



THE WORLD BANK



## Techniques de matching

Human Development  
Network

Middle East and North Africa  
Region

Spanish Impact Evaluation  
Fund

# Le cas d'affectation aléatoire au traitement

- Si l'affectation au traitement est aléatoire dans la population, les résultats potentiels sont tous deux indépendants du statut du traitement

$$Y(1), Y(0) \perp D \quad (1)$$

- Dans ce cas, l'information manquante ne crée pas de problèmes parce que :

$$E\{Y_i(0)|D_i = 0\} = E\{Y_i(0)|D_i = 1\} = E\{Y_i(0)\} \quad (2)$$

$$E\{Y_i(1)|D_i = 0\} = E\{Y_i(1)|D_i = 1\} = E\{Y_i(1)\} \quad (3)$$

- Alors,

$$\delta = E\{\Delta_i | D_i = 1\} \quad (4)$$

$$= E\{Y_i(1) - Y_i(0) | D_i = 1\}$$

$$= E\{Y_i(1)|D_i = 1\} - E\{Y_i(0) | D_i = 1\}$$

$$= E\{Y_i(1)|D_i = 1\} - E\{Y_i(0)|D_i = 0\}$$

$$= E\{Y_i|D_i = 1\} - E\{Y_i|D_i = 0\}.$$

# Le cas d'affectation aléatoire au traitement



- ▣ Le tirage aléatoire assure que le biais de sélection de l'échantillon est de zéro :

$$E\{Y_i(0) \mid D_i = 1\} - E\{Y_i(0) \mid D_i = 0\} = 0 \quad (5)$$

- ▣ Noter que le tirage aléatoire implique que l'information manquante est «manquante complètement au hasard » et que pour cette raison il ne crée pas de problèmes.
- ▣ Si le tirage aléatoire n'est pas possible et que les expérimentations naturelles ne sont pas disponibles, nous devons commencer avec une série différente d'hypothèses.

# *Unconfoundedness* et sélection sur observables



- Laissons  $X$  dénoter une matrice dans laquelle chaque rang est un vecteur de variables observables de prétraitement pour un individu  $i$ .

- **Définition** Unconfoundedness

*L'affectation au traitement n'est pas confondue étant donné les variables de prétraitement si*

$$Y(1), Y(0) \perp D \mid X$$

- Noter qu'assumer l'*unconfoundedness* équivaut à dire que :
  - Dans chaque cellule définie par  $X$  le traitement est aléatoire ;
  - La sélection dans le traitement ne dépend seulement que des observables  $X$ .

# Effets moyens du traitement sur les groupes de bénéficiaires en assumant l'unconfoundedness

- Si nous voulons assumer l'unconfoundedness:

$$E\{Y_i(0)|D_i = 0, X\} = E\{Y_i(0)|D_i = 1, X\} = E\{Y_i(0)|X\} \quad (6)$$

$$E\{Y_i(1)|D_i = 0, X\} = E\{Y_i(1)|D_i = 1, X\} = E\{Y_i(1)|X\} \quad (7)$$

- En utilisant ces expressions, nous pouvons définir pour chaque cellule définie par  $X$  :

$$\begin{aligned} \delta_x &= E\{\Delta_i|X\} & (8) \\ &= E\{Y_i(1) - Y_i(0)|X\} \\ &= E\{Y_i(1)|X\} - E\{Y_i(0)|X\} \\ &= E\{Y_i(1)|D_i = 1, X\} - E\{Y_i(0)|D_i = 0, X\} \\ &= E\{Y_i|D_i = 1, X\} - E\{Y_i|D_i = 0, X\} \end{aligned}$$

# Effets moyens du traitement sur les groupes de bénéficiaires en assumant l'unconfoundedness

- En utilisant la Loi des attentes itérées, l'effet moyen du traitement sur les groupes de bénéficiaires est donné par :

$$\begin{aligned}\delta &= E\{\Delta_i | D_i = 1\} && (9) \\ &= E\{E\{\Delta_i | D_i = 1, X\} | D_i = 1\} \\ &= E\{E\{Y_i | D_i = 1, X\} - E\{Y_i | D_i = 0, X\} | D_i = 1\} \\ &= E\{\delta_x | D_i = 1\}\end{aligned}$$

- Où l'attente extérieure est sur la distribution de  $X | D_i = 1$ .

