

Chapitre 2: Efficacité des politiques microéconomiques – un panorama des méthodes d'évaluation

Supposons que l'on désire évaluer l'efficacité d'un programme public de lutte contre la pauvreté ou d'aide à l'emploi.

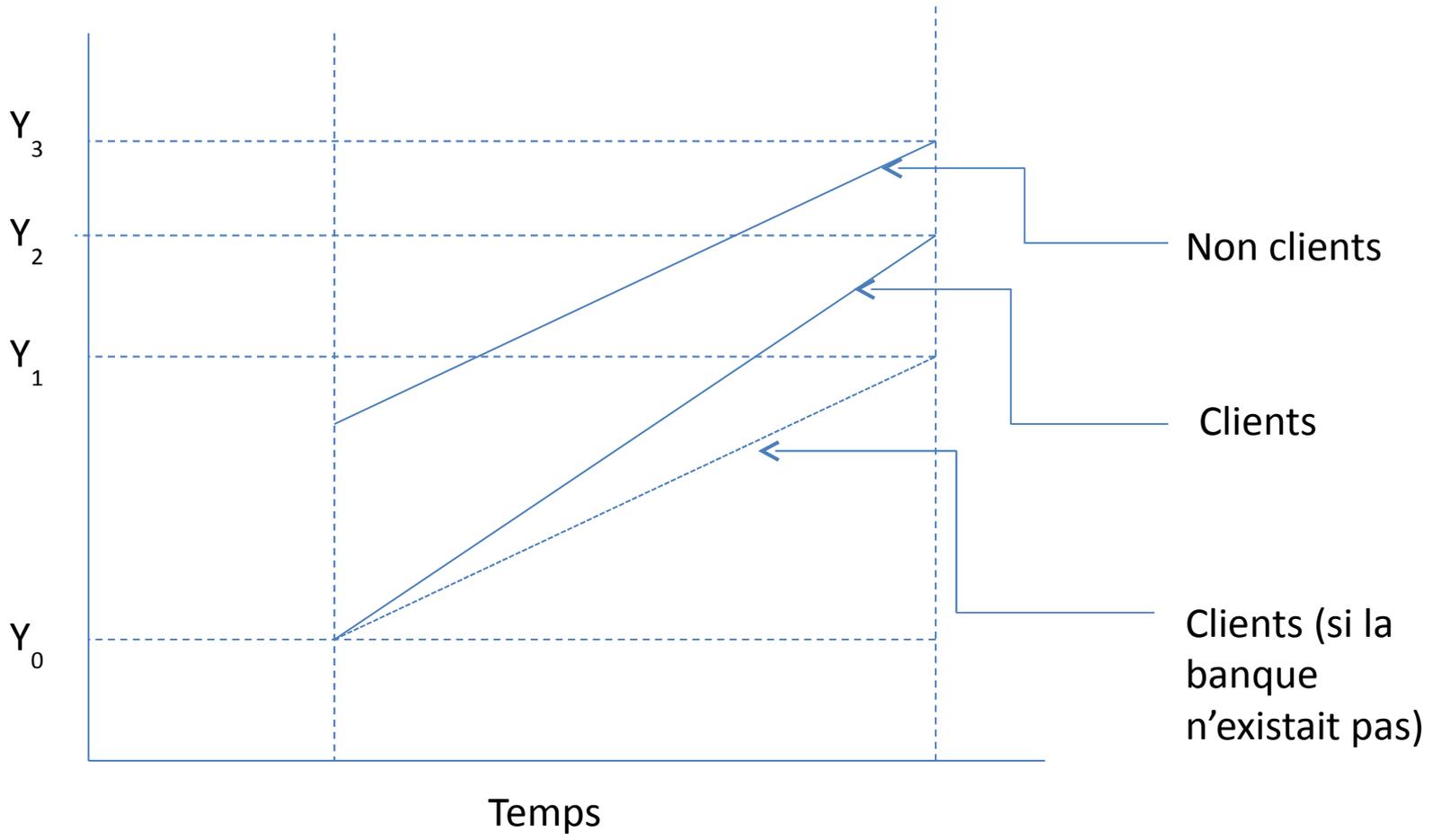
Comme en macro-économie, la grosse difficulté est d'obtenir une image précise de ce que serait la situation en l'absence du programme.

Or il est rarement simple d'obtenir cette image, parce que le programme modifie le comportement des agents.

