

# 1. Aspects méthodologiques

## 1.1. Le problème fondamental de l'évaluation et la méthode des différences de différences

Depuis longtemps, on utilise des techniques statistiques pour comparer deux groupes d'individus et déterminer si les caractéristiques de la population dont ils sont issus sont identiques. Cependant les premières méthodes utilisées supposent que l'attribution à un groupe est faite de manière purement exogène ou indépendante du phénomène étudié. On compare ainsi les hommes et les femmes, un pays à un autre, etc. Mais, le problème devient plus complexe lorsque l'appartenance à un groupe est décidée par l'individu ou par un processus qui n'est pas indépendant du résultat possible. Dans ce cas, les individus s'autosélectionnent dans un groupe donné suivant leurs caractéristiques et suivant les résultats escomptés dans chacune des situations. Ce mécanisme d'autosélection est très fréquent dans la vie économique moderne. Il faut en tenir compte pour évaluer les résultats obtenus par les groupes d'individus. C'est par exemple, le cas du choix d'une activité qui procure tel ou tel revenu. Dans notre cas, il s'agit donc du choix de la pratique ou non de l'agriculture biologique. Pour nous, les individus sont des exploitations agricoles et la variable d'intérêt est le volume de travail agricole.

Plusieurs méthodes ont été développées pour chercher à évaluer l'effet d'un « traitement » en tenant compte d'éventuels effets de sélection. 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. Par extension, ce terme est utilisé pour désigner toute pratique discriminante dont on cherche à évaluer l'effet (ici le choix du passage en bio). Ces méthodes s'appuient sur le modèle causal proposé par

Rubin en 1974, qui vise à comparer les performances d'individus qui reçoivent le traitement avec celles d'individus qui ne le reçoivent pas.

Dans ce modèle, on suppose que pour chaque unité  $u$ , le traitement cause l'effet suivant :

$$\delta_u = Y_{u1} - Y_{u0}$$

où  $Y_{u0}$  est le résultat qui aurait été observé en l'absence du traitement  $T$ , et  $Y_{u1}$  est le résultat observé en présence du traitement.

Cette définition d'un effet causal suppose que le statut de traitement d'un individu n'affecte pas les résultats potentiels d'autres individus, et inversement. Le problème fondamental de l'inférence causale est qu'il est impossible d'observer la valeur de  $Y_{u1}$  et  $Y_{u0}$  sur la même unité  $u$ . En effet, soit l'unité  $u$  reçoit le traitement et on connaît  $Y_{u1}$ , soit elle ne le reçoit pas et on connaît  $Y_{u0}$ . Il est donc impossible d'observer directement l'effet de  $T$  sur  $u$ . Autrement dit, nous ne pouvons pas inférer l'effet de traitement parce que nous n'avons pas la preuve contrefactuelle. Pour les individus traités, nous ne savons pas ce qui se serait passé en l'absence de traitement.

Étant donné que l'effet causal pour une certaine unité  $u$  ne peut pas être observé, nous cherchons à identifier l'effet causal moyen pour la population dans son ensemble ou pour certaines sous-populations. En particulier, on peut calculer l'effet moyen sur la population traitée (*ATT, Average Treatment on the Treated, Caliendo & Kopeinig, 2008*) :

$$\begin{aligned} \text{ATT} &= E(Y_{u1}|T_u = 1) - E(Y_{u0}|T_u = 1) \\ &= E(Y_{u1} - Y_{u0}|T_u = 1) \end{aligned}$$

Le problème est le même que pour l'effet causal au niveau individuel. En revanche, si on dispose de données en coupes répétées, des doubles différences aussi appelées différences

de différences peuvent être utilisées pour évaluer l'ATT. Cela correspondra à la différence entre le groupe des traités et des non traités de l'évolution du résultat au cours du temps (avant et après le traitement). Ceci revient à supposer que, en l'absence du traitement, le groupe des traités aurait suivi une tendance parallèle à la tendance observée dans le groupe des non traités. Ceci revient donc à faire l'hypothèse que ce qui différencie les traités et non traités est invariant au cours du temps, de sorte que le biais de sélection s'annule par différenciation. D'autre part, les différences de différences permettent d'annuler les effets individuels fixes de chacune des unités.