# Stratégies de matching et de régression pour l'estimation des effets causaux moyens

- L'unconfoundedness suggère la stratégie suivante pour l'estimation de l'effet de traitement moyen défini dans les équations 8 et 9 :
  - i. Stratifier les données en cellules définies par chaque valeur particulière de  $X$  ;
  - ii. Dans chaque cellule (i.e., conditionnement sur  $X$ ) calculer la différence entre les résultats moyens des groupes de bénéficiaires et des groupes témoins ;
  - iii. Faire la moyenne de ces différences par rapport à la distribution de  $X_i$  dans la population des unités bénéficiaires.
  
- Cette stratégie soulève les questions suivantes :
  - Cette stratégie est-elle différente de l'estimation d'une régression linéaire de  $Y$  sur  $D$  saisissant de manière non paramétrique l'ensemble complet des principaux effets et interactions des covariables  $X$  ?
  - Cette stratégie est-elle réalisable ?

# Le matching est-il réalisable ? Le problème de la dimensionnalité

- Il est évident, toutefois, que l'inclusion dans une régression d'un ensemble complet d'interactions non paramétriques entre toutes les variables observables peut ne pas être réalisable quand l'échantillon est petit, l'ensemble des covariables est grand et que de nombreuses d'entre elles sont multivaluées ou pire, continues.
- Ce problème de la dimensionnalité compromettra probablement également la stratégie de matching décrite dans les équations 8 et 9 :
  - Avec les variables binaires  $K$  le nombre de cellules est  $2^K$  et grandit exponentiellement avec  $K$ .
  - Le nombre de cellules augmente davantage si certaines variables dans  $X$  prennent plus que deux valeurs.
  - Si le nombre de cellules est très grand par rapport à la taille de l'échantillon il est très facile de rencontrer des situations dans lesquelles il y a :
    - Des cellules ne contenant que des groupes de bénéficiaires et/ou
    - Des cellules ne contenant que des témoins.

# Le matching et la régression sont ils réalisables : Le problème de la dimensionnalité



- Ainsi, l'effet de traitement moyen pour ces cellules ne peut être calculé.
- Rosenbaum et Rubin (1983) proposent une stratégie d'estimation équivalente et réalisable basée sur le concept de *score de propension* et sur ses propriétés qui permettent de réduire le problème de la dimensionnalité.
- Il est important de réaliser que la régression avec un modèle non saturé n'est pas une solution et peut mener à des conclusions sérieusement fallacieuses.

# Le matching basé sur le score de propension

- **Définition** Score de propension (Rosenbaum et Rubin, 1983): *Le score de propension est la probabilité conditionnelle de recevoir le traitement étant donné les variables de prétraitement :*

$$p(X) = Pr\{D = 1|X\} = E\{D|X\} \quad (10)$$

- Le score de propension a deux propriétés importantes :
  - **Lemme 1** Equilibrer les variables de prétraitement étant donné le score de propension (Rosenbaum et Rubin, 1983)

*Si  $p(X)$  est le score de propension*

$$D \perp X | p(X) \quad (11)$$

- **Lemme 2** Unconfoundedness étant donné le score de propension (Rosenbaum et Rubin, 1983)

*Supposons que l'affectation au traitement n'est pas confondue, i.e.,*

$$Y(1), Y(0) \perp D | X$$

*Alors l'affectation au traitement n'est pas confondue étant donné le score de propension, i.e.,*

$$Y(1), Y(0) \perp D | p(X) \quad (12)$$

# Effets moyens du traitement et le score de propension

- En utilisant le score de propension et ses propriétés, nous pouvons maintenant appairer les cas et les témoins sur la base du score de propension plutôt que le vecteur multidimensionnel des observables  $X$ .

$$E\{Y_i(0)|D_i = 0, p(X_i)\} = E\{Y_i(0)|D_i = 1, p(X_i)\} = E\{Y_i(0)|p(X_i)\} \quad (13)$$

$$E\{Y_i(1)|D_i = 0, p(X_i)\} = E\{Y_i(1)|D_i = 1, p(X_i)\} = E\{Y_i(1)|p(X_i)\} \quad (14)$$

- En utilisant ces expressions, nous pouvons définir pour chaque cellule définie par  $p(X)$ :

$$\begin{aligned} \delta_{p(x)} &\equiv E\{\Delta_i|p(X_i)\} && (15) \\ &\equiv E\{Y_i(1) - Y_i(0)|p(X_i)\} \\ &\equiv E\{Y_i(1)|p(X_i)\} - E\{Y_i(0)|p(X_i)\} \\ &= E\{Y_i(1)|D_i = 1, p(X_i)\} - E\{Y_i(0)|D_i = 0, p(X_i)\} \\ &= E\{Y_i|D_i = 1, p(X_i)\} - E\{Y_i|D_i = 0, p(X_i)\}. \end{aligned}$$

