text{ et } t = 1, \dots, T \quad (15)$$

Dans cette expression,  $Y_{it}$  représente la variable de résultat pour l'agent  $i$  à la date  $t$ ,  $T_{it}$  est la variable indicatrice du traitement pour l'agent  $i$  à la date  $t$  (qui vaut 1 s'il est traité, 0 sinon),  $\gamma$  est un paramètre à estimer, qui représente l'effet du traitement (ici supposé *constant*),  $\alpha_i$  est un effet fixe individuel,  $\lambda_t$  est un effet temporel *commun* à tous les agents. Les termes  $T_{it}$ ,  $\alpha_i$  et  $\lambda_t$  sont potentiellement corrélés, alors qu' $\varepsilon_{it}$  est un aléa centré, homoscedastique, non corrélé à  $T_{it}$ ,  $\alpha_i$  et  $\lambda_t$ .

Seuls les agents appartenant à un groupe particulier (par exemple, une classe d'âge, les habitants d'une ville ou d'une région donnée, etc.) sont traités à partir de la date  $t = \tau$ . Ceci implique que :

$$T_{it} = 1 \text{ pour } i \in \mathbf{Tr} \text{ et } t \geq \tau$$

$$T_{it} = 0 \text{ pour } i \in \mathbf{C} \text{ ou } t < \tau$$

où  $\mathbf{Tr}$  est le groupe de traitement et  $\mathbf{C}$  le groupe de contrôle. Le principe de la méthode est d'éliminer les effets fixes et temporels à l'aide de deux différences successives. La première différence permet d'éliminer les effets fixes :

$$\Delta Y_{it} = \gamma \Delta T_{it} + \Delta \lambda_t + \Delta \varepsilon_{it}$$

où  $\Delta$  est l'opérateur de différences premières défini par :

$$\Delta Y_{it} = Y_{it} - Y_{it-1}, \Delta \lambda_t = \lambda_t - \lambda_{t-1}, \Delta \varepsilon_{it} = v_{it} = \varepsilon_{it} - \varepsilon_{it-1}$$

Posons maintenant  $t = \tau$  et  $t - I = \tau - I$  (ou  $t \geq \tau$  et  $t - I \leq \tau - I$ ).

Si  $i \in \mathbf{Tr}$ ,  $\Delta T_{it} = I$ , ce qui implique que  $\Delta Y_{it}^{Tr} = \gamma + \mu_t^{Tr} + v_{it}^{Tr}$ .

Si  $i \in \mathbf{C}$ ,  $\Delta T_{it} = 0$ , ce qui implique que  $\Delta Y_{it}^C = \mu_t^C + v_{it}^C$ .

La deuxième différence élimine les effets temporels communs :

$$\gamma = E(\Delta Y_{it}^{Tr}) - E(\Delta Y_{it}^C)$$

puisque  $\mu_t^{Tr} = \mu_t^C$  et  $E(v_{it}^{Tr}) = E(v_{it}^C) = 0$ . L'estimateur des doubles différences est alors défini comme :

$$\hat{\gamma} = \left( \overline{\Delta Y_{it}^{Tr}} \right) - \left( \overline{\Delta Y_{it}^C} \right) \quad (16)$$

avec

$$\overline{\Delta Y_{it}^k} = \frac{1}{N_k} \sum_{i=1}^{N_k} (Y_{it}^k - Y_{it-1}^k), k \in \{Tr, C\}$$

L'extension au cas avec covariables se fait en posant :

$$Y_{it} = \gamma T_{it} + Z_{it} \delta + \alpha_i + \lambda_t + \varepsilon_{it} \quad \text{avec } i = 1, \dots, N \text{ et } t = 1, \dots, T \quad (17)$$

L'estimateur des doubles différences est équivalent à l'estimateur « *within* » dans le modèle projeté sur l'espace orthogonal aux effets fixes et temporels :

$$WY = WX\beta + W\varepsilon$$

où  $X_{it} = [T_{it}, Z_{it}]$ ,  $\beta = (\gamma, \delta)'$ ,  $W = I_{NT} - G - B_N - B_T$ ,  $G = J_{NT}/NT$ ,

