





## TECHNIQUES D'APPARIEMENT

### Session Technique VI

**Emanuela Galasso**  
**La Banque mondiale**

These slides were developed by Christel Vermeersch and modified by Emanuela Galasso for the purpose of this workshop

## Quand faut-il utiliser l'appariement?

Et si l'assignation au traitement n'était pas aléatoire, mais sur la base d'observables?

C'est alors que surviennent les méthodes de d'appariement! Les méthodes d'appariement vous permettent de construire des groupes de comparaison lorsque l'assignation au traitement est faite sur la base de variables observables.



## When can we use matching?

- **Intuition:** il est nécessaire que le groupe de comparaison soit aussi identique que possible au groupe de traitement, en termes d'observables avant le lancement du traitement.
- La méthode **suppose** qu'il n'existe plus de différences inobservables entre les groupes de traitement et témoin.



## Question-clé

Quel est l'effet de traitement sur les unités traitées quand l'assignation est faite sur la base de variables observables?

## Unconfoundedness et sélection sur observables

- Laissons  $X$  dénoter une matrice dans laquelle chaque rang est un vecteur de variables observables prétraitées pour un individu  $i$ .
- **Unconfoundedness:** l'assignation au traitement est non-confondu étant donné les variables  $X$  si

$$Y_1, Y_0 \perp D \mid X$$

Unconfoundedness équivaut à dire que:

- (1) Dans chaque cellule définie par  $X$ : le traitement est aléatoire
- (2) La sélection au traitement ne dépend que des observables  $X$



## Les effets moyens du traitement sur les unités traitées

supposant unconfoundedness,  $X$  donné

- **Intuition**
  - Estimer l'effet traitement dans chaque cellule définie par  $X$
  - Considérer la moyenne sur les différentes cellules
- **Math**  
dans votre documentation en *annexe 1*

## Stratégie de l'estimation de l'effet moyen de traitement sur les unités traitées sélection sur observables

- Unconfoundedness suggère la **stratégie** suivante pour l'estimation de l'effet moyen de traitement  $\delta$ 
  - Stratifier les données dans les cellules définies par chaque valeur particulière de  $X$
  - Dans chaque cellule (i.e. conditionnement sur  $X$ ) calculer la différence entre les réalisations moyennes des unités traitées et témoins
  - Faire la moyenne de ces différences en respectant la distribution de  $X$  dans la population des unités traitées.
- Cette stratégie est-elle **réalisable**?



## Le problème de dimensionnalité

- Exemples
  - Combien avons-nous de cellules avec 2 variables binaires  $X$ ? et avec 3 variables binaires  $X$  et avec  $K$  variables binaires  $X$ ?
  - Et si nous avons 2 variables qui prenaient chacune 7 valeurs?
- À mesure que le nombre de cellules grandit, nous obtenons "**une perte de support commun**"
  - Cellules contenant seulement les observations traitées
  - Cellules contenant seulement les témoins.



## Notre stratégie est-elle réalisable? Le problème de dimensionnalité

- Ne peut être réalisable quand
  - L'échantillon est petit
  - L'ensemble des covariables est grand
  - Plusieurs covariables ont plusieurs valeurs, ou sont continues
- C'est cela que nous appelons...  
**Le problème de dimensionnalité**

## Une alternative pour résoudre le problème de dimensionnalité



Le score de propension permet de convertir un dispositif multidimensionnel d'appariement en un dispositif unidimensionnel. Comme suit, il permet de réduire le problème de dimensionnalité.

*Rosenbaum et Rubin*

Rosenbaum et Rubin (1983) proposent une stratégie d'estimation équivalente et réalisable sur la base du concept de *score de propension*



## Appariement basé sur le score de propension

### Définition

Le score de propension est la probabilité conditionnelle de recevoir le traitement étant donné les variables de prétraitement:

$$p(X) = Pr\{D = 1|X\} = EX\{D|X\}$$

### Lemma 1

si  $p(X)$  est le score de propension, alors  $D \perp X | p(X)$   
"à score de propension donné les variables de prétraitement sont équilibrées entre bénéficiaires et non-bénéficiaires"

### Lemma 2

"Supposons que l'assignation au traitement est non-confondue étant donné les variables de prétraitement  $X$ , alors l'assignation au traitement est non-confondue étant donné le score de propension  $p(X)$ ."