Dans le cas de deux périodes, notées respectivement  $t'$  et  $t$ , on peut donc estimer l'effet moyen du traitement comme suit (Khandker *et al.*, 2010) :

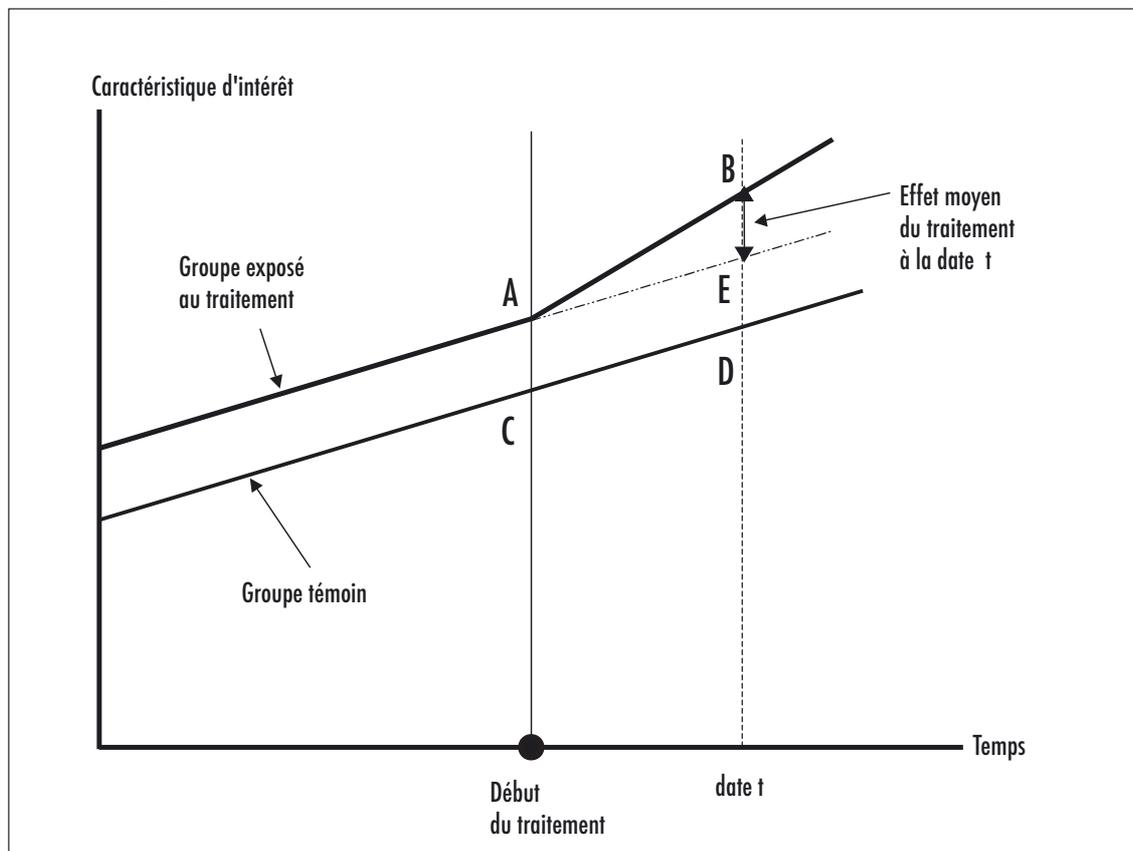
$$ATT = \frac{1}{N_T} \sum_{i \in T} (Y_{it'} - Y_{it}) - \frac{1}{N_C} \sum_{j \in C} (Y_{jt'} - Y_{jt})$$

où  $N_T$  est le nombre d'observations dans le groupe des traités  $T$ , et  $N_C$  est le nombre d'observations dans le groupe de contrôle  $C$  (c'est-à-dire, les non traités).

## 1.2. Combiner différences de différences et appariement sur le score de propension

Idéalement, pour que la double différenciation permette d'estimer l'effet moyen d'un traitement à une date  $t$ , il faudrait que le groupe traité et le groupe de contrôle aient une même tendance d'évolution avant le début du traitement. Le graphique 1 illustre le cas favorable pour lequel la méthode de différence de différence permet d'estimer sans biais l'effet moyen du traitement. Néanmoins, dans notre cas, nous ne disposons pas d'information avant la date de passage au bio. Il ne nous est donc pas possible de déterminer un groupe de contrôle pour lequel on pourrait vérifier directement l'hypothèse de tendance commune.

Graphique 1  
illustration graphique de la double différenciation



Lecture : il est impossible d'observer le point E, c'est-à-dire de savoir quelle valeur aurait pris notre caractéristique d'intérêt à la date  $t$  si le groupe « exposé » n'avait pas reçu de traitement. Mais l'évolution C-D observée sur le groupe témoin permet d'estimer ce point E et donc l'effet moyen du traitement.

