

# Modélisation et identification causale

## Séance 5 – Comment contrôler les caractéristiques observables ?

Pierre Pora

2022-12-08

## Objet de la séance

## Rappel de la semaine dernière

- ▶ Dans certaines situations, on peut justifier que l'on sait identifier les **effets causaux moyens** d'une **intervention** sous deux hypothèses
  - ▶ **Indépendance conditionnelle** : à l'intérieur de strates définies par les valeurs de caractéristiques observables – les variables de conditionnement – tout se passe comme si on regardait une expérience aléatoire contrôlée ou une expérience naturelle
  - ▶ **Support commun** : il n'y a pas de strates dans laquelle il n'y a que des individus ayant fait l'objet de l'intervention ou que des individus n'ayant pas fait l'objet de l'intervention
- ▶ En fait presque toute la fin de la séance était consacrée à la possibilité de **justifier théoriquement** la première hypothèse

## Et après ?

- ▶ Sous ces deux hypothèses, les effets causaux moyens de l'intervention dans la population sont **la moyenne des effets causaux moyens dans chacune des strates**, avec des poids proportionnels à la taille de chaque strate
- ▶ Comme à l'intérieur d'une strate tout se passe comme si on regardait une **expérience aléatoire contrôlée** ou une **expérience naturelle** on peut identifier ces effets causaux moyens spécifiques à chaque strate en faisant de **simples comparaisons de moyennes à l'intérieur de chaque strate**

## Le reste de la séance

- ▶ Tout cela suggère une **stratégie très simple** pour estimer les effets causaux moyens de l'intervention :
  - ▶ On estime à *l'intérieur de chaque strate*
  - ▶ Puis on *agrège*
- ▶ Quand c'est simple, c'est vraiment simple
- ▶ L'objet de la séance :
  - ▶ Montrer qu'en fait **ce n'est (presque) jamais simple**
  - ▶ Proposer des façons de s'en sortir quand ce n'est pas simple
  - ▶ Les implémenter sur des exemples empiriques
  - ▶ (En fait ces solutions sont finalement assez simples)

L'esprit dans lequel j'ai construit la séance : c'est simple mais c'est compliqué mais c'est simple

- ▶ Mon idée générale sur toutes ces approches :
  - ▶ L'intuition de départ est quand même assez simple
  - ▶ Mais quand on rentre dans les détails on voit que ça devient vite compliqué
  - ▶ Mais les raisons pour lesquelles c'est compliqué sont finalement assez simples
    - ▶ Tout tourne autour du fait que **ce n'est pas si facile** que ça d'utiliser l'**hypothèse de support commun**
  - ▶ Et on a des façons raisonnables de lever ces difficultés

## L'esprit dans lequel j'ai construit la séance : c'est simple mais c'est compliqué mais c'est simple

- ▶ Comment ces techniques sont parfois enseignées :
  - ▶ L'intuition de départ qui n'est pas très compliquée
  - ▶ La façon dont on recommande l'implémentation
- ▶ Le problème d'après moi :
  - ▶ On **évacue les difficultés** auxquelles l'implémentation courante répond
  - ▶ Quand celle-ci tombe elle-même dans des problèmes (ça arrive) les anciens étudiants n'ont pas en tête les difficultés auxquelles cette approche vise à répondre
  - ▶ Et mettent donc en place des solutions intuitives qui sont en fait celles dont les dysfonctionnements motivent l'implémentation courante
  - ▶ Problème additionnel de communication

L'esprit dans lequel j'ai construit la séance : c'est simple mais c'est compliqué mais c'est simple

- ▶ Comment je voudrais faire **avec un temps illimité** :
  - ▶ Vous pousser systématiquement à développer des solutions que je sais **condamnées à l'échec**
    - ▶ Bon pour la compréhension de voir pourquoi elles ne marchent pas et de constater empiriquement que ça ne marche pas
    - ▶ Bon pour la pratique informatique aussi
- ▶ **Le temps nous est compté !**
  - ▶ Parfois je vous montrerai moi-même que ça ne marche pas, et pourquoi, mais sans vous laisser le temps de regarder vous-même
- ▶ Rappel : tout le code + des exemples et des détails additionnels sur mon site

## L'esprit dans lequel j'ai construit la séance : implémentation informatique

- ▶ Il existe de nombreux packages disponibles pour R pour implémenter ces techniques (e.g. Matching, cobalt, causalweight)
- ▶ **C'est très bien d'apprendre à s'en servir !**
- ▶ Mais ce n'est **pas l'objet de la séance**
  - ▶ ne permet pas de comprendre ce que font exactement ces techniques
  - ▶ ni de voir pourquoi ça marche
  - ▶ ou de les comparer entre elles
- ▶ Je privilégie ici des implémentations un peu plus pédestres mais (j'espère !) un peu plus transparentes

## Le premier exemple sur lequel repose la séance : Gerber and Green (2000) repris par Imai (2005)

- ▶ Gerber and Green (2000) : **expérience aléatoire contrôlée** visant à étudier plusieurs déterminants de la participation électorale à New Haven pour les élections de 1998
  - ▶ Conduit à quelques résultats étonnantes
  - ▶ En particulier recevoir un appel téléphonique incitant à voter réduirait substantiellement la participation électorale
- ▶ Réévaluation par Imai (2005) qui montre que **la randomisation était défectueuse** et propose des façons de corriger
- ▶ Une extraction des données utilisées par Imai (2005) pour évaluer les effets de l'appel téléphonique est disponible à partir du package Matching

Le premier exemple sur lequel repose la séance : Gerber and Green (2000) repris par Imai (2005)

- ▶ Chargez les données GerberGreenImai du package Matching
- ▶ Pour chaque quartier de New Haven (variable WARD) estimatez la probabilité de recevoir un appel téléphonique incitant à voter (variable PHN.C1)
- ▶ Que peut-on en conclure ?

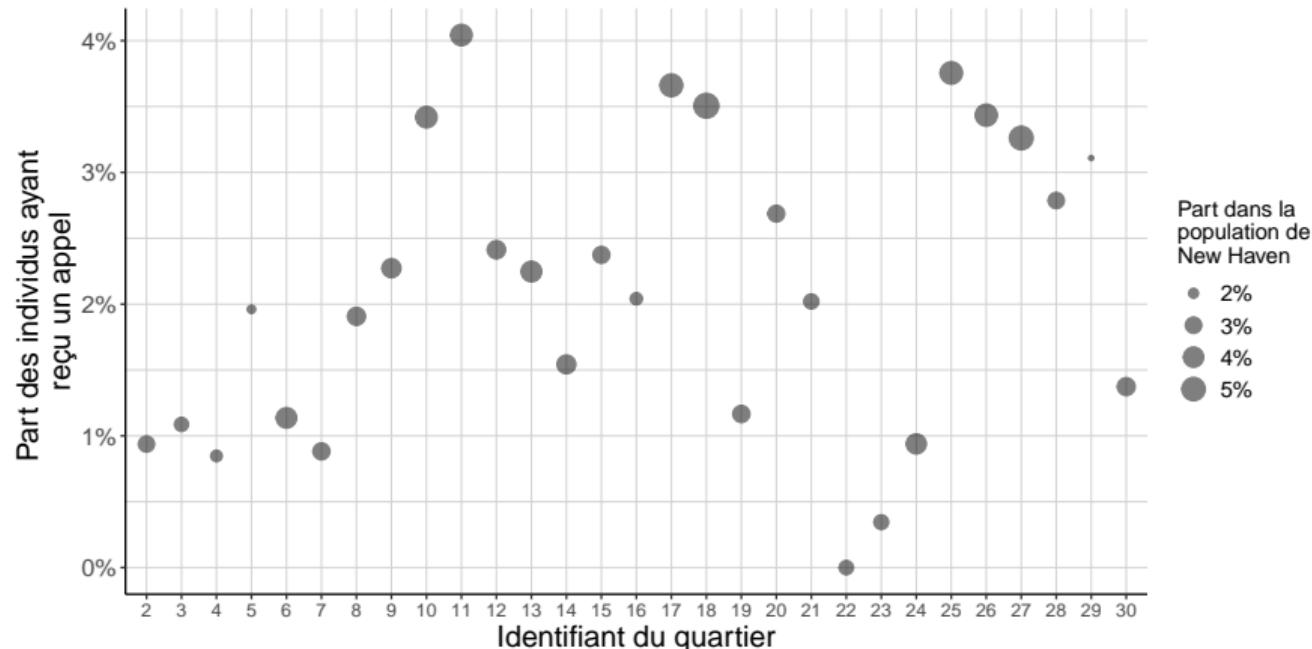
Le premier exemple sur lequel repose la séance : Gerber and Green (2000) repris par Imai (2005)

```
library(data.table)
library(Matching)

#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven
proba_appel<-GerberGreenImai[,  
                           list(prob_phone_call=mean(PHN.C1),  
                                taille=.N),  
                           by=c("WARD")]
```

# La part d'individus ayant reçu l'appel téléphonique diffère largement d'un quartier à l'autre



Le second exemple sur lequel repose la séance : LaLonde (1986), Dehejia and Wahba (1999), Dehejia and Wahba (2002)

- ▶ **Exemple historiquement important**
- ▶ Travaux consacrés à l'évaluation des effets du National Supported Work Demonstration
  - ▶ Un programme d'emploi aidé étastunien mis en place dans la seconde moitié des années 1970
  - ▶ Qui a fait l'objet d'une **expérience aléatoire contrôlée**
- ▶ Proposition de LaLonde (1986) : utiliser les résultats de l'expérience aléatoire contrôlée comme étalon pour **évaluer la qualité des techniques d'évaluations** économétriques usuelles à l'époque :
  - ▶ On garde seulement les données sur les participants au programme, en jetant les volontaires qui du fait de l'aléas n'ont pas été retenus
  - ▶ On utilise des données d'enquête pour constituer un groupe d'individus éligibles

Le second exemple sur lequel repose la séance : LaLonde (1986), Dehejia and Wahba (1999), Dehejia and Wahba (2002)

- ▶ Parvient-ton avec des techniques de régression appliquées à ces données à **retrouver les résultats de l'expérience aléatoire contrôlée ?**
- ▶ **La réponse est négative** pour un très large panel de méthodes
  - ▶ Résultat qui contribue à la préoccupation de toute une génération d'économètres appliqués de proposer des approches plus crédibles et plus transparentes d'évaluation des politiques publiques
- ▶ Dehejia and Wahba (1999) et Dehejia and Wahba (2002) : nouvel examen des données de LaLonde (1986) en utilisant d'autres techniques d'estimation
  - ▶ Résultats plutôt satisfaisants !
- ▶ Beaucoup de travaux consacrés aux techniques d'estimation disponibles sous l'hypothèse d'indépendance conditionnelle mobilisent ces données comme un **terrain de jeu pour évaluer empiriquement la qualité de l'estimation**

## Le second exemple sur lequel repose la séance : LaLonde (1986), Dehejia and Wahba (1999), Dehejia and Wahba (2002)

- ▶ On va utiliser un extrait des données non-expérimentales de LaLonde (1986) mobilisées par Dehejia and Wahba (1999)
  - ▶ Un groupe d'individus qui ont été tirés au sort pour participer au NSW dans le cadre expérimental
  - ▶ Un groupe d'individus éligibles au programme, constitué par LaLonde (1986) à partir du Current Population Survey (~ équivalent étatsunien de l'enquête Emploi)
- ▶ Données disponibles à partir du package cobalt sous le nom `lalonde`
  - ▶ Attention le package Matching propose aussi des données sous ce nom mais cette fois ce sont les données issues de l'expérience aléatoire contrôlée !