$J$  étant une matrice carrée composée de 1,  $B_N = [(I_N \otimes J_T)/T] - G$

et  $B_T = [(J_N \otimes I_T)/N] - G$ . L'estimateur *within* du vecteur de paramètres est alors :

$$\hat{\beta}_{\text{cov}} = (X'WX)^{-1} (X'WY) \quad (18)$$

et

$$V(\hat{\beta}_{\text{cov}}) = \sigma^2 (X'WX)^{-1}$$

Les hypothèses cruciales pour la mise en œuvre de l'estimateur des doubles différences sont les suivantes :

1. les effets temporels sont supposés être communs aux deux groupes de traitement et de contrôle ( $\lambda_t^{Tr} = \lambda_t^C$ ) au moins en  $t = \tau$  et  $t - I = \tau - I$ ;
2. il ne peut y avoir ni attrition, ni sélection endogène entre  $(\tau - I)$  et  $\tau$ ; ainsi, cette méthode ne peut être appliquée à une réforme fiscale (telle que la prime pour l'emploi) qui peut influencer la participation au marché du travail, variable par nature discrète; en ce cas, certaines personnes peuvent être sans emploi en  $(\tau - I)$  mais employées en  $\tau$  (ou l'inverse);
3. les termes d'erreur et les covariables sont supposés non autocorrélés; dans le cas contraire, l'écart-type de l'effet du traitement est systématiquement sous-estimé, comme l'ont montré Bertrand, Duflo et Mullainathan (2004); en ce cas, l'hypothèse nulle d'absence d'effet du traitement, *i. e.*  $H_0: \gamma = 0$ , est trop souvent rejetée.



## ***Exemples***

Card et Krueger (1994) ont appliqué cette méthode pour estimer les effets du salaire minimum sur l'emploi. Ils utilisent pour cela le fait qu'en 1992, le salaire minimum a été augmenté de 80 % au New Jersey. Leur idée consiste à mesurer l'évolution de l'emploi dans les fast-foods situés des deux côtés de la frontière entre les états du New Jersey et de Pennsylvanie. Appliquant la méthode des doubles différences aux deux années 1991 et 1993, ils trouvent de manière inattendue que la hausse de l'emploi (en équivalent à temps complet) a été de 0,47 point dans le New Jersey, alors que l'emploi (en équivalent à temps complet) a baissé de 2,28 points en Pennsylvanie dans le même temps. Dans un article antérieur, Card (1990) a mesuré les effets de l'immigration sur l'emploi non qualifié en mettant en œuvre une fois encore cette méthode. Il utilise comme « *expérience naturelle* » l'arrivée massive d'immigrés cubains non qualifiés à Miami en 1979. Il estime l'évolution du taux de chômage des travailleurs non qualifiés blancs, noirs et hispaniques à Miami, mais aussi dans quatre autres villes américaines (Atlanta, Houston, Los Angeles et Tampa) qui n'ont pas connu de flux migratoires significatifs au cours de la même période (1979-1981). Il constate une hausse de 1,3 point du taux de chômage des noirs à Miami, mais une hausse de 2,6 points dans les autres villes durant ces années.

Plus récemment, Draca, Machin et Van Reenen (2005) ont estimé les effets de l'introduction d'un salaire minimum en 1999 au Royaume-Uni. Les effets de cette mesure étaient *a priori* différents pour les entreprises employant beaucoup de salariés au niveau du salaire minimum avant la réforme (les entreprises « traitées ») et celles en employant peu (les « contrôles »). Draca *et al.* (2005) trouvent un effet significatif du salaire minimum sur la profitabilité des entreprises : la marge de profit, définie comme le rapport du profit brut aux ventes, a été réduite de 11 à 8 % après introduction du salaire minimum. Mais ils ne trouvent pas d'effet significatif sur la probabilité de défaillance des entreprises.

## ***L'estimateur des doubles différences avec appariement (difference in difference kernel matching)***

La validité de la condition d'identification  $Y_0 \perp\!\!\!\perp T \mid X$ , nécessaire pour la mise en œuvre des techniques d'estimation par appariement, peut être légitimement mise en cause par la présence d'un effet fixe individuel inobservable, noté  $u$ , affectant à la fois les variables potentielles de résultat et la variable de traitement. Une hypothèse plus vraisemblable pourrait donc être  $Y_0 \perp\!\!\!\perp T \mid X, u$ . Cette hypothèse stipule donc qu'une partie des variables nécessaires à la propriété d'indépendance est inobservée.