Deux bases de comparaison potentiellement sources d'erreurs

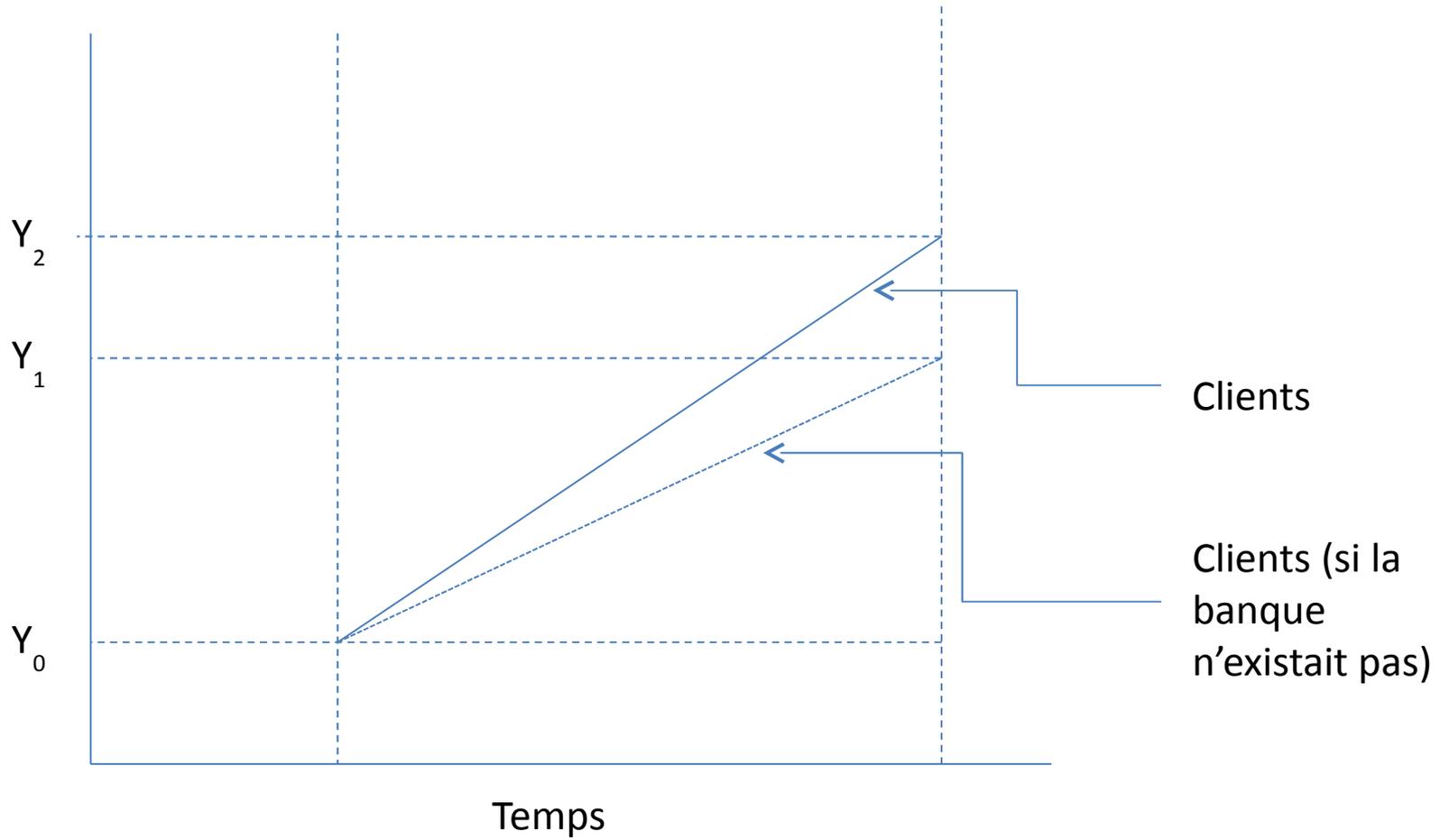
- Comparaison « avec » et « sans »:

Supposons que l'on désire évaluer l'efficacité d'une banque de micro-crédit. Une telle banque prête de l'argent en priorité aux gens pauvres. Supposons que l'on compare les revenus des individus clients de la banque avec ceux d'autres individus qui n'en utilisent pas les services. On trouve que les revenus des clients sont plus faibles que ceux des non clients. Echec de la banque ?