Une estimation apparemment facile

## Quand tout se passe bien

- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le quartier de résidence dans New Haven (variable WARD)
- ▶ Estimez les effets causaux moyens de l'appel (variable PHN.C1) sur la participation aux élections de 1998 (variable VOTED98) sous cette hypothèse

## Quand tout se passe bien

```
library(data.table)
library(ggplot2)
library(Matching)

#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven
proba_appel<-GerberGreenImai[,
                           list(prob_phone_call=mean(PHN.C1),
                                taille=.N),
                           by=c("WARD")]
```

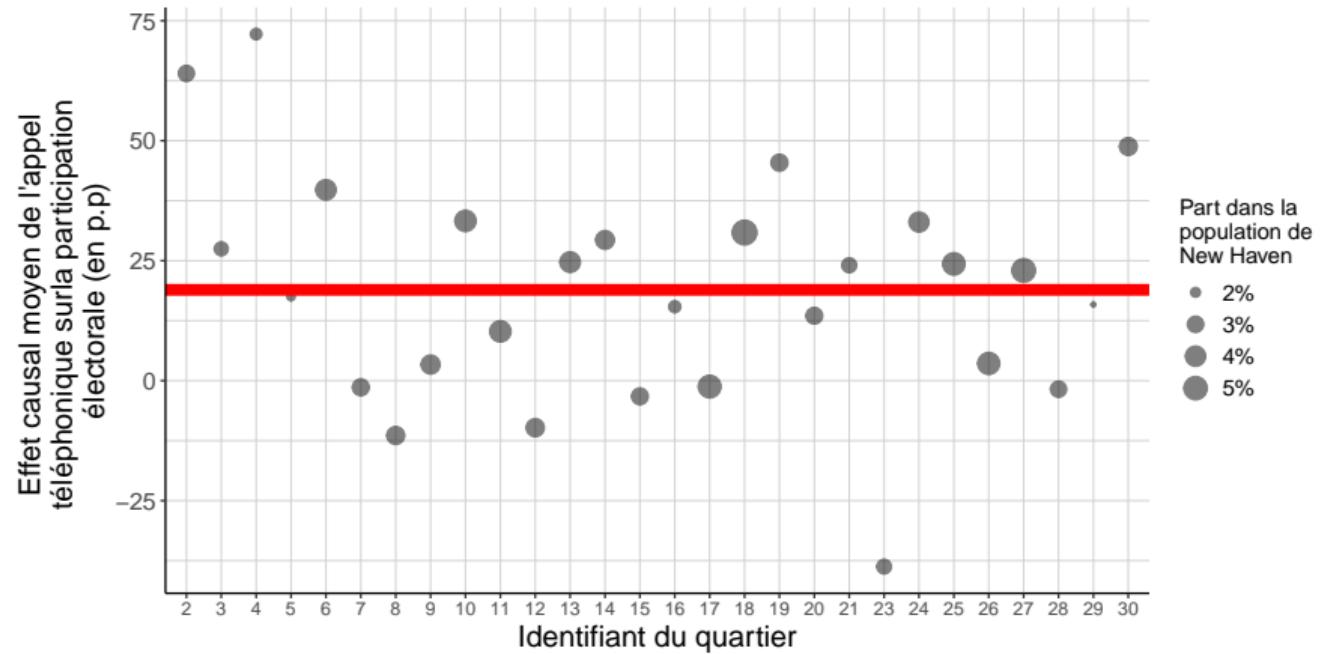
## Quand tout se passe bien

```
#On estime les probabilités de vote, selon que l'on a reçu un appel ou non,  
# quartier par quartier  
proba_vote<-GerberGreenImai[,  
                                list(prob_vote_traite=sum(VOTED98*PHN.C1)/  
                                    sum(PHN.C1),  
                                    prob_vote_controle=sum(VOTED98*(1-PHN.C1))/  
                                    sum(1-PHN.C1),  
                                    taille=.N),  
                                by=c("WARD")]  
  
#Comme à l'intérieur de chaque quartier, on suppose que la randomisation est  
# correcte, on peut en tirer un effet moyen spécifique à chaque quartier  
proba_vote[,  
          effet_moyen:=prob_vote_traite-prob_vote_controle]
```

## Quand tout se passe bien

```
#L'effet causal moyen sur tout New Haven est la moyenne des effets moyens dans
# chaque quartier (quand on pondère par la taille du quartier)
ATE_direct<-proba_vote[!is.na(effet_moyen),
                         sum(effet_moyen*taille)/
                         sum(taille)]
```

# Quand tout se passe bien



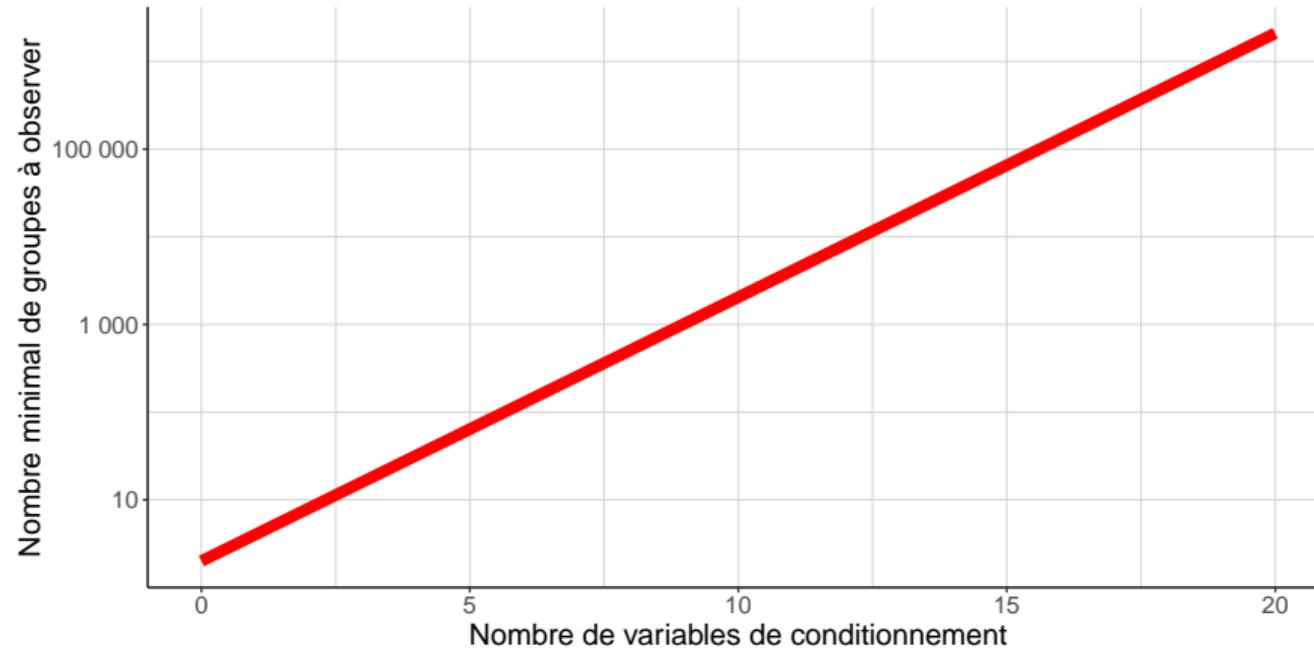
Pourquoi les choses ne sont (presque) jamais aussi simples

Fléau de la dimension

## Que se passe-t-il lorsque l'on considère davantage de variables de conditionnement ?

- ▶ Chaque variable de conditionnement défini *au moins* deux strates
- ▶ Donc chaque fois que l'on rajoute une variable de conditionnement on doit *au moins* doubler le nombre de strates à considérer
- ▶ Le **nombre de strates** dans lesquelles on veut faire la comparaison **croît de façon exponentielle** avec le **nombre de variables** de conditionnement

Le nombre minimal de groupes à observer pour pouvoir faire la comparaison à l'intérieur de chaque strate augmente très rapidement avec le nombre de variables de conditionnement



## Fléau de la dimension

- ▶ Le nombre de groupes sur lesquels on doit recueillir des données augmente de façon explosive avec le nombre de variables de conditionnement
- ▶ On travaille pourtant sur un échantillon de taille finie !
  - ▶ L'enquête Emploi en continu = une grosse enquête → 92000 logements par trimestre
  - ▶ Dès qu'on va avoir besoin de plus de 15 variables **il est certain que l'on va rencontrer des groupes vides dans les données**
- ▶ On appelle la croissance exponentielle du nombre de groupe avec la dimension des variables de conditionnement **fléau de la dimension**

## Et avec une variable de conditionnement continue ?

- ▶ Chargez les données de LaLonde (1986) à partir du package cobalt
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le niveau des revenus du travail en 1974 (variable `re74`)
- ▶ Pourvez-vous estimer les effets causaux moyens de la participation au NSW (variable `treat`) sur les revenus du travail en 1978 (variable `re78`) sous cette hypothèse ? Pourquoi ?

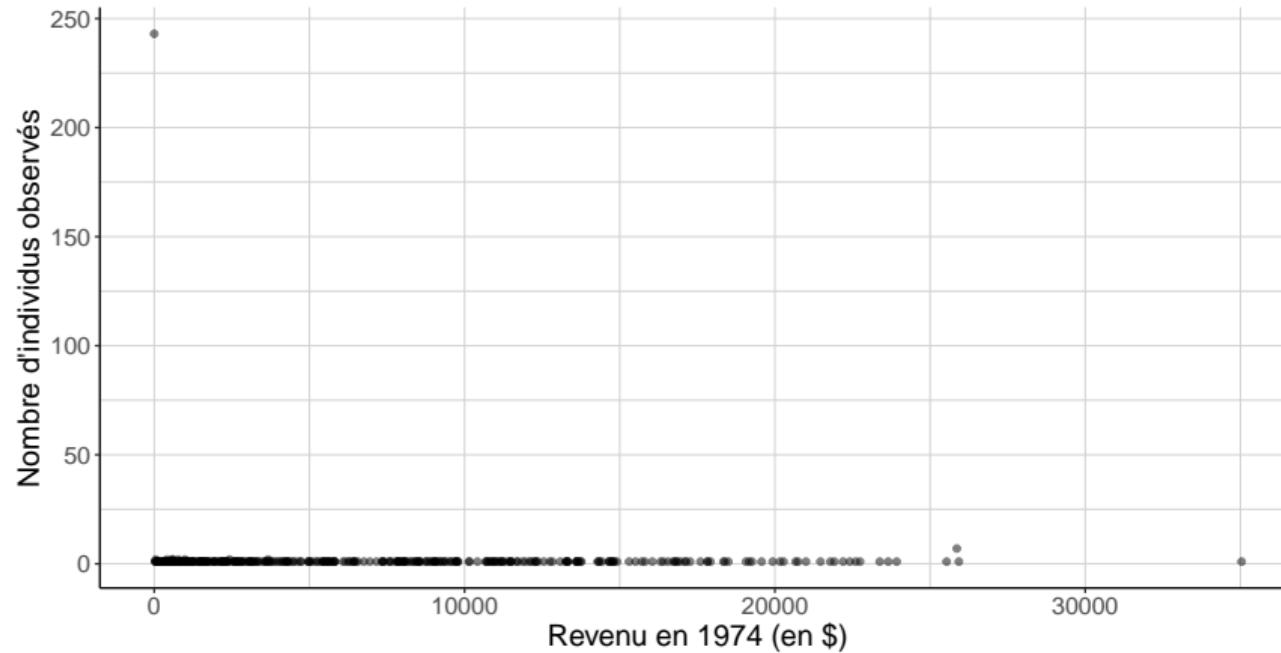
## Et avec une variable de conditionnement continue ?

```
library(cobalt)
library(data.table)

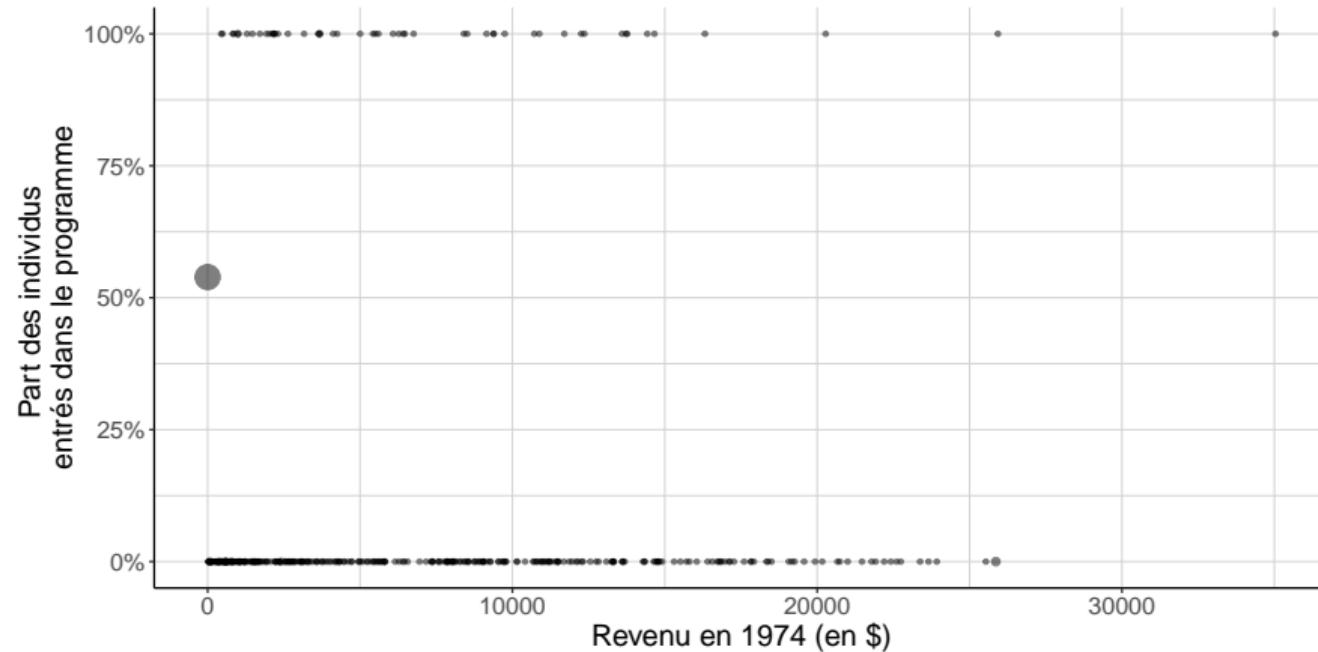
data("lalonde")
lalonde<-data.table(lalonde)

prob_traitement<-lalonde[,
                           list(prob_treat=mean(treat),
                                taille=.N),
                           by=c("re74")]
```

