Économétrie 101

Ce document présente un traitement informel du modèle causal de Rubin et des méthodes d'économétrie. Deux références sur le sujet sont Mastering Metrics et Mostly Harmless Econometrics, de J. Angrist et J.-S. Pischke.

La plupart des travaux empiriques d'économie cherchent à mesurer l'effet que cause une variable (que l'on nomme *traitement*) sur une autre (que l'on appelle généralement *variable d'intérêt*).

Qu'entend-on par *effet du traitement* ? Il s'agit de la différence du niveau de la variable d'intérêt avec ou sans traitement, toutes choses égales par ailleurs.

Par exemple (chiffres fictifs) si le taux de chômage à Trantor est de 10% avec une semaine de travail de 35 heures alors qu'il aurait été de 8% avec une semaine de travail de 39 heures, l'effet du traitement « passer d'une durée de la semaine de travail de 39 heures à 35 heures »sur la variable d'intérêt « taux de chômage » est de $10\% - 8\% = 2$ points de pourcentage.

Soit T_{Trantor} une variable qui est égale à 1 si la Trantor est « traitée » (dans notre exemple, passage aux 35 heures) et 0 si elle ne l'est pas.

On note Y_{Trantor} notre variable d'intérêt. Elle peut prendre deux valeurs possibles : $Y_{\text{Trantor}}(1)$, si $T_{\text{Trantor}} = 1$ et $Y_{\text{Trantor}}(0)$, si $T_{\text{Trantor}} = 0$.

Autrement dit,

$$Y_{\text{Trantor}} = \begin{cases} Y_{\text{Trantor}}(1) & \text{si } T_{\text{Trantor}} = 1 \\ Y_{\text{Trantor}}(0) & \text{si } T_{\text{Trantor}} = 0 \end{cases}$$

Dans notre exemple,

$$Y_{\text{Trantor}}(1) = 10\%$$

$$Y_{\text{Trantor}}(0) = 8\%$$

$$Y_{\text{Trantor}} = Y_{\text{Trantor}}(1) = 10\%$$

L'effet causal Δ du traitement est donc :

$$\Delta_{\text{Trantor}} = Y_{\text{Trantor}}(1) - Y_{\text{Trantor}}(0)$$

Ou plus généralement, pour une observation i :

$$\Delta_i = Y_i(1) - Y_i(0)$$

Est-il possible de mesurer à la fois $Y_{\text{Trantor}}(1)$ et $Y_{\text{Trantor}}(0)$?

Non : il n'est possible de mesurer que des évènements qui se sont effectivement réalisés, autrement dit Y_{Trantor}.

Soit $T_{\text{Trantor}} = 1$ et il est possible d'observer $Y_{\text{Trantor}} = Y_{\text{Trantor}}(1)$ mais pas $Y_{\text{Trantor}}(0)$; soit $T_{\text{Trantor}} = 0$ et il est possible d'observer $Y_{Y_{\text{Trantor}}=\text{Trantor}}(0)$ mais pas $Y_{\text{Trantor}}(1)$.

Quelque soit le statut du traitement (traité ou non traité), il manquera toujours la mesure du *contrefactuel*, c'est à dire de l'évènement non réalisé pour calculer l'effet du traitement.

L'observation reçoit-elle le traitement ?	Réalité observable	Contrefactuel inobservable
Oui : $T_i = 1$	$Y_i(1)$	$Y_i(0)$
Non : $T_i = 0$	$Y_i(0)$	$Y_i(1)$

Implicitement ou explicitement, toute calcul d'effet de traitement (tel que défini dans ce document) repose sur des hypothèses faites sur le ou les contrefactuels.

Pour reprendre notre exemple, une première solution à notre problème d'impossibilité de mesurer le contrefactuel $Y_{\text{Trantor}}(0)$ serait de comparer Trantor à une autre région non traitée, i.e. dont la semaine de travail est de 39 heures, par exemple Helicon.

On pourrait alors calculer

$$\begin{aligned}\Delta' &= Y_{\text{Trantor}} - Y_{\text{Helicon}} \\ &= Y_{\text{Trantor}}(1) - Y_{\text{Helicon}}(0)\end{aligned}$$

Les deux grandeurs, $Y_{\text{Helicon}}(0)$ et $Y_{\text{Trantor}}(1)$ sont bien mesurables.

Est-ce que $\Delta' = \Delta$? Autrement dit, notre comparaison Helicon—Trantor permet-elle de mesurer l'effet du traitement à Trantor ?

Non, car la différence entre les deux régions est composée à la fois de l'effet du traitement *et* des différences éventuelles entre Helicon et Trantor qui ne sont pas liées au traitement. Il est tout à fait possible que, même en l'absence de traitement, les deux régions aient un taux de chômage différent.