Ce cadre peut se ramener au cadre développé précédemment dès lors qu'il est possible d'éliminer l'information contenue dans l'élément inobservé par une transformation adaptée de celles-ci. En partant de l'hypothèse

d'indépendance conditionnellement à des observables et à des inobservables, on peut en effet montrer le résultat suivant :

$$Y_0 \perp\!\!\!\perp T \mid X, u \text{ et } g(Y_0, X) \perp\!\!\!\perp u \mid X \Rightarrow g(Y_0, X) \perp\!\!\!\perp T \mid X \quad (19)$$

Un cas intéressant est celui dans lequel la liste des variables de conditionnement comprend les observations des variables de résultat à une date antérieure au traitement, notées  $Y^p$ , avec  $Y^p \subset X$ . Alors, par soustraction, il est possible d'éliminer l'effet individuel :  $(Y_0 - Y^p) \perp\!\!\!\perp u \mid X$ , à l'instar de ce qui est communément effectué avec les méthodes de différences présentée ci-dessus<sup>1</sup>. On obtient alors la propriété d'indépendance conditionnellement à des observables pour les évolutions des variables de résultat, propriété qui s'écrit  $(Y_0 - Y^p) \perp\!\!\!\perp T \mid X$ . Ceci a conduit Heckman, Ichimura et Todd (1998) à une généralisation de l'estimateur par différences de différences, largement utilisé dans le cas de l'évaluation des politiques publiques. La forme de l'estimateur de l'effet moyen du traitement sur les traités est alors :

$$\hat{\Delta}_{DDKM}^{IT} = \frac{1}{N_1} \sum_{i \in I_1} \left\{ \Delta y_i - \sum_{j \in I_0} \frac{K[(P(x_j) - P(x_i))/h]}{\sum_{j \in I_0} K[(P(x_j) - P(x_i))/h]} \Delta y_j \right\} \quad (20)$$

où  $\Delta Y = Y - Y^p$ .

Even et Klein (2008) ont appliqué la méthode de l'estimateur des doubles différences avec appariement à l'évaluation de certains dispositifs d'emploi aidé introduits en France (contrats initiative emploi, contrats emploi solidarité, stages d'insertion et de formation à l'emploi).

## ■ La régression avec discontinuité

La méthode de la régression avec discontinuité est appliquée en général dans les cas où la probabilité de traitement est une *fonction discontinue* de la valeur prise par une covariable autre que la variable de résultat (par exemple, l'âge ou le revenu de l'individu, l'effectif de l'entreprise, etc.).

Deux situations doivent être distinguées :

- le cas avec discontinuité nette (*sharp design*) : ici l'accès au traitement  $T$  dépend de manière déterministe des valeurs prises par un vecteur de variables observables  $Z$ , soit  $T = f(Z)$ , et le point  $z_0$ , auquel il y a discontinuité, est connu avec certitude. Par exemple, si  $Z$  est de dimension 1 :

$$T = 1 \text{ si } Z > z_0 \text{ et } T = 0 \text{ si } Z < z_0 \text{ (ou l'inverse)}$$

- le cas avec discontinuité floue (*fuzzy design*) : ici l'accès au traitement  $T$  est une variable aléatoire conditionnée par  $Z$ , et la probabilité conditionnelle

1. L'article pionnier en la matière est celui d'Ashenfelter (1978).

$$\Pr(T = 1 \mid Z = z) = E(T \mid Z = z) \equiv f(z)$$

est discontinue en  $Z = z_0$ .

Le problème provient du fait que  $Z$  peut être corrélé avec la variable de résultat. Une comparaison des résultats des agents ayant reçu le traitement et de ceux n'y ayant pas eu accès, est, de ce fait, généralement biaisée. Toutefois, il est vraisemblable que les agents proches du point de discontinuité (situés des deux côtés de la limite ou du seuil) soient très semblables.

Posons  $Y = Y_0 + (Y_1 - Y_0)T = \alpha + \beta T$ . Dans le cas d'une discontinuité nette (*sharp design*), et si  $\varepsilon$  est un nombre arbitrairement faible, alors :

$$E(Y \mid Z = z_0 + \varepsilon) - E(Y \mid Z = z_0 - \varepsilon) = E(\beta \mid Z = z_0 + \varepsilon) + E(\alpha \mid Z = z_0 + \varepsilon) - E(\alpha \mid Z = z_0 - \varepsilon)$$

Si les agents proches du seuil sont identiques, on devrait avoir :

$$E(\alpha \mid Z = z_0 + \varepsilon) \cong E(\alpha \mid Z = z_0 - \varepsilon)$$

Les conditions d'identification de la régression avec discontinuité sont donc :

1. **(C1)**  $E(\alpha \mid Z = z)$  est continue en  $Z = z_0$
2. **(C2)** La limite  $\lim_{\varepsilon \rightarrow 0} E(\beta \mid Z = z_0 + \varepsilon)$  existe et est définie.