Les valeurs d'une variable continue défissent une quasi-infinité de strates dans lesquelles on ne trouve qu'un individu



Et dans les strates avec un seul individu, la part de ceux qui font l'objet de l'intervention ne peut être que 0 ou 100% : la comparaison est impossible



Il faut extrapoler !

## Que fait-on ici et que pourrait-on essayer de faire à la place ?

- ▶ On interprète ici **très rigoureusement** l'hypothèse d'indépendance conditionnelle en considérant que les strates sont définies par les valeurs exactes des variables de conditionnement
- ▶ Ce qui se passe pour l'individu qui a gagné \$5424 en 1974 ne nous dit rien d'utile sur ce qui aurait pu raisonnablement arriver à celui qui a gagné \$5454
- ▶ On peut peut-être être un peu moins littéral que ça et utiliser l'information des individus avec des revenus proches mais pas strictement égaux pour boucher les trous...

## Boucher les trous en utilisant les voisins

```
library(cobalt)
library(data.table)

data("lalonde")
lalonde<-data.table(lalonde)

table_extrapolation<-
  setorder(lalonde[,
    list(revenu78_observe_traite=
        sum(re78*treat)/
        sum(treat),
        revenu78_observe_nontraite=
        sum(re78*(1-treat))/
        sum(1-treat)),
    by=c("re74")],
  re74)
```

## Boucher les trous en utilisant les voisins

```
table_extrapolation[re74>=5250  
                  & re74<=5750]
```

```
##          re74 revenu78_observe_traite revenu78_observe_nontraite  
## 1: 5260.631                 NaN             9253.5240  
## 2: 5374.269                 NaN            0.0000  
## 3: 5424.485                 6788.4630           NaN  
## 4: 5454.599                 NaN            0.0000  
## 5: 5460.477                 NaN            7539.3610  
## 6: 5506.308                 671.3318           NaN  
## 7: 5523.173                 NaN            5040.5250  
## 8: 5597.625                 NaN            122.6513  
## 9: 5605.852                 0.0000           NaN  
## 10: 5607.422                NaN            94.5745  
## 11: 5683.833                NaN            4742.0250  
## 12: 5699.507                NaN            8844.1940
```

## Accepter d'extrapoler

- ▶ Il faut interpréter l'hypothèse d'indépendance conditionnelle\* **un peu moins strictement** que ce que l'on a fait auparavant !
- ▶ Pour cela on va extrapolier : utiliser de l'information qui n'est pas *exactement* celle que l'on voudrait, que l'on obtient en utilisant une autre information un peu en dehors de la région sur laquelle on est sûrs qu'elle est valable
  - ▶ Cette information c'est celle des individus qui n'ont pas exactement les mêmes valeurs des variables de conditionnement, mais ne sont pas trop loin quand même
- ▶ Il faut donc se doter d'un **concept de proximité entre individus** ou (ça revient au même) entre strates définies par les valeurs des variables de conditionnement
  - ▶ L'idée (en gros) : cela va nous permettre de définir des **strates plus grosses** et de ne plus avoir les mêmes problèmes

## Notion de score de propension

## Qu'est-ce que le score de propension ? Quel est son intérêt ?

- ▶ **Le problème que l'on a** : les valeurs des variables de conditionnement définissent des **strates trop fines** pour qu'on puisse les utiliser avec des échantillons de taille finie et tirer parti de l'hypothèse de support commun
- ▶ **Ce que l'on voudrait** : des **strates plus grosses** dans lesquelles la comparaison entre ceux qui font l'objet de l'intervention et les autres ait encore une interprétation causale
- ▶ **Les valeurs du score de propension définissent les plus grosses strates possibles (au sens de l'inclusion) telles que la comparaison ait une interprétation causale**

## Comment définir le score de propension

- ▶ C'est une **fonction des seules caractéristiques observables** définies par les variables de conditionnement
- ▶ Quand on considère une strate définie par la valeur exacte des variables de conditionnement, le score de propension a donc **la même valeur pour tous les individus de la strate**
  - ▶ Une strate définie par une valeur des variables de conditionnement est forcément incluse dans la strate définie par la valeur du score de propension
- ▶ C'est **la probabilité dans la population** ( $\neq$  la part dans l'échantillon) de faire l'objet de l'intervention pour les individus de cette strate

$$p(X_i) := \mathbb{P}(D_i = 1 | X_i)$$

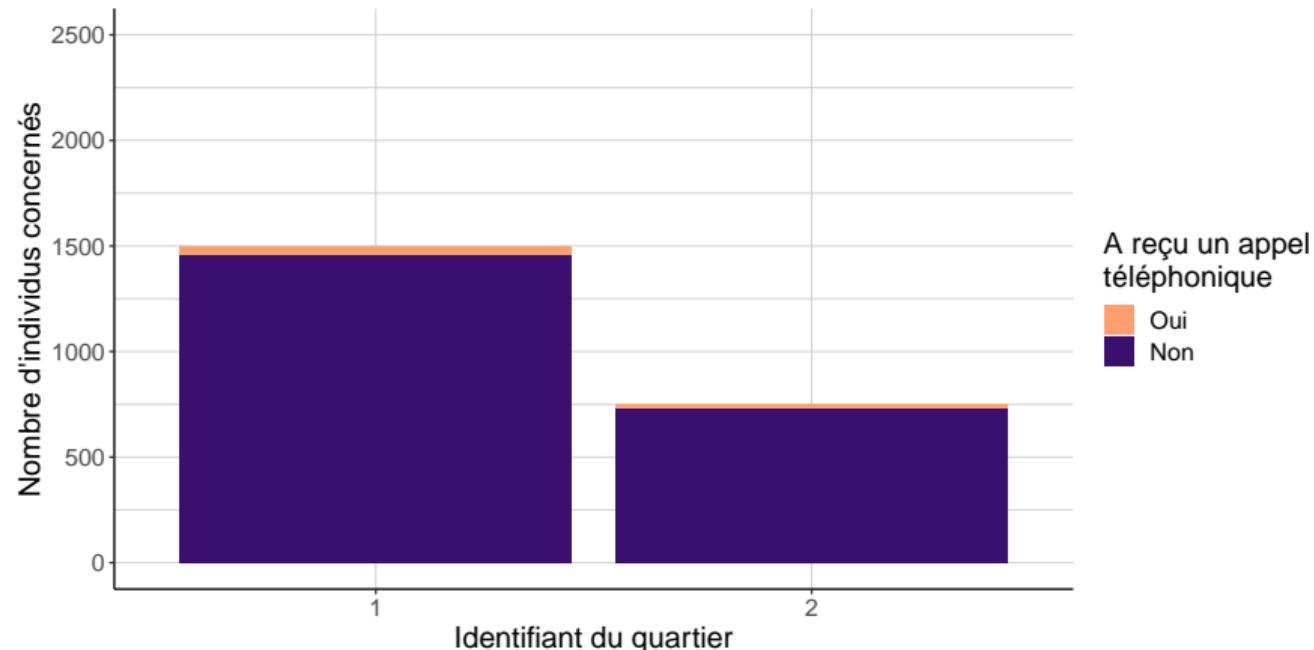
- ▶ Ou encore puisque  $D_i$  est dichotomique

$$p(X_i) := \mathbb{E}[D_i | X_i]$$

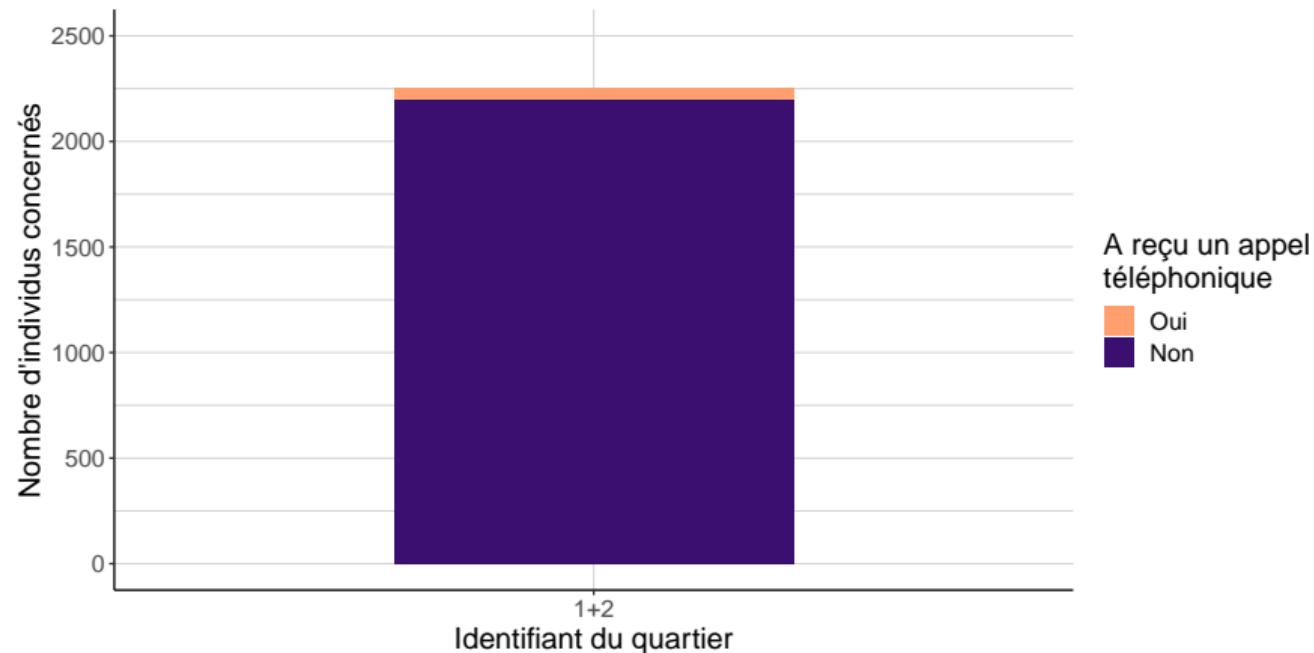
## Propriétés fondamentales du score de propension

