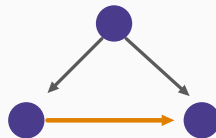


Inférence causale et machine learning

Chapitre 8 : Les variables instrumentales, revisitées



Le principe

L'estimateur

Sur qui porte le résultat

Les instruments faibles

La condition d'exclusion

Ce qu'il faut retenir

Le principe

Instrument

Un **instrument** Z est une variable qui influence le traitement T sans affecter le résultat Y autrement que **par** T .

Les trois conditions

Condition	Énoncé	Testable ?
Pertinence	Z agit sur T	oui
Exclusion	Z n'agit sur Y que par T	non
Indépendance	Z est comme attribué au hasard	non

Intuition

On ne peut pas randomiser le traitement. Mais si quelque chose d'extérieur pousse certains à le suivre et pas d'autres, **cette poussée** est exploitable.

On n'utilise alors que la part de variation du traitement **imputable à l'instrument** — celle dont on peut soutenir qu'elle est indépendante des résultats potentiels.

L'estimateur

L'estimateur de Wald

Pour un instrument binaire :

$$\hat{\tau} = \frac{\overbrace{E[Y \mid Z=1] - E[Y \mid Z=0]}^{\text{forme réduite}}}{\underbrace{E[T \mid Z=1] - E[T \mid Z=0]}_{\text{premier stade}}}$$

Une loterie d'accès au programme

Les places étant limitées, l'accès prioritaire est tiré au sort. $Z = 1$ pour les gagnants.

Quantité	$Z = 1$	$Z = 0$	Écart
Salaire moyen	60	50	10
Part de formés	0,6	0,2	0,4

Sur qui porte le résultat

L'effet local sur les conformes

Les quatre types

Type	Comportement
Conformes	suivent le traitement si et seulement si $Z = 1$
Toujours preneurs	le suivent quoi qu'il arrive
Jamais preneurs	ne le suivent jamais
Contrariants	font l'inverse de l'incitation

La monotonie

On suppose qu'il n'existe **pas de contrariants** : l'instrument pousse tout le monde dans le même sens.

Sous cette hypothèse, l'estimateur identifie l'effet moyen **sur les seuls conformes** — le LATE.

Une population qu'on ne peut pas nommer

On sait que les conformes existent et on peut estimer leur **proportion** — ici 0,4. On ne peut pas dire **qui** ils sont : l'appartenance dépend de comportements contrefactuels non observés.

Un résultat de variable instrumentale porte donc sur une sous-population non identifiable.

Les instruments faibles

Le problème le plus fréquent

Instrument faible

Un instrument est **faible** lorsque son effet sur le traitement est petit — le dénominateur de Wald est proche de zéro.

Pourquoi c'est grave

Diviser par un petit nombre **amplifie tout** : le bruit d'estimation, et surtout la moindre violation de l'exclusion.

Un instrument faible et légèrement invalide produit un biais **supérieur** à celui de la régression naïve qu'il prétendait corriger. Le remède devient pire que le mal.

Le diagnostic

On examine la statistique F du premier stade. Une valeur inférieure à 10 est traditionnellement jugée préoccupante; les travaux récents suggèrent des seuils nettement plus élevés.

Le premier stade doit toujours être présenté. Une étude qui ne le montre pas est incomplète.

La condition d'exclusion

L'hypothèse qu'on ne peut pas tester

Là où tout se joue

La pertinence se vérifie. L'exclusion, non — et c'est elle qui porte l'identification.

Sur notre loterie : gagner le tirage pourrait affecter le salaire **autrement** que par la formation — par un effet de confiance, par un signal envoyé aux employeurs, ou parce que les perdants reçoivent une compensation.

Chacune de ces voies invalide l'instrument. **Aucun test ne les détecte**; seul l'argument institutionnel permet de les écarter.

Ce qui fait un bon instrument

Une source de variation dont on peut **raconter précisément** l'origine, et dont on voit pourquoi elle ne peut pas agir autrement.

Les meilleurs viennent de règles administratives, de tirages au sort, de contraintes géographiques ou de calendrier. Les plus contestables sont ceux qu'on justifie par une corrélation observée — ce qui n'est pas un argument d'exclusion.

Ce qu'il faut retenir

Cinq points

1. Un **instrument** agit sur T et n'agit sur Y que par T .
2. **Wald** : forme réduite divisée par premier stade — ici $10/0,4 = \mathbf{25}$.
3. Sous **monotonie**, on identifie l'effet sur les seuls **conformes** — une population non nommable.
4. Un instrument **faible** amplifie bruit et violations : le premier stade doit toujours être montré.
5. L'**exclusion** ne se teste pas; elle se défend par le contexte institutionnel.

La suite

Les chapitres 6 à 8 ont exploité le temps, un seuil, un instrument. Le chapitre 9 traite le cas où l'on ne dispose que d'**une seule** unité traitée.