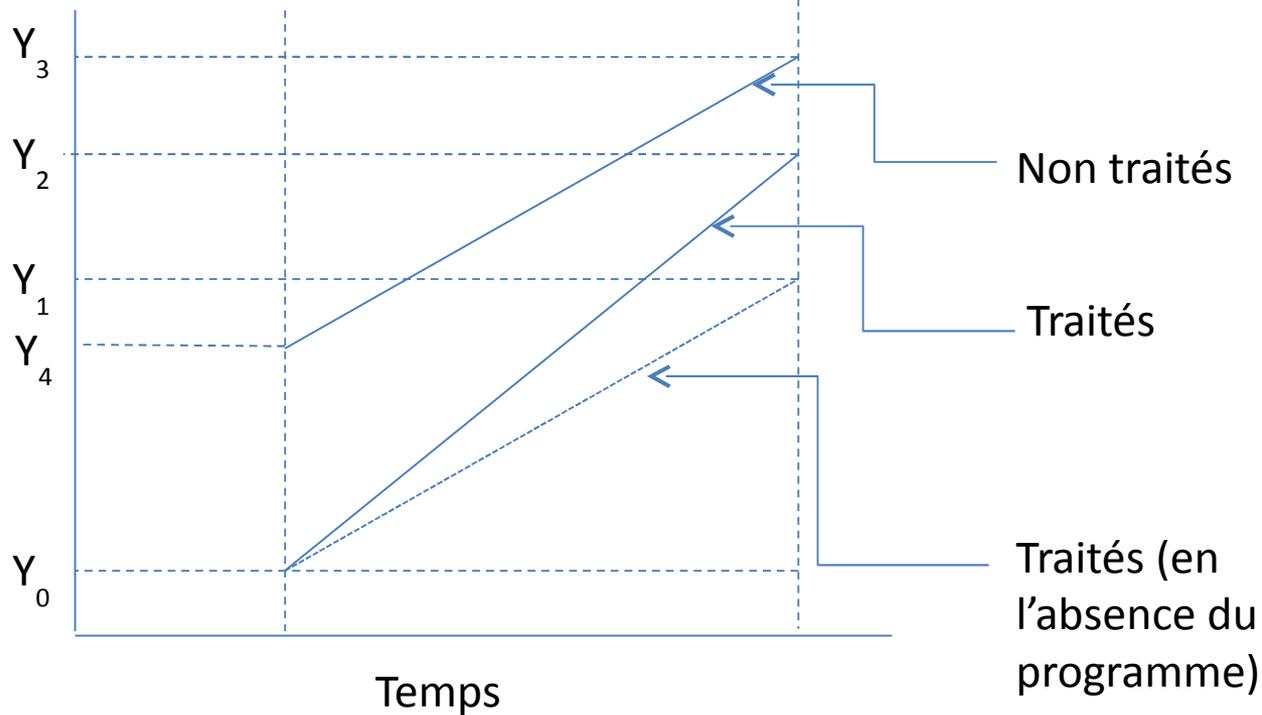
Comparaison « avec et sans »: $Y_3 - Y_2$ – Evaluation correcte: $Y_2 - Y_1$

- Comparaison « avant » - « après »: on peut aussi imaginer d'estimer l'efficacité du programme en comparant la valeur de la variable d'intérêt – disons le revenu – avant et après l'intervention. Dans l'exemple qui nous intéresse: avant et après la prise d'un crédit. Mais c'est négliger la possibilité qu'entre les deux dates d'observation interviennent d'autres phénomènes qui modifient le revenu



Comparaison « avant - après »: $Y_2 - Y_0$ – Evaluation correcte: $Y_2 - Y_1$

- Ces deux exemples montrent l'intérêt à identifier un « groupe de contrôle » qui n'est pas affecté par le programme évalué et qui va pouvoir servir de base de comparaison.



La comparaison « avec et sans » devient pertinente si le groupe de contrôle est bien identifié et si l'on peut tenir compte du fait qu'il est plus riche toutes choses égales par ailleurs: $Y_2 - Y_1 = Y_2 - Y_3 - (Y_1 - Y_3) = Y_2 - Y_3 - (Y_0 - Y_4)$ qui n'est autre que l'estimateur de la différence de différences.

La comparaison « avant-après » est facilitée si l'observation du groupe de contrôle permet de mesurer l'évolution de Y en dehors de l'effet du programme: $Y_2 - Y_1 = Y_2 - Y_0 - (Y_1 - Y_0) = Y_2 - Y_1 - (Y_3 - Y_4)$.

Le problème du biais de sélection

- Toute la difficulté provient de ce que l'on ne peut pas observer ce qu'il serait advenu des personnes concernées par le programme testé en l'absence de celui-ci.
- Pour s'en sortir il faut avoir un « bon » groupe de contrôle, c'est-à-dire un groupe d'individus qui ont des caractéristiques aussi proches que possible que les individus traités, mais qui ne reçoivent pas le traitement.
- *A priori* un tel groupe n'est pas simple à constituer parce que le groupe des individus traités est rarement déterminé selon un processus aléatoire.

- L'équation ci-dessous modélise le problème de base de l'évaluation de programme:

$$Y_i = \alpha X_i + \beta T_i + \epsilon_i$$

où T_i est une indicatrice prenant la valeur 1 si l'individu i est traité et 0 sinon ; X_i est un ensemble de variables de contrôle, caractéristiques de l'individu i et ϵ_i est un terme d'erreur reflétant l'hétérogénéité non observée entre les individus et qui explique également la valeur prise par Y .

- Le problème du biais de sélection provient du fait que le terme d'erreur a de très bonnes chances d'être corrélé avec la valeur prise par l'indicatrice T . En effet, les individus qui participent au programme ne sont en général pas choisis par hasard:
 - Soit ils ont été choisis comme bénéficiaires de façon délibérée par les initiateurs du programme;
 - Soit ils se sont auto-sélectionnés.

- La sélection dans le programme peut reposer sur des caractéristiques observées ou non observées. Lorsqu'elles sont non observées, alors on peut craindre qu'elles déterminent également la variable d'intérêt, Y . On alors: $cov(\varepsilon, T) \neq 0$ et l'estimation de β par les moindres carrés ordinaires conduit à un biais.