- ▶ Au sein d'une strate définie par la valeur exacte  $p_0$  du score de propension, **tout se passe comme si on regardait une expérience aléatoire contrôlée ou une expérience naturelle** :
  - ▶ La distribution des caractéristiques observables définies par les variables de conditionnement est la même pour les individus qui ont fait l'objet de l'intervention et pour ceux qui n'en ont pas fait l'objet
  - ▶ La distribution des valeurs potentielles de la variable d'intérêt est la même dans les deux groupes
- ▶ On peut donc toujours simplement **comparer les individus selon qu'ils ont fait ou pas l'objet de l'intervention**
- ▶ Et en déduire la valeur des **effets causaux moyens de l'intervention** dans cette grosse strate
- ▶ On parle de **propriété équilibrante du score de propension**

# Pourquoi ça marche ? Une expérience de pensée fondée sur Imai (2005)



## Pourquoi ça marche ? Une expérience de pensée fondée sur Imai (2005)



Trois grandes familles de techniques fondées sur le score de  
propension

## Stratification

## Stratification sur le score de propension

- ▶ C'est la technique la plus **intuitive**
- ▶ Le score de propension définit les strates les plus grosses (au sens de l'inclusion) au sein desquelles la situation est assimilable à une expérience aléatoire contrôlée ou à une expérience naturelle
- ▶ **On se met dans ces grosses strates** et on fait la comparaison entre ceux qui ont et ceux qui n'ont pas fait l'objet de l'intervention
- ▶ La différence est égale aux effets causaux moyens spécifiques à ces grosses strates
- ▶ On peut les **réagréger avec des poids égaux à la taille de ces grosses strates** pour récupérer les effets causaux moyens dans toute la population

## Stratification sur le score de propension : un exemple empirique

- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le quartier de résidence dans New Haven (variable WARD)
- ▶ On suppose aussi que la part d'individus ayant reçu l'appel téléphonique incitant à voter dans chaque quartier est un estimateur raisonnable du score de propension dans ce contexte
- ▶ Estimez le score de propension et estimez par stratification les effets causaux de l'appel téléphonique sur la participation électorale sous ces hypothèses

## Stratification sur le score de propension : un exemple empirique

```
#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven : c'est le score de propension
proba_appel<-GerberGreenImai[,  
                                list(score_propension=mean(PHN.C1),  
                                     taille=.N),  
                                by=c("WARD")]

#On remet ça dans les données
GerberGreenImai<-merge(GerberGreenImai,  
                       proba_appel,  
                       by=c("WARD"))
```

## Stratification sur le score de propension : un exemple empirique

## Stratification sur le score de propension : un exemple empirique

```
#On réagrège cette effets moyens par grande strate en faisant la moyenne
# avec des poids proportionnels à la taille de chaque grande strate
ATE_stratif<-effet_par_score_propension[,  
                                list(ATE=  
                                     sum(taille*effet_causal)/  
                                     sum(taille))]  
  
ATE_stratif
```

```
##          ATE  
## 1: 0.1892938
```

Appariement

## Appariement sur le score de propension

- ▶ **Procédure d'appariement :**
  - ▶ Chaque fois que l'on rencontre un individu qui a reçu l'appel téléphonique (qui ne l'a pas reçu)
  - ▶ On lui **associe un ou des individus** qui n'ont pas reçu l'appel téléphonique (qui l'ont reçu) et qui ont la même valeur du score de propension → appariement sur le score de propension
  - ▶ Par différence on peut calculer un **pseudo-effet causal individuel** : sa participation moins celle de l'individu qu'ont lui a apparié, ou la moyenne de celles des individus qu'on lui a apparié s'ils sont plusieurs (l'opposé de cette différence si l'individu de départ n'a pas reçu l'appel téléphonique)
  - ▶ Ensuite on fait la **moyenne de ces pseudo-effets causaux individuels** sur toute la population

## Appariement sur le score de propension : un exemple empirique

- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le quartier de résidence dans New Haven (variable WARD)
- ▶ On suppose aussi que la part d'individus ayant reçu l'appel téléphonique incitant à voter dans chaque quartier est un estimateur raisonnable du score de propension dans ce contexte
- ▶ Estimez le score de propension et estimez, par appariement sur le score de propension, en appariant à chaque individu tous ceux qui ont le même score de propension que lui, les effets causaux moyens de l'appel téléphonique sur la participation électorale

## Appariement sur le score de propension : un exemple empirique

```
#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven : c'est le score de propension
proba_appel<-GerberGreenImai[,  
                                list(score_propension=mean(PHN.C1),  
                                     taille=.N),  
                                by=c("WARD")]

#On remet ça dans les données
GerberGreenImai<-merge(GerberGreenImai,  
                        proba_appel,  
                        by=c("WARD"))
```

## Appariement sur le score de propension : un exemple empirique

```
#On associe à chaque individu des couples avec :  
# tous les individus ayant le même score de propension que lui  
# mais qui n'ont pas reçu l'appel téléphonique si l'individu de départ l'a  
# reçu, et inversement s'il ne l'a pas reçu  
  
#On met un identifiant individuel dans les données  
GerberGreenImai[,  
    indiv_id:=as.character(.I)]  
  
#On veut apparter chaque individu à seulement à des individus dont le statut  
# vis-à-vis de l'appel téléphonique est le contraire du sien, on crée une  
# variable pour cela  
GerberGreenImai[,  
    oppose_statut_appel:=1-PHN.C1]
```

## Appariement sur le score de propension : un exemple empirique

```
#On peut créer des couples
creation_couple<-merge(GerberGreenImai[score_propension!=0,
                                         c("indiv_id",
                                           "score_propension",
                                           "PHN.C1",
                                           "VOTED98")],
                         GerberGreenImai[score_propension!=0,
                                         c("score_propension",
                                           "oppose_statut_appel",
                                           "VOTED98")],
                         by.x=c("score_propension",
                               "PHN.C1"),
                         by.y=c("score_propension",
                               "oppose_statut_appel"),
                         all.X=TRUE,
                         allow.cartesian=TRUE)
```

## Appariement sur le score de propension : un exemple empirique

```
#On calcule le pseudo effet individuel à l'intérieur de chaque couple
creation_couple[,
  effet_indiv:=PHN.C1*(VOTED98.x-VOTED98.y)+  

  (1-PHN.C1)*(VOTED98.y-VOTED98.x)]  
  

#On calcule l'effet causal moyen comme la moyenne de ces pseudo-effets
# individuels
ATE_matching<-creation_couple[,  

  list(moyenne_indiv=mean(effet_indiv)),
  by=c("indiv_id"))][,
  list(ATE=mean(moyenne_indiv))]  
  

ATE_matching  
  

##          ATE
## 1: 0.1892938
```

Repondération

## Repondération : qu'est-ce que c'est ?

- ▶ Résultat sur lequel tout repose : il existe une **pondération telle que lorsqu'on l'utilise, la différence de comportement moyen entre ceux qui ont reçu l'appel téléphonique et les autres identifie les effets causaux moyens de l'appel téléphonique**
- ▶ Analogie avec la **statistique d'enquête** : une expérience aléatoire ~ un sondage aléatoire → la différence entre ceux qui font l'objet de l'intervention et ceux qui n'en font pas l'objet coïncide avec les effets causaux moyens de l'intervention parce que :
  - ▶ Comme on les a choisis au hasard les individus qui font l'objet de l'intervention sont représentatifs de la population dans son ensemble
  - ▶ Idem pour ceux qui n'en font pas l'objet

## Repondération : qu'est-ce que c'est ?

- ▶ Là c'est exactement la même chose sauf que **la probabilité de tirage n'est pas la même pour tout le monde**
  - ▶ Il faut une pondération qui redresse le groupe de ceux qui ont fait l'objet de l'intervention de sorte qu'ils soient représentatifs de la population dans son ensemble
  - ▶ Idem pour ceux qui n'en font pas l'objet

## Repondération : quelle est la valeur de ces poids ?

- ▶ Approche classique de statistique d'enquête pour les sondages aléatoires stratifiés : il faut des **poids inversement proportionnels à la probabilité de rentrer dans l'échantillon**
- ▶ Par analogie il faut des poids :
  - ▶ inversement proportionnels au score de propension  $p(X_i)$  pour les individus qui ont fait l'objet de l'intervention
  - ▶ inversement proportionnels au complément du score de propension  $1 - p(X_i)$  pour les individus qui n'ont pas fait l'objet de l'intervention
- ▶ Pour avoir des poids égaux à 1 en moyenne : mettre la probabilité de faire l'objet de l'intervention (de ne pas faire l'objet de l'intervention) au numérateur

## Repondération : quelles sont les propriétés des poids ?

- ▶ Avec des tels poids **tout se passe comme si on faisait face à une expérience aléatoire contrôlée ou une expérience naturelle**
  - ▶ La différence moyenne entre le groupe qui fait l'objet de l'intervention et les autres identifie les effets causaux moyens de l'intervention
  - ▶ La distribution des caractéristiques observables dans chacun des deux groupes est la même que dans la population toute entière
    - ▶ et donc en particulier la même dans les deux groupes

## Repondération : un exemple empirique

- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le quartier de résidence dans New Haven (variable WARD)
- ▶ On suppose aussi que la part d'individus ayant reçu l'appel téléphonique incitant à voter dans chaque quartier est un estimateur raisonnable du score de propension dans ce contexte
- ▶ Estimez le score de propension et estimez, par répondération les effets causaux moyens de l'appel téléphonique sur la participation électorale

## Repondération : un exemple empirique

```
#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven : c'est le score de propension
proba_appel<-GerberGreenImai[,  
                                list(score_propension=mean(PHN.C1),  
                                     taille=.N),  
                                by=c("WARD")]

#On remet ça dans les données
GerberGreenImai<-merge(GerberGreenImai,  
                        proba_appel,  
                        by=c("WARD"))
```

## Repondération : un exemple empirique

```
#On va aussi avoir besoin de la part totale des individus
# qui ont reçu l'appel téléphonique
proba_appel_globale<-mean(GerberGreenImai[score_propension!=0]$PHN.C1)

#On calcule le jeu de poids dont on a besoin
GerberGreenImai[,  
  poids:=fcase(  
    PHN.C1==1,  
    proba_appel_globale/score_propension,  
    PHN.C1==0,  
    (1-proba_appel_globale)/  
      (1-score_propension)  
  )]
```

## Repondération : un exemple empirique

```
#On calcule la différence des taux de participation électorale
# entre ceux qui ont et qui n'ont pas reçu l'appel téléphonique
# avec ce nouveau jeu de poids
ATE_reweighting<-GerberGreenImai[score_propension!=0,
                                    list(ATE=sum(PHN.C1*poids*VOTED98)/
                                         sum(PHN.C1)-
                                         sum((1-PHN.C1)*poids*VOTED98)/
                                         sum(1-PHN.C1))]

ATE_reweighting

##          ATE
## 1: 0.1892938
```

Pour utiliser le score de propension, il faut d'abord l'estimer !

Une estimation estimation apparemment facile (de nouveau !)

## Estimation naïve du score de propension

- ▶ Dans tout ce qui précède, on a eu recours à une **estimation naïve du score de propension** :
  - ▶ L'ensemble des variables de conditionnement se résumait au quartier de résidence dans New Haven
  - ▶ Les strates définies par les valeurs exactes des variables de conditionnement sont donc exactement la même chose que les quartiers de New Haven
  - ▶ Le score de propension est la probabilité de recevoir l'appel dans chaque quartier
  - ▶ On l'a estimé comme la part d'individus ayant reçu l'appel dans le quartier
- ▶ C'est la façon intuitive, mais naïve de faire !
- ▶ Mais pourquoi naïve exactement ?