# Effets moyens du traitement et le score de propension

- En utilisant la Loi des attentes itérées, l'effet moyen du traitement sur les groupes de bénéficiaires est donné par :

$$\begin{aligned}\tau &= E\{\Delta_i | D_i = 1\} \quad (29) \\ &= E\{E\{\Delta_i | D_i = 1, p(X_i)\} | D_i = 1\} \\ &= E\{E\{Y_i(1) | D_i = 1, p(X_i)\} - E\{Y_i(0) | D_i = 0, \\ &\quad , p(X_i)\} | D_i = 1\} \\ &= E\{p(x) | D_i = 1\}\end{aligned}$$

- Où l'attente extérieure est sur la distribution de  $p(X_i) | D_i = 1$ .

# Exécution de la stratégie d'estimation



- Pour exécuter la stratégie d'estimation suggérée par le score de propension et ses propriétés, deux étapes consécutives sont requises.

*i. Estimation du score de propension*

Cette étape est nécessaire parce que le « véritable » score de propension est inconnu et par conséquent le score de propension doit être estimé.

*ii. Estimation de l'effet moyen de traitement étant donné le score de propension*

En bonne logique, dans cette étape nous aimerions

- appairer les cas et les témoins avec exactement le même score de propension (estimé) ;
- Calculer l'effet de traitement pour chaque valeur du score de propension (estimé) (voir équation 28).
- Obtenir la moyenne de ces effets conditionnels comme dans l'équation 29.

# Exécution de la stratégie d'estimation

- ❑ Ceci n'est pas réalisable dans la pratique parce qu'il est rare de trouver deux unités avec exactement le même score de propension.
- ❑ Il y a cependant plusieurs procédures alternatives et réalisables pour réaliser cette étape :
  - Stratification sur le score ;
  - Voisin le plus proche matching sur le score ;
  - Radius matching sur le score ;
  - Kernel matching sur le score ;
  - Pondération sur la base du score.

# Estimation du score de propension

- Apparemment, le même problème de dimensionnalité qui empêche l'estimation des effets de traitement devrait également empêcher l'estimation des scores de propension.
- Toutefois, ceci n'est pas le cas grâce à la *propriété d'équilibre* du score de propension (Lemme 1) selon lequel :
  - Les observations avec le même score de propension ont la même distribution des covariables observables indépendamment du statut du traitement ;
  - Pour un score de propension donné, l'affectation au traitement est aléatoire et par conséquent créée et les unités témoins sont en moyenne identiques sur le plan de l'observation.

# Estimation du score de propension

- Donc, tout modèle de probabilité standard peut être utilisé pour estimer le score de propension, e.g., 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 covariables avec des conditions linéaires et d'ordre supérieur.

- Le choix des conditions d'ordre supérieur à inclure est déterminé uniquement par le besoin d'obtenir une estimation du score de propension qui satisfait la *propriété d'équilibre*.

# Estimation du score de propension

- Pour autant que la spécification de  $h(X_i)$  qui satisfait la *propriété d'équilibre* est plus parcimonieuse que l'ensemble complet des interactions requises pour appariar les cas et les témoins sur la base des observables (comme dans les équations 8 et 9), le score de propension réduit la dimensionnalité du problème d'estimation.
- Noter qu'en raison de cet objectif, l'estimation du score de propension n'a pas besoin d'une interprétation de comportement.

# Un algorithme pour estimer le score de propension

- i. Commencer avec une fonction parcimonieuse de logit ou probit pour estimer le score.
- ii. Trier les données conformément au score de propension estimé (du plus bas au plus haut).
- iii. Stratifier toutes les observations en blocs de telle sorte que dans chaque bloc les scores de propension estimés pour les groupes de bénéficiaires et les témoins ne sont pas statistiquement différents :
  - (a) Commencer avec cinq blocs de type de score égal  $\{0 - 0.2, \dots, 0.8 - 1\}$ ;
  - (b) Tester si les moyennes des scores pour les groupes de bénéficiaires et les témoins sont statistiquement différents dans chaque bloc ;
  - (c) Si oui, augmenter le nombre de blocs et tester à nouveau ;
  - (d) Sinon, passez à l'étape suivante.

(suite sur la diapo suivante...)