## L'approche du score de propension résoud-il le problème de dimensionnalité?

**OUI!**

- La condition d'équilibre du score de propension (Lemme 1) garantit que:
  - Les observations avec le même score de propension ont la même distribution de covariables observables indépendamment de l'état de traitement; et
  - Pour un score de propension donné: l'assignation au traitement est aléatoire et par conséquent les unités de traitement et témoins sont en moyenne identiques en observation.



## Exécution de la stratégie d'estimation

Ceci suggère la **stratégie** suivante pour l'estimation de l'effet moyen de traitement  $\delta$

### Phase 1

Estimer un logit (ou probit) modèle de participation au programme. Les valeurs prédites sont les « scores de propension ».

Ex. avec une fonction Logit, voir Annexe 3

Cette phase est nécessaire car le „véritable” score de propension est inconnu et par conséquent le score de propension doit être estimé.



## Quand l'appariement du score de propension est-il approprié?

- **Idée derrière l'appariement** du score de propension: l'estimation des effets de traitement requiert un appariement prudent des unités traitées et témoins.
- Si les unités traitées et témoins sont très différents en termes d'observables cet appariement n'est pas suffisamment proche et crédible. Il peut aussi s'avérer impossible.
- La comparaison des résultats de propension estimés des unités traitées et des témoins fournit un outil utile de diagnostic pour évaluer le degré de similarité entre unités traitées et témoins, et de fait d'évaluer la crédibilité de la stratégie d'estimation.

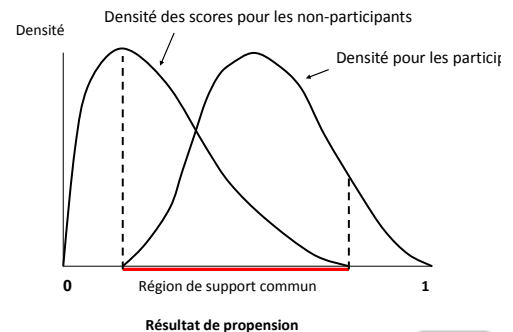


## Donc vous voulez que les résultats de propension soient les "mêmes" pour les unités traitées et les témoins...

- L'amplitude de variation des résultats de propension devraient être la même pour les unités traitées et les témoins.
  - Comptez combien d'unités témoins ont un score de propension plus bas que le minimum ou plus élevé que le maximum des résultats de propensions des unités traitées.
  - Et vice versa.
- La fréquence des résultats de propension est la même pour les unités traitées et témoins.
  - Tracer un histogramme des résultats de propension estimés pour les unités traitées et témoins.
  - Les paniers correspondent aux blocs construits pour l'estimation des résultats de propension.



## Problèmes de support commun



## Exemple: problèmes de support commun

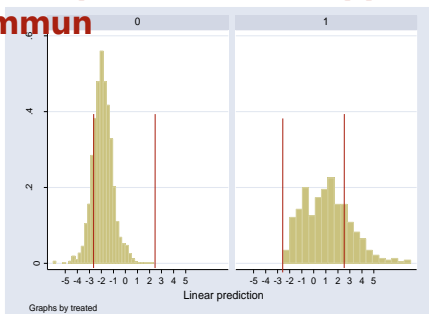


Figure A1: résultats de propension au niveau des écoles pour EIC (Excellence In Cities) Phase 1 et non-EIC  
Source Machin, McNally, Meghir. EXCELLENCE IN CITIES: ÉVALUATION D'UNE POLITIQUE D'ENSEIGNEMENT DANS LES SECTEURS DÉFAVORISÉS

## Exécution de la stratégie d'estimation

Rappelez-vous que nous traitons d'une stratégie pour l'estimation de l'effet moyen de traitement sur les unités traitées, appelé  $\delta$

### Phase 1

phase 1: estimer le score de propension (voir annexe 3)

### Step 2

Restreindre l'analyse à la région de support commun (principale source de biais dans les études d'observation)



### Phase 3: estimer l'effet moyen de traitement étant donné le score de propension

- Pour chaque participant, trouvez un échantillon non-participant ayant "les mêmes" scores de propension
- Comparer l'indicateur de résultat pour chaque participant et son groupe de comparaison.
- Calculer les moyennes de ces gains individuels afin d'obtenir le gain moyen général.