Pourquoi ça ne marche pas ?

## Estimation naïve du score de propension : retour sur LaLonde (1986)

- ▶ Chargez les données de LaLonde (1986) à partir du package cobalt
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le niveau des revenus du travail en 1974 (variable re74)
- ▶ Quel est l'estimateur naïf du score de propension ici ? Implémentez-le
- ▶ Pourquoi ne peut-on pas l'utiliser pour estimer les effets causaux moyens de la participation au NSW sur les revenus du travail en 1978 ?

## Estimation naïve du score de propension : retour sur LaLonde (1986)

```
library(cobalt)
library(data.table)

data("lalonde")
lalonde<-data.table(lalonde)

prob_traitement<-lalonde[,
                           list(score_propension_naif=mean(treat),
                                taille=.N),
                           by=c("re74")]
```

L'estimation naïve du score de propension permet de passer de 358 strates à 3 strates beaucoup plus grosses

```
nrow(prob_traitement)  
## [1] 358  
length(levels(factor(prob_traitement$score_propension_naif)))  
## [1] 3
```

Mais parce que 2 de ces strates sont définies à partir des valeurs 0 et 1 du score de propension estimé, elles sont inutilisables, et la troisième coïncide avec une strate que l'on a déjà sans passer par le score de propension

```
strate_score_propension_naif<-setorder(prob_traitement[,  
                                         list(taille=sum(taille),  
                                              min_re74=min(re74),  
                                              max_re74=max(re74),  
                                              nbr_strates_init=.N),  
                                         by=c("score_propension_naif")],  
                                         taille)  
  
strate_score_propension_naif  
  
##   score_propension_naif taille  min_re74 max_re74 nbr_strates_init  
## 1:      1.0000000     54 445.17040 35040.07          52  
## 2:      0.5390947    243  0.00000     0.00           1  
## 3:      0.0000000    317 33.30754 25862.32         305
```

## Estimation naïve du score de propension : et quand ça marche ? Retour sur Imai (2005)

- ▶ Reprenez les valeurs estimées des effets causaux moyens de l'appel téléphonique incitant à voter sur la participation aux élections de 1998 à New Haven, sous l'hypothèse que le conditionnement sur le quartier suffit
- ▶ Comparez ces valeurs entre elles les valeurs obtenues avec les différentes techniques fondées sur le score de propension estimé de façon naïve, et comparez les également avec l'estimation directe effectuée en début de séance
- ▶ Que peut-on en déduire ?

Quand l'estimation naïve du score de propension permet d'estimer les effets causaux moyens de l'intervention, elle est exactement équivalente à l'estimation directe de ces effets causaux moyens et le passage par le score de propension n'a aucun intérêt

```
all.equal(ATE_direct,
          as.numeric(ATE_stratif))

## [1] TRUE

all.equal(ATE_direct,
          as.numeric(ATE_matching))

## [1] TRUE

all.equal(ATE_direct,
          as.numeric(ATE_reweighting))

## [1] TRUE
```

On ne va pas faire grand chose avec l'estimation naïve du score de propension... Mais pourquoi au juste ?

- ▶ On veut passer par le score de propension parce qu'on assume la **nécessité de se placer sur de plus grosses strates quitte à extrapoler...**
- ▶ Mais en fait en utilisant l'estimation naïve du score de propension **on refuse totalement d'extrapoler**
  - ▶ On n'utilise pour cette estimation que ce qui se passe dans les strates définies par les valeurs exactes des variables de conditionnement
  - ▶ Sans faire appel à ce qui se passe pas trop loin
- ▶ Une forme d'incohérence dans l'approche
- ▶ Il faut assumer clairement la nécessité d'extrapoler !

Comment s'en sortir en pratique

## Mettre de la structure sur l'estimation du score de propension

- ▶ Pour extrapolier on va mettre des contraintes beaucoup plus importantes sur l'estimation du score de propension
- ▶ **Approche paramétrique** : on va utiliser des modèles de régression dichotomiques
  - ▶ Cela revient à chercher  $p(x)$  sous la forme  $f(x'\beta)$ 
    - ▶ La fonction de lien  $f$  est choisie *a priori* — **logit, probit etc.**
    - ▶ Le vecteur  $\beta$  est de dimension finie (*choisie a priori*)
    - ▶ Estimer le score de propension équivaut à estimer  $\beta$

## Mettre de la structure sur l'estimation du score de propension : des choix à faire !

- ▶ Il faut **choisir la fonction de lien  $f$** 
  - ▶ Veut-on estimer un logit, un probit etc.
- ▶ Il faut choisir ce que l'on met dans le modèle de régression — **choix de spécification**
  - ▶ On a dit par exemple qu'on mettait le revenu du travail en 1974 dans les variables de conditionnement
  - ▶ Mais que met-on dans la régression ? le revenu ? son logarithme ? des termes d'ordre  $>1$  ? des interactions ?

## Mettre de la structure sur l'estimation du score de propension : des choix à faire !

- ▶ Il n'y a **aucune règle *a priori* pour guider ces choix**
- ▶ On navigue entre **deux écueils**
  - ▶ Une spécification trop riche et trop souple → on retombe sur les problèmes de l'estimation naïve
  - ▶ Une spécification trop pauvre et trop rigide → le score de propension estimé risque d'être biaisé et loin du vrai score de propension dont on a théoriquement besoin

## Mettre de la structure sur l'estimation du score de propension : des choix à faire !

- ▶ En apparence : une mission presque impossible !
- ▶ **En pratique : on peut s'en sortir**
  - ▶ Lorsque la spécification est suffisamment riche, le choix de fonction de lien n'importe pas
  - ▶ On a des éléments empiriques que l'on peut regarder pour détecter les spécifications pas suffisamment riches

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le quartier de résidence dans New Haven (variable WARD)
- ▶ Estimez le score de propension avec un modèle logit, un modèle probit, un modèle cauchit, par les MCO, et avec l'estimateur naïf.
- ▶ Que pouvez-vous remarquer ?

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

```
#On récupère les données et on convertit en data.table
data(GerberGreenImai)
GerberGreenImai<-data.table(GerberGreenImai)

#On estime la probabilité d'avoir reçu un appel téléphonique pour chaque
# quartier de New Haven, c'est-à-dire le score de propension
# avec l'estimateur naïf
stat_niveau_quartier<-GerberGreenImai[,
                           list(score_propension_naif=mean(PHN.C1)),
                           by=c("WARD")]

#On remet ça avec les données initiales
GerberGreenImai<-merge(GerberGreenImai,
                       stat_niveau_quartier,
                       by=c("WARD"))
```

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

```
#On fait la même estimation avec un logit, et on récupère
# les probabilités prédites, c'est-à-dire l'estimation du score
# de propension
logit_propensity_score<-glm(PHN.C1 ~ WARD,
                               data=GerberGreenImai,
                               family = binomial(link = logit))$fitted.values
```

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

```
#On fait la même estimation avec un probit, et on récupère
# les probabilités prédites, c'est-à-dire l'estimation du score
# de propension
probit_propensity_score<-glm(PHN.C1 ~ WARD,
                                data=GerberGreenImai,
                                family = binomial(link = probit))$fitted.values
```

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

```
#On fait la même estimation avec un cauchit, et on récupère
# les probabilités prédites, c'est-à-dire l'estimation du score
# de propension
cauchit_propensity_score<-glm(PHN.C1 ~ WARD,
                                 data=GerberGreenImai,
                                 family = binomial(link = cauchit))$fitted.values
```

Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

## Le choix de fonction de lien importe-t-il ? Retour sur Imai (2005)

```
#On regroupe ces quatre estimations avec les données initiales  
GerberGreenImai<-cbind(GerberGreenImai,  
                           logit_propensity_score,  
                           probit_propensity_score,  
                           cauchit_propensity_score,  
                           MC0_propensity_score)
```

Pour le modèle saturé, lorsque les variables de conditionnement décrivent des groupes en nombre fini, tous les choix de fonction de lien sont équivalents

```
#On vérifie que les quatre estimations sont égales entre elles
all(all.equal(logit_propensity_score,
              probit_propensity_score),
    all.equal(logit_propensity_score,
              cauchit_propensity_score),
    all.equal(logit_propensity_score,
              MC0_propensity_score,
              tolerance=1e-7))
```

```
## [1] TRUE
```

*Pour les MC0 il faut autoriser des très légères différences qui tiennent à la façon dont glm procède à l'estimation*

En fait, ils sont équivalents à l'estimateur naïf du score de propension

```
#On vérifie que ces estimations sont égales à l'estimation
# naïve du score de propension
#L'erreur ne tient qu'à certaines approximations inhérentes
# à l'estimation par glm
all.equal(GerberGreenImai$score_propension_naif,
          GerberGreenImai$logit_propensity_score,
          tolerance=1e-7)

## [1] TRUE
```

## Et avec une variable de conditionnement continue ? Retour sur LaLonde (1986)

- ▶ Chargez les données de LaLonde (1986) à partir du package cobalt
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le niveau des revenus du travail en 1974 (variable re74)
- ▶ Estimez le score de propension avec toutes les fonction de lien précédentes lorsque l'on n'inclut qu'un terme linéaire en revenus du travail en 1974 dans les régressions
- ▶ Recommencez le même exercice en incluant en plus une variable dichotomique qui vaut 1 si ces revenus sont nuls et 0 sinon
- ▶ Que remarque-t-on ?

## Et avec une variable de conditionnement continue ? Retour sur LaLonde (1986)

```
library(cobalt)
library(data.table)

#On récupère les données
data("lalonde")
lalonde<-data.table(lalonde)

#On crée la variable qui indique des revenus du travail nul en 1974
lalonde[,  
    u74:=as.numeric(re74==0)]
```

## Et avec une variable de conditionnement continue ? Retour sur LaLonde (1986)

```
#Fonction qui renvoie les probabilités prédictes d'entrée dans le programme
#pour chaque observation, pour un choix de spécification et une
#fonction de lien dans un modèle dichotomique
estimationparamdichot<-function(specification,
                                    link_function){

  glm(as.formula(paste0("treat ~",
                        specification)),
      data = lalonde,
      family = binomial(link = link_function))$fitted.values

}
```

## Et avec une variable de conditionnement continue ? Retour sur LaLonde (1986)

```
#Fonction qui renvoie les probabilités prédictes d'entrée dans le programme
#pour chaque observation, pour un choix de spécification et une estimation
#par les moindres carrés ordinaires
estimationparamMC0<-function(specification){

  lm(as.formula(paste0("treat ~",
                       specification)),
  data=lalonde)$fitted.values

}
```

