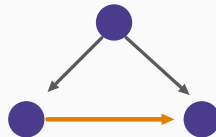


Inférence causale et machine learning

Chapitre 6 : Les doubles différences



L'idée

L'hypothèse

Les difficultés

Ce qu'il faut retenir

L'idée

Deux différences valent mieux qu'une

Pourquoi une seule différence ne suffit pas

Comparer **avant et après** chez les traités attribue au programme tout ce qui a changé entre-temps — conjoncture, saison, tendance.

Comparer **traités et témoins après** attribue au programme les écarts qui préexistaient.

Chaque comparaison a un défaut; **les combiner les annule.**

L'estimateur

$$\text{DiD} = \underbrace{(\bar{Y}_{T,\text{après}} - \bar{Y}_{T,\text{avant}})}_{\text{évolution des traités}} - \underbrace{(\bar{Y}_{C,\text{après}} - \bar{Y}_{C,\text{avant}})}_{\text{évolution des témoins}}$$

Exemple

Taux d'emploi, en pourcentage :

	Avant	Après	Évolution
Région traitée	20	32	+12
Région témoin	25	29	+4
			$DiD = 12 - 4 = 8$

La région traitée a gagné douze points. Mais quatre auraient été gagnés de toute façon, comme le montre la région témoin.

L'effet du programme est de huit points, non de douze.

Intuition

Il ne sert pas de point de comparaison en niveau — les deux régions diffèrent au départ, et cela n'a pas d'importance.

Il fournit une **estimation de la trajectoire contrefactuelle** : ce qu'auraient fait les traités sans le programme.

L'hypothèse

L'hypothèse d'identification

$$E[Y_{\text{après}}(0) - Y_{\text{avant}}(0) \mid T=1] = E[Y_{\text{après}}(0) - Y_{\text{avant}}(0) \mid T=0]$$

En l'absence de traitement, les deux groupes auraient évolué en parallèle.

Ce qu'elle exige, et ce qu'elle n'exige pas

Elle **n'exige pas** que les groupes soient semblables : les niveaux peuvent différer autant qu'on veut. C'est ce qui rend la méthode si commode.

Elle exige que leurs **trajectoires** l'auraient été. C'est une hypothèse sur un contrefactuel — donc invérifiable directement.

Définition	
Vérification	Ce qu'elle apporte
Tendances pré-traitement	plusieurs périodes avant : les courbes étaient-elles parallèles ?
Test placebo	appliquer la méthode à une période sans traitement — l'effet doit être nul
Groupe témoins alternatifs	le résultat change-t-il selon le témoin retenu ?
Résultat placebo	une variable que le programme ne peut affecter ne doit rien montrer

Application en gestion

Des tendances parallèles avant le traitement rendent l'hypothèse **crédible**. Elles ne la démontrent pas : rien n'interdit qu'elles aient divergé précisément au moment du programme.

C'est le meilleur argument disponible, et il reste un argument.

Les difficultés

Trois menaces

1. **L'anticipation** — les agents modifient leur comportement avant la mise en œuvre, dès l'annonce.
2. **Un choc concomitant** — un autre événement frappe le groupe traité au même moment.
3. **La sélection sur les tendances** — le programme a été implanté là où la situation se dégradait, ou s'améliorait.

La troisième est la plus insidieuse

Un dispositif est souvent déployé là où les besoins sont les plus criants — c'est-à-dire là où la trajectoire était déjà mauvaise.

Les tendances parallèles sont alors fausses **par construction du dispositif lui-même**, et le biais va dans un sens prévisible : l'effet est sous-estimé.

Comprendre comment le traitement a été attribué est aussi important que la technique d'estimation.

Le problème des traitements décalés

Lorsque les unités sont traitées à des dates différentes, la régression à effets fixes usuelle ne donne plus l'effet moyen recherché.

Elle compare implicitement des unités traitées tôt à des unités **déjà traitées**, et peut affecter des poids **négatifs** à certaines comparaisons.

Le résultat qui a surpris la profession

Ce défaut est resté longtemps ignoré, alors que la spécification concernée était l'une des plus employées en économie appliquée.

Des estimateurs corrigés existent aujourd'hui, qui reconstruisent explicitement les comparaisons légitimes. Devant une adoption échelonnée, **l'estimateur classique ne convient pas** — et c'est un point de méthode récent qu'il faut connaître.

Ce qu'il faut retenir

Cinq points

1. La **double différence** retrace à l'évolution des traités celle des témoins.
2. Sur notre programme : $12 - 4 = 8$ points d'effet, et non 12.
3. L'hypothèse est celle des **tendances parallèles** — pas des niveaux comparables.
4. On la défend par les tendances **pré-traitement**, des **placebos**, des témoins alternatifs.
5. **Anticipation**, chocs concomitants et **sélection sur les tendances** la menacent; l'adoption **échelonnée** exige des estimateurs corrigés.

La suite

Le chapitre 7 exploite une source de variation encore plus locale, et souvent plus crédible : un **seuil administratif**.