En effet,

$$
\begin{aligned}
\Delta' = \quad & Y_{\text{Trantor}}\left(T_{\text{Trantor}} = 1\right) - Y_{\text{Helicon}}\left(T_{\text{Helicon}} = 0\right) \\
= \quad & Y_{\text{Trantor}}\left(T_{\text{Trantor}} = 1\right) \\
& - Y_{\text{Trantor}}\left(T_{\text{Trantor}} = 0\right) + Y_{\text{Trantor}}\left(T_{\text{Trantor}} = 0\right) \\
& - Y_{\text{Helicon}}\left(T_{\text{Helicon}} = 0\right) \\
= \quad & \underbrace{\Delta}_{\text{Effet du traitement}} + \underbrace{Y_{\text{Trantor}}\left(T_{\text{Trantor}} = 0\right) - Y_{\text{Helicon}}\left(T_{\text{Helicon}} = 0\right)}_{\text{Différence entre Helicon et Trantor en l'absence de traitement}}
\end{aligned}
$$

Affirmer que $\Delta' = \Delta$, c'est supposer qu'en l'absence de traitement à Trantor, Trantor et Helicon auraient le même taux de chômage, i.e. que l'on se trouve dans le cas particulier où $Y_{\text{Trantor}}(0) - Y_{\text{Helicon}}(0) = 0 \Rightarrow Y_{\text{Trantor}}(0) = Y_{\text{Helicon}}(0)$. Cette hypothèse n'est pas directement vérifiable (puisque l'on ne peut observer $Y_{\text{Trantor}}(0)$).

Le même problème se retrouve lorsque l'on cherche à calculer un effet moyen du traitement sur un ensemble d'observations. En notant $\mathbf{E}[Y|X]$, la moyenne de Y pour le groupe X, l'effet moyen du traitement (que l'on notera $\boldsymbol{\Delta}$) est :

$$
\begin{aligned}
\boldsymbol{\Delta} = \mathbf{E}[\Delta_i] &= \mathbf{E}[Y_i(1) - Y_i(0)] \\
&= \mathbf{E}[Y_i(1)] - \mathbf{E}[Y_i(0)]
\end{aligned}
$$

Comme pour notre comparaison Trantor — Helicon, on pourrait être tenté de comparer la moyenne de Y parmi les régions traitées à la moyenne des Y dans les régions non traitées. Cela permet-il de calculer l'effet moyen du traitement ? Non, pour la même raison que précédemment : les régions traitées et non traitées diffèrent certes du fait de l'effet du traitement, mais peuvent également être différentes même en l'absence de traitement :

$$\begin{aligned}
\mathbf{\Delta'} = & \underbrace{\mathbf{E}\left[Y_i | T(i)=1\right]}_{\text{Moyenne de la variable d'intérêt chez les traités}} - \underbrace{\mathbf{E}\left[Y_i | T(i)=0\right]}_{\text{Moyenne de la variable d'intérêt chez les non traités}} \\
= & \underbrace{\mathbf{E}\left[Y_i(1) | T(i)=1\right]}_{\text{Moyenne de la variable d'intérêt chez les traités}} - \underbrace{\mathbf{E}\left[Y_i(0) | T(i)=0\right]}_{\text{Moyenne de la variable d'intérêt chez les non traités}} \\
= & \mathbf{E}\left[Y_i(1) | T(i)=1\right] \underbrace{-\mathbf{E}\left[Y_i(0) | T(i)=1\right] + \mathbf{E}\left[Y_i(0) | T(i)=1\right]}_{=0} - \mathbf{E}\left[Y_i(0) | T(i)=0\right] \\
= & \underbrace{\mathbf{E}\left[Y_i(1) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=1\right]}_{=\mathbf{\Delta}} + \mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right] \\
= & \underbrace{\mathbf{\Delta}}_{\text{Effet du traitement}} + \underbrace{\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right]}_{\text{Différence entre traités et non traités en l'absence de traitement}}
\end{aligned}$$

Une comparaison naïve entre régions traitées et régions non traitées ne permet pas de distinguer l'effet du traitement $\mathbf{\Delta}$ des différences pré-existantes entre traités et non traités $\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right]$.

En effet, si — par exemple — $\mathbf{\Delta'} = 0.05$, comment savoir si :

— $\mathbf{\Delta} = 0.03$ et $\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right] = 0.02$
— ou si $\mathbf{\Delta} = -0.04$ et $\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right] = 0.09$
— ou encore $\mathbf{\Delta} = 0$ et $\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right] = 0.05$?

Il faut à nouveau supposer que $\mathbf{E}\left[Y_i(0) | T(i)=1\right] - \mathbf{E}\left[Y_i(0) | T(i)=0\right] = 0$. Hypothèse tout aussi invérifiable que précédemment (si $T(i) = 1$, alors $Y_i = Y_i(1)$ et $Y_i(0)$ n'est pas observé et donc $\mathbf{E}\left[Y_i(0) | T(i)=1\right] = 0$ non plus).