Sous ces conditions,

$$\lim_{\varepsilon \rightarrow 0} [E(Y \mid Z = z_0 + \varepsilon) - (E(Y \mid Z = z_0 - \varepsilon))] = E(\beta \mid Z = z_0)$$

Remarquons que l'effet moyen du traitement sur les traités, n'est identifié que pour les agents proches du seuil (sauf dans le cas où cet effet est constant dans la population). Toutefois, dans le cas de points de discontinuité multiples, il est possible d'approcher l'identification de cet effet « pour tout  $Z$  » : il est donc possible de tester l'hypothèse d'effet constant du traitement.

Considérons maintenant le modèle plus général :

$$Y = \alpha + \beta T = g(Z) + \beta T + u \tag{21}$$

où  $g(Z) = E(\alpha \mid Z)$  est une fonction de contrôle flexible pour la régression de  $Y$  sur  $T$ . Dans le cas d'une discontinuité floue (*fuzzy design*), Van der Klaauw (2001) propose de remplacer  $T$  par un estimateur de première étape de  $E(T \mid Z)$ . Dans le cas général (*sharp* ou *fuzzy design*), le rapport :

$$\frac{\lim_{\varepsilon \rightarrow 0^+} E(Y \mid Z = z_0 + \varepsilon) - \lim_{\varepsilon \rightarrow 0^+} E(Y \mid Z = z_0 - \varepsilon)}{\lim_{\varepsilon \rightarrow 0^+} E(T \mid Z = z_0 + \varepsilon) - \lim_{\varepsilon \rightarrow 0^+} E(T \mid Z = z_0 - \varepsilon)} \tag{22}$$

identifie l'effet du traitement en  $Z = z_0$

## ■ Conclusion : quelle méthode adopter ?

Il y a donc plusieurs façons de traiter le problème de la sélectivité. Chacune correspond à des hypothèses bien spécifiques. En pratique, il est important de savoir laquelle de ces hypothèses est la mieux adaptée. Les travaux d'Heckman, Ichimura et Todd (1997, 1998) apportent quelques éléments de réponse à cette question. Ils utilisent les données du programme de formation *JTPA (Job Training Partnership Act)* mis en place aux États-Unis. Ce programme a été conçu comme une expérience contrôlée. Il est donc possible d'obtenir une évaluation de son effet en utilisant le groupe de contrôle expérimental. En utilisant un groupe de contrôle non expérimental, il est par ailleurs possible de mettre en œuvre les différents estimateurs non expérimentaux que nous venons de présenter. Il s'agit ensuite d'examiner ceux qui se rapprochent le plus de l'évaluation sur données expérimentales.

Les enseignements de cet exercice sont les suivants :

1. Les méthodes d'appariement sur les niveaux des variables de résultat (plutôt que sur leurs différences premières) sont mal adaptées : il y a persistance d'effets individuels inobservés, même lorsque l'on introduit un grand nombre de caractéristiques individuelles observables.
2. Les méthodes d'appariement appliquées aux variations des variables de résultat, telles que l'estimation par « différences de différences », marchent bien, mais leur bonne performance dépend néanmoins de la richesse des variables de conditionnement. Ces méthodes ne sont pas pour autant la panacée : il subsiste en général un biais qui, bien qu'il ne soit pas le plus important, représente une fraction significative de la valeur de l'effet du traitement. Ceci signifie que les méthodes par appariement ne doivent pas être utilisées lorsqu'il n'est pas possible d'observer l'évolution de la situation des individus, ou lorsqu'il n'est pas possible d'éliminer les effets individuels par simple application des différences premières.
3. Les problèmes relatifs au support des scores de propension ont une réelle importance dans la pratique. Ils reflètent l'hétérogénéité de l'effet du traitement dans la population et mettent l'accent sur les difficultés d'extrapolation des résultats obtenus.
4. L'étude réalisée par Mueser, Troske et Gorislavsky (2003) à l'aide de données non expérimentales complète les résultats obtenus par Heckman, Ichimura et Todd (1997, 1998) ; en outre, elle montre que les choix techniques liés aux méthodes d'appariement (choix de la fonction noyau, choix de la fenêtre, etc.) donnent des résultats assez peu différents les uns des autres, mais aussi de ceux obtenus avec des méthodes plus classiques, telles que les régressions à variables de contrôle ; toutefois, ces choix affectent la précision des estimations.