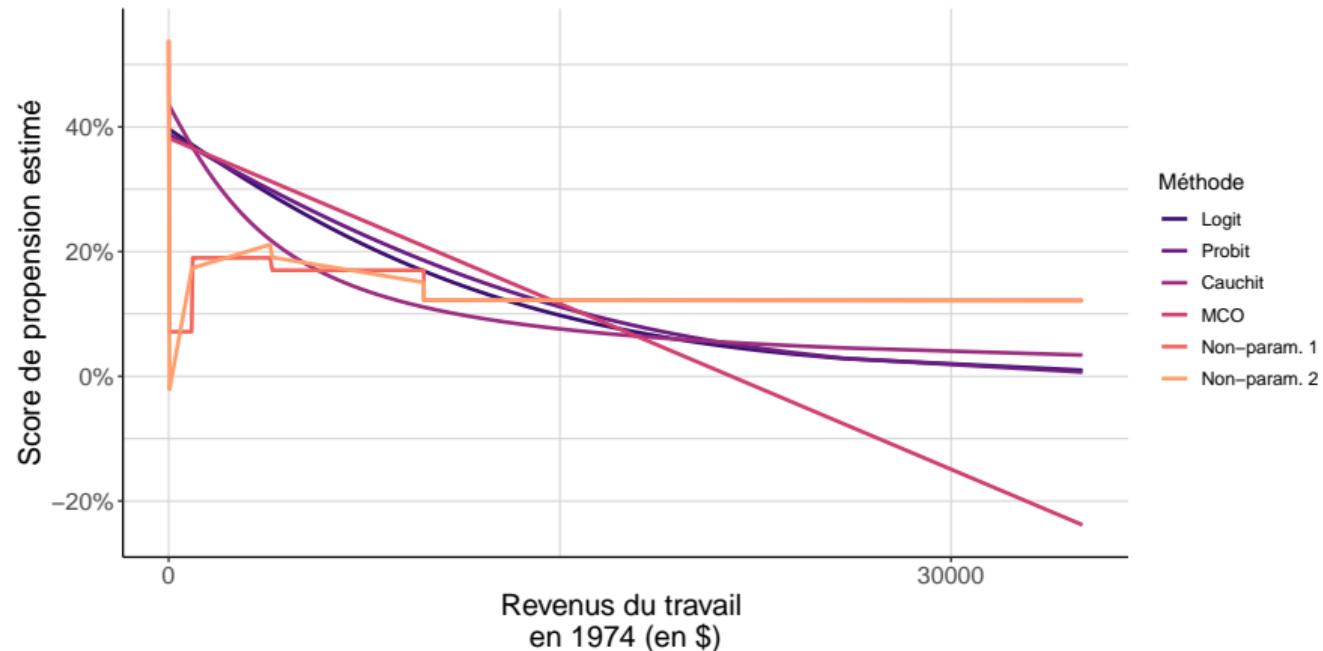
Lorsque la spécification est trop simple, le choix de fonction de lien n'est pas sans importance

```
specification_simple<-comparison_estimation_method(specification = "re74",
                                                    taillepseudostrates = 100)

cor(specification_simple[,-c("re74")])

##              logit      probit     cauchit       MCO Non-param. 1 Non-param. 2
## logit      1.0000000 0.9987608 0.9681543 0.9721981    0.6596966 0.6563494
## probit     0.9987608 1.0000000 0.9558300 0.9819866    0.6442591 0.6413123
## cauchit    0.9681543 0.9558300 1.0000000 0.8971515    0.7448040 0.7371257
## MCO        0.9721981 0.9819866 0.8971515 1.0000000    0.5867835 0.5846558
## Non-param. 1 0.6596966 0.6442591 0.7448040 0.5867835    1.0000000 0.9952847
## Non-param. 2 0.6563494 0.6413123 0.7371257 0.5846558    0.9952847 1.0000000
```

Lorsque la spécification est trop simple, le choix de fonction de lien n'est pas sans importance

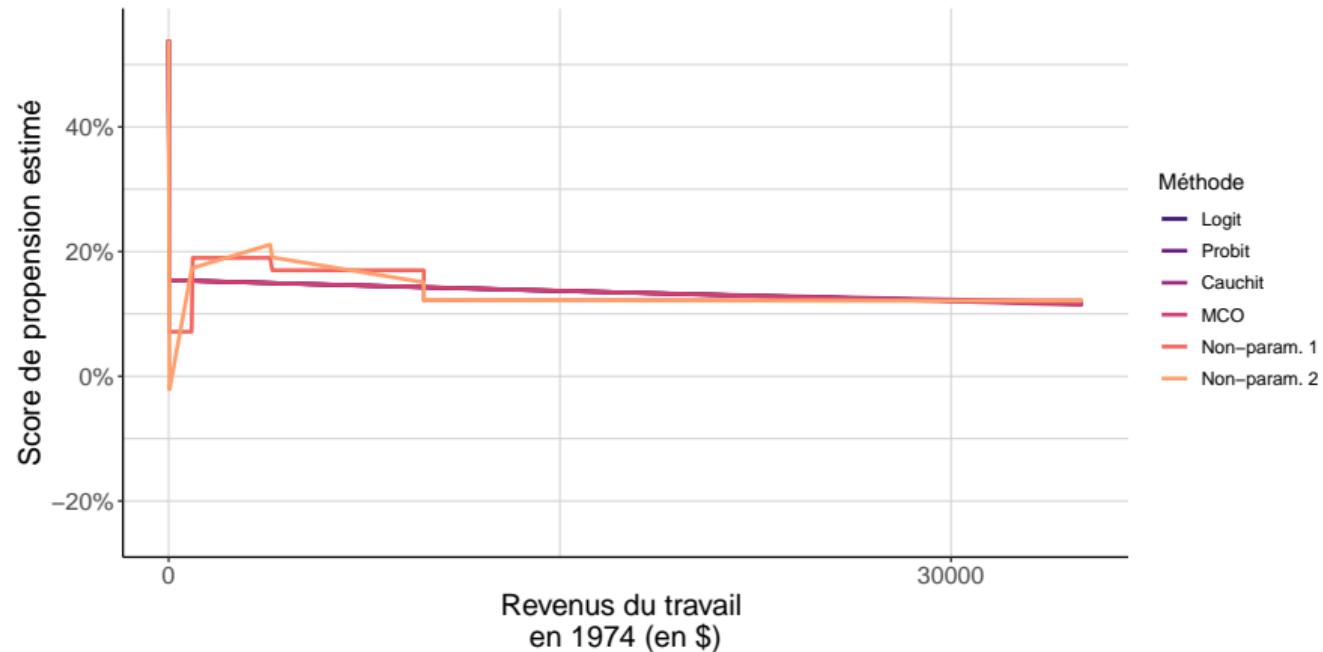


Avec une fonction de lien à peine plus riche, le choix de fonction de lien est tout de suite beaucoup moins important

```
specification_souple<-comparison_estimation_method(specification = "re74 + u74",
                                                    taillepseudostrates = 100)
cor(specification_souple[,-c("re74")])
```

	logit	probit	cauchit	MCO	Non-param.	1	Non-param.	2
## logit	1.0000000	1.0000000	0.9999994	0.9999996	0.9868464		0.9822143	
## probit	1.0000000	1.0000000	0.9999993	0.9999998	0.9868535		0.9822219	
## cauchit	0.9999994	0.9999993	1.0000000	0.9999983	0.9868023		0.9821655	
## MCO	0.9999996	0.9999998	0.9999983	1.0000000	0.9868808		0.9822516	
## Non-param.	0.9868464	0.9868535	0.9868023	0.9868808	1.0000000		0.9952847	
## Non-param.	0.9822143	0.9822219	0.9821655	0.9822516	0.9952847		1.0000000	

Avec une fonction de lien à peine plus riche, le choix de fonction de lien est tout de suite beaucoup moins important



## Une propriété essentielle d'un bon travail empirique : la robustesse

- ▶ *A priori* il n'y a **aucun choix vraiment pertinent** de fonction de lien
- ▶ Donc si des résultats dépendent de choisir une fonction de lien plutôt qu'une autre...
- ▶ C'est que ces résultats n'ont pas vraiment de valeur
  - ▶ La suspicion liée au choix d'une fonction de lien particulière **s'étend à tous les autres choix**
  - ▶ Exemple : si des résultats diffèrent beaucoup lorsque l'on choisit de passer par un logit alors les résultats obtenus avec le logit sont suspects
  - ▶ Mais c'est aussi le cas de ceux obtenus sur le même cas empirique avec les autres fonctions de lien (qui n'ont pas *a priori* plus de justification que le logit)

## Une propriété essentielle d'un bon travail empirique : la robustesse

- ▶ Un travail empirique crédible doit
  - 1. être robuste
  - 2. mettre en scène sa robustesse
    - ▶ il faut documenter et systématiquement mettre en évidence la robustesse de ses résultats à des modifications raisonnables sur toutes les dimensions où le choix n'est pas évident *a priori* et pas motivé théoriquement ou empiriquement
    - ▶ c'est de la mise en évidence de cette robustesse que dépend une bonne partie de la crédibilité du travail empirique

En dehors de la robustesse, qu'est-ce qui caractérise une bonne estimation du score de propension ?

- ▶ On sait que le vrai score de propension, inconnu, satisfait deux contraintes :
  - ▶ **Respecter l'hypothèse de support commun** : on doit trouver des individus ayant fait et n'ayant pas fait l'objet de l'intervention dans toutes les strates
  - ▶ **Propriété équilibrante du score de propension** : au sein des grosses strates définies par le score de propension, la distribution des variables de conditionnement ne dépend pas de l'assignation à l'intervention
- ▶ Utiliser astucieusement ces propriétés pour évaluer la qualité de l'estimation
  - ▶ Le principe simple : **si une estimation du score de propension ne satisfait pas ces contraintes, alors ce n'est pas une bonne estimation**

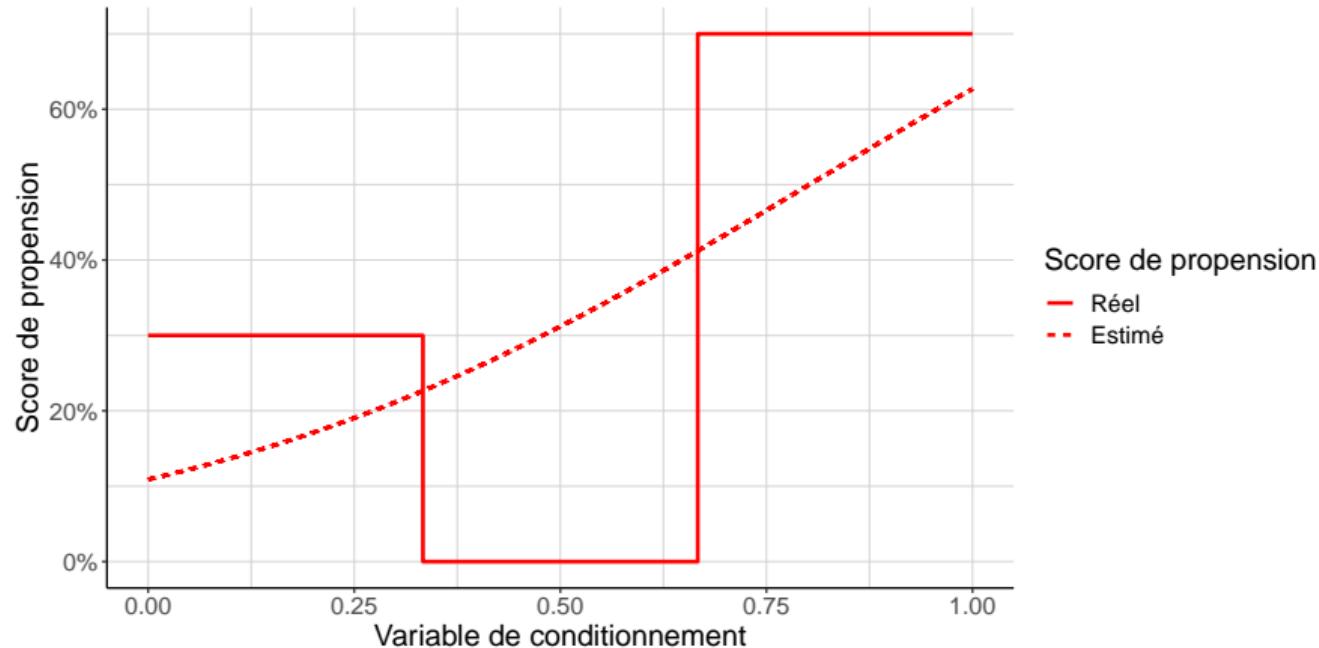
En dehors de la robustesse, qu'est-ce qui caractérise une bonne estimation du score de propension ?