# Un algorithme pour estimer le score de propension

- vi. Tester que la *propriété d'équilibre* tient dans tous les blocs pour toutes les covariables :
- (a) pour chaque covariable, tester si les moyennes ( et si possible les conditions d'ordre supérieur) pour les groupes de bénéficiaires et les témoins sont statistiquement différents dans tous les blocs ;
  - (b) si une covariable n'est pas équilibrée dans un bloc, coupez le bloc et tester à nouveau dans chaque bloc plus fin ;
  - (c) si une covariable n'est pas équilibrée dans tous les blocs, modifier l'estimation logit du score de propension en ajoutant plus d'interaction et des conditions d'ordre supérieur et tester à nouveau.
- Noter que dans toute la procédure le résultat n'a pas de rôle.
- Voir the STATA program `pscore.ado` téléchargeable à <http://www.iue.it/Personal/Ichino/Welcome.html>

# Quelques outils utiles de diagnostic

- ❑ Comme nous en avons discuté au début de cette section, les méthodes de score de propension sont basées sur l'idée que l'estimation des effets de traitement requiert un matching soigneux des cas et des témoins.
- ❑ Si les cas et les témoins sont très différents en termes d'observables ce matching n'est pas suffisamment proche et fiable ou il peut même être impossible.
- ❑ La comparaison des scores de propension estimés entre groupes de bénéficiaires et groupes témoins offre un outil utile de diagnostic pour évaluer la similarité entre les cas et les témoins, et par conséquent la façon dont la stratégie d'estimation est fiable.

# Quelques outils utiles de diagnostic

- Plus précisément, il est conseillable de :
  - Compter combien de témoins ont un score de propension plus bas que le minimum ou plus élevé que le maximum des scores de propension des groupes de bénéficiaires.
    - De préférence nous aimerions que l'ordre de variation des scores de propension soit le même dans les deux groupes.
  - Générer des histogrammes des scores de propension estimés pour les groupes de bénéficiaires et les groupes témoins avec des compartiments correspondants aux strates construites pour l'estimation des scores de propension.
    - De préférence nous aimerions une fréquence égale de groupes de bénéficiaires et de groupes témoins dans chaque compartiment.

Noter que ces indicateurs fondamentaux de diagnostic ne sont pas calculés dans une analyse de régression standard, même s'ils étaient également utiles pour cette analyse. (Voir Dehejia et Wahba, 1999).

# Commentaires et extensions

- *Témoins hors de cause*

Si le but est d'estimer l'effet de traitement sur les groupes de bénéficiaires la procédure devrait être appliquées après avoir écarté tous les témoins avec un score de propension plus élevé que le maximum ou plus bas que le minimum des scores de propension des groupes de bénéficiaires.

- *Pénalité pour un nombre inégal de groupes de bénéficiaires et de groupes témoins dans un bloc*

- Noter que s'il y a un bloc dans lequel le nombre de témoins est plus petit que le nombre de groupes de bénéficiaires, la variance augmente et la pénalité est plus grande, plus grande est la fraction de groupes de bénéficiaires dans ce bloc. Si  $N_q^T = N_q^C$  la variance se simplifie à :

$$Var(\tau^S) = \frac{1}{N^T} [Var(Y_i^T) + Var(Y_j^T)] \quad (20)$$

# Commentaires et extensions

- *Alternatives pour l'estimation des résultats moyens dans les blocs*

Dans les expressions ci-dessus, le résultat dans le cas de traitement dans un bloc a été estimé comme le résultat moyen des groupes de bénéficiaires dans ce bloc (et de même pour les témoins).

Une autre possibilité est d'obtenir ces résultats comme des valeurs prédites de l'estimation des fonctions linéaires (ou plus sophistiquées des scores de propension.

Les gains de l'utilisation de ces techniques plus sophistiquées ne semblent pas grands. (Voir Dehejia et Wahba, 1996.)

# Estimation de l'effet de traitement par plus proche voisin, matching Radius et matching Kernel



- ❑ De préférence, nous aimerions apparier chaque unité bénéficiaire avec une unité témoin ayant exactement le même score de propension et vice versa.
- ❑ Ce matching exact est toutefois impossible dans la plupart des applications.
- ❑ Le plus près que nous pouvons arriver à un matching exact est d'apparier chaque unité bénéficiaire avec le témoin *le plus proche* en termes de score de propension.
- ❑ Ceci soulève toutefois la question de savoir quoi faire avec les unités pour lesquelles le match le plus proche a déjà été utilisé..
- ❑ Nous décrivons ici trois méthodes pour régler ce problème.
  - Matching par plus proche voisin avec remplacement ;
  - Matching Radius avec remplacement ;
  - Matching Kernel

# Références

---

- Dehejia, R.H. and S. Wahba (1999), “Causal Effects in Nonexperimental 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 Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, Harvard University, Mimeo.
  
- Hahn, Jinyong (1998), “ ON the role of the propensity score in efficient semiparametric estimation of average treatment effects”, *Econometrica*, 66,2,315-331.

# Références

---

- ❑ 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.