$$ATT = \frac{1}{P} \sum_{j=1}^P (Y_{j1} - \sum_{i=1}^{NP} W_{ij} Y_{i0}) / P$$



### Phase 3: estimer l'effet moyen de traitement étant donné le score de propension

- "Similaire" peut être défini de plusieurs manières. A chaque définition correspond une certaine méthode d'appariement:
  - Stratification sur le résultat
  - Appariement au voisin le plus proche sur la base du résultat
  - Appariement Radius sur la base de résultat
  - Appariement Kernel sur la base de résultat
  - Pondération sur la base du résultat



### Résumé:

- L'appariement est l'analogue d'observation d'une expérience dans laquelle le placement est indépendante des résultats
- La principale différence est que l'expérimentation pure ne nécessite pas l'hypothèse d'indépendance conditionnelle invérifiable sur les observables.
- PSM exige de bonnes données
- Souvent combiné avec des méthodes de différence dans la différence (contrôle pour la sélection basée sur le temps- caractéristiques invariantes non observées)



### Références

- Dehejia, R.H. and S. Wahba (1999), "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, 448, 1053-1062.
- Dehejia, R.H. and S. Wahba (1996), "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs", Harvard University, Mimeo.
- Hahn, Jinyong (1998), "On the role of the propensity score in efficient semi-parametric estimation of average treatment effects", *Econometrica*, 66, 2, 315-331.
- Heckman, James J. H. Ichimura, and P. Todd (1998), "Matching as an econometric evaluation estimator", *Review of Economic Studies*, 65, 261-294.
- Hirano, K., G.W. Imbens and G. Ridder (2000), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score", mimeo.
- Rosenbaum, P.R. and D.B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, 1, 41-55.
- Vinha, K. (2006) "A primer on Propensity Score Matching Estimators" Documento CEDE 2006-13, Universidad de los Andes



JE VOUS REMERCIE

Q&R



**Annexe 1: effets moyens de traitement sur les unités traitées supposant l'unconfoundedness donné X**

If we are willing to assume unconfoundedness:  
 $E_i \{Y_0(u_i) | D_i=0, X\} = E_i \{Y_0(u_i) | D_i=1, X\} = E_i \{Y_0(u_i) | X\}$   
 $E_i \{Y_1(u_i) | D_i=0, X\} = E_i \{Y_1(u_i) | D_i=1, X\} = E_i \{Y_1(u_i) | X\}$

Using these expressions, we can define for each cell defined by X  
 $\delta_x$  = average treatment effect on the treated in cell defined by X  
 $= E_i \{\Delta_i | D_i = 1, X\}$   
 $= E_i \{Y_1(u_i) - Y_0(u_i) | D_i = 1, X\}$   
 $= \underbrace{E_i \{Y_1(u_i) | D_i = 1, X\}}_{\text{can measure sample analog}} - \underbrace{E_i \{Y_0(u_i) | D_i = 1, X\}}_{\text{can NOT measure sample analog}}$   
 $= E_i \{Y_1(u_i) | D_i = 1, X\} - \underbrace{E_i \{Y_0(u_i) | D_i = 0, X\}}_{\text{can measure sample analog}}$



**Annexe 1: effets moyens de traitement sur les unités traitées supposant l'unconfoundedness donné X**

Now what is the relation between  
 $\delta$  "average treatment effect on the treated"... and...  
 $\delta_x$  "average treatment effect on the treated within cell defined by X"?"

$\delta$  = average treatment effect on the treated  
 $= E_i \{\Delta_i | D_i = 1\}$   
 $\Downarrow$  by the law of iterated expectations  
 $= E_i \{E_x \{\Delta_i | D_i = 1, X\}\}$   
 $= E_x \{E_i \{\Delta_i | D_i = 1, X\}\}$   
 $= E_x \{\delta_x\}$   
 $= E_x \{\text{average treatment effect on the treated within cell defined by X}\}$



**Annexe 2: effets moyens de traitement et le score de propension**



So let's match treatments and controls on the basis of the propensity score  $p(X)$  instead of X.  
 $E_i \{Y_0(u_i) | D_i=0, p(X_i)\} = E_i \{Y_0(u_i) | D_i=1, p(X_i)\} = E_i \{Y_0(u_i) | p(X_i)\}$   
 $E_i \{Y_1(u_i) | D_i=0, p(X_i)\} = E_i \{Y_1(u_i) | D_i=1, p(X_i)\} = E_i \{Y_1(u_i) | p(X_i)\}$