Exemple:

On veut mesurer l'efficacité du microcrédit. La difficulté est que les clients de la banque se sont autosélectionnés, selon des caractéristiques non observables et corrélées avec le revenu.

Une telle caractéristique peut être par exemple la motivation de l'individu à sortir d'une situation de pauvreté, dont on peut penser qu'elle n'agit pas que sur la probabilité de solliciter un microcrédit.

Dans un tel cas de figure, imaginons que l'on trouve un effet positif de l'accès au microcrédit sur la variable d'intérêt en employant une méthode qui ne tient pas compte de l'auto-sélection. A quoi est dû cet effet ? Est-ce parce que le microcrédit est vraiment efficace, ou est-ce parce que ceux qui ont sollicité ce type de crédit sont aussi ceux qui ont le plus d'aptitude à obtenir une valeur élevée de Y ?

Présentation générale du problème

Supposons que l'on désire évaluer les bénéfices d'un programme social ou d'aide à l'emploi. On suppose que deux types d'individus sont observés: ceux qui participent, pour lesquels l'indicatrice T prend la valeur 1 et ceux qui ne participent pas, pour lesquels T prend la valeur zéro.

Soit Y la variable mesurant le "résultat" du programme (mesure de bien être, revenu, probabilité de retrouver un emploi etc.). Notons $Y_i(0)$ la valeur de Y si l'individu i ne participe pas au programme et $Y_i(1)$ la valeur s'il participe. Notons enfin Y_i la valeur observée de Y . Pour évaluer l'efficacité du programme il faut calculer la quantité suivante:

$$TTE = E(Y_i(1) | T_i = 1) - E(Y_i(0) | T_i = 1)$$

Le premier terme est la moyenne de Y sur la population des traités et le second la moyenne de Y sur cette même population, si elle n'avait pas subi le traitement.

- TTE est l'effet moyen du traitement sur les traités. Il est différent de l'effet moyen du traitement dans la population qui s'écrit:

$$\begin{aligned}
 ATE &= E(Y_i(1)) - E(Y_i(0)) \\
 &= E((Y_i(1) - Y_i(0)) | T_i = 1) \cdot P(T_i = 1) + E((Y_i(1) - Y_i(0)) | T_i = 0) \cdot P(T_i = 0) \\
 &= TTE \cdot P(T_i = 1) + E((Y_i(1) - Y_i(0)) | T_i = 0) \cdot P(T_i = 0)
 \end{aligned}$$

- L'effet moyen du traitement dans la population est une moyenne pondérée de l'effet du traitement sur les traités et de l'effet qu'il pourrait avoir sur les non traités, autrement dit l'ensemble du reste de la population. Il n'est pas évident que l'effet du traitement sur les traités, même bien évalué, soit identique à l'effet du traitement sur les non traités (problème de la validité externe des résultats).

- La difficulté est que l'on observe pas la valeur de $Y(0)$ dans la population des individus traités. On ne sait donc pas calculer TTE directement. Ce que l'on sait calculer est:

$$\begin{aligned}
 D &= E(Y_i(1) | T_i=1) - E(Y_i(0) | T_i=0) \\
 &= TTE + E(Y_i(0) | T_i=0) - E(Y_i(0) | T_i=1) \\
 &= TTE + B
 \end{aligned}$$

B est le biais qui résulte de la sélection non aléatoire dans le programme. Puisque l'on ne peut pas calculer la valeur de Y pour les participants, s'ils n'avaient pas participé ($E(Y_i(0) | T_i=1)$), on ne sait pas a priori quelle est l'importance de ce biais.

- L'objectif de l'évaluateur est de trouver un moyen d'éliminer ce biais.

- Exemple: supposons que l'on veut évaluer l'effet de la distribution de livres non scolaires dans les écoles. Comme les fonds sont limités, toutes les écoles ne reçoivent pas les livres. Seules celles qui en font la demande peuvent bénéficier du programme. Le programme est évalué en mesurant les connaissances des élèves.
- On peut alors penser que les écoles qui font la demande sont celles qui accordent le plus d'importance à la lecture. Il est alors probable que:

$$E(Y_i(0) | T_i=1) > E(Y_i(0) | T_i=0),$$

de sorte que le biais B est positif et que le calcul de la différence $E(Y_i(1) | T_i=1) - E(Y_i(0) | T_i=0)$ surestime le véritable impact du programme.

Quelles méthodes sont disponibles ?

- Randomisation
- Méthodes d'appariement (Propensity Score Matching – PSM)
- Méthodes de la Double Différence (DD)
- Méthodes instrumentales
- Régressions sur la Discontinuité (RD)
- Méthode du Pipeline

Chapitre 3: La randomisation

- Cette méthode peut être employée lorsqu'il est possible d'intervenir au moment de la mise en place du programme.
- Le cas le plus simple est celui où on peut tirer au hasard les personnes traitées dans la population. Il y a alors indépendance entre le fait d'être dans le groupe de contrôle ou de traitement et la valeur de Y:

$$E(Y_i(1) | T_i=1) = E(Y_i(1) | T_i=0)$$

$$E(Y_i(0) | T_i=1) = E(Y_i(0) | T_i=0)$$

autrement dit, la valeur prise par Y, que l'on ait ou non reçu le traitement, ne dépend pas de l'appartenance au groupe.