Supposons par exemple qu'il y ait deux types de régions, des régions A pour lesquelles $Y(0) = 5\%$ et $Y(1) = 10\%$ et des régions B pour lesquelles $Y(0) = 11\%$ et $Y(1) = 16\%$. Pour les deux types de régions, $\mathbf{\Delta} = 5\%$.

Si toutes les régions traitées sont de type A et toutes les régions non traitées sont de type

B,

$$\begin{aligned}
\mathbf{\Delta'} &= \mathbf{E}\left[Y_i | T\left(i\right) = 1\right] - \mathbf{E}\left[Y_i | T\left(i\right) = 0\right] \\
&= \mathbf{E}\left[Y_i\left(1\right)|i \in A\right] - \mathbf{E}\left[Y_i\left(0\right)|i \in B\right] \\
&= 10\% - 11\% \\
&= -1\% \\
&\neq \mathbf{\Delta}
\end{aligned}$$

Reprenons la décomposition précédente :

$$\begin{aligned}
\mathbf{\Delta'} &= \underbrace{\mathbf{\Delta}}_{\text{Effet du traitement}} + \underbrace{\mathbf{E}\left[Y_i\left(0\right)|T\left(i\right) = 1\right] - \mathbf{E}\left[Y_i\left(0\right)|T\left(i\right) = 0\right]}_{\text{Différence entre traités et non traités en l'absence de traitement}} \\
&= \mathbf{\Delta} + \mathbf{E}\left[Y_i\left(0\right)|i \in A\right] - \mathbf{E}\left[Y_i\left(0\right)|i \in B\right] \\
&= \mathbf{\Delta} + 5\% - 11\% \\
&= \mathbf{\Delta} - 6\%
\end{aligned}$$

Notre comparaison traités — non traités ne permet pas de retrouver le véritable effet du traitement $\mathbf{\Delta}$, puisque la différence entre traités et non traités en l'absence de traitement est non-nulle et inconnue.

Dans quel cas l'hypothèse $\mathbf{E}\left[Y_i\left(0\right)|T\left(i\right) = 1\right] - \mathbf{E}\left[Y_i\left(0\right)|T\left(i\right) = 0\right] = 0$ n'est-elle pas aberrante ?

Si le traitement est assigné aléatoirement, il n'y a pas, en moyenne, de différence entre le groupe traité et le groupe non traité (dit groupe témoin).

Si le traitement est assigné aléatoirement, cela implique que la probabilité d'être traitré est la même pour toutes les observations dans l'échantillon, et ce quelque soient leurs charactéristiques : autrement dit, $\forall x, p\left(T|x\right) = p\left(T\right)$.

Intuitivement, cela implique que la proportion de chaque sous-groupe sera la même dans l'échantillon entier, parmi les traités et parmi les non traités.

Supposons par exemple que la probabilité d'être traité soit de 25%, et pour reprendre notre exemple précédent, qu'il y ait 40 régions de type A et 60 de régions de type B. Il y aura donc $25\% \times 40 = 10$ régions de type A traitées, $(1 - 25\%) \times 40 = 30$ régions de type

A non traitées ; $25\% \times 60 = 15$ régions de type B traitées, $(1 - 25\%) \times 60 = 45$ régions de type B non traitées.

La proportion de régions A parmi les traités est donc de $\frac{10}{10+15} = 40\%$ et la proportion de régions B parmi les traités est de $\frac{15}{10+15} = 60\%$.

La proportion de régions A parmi les non traités est donc de $\frac{30}{30+45} = 40\%$ et la proportion de régions B parmi les traités est de $\frac{45}{30+45} = 60\%$.

Les deux groupes, non traités et traités, ont donc la même composition, et donc la même moyenne de $Y(0)$ et de $Y(1)$).

On a :

$$\mathbf{E}\left[Y_i(0)\,|T(i)=1\right] = Y_{0,A} \times p\left(A|T=1\right)\right) + Y_{0,B} \times p\left(B|T=1\right)\right) = 5\% \times 40\% + 11\% \times 60\% = 8,6\%$$

et :

$$\mathbf{E}\left[Y_i(0)\,|T(i)=0\right] = Y_{0,A} \times p\left(A|T=0\right)\right) + Y_{0,B} \times p\left(B|T=0\right)\right) = 5\% \times 40\% + 11\% \times 60\% = 8,6\%$$

C'est à dire

$$\mathbf{E}\left[Y_i(0)\,|T(i)=1\right] - \mathbf{E}\left[Y_i(0)\,|T(i)=0\right] = 0$$

Les essais contrôlé aléatoires ou *randomized controlled trial* (RCT) permettent donc d'« équilibrer » le groupe traité et le groupe témoin sur l'ensemble de leurs caractéristiques, observées ou non observées.