- ▶ Un point important : on n'a **pas besoin pour cela de faire appel à des données sur la variable d'intérêt !**
  - ▶ L'évaluation de la qualité de l'estimation du score de propension *ne doit pas* être faite en regardant les effets estimés de l'intervention qui s'en déduisent
- ▶ Il suffit d'avoir les **données sur les variables de conditionnement et l'intervention**
- ▶ Important car cela évite de distordre l'estimation des effets de l'intervention à partir de l'idée que l'on se fait des résultats qu'on doit obtenir

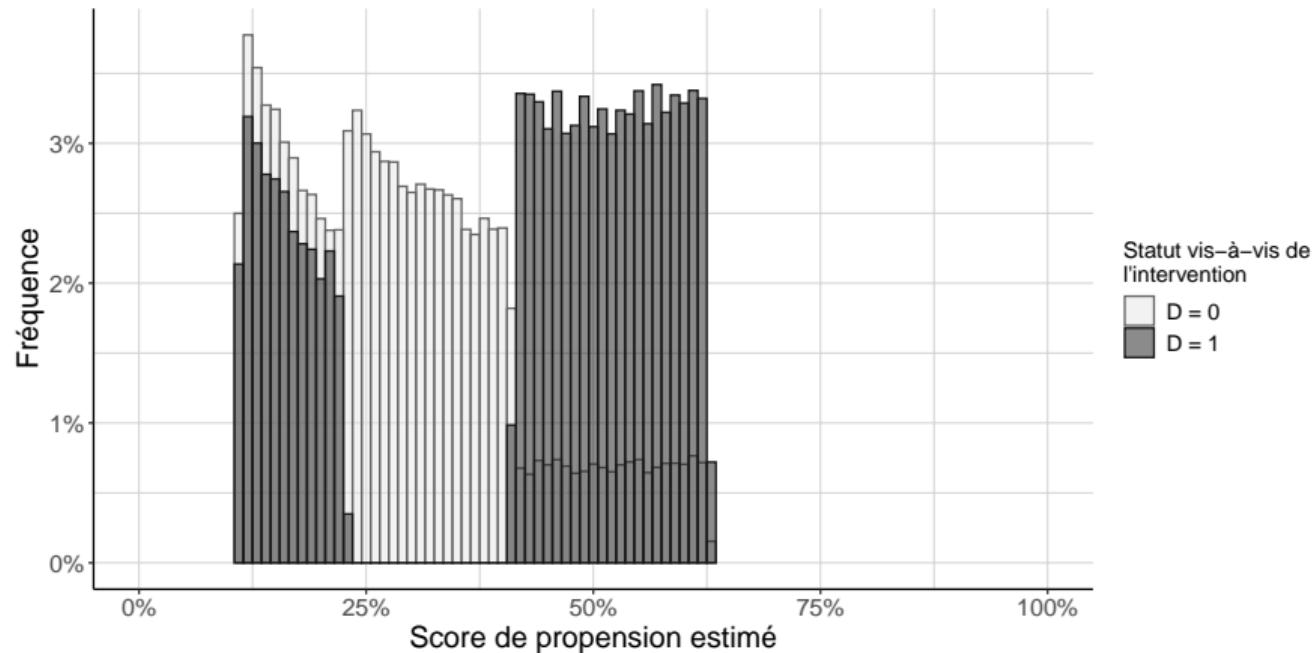
## Le score de propension estimé respecte-t-il l'hypothèse de support commun ?

- ▶ Si on connaissait le vrai score de propension, il suffirait de vérifier que celui-ci ne vaut jamais 0 ou 1
  - ▶ Parce que le vrai score de propension est la probabilité de faire l'objet de l'intervention pour une certaine valeur des covariables, et que l'hypothèse de support commun dit simplement que celle-ci ne vaut jamais 0 ou 1
- ▶ Pour le score de propension estimé c'est plus compliqué !
- ▶ Mais absolument pas désespéré

## Le score de propension estimé respecte-t-il l'hypothèse de support commun ? Un exemple sur données simulées



Représenter la distribution du score de propension séparément pour les deux groupes permet de repérer les violations de l'hypothèse de support commun, et dans le cas présent une mauvaise spécification



## Mettre à l'épreuve la propriété équilibrante du score de propension estimé : retour sur LaLonde (1986)

- ▶ Chargez les données de LaLonde (1986) à partir du package cobalt
- ▶ Pour simplifier, on va supposer que la seule variable de conditionnement dont on a besoin est le niveau des revenus du travail en 1974 (variable re74)
- ▶ Reprenez les estimations du score de propension que l'on a faites précédemment
- ▶ Utilisez ces estimations du score de propension pour estimer la différence moyenne de revenu du travail en 1974 entre le groupe qui a participé au programme et l'autre après repondération, et comparez cette différence à celle estimée avant repondération
- ▶ Quelle devrait être la valeur de la différence après repondération si l'on connaissait les vraies valeurs du score de propension ?
- ▶ Que peut-on conclure ?

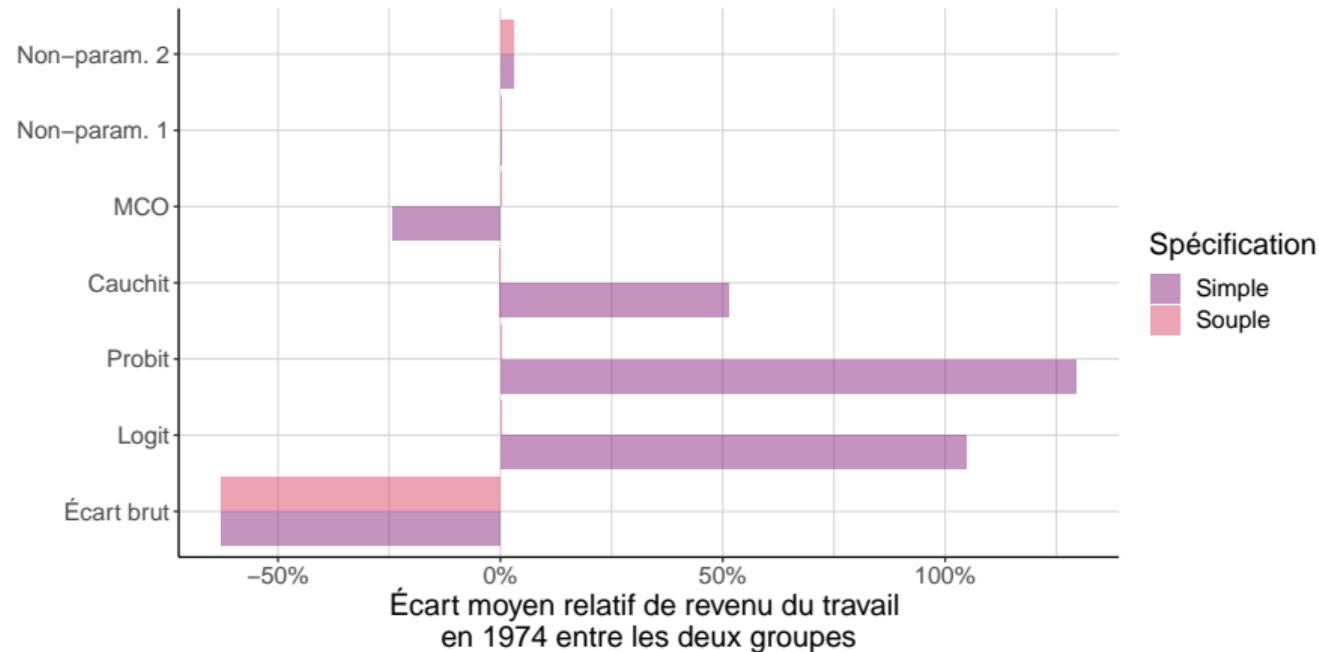
## Mettre à l'épreuve la propriété équilibrante du score de propension estimée : retour sur LaLonde (1986)

```
#On crée le jeu de poids pour chaque estimation du score de propension
estimations[,  
           poids:=treat*(sum(treat)/.N)*(1/pscore)+  
           (1-treat)*(sum(1-treat)/.N)*(1/(1-pscore)),  
           by=list(specification,  
                    link_function)]
```

## Mettre à l'épreuve la propriété équilibrante du score de propension estimé : retour sur LaLonde (1986)

```
balancing_test<-estimations[,  
                                list(balancing_test=  
                                      (sum(treat*poids*re74)/  
                                       sum(treat*poids)-  
                                       sum((1-treat)*poids*re74)/  
                                       (sum((1-treat)*poids))))/  
                                      (sum((1-treat)*poids*re74)/  
                                       sum((1-treat)*poids))),  
                                by=c("specification",  
                                     "link_function")]
```

Mettre à l'épreuve la propriété équilibrante du score de propension estimé permet de repérer les mauvaises spécifications



Après l'estimation, l'utilisation pratique du score de propension

## Implémenter les différentes approches fondées sur le score de propension

- ▶ On a maintenant **presque tous les ingrédients nécessaires** pour se lancer !
- ▶ On sait ce qu'est le score de propension, on sait l'estimer, et on sait s'assurer que l'on en a une estimation raisonnable
- ▶ Il reste maintenant la dernière étape : **utiliser le score de propension estimé**

## Implémenter les différentes approches fondées sur le score de propension : retour sur Gerber and Green (2000) et Imai (2005)

- ▶ Préalable à tout ce qui suit : on veut un score de propension estimé !
- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Estimez le score de propension en suivant cette fois la spécification et la fonction de lien retenue par Imai (2005) :
  - ▶ modèle logistique
  - ▶  $\text{PHN.C1} \sim \text{PERSONS} + \text{WARD} + \text{AGE} + \text{MAJORPTY} + \text{VOTE96.0} + \text{VOTE96.1} + \text{NEW} + \text{AGE2} + \text{PERSONS} * \text{VOTE96.0} + \text{PERSONS} * \text{NEW}$
- ▶ Représentez la distribution du score de propension estimé pour les deux groupes. L'hypothèse de support commun est-elle raisonnable ?

On estime le score de propension par logit avec la spécification retenue par Imai (2005)

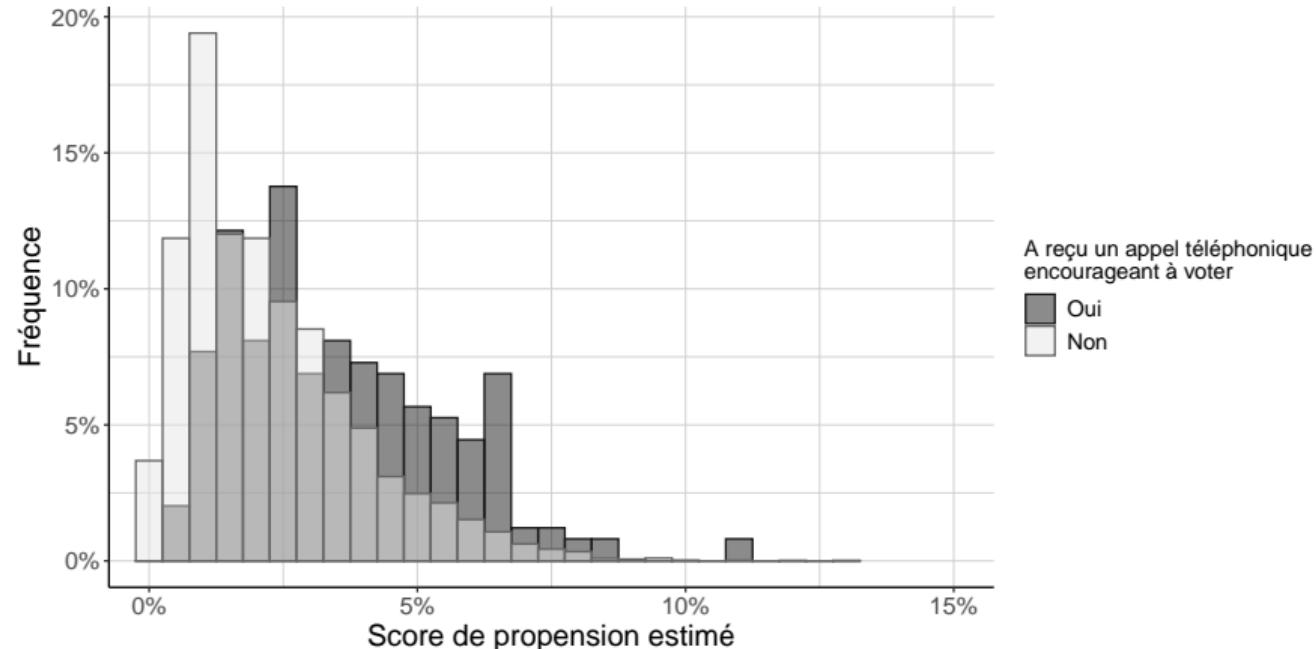
#On estime le score de propension avec un modèle logit

```
pscore_estimates<-glm(PHN.C1 ~ PERSONS +
                         WARD +
                         AGE +
                         MAJORPTY +
                         VOTE96.0 +
                         VOTE96.1 +
                         NEW +
                         AGE2 +
                         PERSONS*VOTE96.0 +
                         PERSONS*NEW,
                         data=GerberGreenImai,
                         family=binomial(link=logit))$fitted.values
```