- Dans ce cas de figure, le biais de sélection n'existe pas et on a de plus TTE=ATE.
- L'ATE peut être estimé par les MCO en effectuant la régression:

$$Y_i = \alpha + \beta T_i + \epsilon_i$$

La randomisation en pratique

- La validité de l'évaluation doit être assurée à deux niveaux:
 - En *interne*: les estimations de l'impact du programme doivent fournir une valeur sans biais de l'effet du traitement sur les traités
 - En *externe*: l'estimation doit avoir une portée générale, qui va au delà de l'effet estimé sur les traités.
- A cette fin l'échantillon des traités et l'échantillon de contrôle doivent être constitués en suivant un certain nombre de règles.

Un échantillonnage en deux étapes

- Afin d'assurer la validité externe des estimations, on commence par tirer au hasard un échantillon représentatif de la population.
- Ensuite, afin d'assurer la validité interne, on tire dans cet échantillon le sous-échantillon des individus traités. Le reste constitue le sous-échantillon de contrôle.

- Ce protocole est souvent difficile à mettre en place pour les raisons juridiques, pratiques et éthiques évoquées au chapitre 1.
- Lorsque l'on ne peut y parvenir, il est parfois possible de contourner l'obstacle et d'exploiter certains aspects du programme afin de se placer dans un cadre similaire à celui obtenu lorsque le tirage en deux étapes mentionné précédemment peut être réalisé.

- Exploiter le rationnement des ressources (oversubscription): si le nombre de personnes qui désirent participer au programme dépasse les ressources disponibles, il peut alors être nécessaire de tirer au sort parmi les volontaires. Le groupe des non participants forme alors un bon échantillon de contrôle puisqu'ils partagent les mêmes caractéristiques moyennes que le groupe des participants.
- Soit R une indicatrice prenant la valeur 1 si l'individu est tiré au sort et 0 sinon. On peut écrire:

$$\begin{aligned} TTE &= E(Y_i(1) | T_i=1, R_i=1) - E(Y_i(0) | T_i=1, R_i=0) \\ &= E(Y_i(1) | T_i=1) - E(Y_i(0) | T_i=1) \end{aligned}$$

- D'autres méthodes exploitent la mise en place progressive du programme pour créer un groupe traité et un groupe de contrôle à partir des observations sur les individus qui bénéficient déjà du programme (groupe traité) et sur ceux qui en bénéficieront plus tard (groupe de contrôle).
- Enfin, plutôt que de rendre le traitement aléatoire, on peut décider de randomiser l'information sur l'existence du programme. Dans ce cadre on tire au hasard les personnes qui sont informées de l'existence du programme. Certaines vont participer, d'autres non. Il y aura également des participants qui n'ont pas été informés directement. L'idée est alors d'utiliser le fait d'avoir ou non été informé – officiellement – de l'existence du programme comme instrument de la participation.

Difficultés

- La randomisation, bien que permettant d'aplanir beaucoup de difficultés n'est pas sans limite.
 - Effet Hawthorne
 - Contamination de l'échantillon de contrôle
 - Attrition sélective
 - Bouclage macroéconomique

Conséquences de l'effet Hawthorne

- Il est possible que le fait même d'être sélectionné ait un impact sur la variable d'intérêt et ce indépendamment du traitement.
- Dans ce cas il peut arriver que:

$E(Y_i(1) | T_i=1) \neq E(Y_i(1) | T_i=0)$ et la seule chose que la randomisation puisse réellement garantir est que: $E(Y_i(0) | T_i=1) = E(Y_i(0) | T_i=0)$, puisque $Y_i(0)$ est mesuré avant la mise en place du programme et la sélection des échantillons.

- Seul le TTE, $E(Y_i(1) - Y_i(0) | T_i=1)$, peut alors être estimé et de plus il est possible que la différence entre $Y_i(1)$ et $Y_i(0)$ ne résulte pas du programme lui-même, mais de la sélection dans l'échantillon.

Contamination de l'échantillon de contrôle

- Certaines mesures, si elles produisent des effets d'externalité, peuvent avoir un impact au delà de l'échantillon des individus traités.
- Exemple: dans de très nombreux pays les enfants souffrent de vers intestinaux. On voudrait tester l'efficacité d'un traitement médical. Comment procéder ?

- Première méthode: dans les écoles on donne un remède à certains enfants et un placebo à d'autres et on observe ce qui se passe.
- L'expérience conduit à des résultats décevants: les enfants qui reçoivent le médicament n'ont pas moins de vers que les autres.