Au total, il existe aujourd'hui des méthodes alternatives à la démarche expérimentale, souvent jugée coûteuse du double point de vue juridique et politique. Néanmoins, l'utilisation des méthodes non expérimentales dépend

fortement du contexte propre à l'observation. Ce contexte détermine en particulier le recours à tel ou tel type de méthode dans le but de réduire les biais spécifiques au problème de l'évaluation. Mais il conditionne aussi la précision des estimations et donc leur caractère informatif et opérationnel. Enfin, il faut rappeler que les méthodes microéconométriques d'évaluation s'attachent principalement à la mesure de l'effet d'une politique publique sur la situation de certains groupes d'individus bien identifiés (pour l'essentiel, les bénéficiaires et les non-bénéficiaires éligibles). De ce fait, elles ne permettent que rarement d'appréhender les effets plus généraux, par exemple les effets induits à moyen terme, de cette politique. Pour autant, cette limite ne peut, selon nous, disqualifier l'apport de ces méthodes.

## Bibliographie

- ASHENFELTER O. (1978), "Estimating the Effects of Training Programs on Earnings", *Review of Economics and Statistics*, vol. 60, p. 47-57.
- BERTRAND M., DUFLO E., MULLAINATHAN S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?", *Quarterly Journal of Economics*, vol. 119, p. 249-275.
- BONNAL L., FOUGÈRE D., SÉRANDON A. (1997), "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories", *Review of Economic Studies*, vol. 64, p. 683-713.
- BRODATY T., CRÉPON B., FOUGÈRE D. (2001), "Using Kernel Matching Estimators to Evaluate Alternative Youth Employment Programs: Evidence from France, 1986-1988", in LECHNER M., PFEIFFER F. (eds), *Econometric Evaluations of Labour Market Policies*, Physica Verlag, Heidelberg, p. 85-124.
- BRODATY T., CRÉPON B., FOUGÈRE D. (2002), "Do Long-Term Unemployed Workers Benefit from Active Labour Market Policies? Evidence from France 1986-1998", *mimeo*, CREST-INSEE.
- BRODATY T., CRÉPON B., FOUGÈRE D. (2007), « Les méthodes microéconométriques d'évaluation et leurs applications aux politiques actives de l'emploi », *Économie et Prévision*, n° 177, p. 91-118.
- CARD D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market", *Industrial and Labor Relations Review*, vol. 43, p. 245-257.
- CARD D., KRUEGER A. (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", *The American Economic Review*, vol. 84, p. 772-793.
- CAVACO S., FOUGÈRE D., POUGET J. (2008), "Estimating the Effect of a Retraining Program on the Re-Employment Rate of Displaced Workers", *CEPR, Discussion Paper N° 7094*, Londres, à paraître dans *Empirical Economics*.
- DEATON A. (2009), *Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development*, Working Paper, Center for Health and Wellbeing, Princeton University.
- DEHEJIA R., WABBA S. (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, vol. 94, p. 1053-1062.
- DRACA M., MACHIN S., VAN REENEN J. (2005), "Minimum Wages and Firm Profitability", *CEPR Discussion Paper N° 1913*, Londres.
- EVEN K., KLEIN T. (2008), « Les contrats et stages aidés : un profit à moyen terme pour les participants ? Les exemples du CIE, du CES et du Sife », *Économie et Statistique*, n° 408-409, p. 3-32.
- FOUGÈRE D. (2000a), « Expérimenter pour évaluer les politiques d'aide à l'emploi : les exemples anglo-saxons et nord-européens », *Revue française des affaires sociales*, n° 1, janvier-mars, p. 111-144.
- FOUGÈRE D. (2000b), « Accompagnement des chômeurs et sanctions : leurs effets sur le retour à l'emploi », in *Plein Emploi*, Rapport du Conseil d'analyse économique, n° 30, La Documentation française, Paris, p. 313-340.
- FOUGÈRE D. (2007), « Faut-il encore évaluer les dispositifs d'emplois aidés ? », *Économie et Statistique*, n° 408-409, p. 33-38.