Il est néanmoins possible de constituer un groupe de contrôle « proche » du groupe traité en ayant recours à une méthode d'appariement. Le principe de ces méthodes est de comparer chaque exploitation bénéficiant du traitement à des exploitations n'en bénéficiant pas qui lui sont « similaires » du point de vue de certaines caractéristiques observables. Il convient alors de neutraliser les facteurs qui ont un impact sur la variable d'intérêt. Par exemple, on pressent que la taille de l'exploitation joue sur le volume de travail. Il faudrait raisonnablement comparer, dans les deux groupes, des exploitations de taille semblable. Cela devient difficile à faire si on doit neutraliser un grand nombre de facteurs. Pour une exploitation bio donnée aux nombreuses caractéristiques à neutraliser, il peut devenir difficile, voire impossible, de lui trouver une correspondance dans le groupe de contrôle.

La solution proposée par *Rosenbaum et Rubin* (1983) revient à résumer un ensemble de critères définissant chaque exploitation par une variable unique: le score de propension. Il s'agit de la probabilité estimée de recourir au traitement en fonction de caractéristiques observables données. Dans un premier temps, on modélise le processus de sélection pour déterminer les variables intervenant le plus dans la décision de se convertir au bio. Puis, on estime, pour chaque exploitation, sa probabilité de conversion. L'appariement par score de propension consiste à comparer chaque observation traitée à un sous-ensemble des observations du groupe de contrôle. Il sera constitué des unités ayant une probabilité de recourir au traitement proche de celle de l'observation traitée. Plus spécifiquement, *Heckman et al.* (1997) proposent l'estimateur apparié en doubles différences suivant (« *Difference-in-Differences Matching Estimator* »):

$$A\hat{T}T = \frac{1}{N_T} \sum_{i \in T} \left\{ \left( Y^{T_i} - Y^{C_i} \right) - \sum_{j \in C} \omega_{ij} \left( Y^{C_j} - Y^{C_{ij}} \right) \right\}$$

où  $\omega_{ij}$  est un poids appliqué à l'unité d'observation non traitée  $j$  lorsqu'on construit le contrefactuel de l'observation traitée  $i$ , ce poids variant en fonction de la distance entre les scores de propension des deux observations.

Pour mettre en œuvre l'appariement sur score de propension, les variables de contrôle retenues ne doivent pas, elles-mêmes, être affectées par le recours au traitement (ou par l'anticipation d'y recourir). Pour s'en assurer, il est donc recommandé de retenir des variables qui sont soit fixes dans le temps, comme le secteur d'activité ou la catégorie juridique, soit mesurées avant le début du traitement. Dans notre cas, nous ne disposons pas de variables antérieures à la conversion au bio. Nous avons donc choisi des variables observées à la date du traitement (ici la conversion au bio) mais majoritairement fixes dans le temps. L'appariement n'est pas utilisé directement pour estimer l'effet causal. Il est complété par une double différenciation. Par conséquent, le fait de ne pas disposer de variables antérieures au traitement n'induit pas systématiquement un biais dans l'estimation de l'effet causal. Ainsi, *Chabé-Ferret* (2010) indique que lorsque les variables de contrôle variant dans le temps sont indépendantes des effets fixes individuels, l'estimateur apparié en doubles différences identifie l'effet du traitement sur les traités si l'ensemble des variables de contrôle comprend la valeur des variables à l'époque où le traitement est décidé. Parmi les effets fixes individuels inobservés susceptibles d'influencer la décision de conversion à l'agriculture biologique, on peut penser à la sensibilité de l'exploitant aux questions environnementales. Cette sensibilité pourrait avoir conduit l'exploitant à utiliser des techniques de production plus intensives en travail avant même sa conversion à l'agriculture biologique. La double différenciation permet de neutraliser un tel effet<sup>2</sup>. Si ce dernier est assez vraisemblablement indépendant de nos variables de contrôle, ce n'est pas forcément le cas de tous les effets fixes inobservés et il est donc probable que nos résultats soient affectés par un léger biais.

Cette méthode permet également de s'assurer de la significativité de l'effet mis en évidence. Pour cela, on a recours à la méthode du *bootstrap*. Il s'agit de réestimer un grand nombre de fois l'effet causal moyen sur des échantillons différents tirés de la population initiale. On effectue cette procédure un grand nombre de fois afin d'obtenir une distribution de l'estimateur et calculer ainsi un (ou plusieurs) intervalle(s) de confiance. Si le 0 ne fait pas partie de cet (ces) intervalle(s), on pourra considérer que l'effet causal mis en évidence est significatif.

2. Par rapport au schéma du graphique 1, l'appariement permet de se rapprocher de l'hypothèse de tendance commune tandis que la double différenciation permet de tenir compte du fait qu'avant même le début du traitement, la caractéristique d'intérêt était (dans cet exemple) plus élevée pour le groupe exposé au traitement.