- Seconde méthode: plutôt que de comparer les enfants traités aux enfants non traités au sein d'un établissement, on choisit de traiter certains établissements et d'autres non.
- On examine alors la prévalence moyenne des taux d'infection entre les établissements traités et non traités.
- Les résultats sont cette fois plus encourageants: dans les écoles traitées les taux d'infection sont significativement plus faibles que dans les écoles non traitées. Et ce en dépit du fait que dans les écoles traitées tous les enfants ne reçoivent pas le traitement.

- Comment expliquer cette différence de résultats ?
- Les enfants traités ne portent plus les vers et donc ne peuvent pas les transmettre aux autres enfants de leur école, ce qui conduit à diminuer le taux d'infection chez les enfants qui ne sont pas traités.
- En comparant l'échantillon des traités avec l'échantillon de contrôle selon la première méthode on aboutit à la conclusion que le traitement n'est pas efficace.

- Il existe d'autres sources de contamination:
 - Si le programme est testé dans une zone géographique déterminée, il est possible que celui-ci ait un impact en dehors de la zone (ex. programmes ayant un impact sur l'activité des entreprises)
 - Inversement, il peut arriver que les individus migrent pour bénéficier du programme. Comme cette migration n'est pas aléatoire, les individus qui appartiennent aux échantillons test et de contrôle ont des caractéristiques différentes de celles de la moyenne de la population.

Attrition sélective

- Il peut aussi arriver que certains individus quittent l'un ou l'autre des deux échantillons. Si cette attrition n'est pas aléatoire, elle peut fausser les résultats de l'évaluation.
- Exemple: Evaluation de l'efficacité d'un programme d'aide aux élèves en difficulté en Inde. Au cours du programme, certains élèves ont abandonné l'école. Si dans les écoles traitées, les élèves sont en moyenne ceux qui ont les moins bonnes performances, alors l'efficacité du programme peut être sur-estimée.

Bouclage macroéconomique

- Les expérimentations sont locales par nature. Pour cette raison, leur validité externe est sujette à caution, car les effets du programme une fois étendu à l'ensemble de la population peuvent ne pas être ceux observés localement (ex. un programme qui améliore la formation des travailleurs peut avoir un effet plus élevé au niveau local que celui attendu au niveau macroéconomique. La raison est qu'un programme localisé n'a pas d'impact sur l'équilibre du marché du travail.)

Chapitre 4: Les méthodes d'appariement (Propensity Score Matching - PSM)

Principe

- Lorsqu'un tirage aléatoire des échantillons test et de contrôle n'a pu être réalisé, il faut trouver un autre moyen de constituer l'échantillon de contrôle.
- Le principe des méthodes d'appariement (Matching) est d'utiliser les caractéristiques observées des participants et des non participants afin de choisir, dans le groupe des non participants, un échantillon d'individus qui ont des caractéristiques moyennes identiques à celles des individus traités.

Mise en oeuvre

- La mise en oeuvre est relativement simple:
 - On commence par estimer un modèle de la participation au programme (probit)
 - Ce modèle permet de prédire, pour chaque individu, la probabilité qu'il participe ou non au programme. Cette probabilité est appelée « score de propension ».
 - Chaque individu membre du groupe test (individus traités) est alors apparié aux individus non traités qui ont le score de propension le plus proche de lui.

Théorie

- Trois conditions doivent être respectées pour pouvoir mettre en oeuvre la méthode PSM:
 - Indépendance conditionnelle: conditionnellement à un ensemble de variables observables X , dont la valeur n'est pas affectée par le traitement, la valeur prise par Y selon que l'individu est traité ou non, ne dépend pas de son affectation au groupe de traitement ou de contrôle.

Autrement dit:

$$E(Y_i(1) | T_i=1, X) = E(Y_i(1) | T_i=0, X)$$

$$E(Y_i(0) | T_i=1, X) = E(Y_i(0) | T_i=0, X)$$

- Dans ce cas le biais, B, est nul conditionnellement à X:

$$B_X = E(Y_i(0)|T_i=0, X) - E(Y_i(0)|T_i=1, X) = 0$$

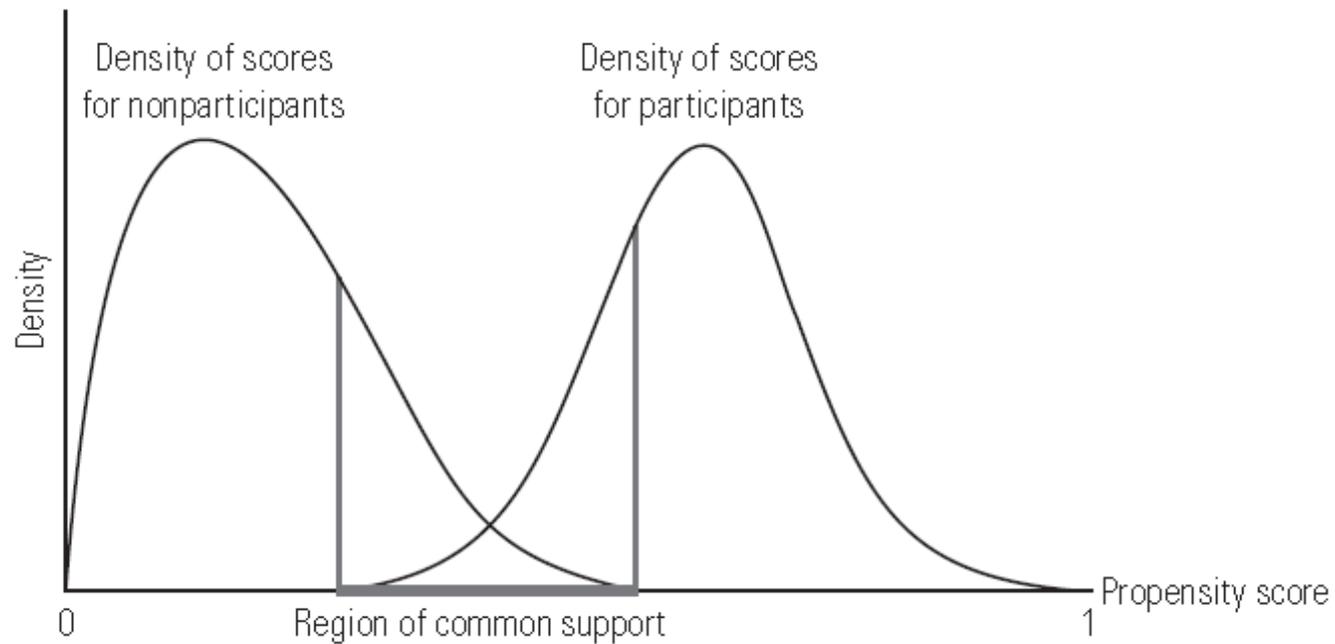
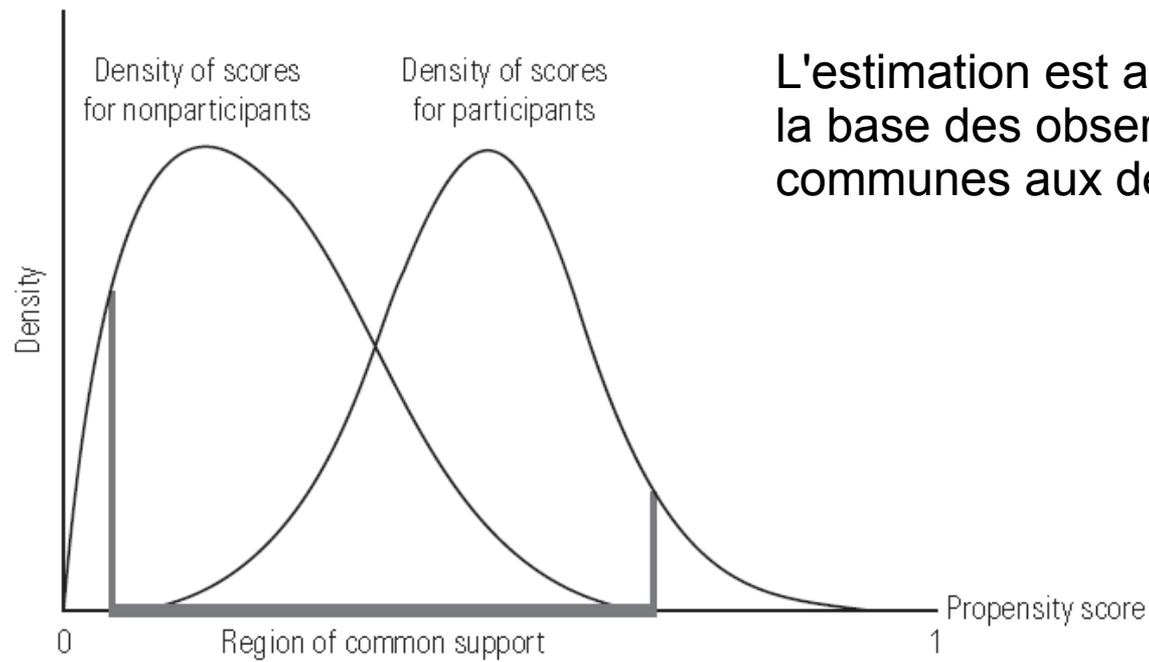
Supposons par exemple que le sexe et l'âge déterminent la probabilité de participer au programme et aient également un effet sur la variable d'intérêt (revenu par ex.). Dans ce cas, la comparaison des revenus moyens des participants et des non participants en l'absence du programme, $Y(0)$, montrerait que les revenus des participants sont différents de ceux des non participants, reflétant la composition par âge et sexe des deux groupes. Conditionner par ces deux variables revient à comparer $Y(0)$ entre individus appartenant à la même classe (âge, sexe) et on ne s'attend alors pas à trouver une différence.

- « Support » commun: cette condition s'écrit de la façon suivante

$$0 < P(T_i = 1 | X) < 1$$

- Elle signifie que les individus traités doivent pouvoir être appariés avec des individus non traités ayant un score de propension proche.