Using these expressions, we can define f cell defined by  $p(X)$   
 $\delta_{p(X)}$  = average treatment effect on the treated in cell defined by  $p(X)$   
 $= E_i \{\Delta_i | D_i = 1, p(X)\}$   
 $= E_i \{Y_1(u_i) - Y_0(u_i) | D_i = 1, p(X)\}$   
 $= \underbrace{E_i \{Y_1(u_i) | D_i = 1, p(X)\}}_{\text{can measure sample analog}} - \underbrace{E_i \{Y_0(u_i) | D_i = 1, p(X)\}}_{\text{can NOT measure sample analog}}$   
 $= E_i \{Y_1(u_i) | D_i = 1, p(X)\} - \underbrace{E_i \{Y_0(u_i) | D_i = 0, p(X)\}}_{\text{can measure sample analog}}$



**Annexe 2: effets moyens de traitement et le score de propension**



Now what is the relation between  
 $\delta$  "average treatment effect on the treated"... and...  
 $\delta_{p(X)}$  "average treatment effect on the treated within cell defined by  $p(X)$ "?"

$\delta$  = average treatment effect on the treated  
 $= E_i \{\Delta_i | D_i = 1\}$   
 $\Downarrow$  by the law of iterated expectations  
 $= E_i \{E_{p(X)} [\Delta_i | D_i = 1, p(X)]\}$   
 $= E_{p(X)} \{E_i [\Delta_i | D_i = 1, p(X)]\}$   
 $= E_{p(X)} \{\delta_{p(X)}\}$   
 $= E_{p(X)} \{\text{treatment effect on the treated within cell defined by } p(X)\}$



**Annexe 3: estimation de score de propension**

Tout modèle de probabilité standard peut être utilisé pour estimer le score de propension, ex. Un modèle Logit :

$$Pr\{D_i | X_i\} = \frac{e^{\lambda h(X_i)}}{1 + e^{\lambda h(X_i)}} \quad (16)$$

Où  $h(X_i)$  est une fonction de covariable avec termes d'ordre supérieur et linéaire.



**Estimation du score de propension**

- Quels termes d'ordre supérieur inclue-t-on dans  $h(X_i)$ ?  
 Ceci est déterminé seulement par la nécessité d'obtenir une estimation de score de propension satisfaisante la propriété d'équilibre.
  - La spécification de  $h(X_i)$  (1) est plus minimale que l'ensemble complet des interactions entre les observables X (2) Bien que pas trop minimale, elle doit satisfaire la propriété d'équilibre.
- Note: l'estimation des scores de propension ne nécessite pas d'interprétation comportementale.



## Un algorithme pour estimer le score propension



1. Commencer par une fonction minimale logit ou probit pour estimer le résultat.
2. Trier les données en raison du score de propension estimé ( du plus bas au plus élevé).
3. Stratifier toutes les observations en blocs de sorte que dans chaque bloc, les résultats de propension estimés pour les unités traitées et témoins ne soient pas statistiquement différentes:
  - a. Commencer avec cinq blocs d'une amplitude de résultat égale  $\{0 - 0.2, \dots, 0.8 - 1\}$
  - b. Tester si les moyennes des résultats pour les unités traitées et témoins sont statistiquement différentes dans chaque bloc
  - c. Si oui, augmenter le nombre de blocs et tester de nouveau
  - d. sinon, passer à l'étape suivante.



## Un algorithme pour estimer le score propension (suite)



4. Tester si la propriété d'équilibre tient dans tous les blocs pour toutes les covariables:
  - a) Pour chaque covariable, tester si les moyennes (et si possible de conditions d'ordre supérieur) pour les unités traitées et témoins sont statistiquement différentes dans tous les blocs;
  - b) Si une covariable n'est pas équilibrée dans un bloc, partager le bloc et tester à nouveau chaque bloc affiné;
  - c) Si une covariable n'est pas équilibrée dans tous les blocs, modifier l'estimation Logit du score de propension en ajoutant plus de conditions d'ordre supérieur et d'interactions et tester à nouveau

*Noter* que dans toute la procédure la réalisation n'a aucun rôle  
Utiliser le programme STATA `pscore.ado`