- FOUGÈRE D., KRAMARZ F., MAGNAC T. (2000), "Youth Employment Policies in France", *European Economic Review*, vol. 44, p. 928-942.
- HÄRDLE W., LINTON O. (1994), "Applied Nonparametric Methods", in ENGLE R. F., MCFADDEN D. (eds), *Handbook of Econometrics*, vol. 4, North-Holland, Amsterdam, p. 2295-2339.
- HECKMAN J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations", *Journal of Human Resources*, vol. 32, p. 441-462.
- HECKMAN J., HOTZ J. (1989), "Choosing Among Alternative Methods of Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, vol. 84, p. 862-874.
- HECKMAN J., ICHIMURA H., SMITH J., TODD P. (2000), "Characterizing Selection Bias Using Experimental Data", *Econometrica*, vol. 66, p. 1017-1098.
- HECKMAN J., ICHIMURA H., TODD P. (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", *Review of Economic Studies*, vol. 64, p. 605-654.
- HECKMAN J., ICHIMURA H., TODD P. (1998), "Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, vol. 65, p. 261-294.
- HECKMAN J., LALONDE R., SMITH J. (1999), "The Economics and Econometrics of Active Labor Market Programs", in ASHENFELTER O., CARD D. (eds), *Handbook of Labor Economics*, vol. III, Amsterdam, North Holland.
- HECKMAN J., ROBB R. (1985), "Alternative Methods for Evaluating the Impact of Interventions: An Overview", *Journal of Econometrics*, vol. 30, p. 239-267.
- HECKMAN J., TOBIAS J., VYTLACIL E. (2000), "Simple Estimators for Treatment Parameters in a Latent Variable Framework With an Application to Estimating the Return to Schooling", *NBER Working Paper* n° 7950.
- HECKMAN J., VYTLACIL E. (2000), "Local Instrumental Variables", *NBER Working Paper* n° T0252.
- ICHIMURA H., TABER C. (2002), "Semiparametric Reduced Form Estimation of Tuition Subsidies", *CeMMAP Working Paper CWP01/02*, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- IMBENS, G. W. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, vol. 87, p. 706-710.
- IMBENS G., ANGRIST J. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, vol. 62, p. 467-476.
- LALONDE R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, vol. 76, p. 604-620.
- LECHNER M. (2001), "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in LECHNER M., PFEIFFER F. (eds), *Econometric Evaluations of Labour Market Policies*, Heidelberg: Physica Verlag, p. 43-58.
- LEVITT S. D., LIST J. A. (2009), "Field Experiments in Economics: The Past, the Present, and the Future", *European Economic Review*, vol. 53, p. 1-18.
- LISE J., SEITZ S. SMITH J. (2005), "Equilibrium Policy Experiments and the Evaluation of Social Programs", *Working Paper*, Department of Economics, University of Maryland.
- MAGNAC T. (2000), "State Dependence and Heterogeneity in Youth Employment Histories", *The Economic Journal*, vol. 110, p. 805-837.

- MUESER P., TROSKE K., GORISLAVSKY A. (2003), "Using State Administrative Data to Measure Program Performance", *IZA Discussion Paper* N° 786.
- ROSENBAUM P., RUBIN D. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, vol. 70, p. 41-55.
- ROSENBAUM P., RUBIN D. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, vol. 39, p. 33-38.
- ROY A. D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, vol. 3, p. 135-146.
- RUBIN D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non Randomized Studies", *Journal of Educational Psychology*, vol. 66, p. 688-701.
- RUBIN D. (1977), "Assignment to Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, vol. 2, p. 1-26.
- SIANESI B. (2004), "An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s", *Review of Economics and Statistics*, vol. 86, p. 133-155.
- TODD P., WOLPIN K. (2005), "Ex Ante Evaluations of Social Programs", *Working Paper*, University of Pennsylvania, à paraître dans *Annales d'économie et de statistique*.
- VAN DER KLAUW W. (2002), "Estimating the Effect of Financial Aid Offers on College Enrolment: A Regression-Discontinuity Approach", *International Economic Review*, vol. 43, p. 1249-1287.