L'estimation est alors effectuée sur la base des observations communes aux deux supports.



- Condition d'équilibre des échantillons: la dernière condition à vérifier pour appliquer la méthode PSM est que les valeurs moyennes de X calculées dans chaque échantillon doivent être proches pour des valeurs proches du score de propension. En effet, même si leur score de propension est identique, deux groupes peuvent avoir des caractéristiques différentes. Lorsque cette condition est violée il est recommandé de reconsidérer la liste des variables X .

Comment effectuer l'appariement ?

- Il est exceptionnel de trouver dans le groupe des participants et des non participants deux individus ayant exactement la même valeur du score de propension. Comment alors effectuer l'appariement ?
- La solution la plus simple consiste à retenir, pour chaque participant, le non participant ayant la valeur du score de propension la plus proche de la sienne (méthode dite du "voisin le plus proche").
L'appariement est effectué en retirant, ou non, les individus appariés du groupe des non participants.
- Ensuite l'effet du programme est évalué en calculant, pour chaque participant, la différence entre la valeur de la variable d'intérêt observée pour le participant et celle observée pour le non participant auquel il est apparié, puis en effectuant la moyenne de ces différences.

Il n'y a pas de raison a priori d'apparier chaque individu du groupe de traitement avec un seul individu du groupe de contrôle. En fait, cette méthode est susceptible de laisser pas mal de place au hasard : même avec un score de propension identique à celui de l'individu traité, le résultat pour l'individu non traité est susceptible de variations incontrôlées qui, si elles n'induisent pas forcément de biais systématique dans l'estimation de l'effet du programme n'en réduisent pas moins la précision. On peut chercher à réduire cette source d'imprécision en appariant chaque individu, i , du groupe de traitement avec plusieurs individus du groupe de contrôle et en évaluant l'effet du programme pour l'individu i en calculant la moyenne des résultats obtenus pour les individus du groupe de contrôle avec lesquels il est apparié.

On peut par exemple choisir d'apparier chaque membre du groupe de traitement avec l'ensemble des membres du groupe de contrôle, mais en donnant un poids plus élevé aux membres "les plus proches" de l'individu traité. Les méthodes dites de "noyau" reposent sur ce principe. De façon plus simple, mais également moins rigoureuse, on peut décider arbitrairement de limiter à quelques-uns le nombre de membres du groupe de contrôle retenus pour l'appariement (par exemple cinq) et effectuer la moyenne arithmétique sur ce groupe.

Autres méthodes

- L'appariement sur la base du score de propension est une solution commode pour éviter la difficulté à trouver deux individus ayant exactement la même valeur de X .
- Il y a cependant d'autres solutions.

- Soient X_i^T et $X_{j(i)}^C$ les valeurs du vecteur X pour un individu i appartenant au groupe de traitement (T) et pour son « jumeau », j, dans le groupe de contrôle (C).
- Comment choisir j ? La façon la plus naturelle consiste à rechercher j tel que $X_{j(i)}^C$ minimise la distance:

$$\|X_i^T - X_{j(i)}^C\|$$

- La distance de Mahalanobis est souvent utilisée:

$$\|X_i^T - X_{j(i)}^C\| = (X_i^T - X_{j(i)}^C)' \Sigma^{-1} (X_i^T - X_{j(i)}^C)$$

où Σ est la matrice de variances-covariances du vecteur X calculée sur les observations du groupe de traitement.

Difficultés

Premièrement, il semble y avoir une contradiction interne à cette méthode: si deux individus ont des caractéristiques identiques pourquoi l'un participe-t-il et l'autre pas ?

La contradiction n'est qu'apparente. En fait, tout se passe comme si, *conditionnellement à X*, la participation ou la non participation au programme était déterminée de façon aléatoire. Le groupe de contrôle a alors toutes les propriétés désirables, qui sont celles d'un groupe de contrôle issu d'une expérience contrôlée.

Mais ceci souligne une difficulté inhérente aux méthodes d'appariement : certains des non participants ont des scores de propension plus faibles que tous les participants. Ces individus doivent être éliminés du groupe de contrôle : en d'autres termes, *le groupe de contrôle doit être constitué de telle sorte que la plage de variation des valeurs du score de propension des non participants retenus dans le groupe soit la même que celle des participants* (condition de support commun).

- Le respect de cette condition peut être problématique si les individus qui sont retirés diffèrent systématiquement de ceux qui appartiennent au support. L'échantillon de contrôle n'est alors pas représentatif de la totalité de la population non traitée, ce qui peut conduire à une estimation biaisée.



L'impact est alors estimé uniquement pour les individus qui ont des caractéristiques observables qui les « placent » dans le support commun.

- L'hypothèse d'indépendance conditionnelle est très forte. Si des variables inobservables déterminent à la fois la participation et la valeur de Y , alors l'estimation par matching sera biaisée. La méthode ne devrait donc être employée que si la sélection sur des variables inobservables est peu probable.