#On remet ce score de propension estimé dans la table initiale

```
GerberGreenImai<-cbind(GerberGreenImai,
                         pscore_estimates)
```

## RéPLICATION de la figure 8 de l'article : l'hypothèse de support commun semble raisonnable



## Implémenter les différentes approches fondées sur le score de propension : la marche à suivre

- ▶ On a maintenant un exemple empirique avec des valeurs estimées du score de propension
- ▶ L'idée : l'utiliser pour essayer les trois approches disponibles pour estimer les effets causaux moyens de l'appel téléphonique incitant à voter sur la participation électorale à New Haven aux élections de 1998

## Implémenter les différentes approches fondées sur le score de propension : la marche à suivre

1. On utiliser d'abord ces techniques pour mettre à l'épreuve la propriété équilibrante du score de propension
  - ▶ Par rapport à ce que l'on a fait juste avant on a une difficulté supplémentaire ! On veut faire ça sur plusieurs variables en même temps
  - ▶ La solution classique est de continuer à regarder la différence moyenne entre les deux groupes
  - ▶ Mais en divisant par une mesure de la dispersion de la variable dans la population entière avant tentative de rééquilibrage
2. Ensuite on l'utilise vraiment pour estimer les effets causaux moyens de l'intervention

## Implémenter les différentes approches fondées sur le score de propension : la marche à suivre

- ▶ Contrainte temporelle : très long de vous laisser tout implémenter
- ▶ Je propose d'en choisir une que je discute et que je vous laisse implémenter
- ▶ Je discute les choses à considérer pour les autres et je montre plus rapidement une proposition d'implémentation sur le même exemple
- ▶ On va tout comparer à la fin

## Stratification

## Stratification sur le score de propension : en pratique

- ▶ **Ce que l'on voudrait faire :**
  - ▶ Définir des strates qui coïncident avec les valeurs exactes du vrai score de propension
  - ▶ C'est impossible
- ▶ **Ce que l'on va faire :**
  - ▶ Définir des strates qui correspondent à des plages de valeurs du score de propension estimé
  - ▶ Les différences :
    - ▶ Score de propension estimé vs. vrai score de propension
    - ▶ On met dans la même strate des individus avec des valeurs du score de propension estimé pas rigoureusement égales

## Stratification sur le score de propension : en pratique

- ▶ La première question à se poser : **combien de strates** et comment les définir ?
  - ▶ La solution courante de facilité : 5 strates définies par les quintiles de la distribution du score de propension estimé dans la population (Cochran (1968))
  - ▶ Il y a des solutions davantage fondées sur les données (Imbens (2015))
    - ▶ En particulier pour que le nombre de strates augmente avec la taille d'échantillon
- ▶ Une seconde question à se poser : ajoute-t-on à tout ça une **correction par régression linéaire** ?

## Stratification sur le score de propension : à votre tour

- ▶ Utilisez le score de propension que l'on vient d'estimer pour définir 5 strates dans la population de New Haven
- ▶ Les valeurs moyennes des variables de conditionnement sont-elles les mêmes selon qu'on regarde les individus qui ont ou qui n'ont pas reçu l'appel téléphonique, à l'intérieur de chacune de ces strates ?
- ▶ Estimez les effets causaux moyens de l'appel téléphonique sur la participation électorale par des comparaisons simples
- ▶ Correction pa régression : estimez dans chaque strate une régression linéaire de la participation électorale sur l'appel téléphonique et les variables de conditionnement. Comment interpréter la moyenne des coefficients portant sur l'appel téléphonique ?

## Stratification sur le score de propension : définir les strates

```
#On définit les strates à partir des quintiles de la distribution du score de
# propension
quintiles<-quantile(GerberGreenImai$pscore_estimates,
                      probs=0.2*c(1:4))
GerberGreenImai[,  
  strates_pscore:=findInterval(x=pscore_estimates,  
                                vec=qintiles)]
```

## Stratification sur le score de propension : mise à l'épreuve de la propriété équilibrante du score de propension

```
#Sans stratification : on calcule simplement la différence moyenne
# pour toutes les variables de conditionnement
contrastes_bruts<-GerberGreenImai[,
                           lapply(X=.SD,
                                   FUN=function(x){
                                       list(
                                           diff_moy(
                                               variable=x,
                                               groupe_intervention=PHN.C1),
                                           sqrt(0.5*var(x)))
                                       }),
                           .SDcols=variables_conditionnement]
contrastes_bruts[,  
               stat:=c("diff_moy",
                      "spread")]
```

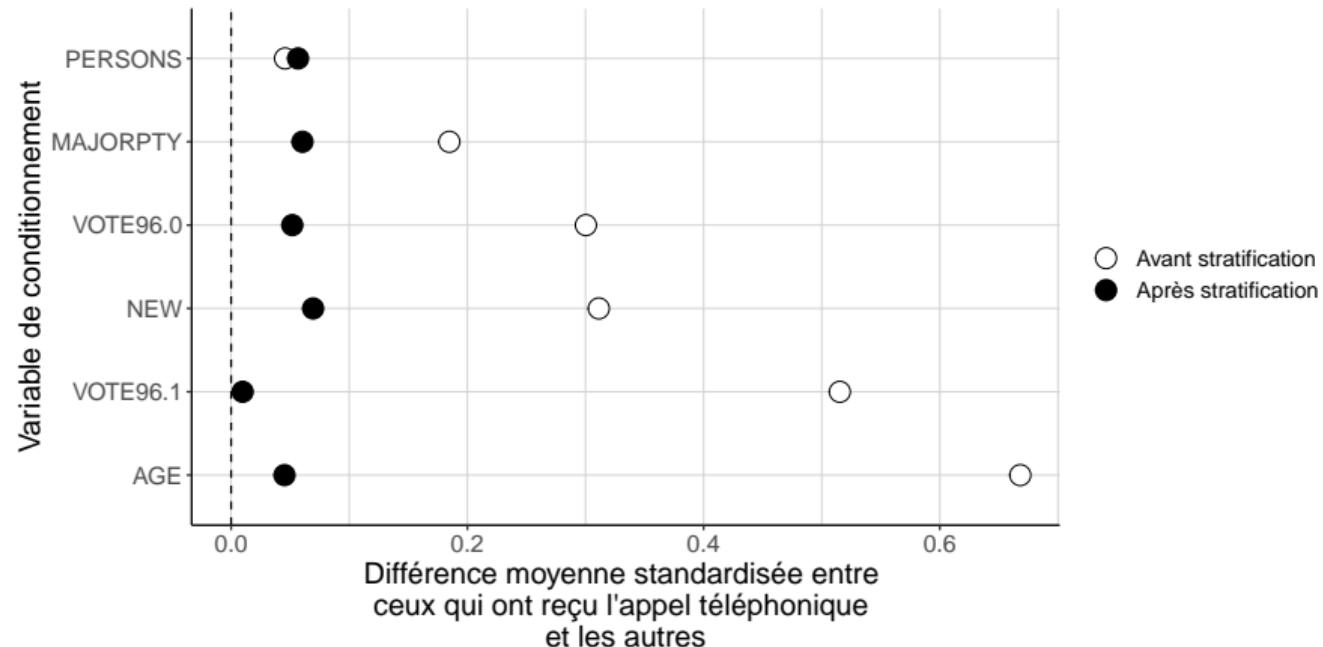
## Stratification sur le score de propension : mise à l'épreuve de la propriété équilibrante du score de propension

```
#Avec stratification
#On commence par calculer la différence moyenne dans chaque strate
contrastes_par_strates<-GerberGreenImai[,
                           lapply(
                               X=.SD,
                               FUN=function(x){
                                   diff_moy(variable=x,
                                             groupe_intervention=
                                               PHN.C1)
                               }),
                           .SDcols=variables_conditionnement,
                           by=c("strates_pscore")
                           ]
```

## Stratification sur le score de propension : mise à l'épreuve de la propriété équilibrante du score de propension

```
#Ensuite on prend la moyenne de ces différences moyennes (comme on a défini  
# les strates à partir des quintiles, elles ont toutes le même poids dans la  
# population et on n'a pas besoin de spécifier les poids)  
contrastes_agreges_stratifies<-  
  contrastes_par_strates[,  
    lapply(X=.SD,  
           mean),  
    .SDcols=variables_conditionnement]  
contrastes_agreges_stratifies[,  
  stat:="diff_moy"]  
contrastes_agreges_stratifies<-  
  melt(contrastes_agreges_stratifies,  
        id.vars="stat",  
        value.name="valeur",  
        variable.name="variable")
```

# Stratification sur le score de propension : mise à l'épreuve de la propriété équilibrante du score de propension



## Stratification sur le score de propension : l'approche directe

#On commence par estimer la différence de taux de participation électorale  
# à l'intérieur de chaque strate

```
effet_par_strate<-GerberGreenImai[,
                           lapply(X=.SD,
                                  FUN=function(x){
                                      diff_moy(
                                          groupe_intervention = PHN.C1,
                                          variable = x)
                                  }),
                           .SDcols="VOTED98",
                           by=c("strates_pscore")]
```

#On n'a plus qu'à agréger ces effets spécifiques à chaque strate

```
ATE_stratification<-effet_par_strate[,,
                                         list(ATE=mean(VOTED98))]
```

```
ATE_stratification
```

```
##          ATE
## 1: 0.114577
```

## Stratification sur le score de propension : avec la correction par régression linéaire

```
#On commence par estimer la différence de taux de participation électorale
# à l'intérieur de chaque strate avec une régression linéaire
effets_par_strates_OLS<-sapply(
  X=levels(as.factor(GerberGreenImai$strates_pscore)),
  FUN=function(x){
    lm(VOTED98 ~ PHN.C1 + PERSONS + WARD + AGE + MAJORPTY + VOTE96.0 + VOTE96.1 +
      data=
        GerberGreenImai[
          strates_pscore==x])$coefficients[
        "PHN.C1"]
  })
})
```

## Stratification sur le score de propension : avec la correction par régression linéaire

*#On n'a plus qu'à agréger ces estimations spécifiques à chaque strate*

```
ATE_stratification_OLS<-data.table(mean(effets_par_strates_OLS))
colnames(ATE_stratification_OLS)<-c("ATE")
ATE_stratification_OLS
```

```
##          ATE
## 1: 0.1247301
```

Appariement

## Appariement sur le score de propension : en pratique

- ▶ **Ce que l'on voudrait faire idéalement :**
  - ▶ appairer à chaque individu de l'échantillon un ou des individus de l'autre groupe qui ont exactement la même valeur du vrai score de propension
  - ▶ c'est impossible
- ▶ **Ce que l'on fait à la place :** on va appairer chaque individu à un ou des individus de l'autre groupe qui sont les plus proches de lui du point de vue du score de propension
  - ▶ Les différences :
    - ▶ Score de propension estimé vs. vrai score de propension
    - ▶ On apparie à des **voisins** et non à des individus **rigoureusement identiques** (*nearest neighbour matching* plutôt que *exact matching* en anglais)

# Appariement sur le score de propension : en pratique

- ▶ Les questions à se poser :
  - ▶ **Appariement avec ou sans remise ?**
    - ▶ Sans remise l'estimation dépend de l'ordre d'appariement et peut reposer sur des couples assez mal appariés
    - ▶ Mais avec remise on peut donner une importance démesurée à certains individus
  - ▶ **Combien d'individus apparie-t-on à chaque individu de départ ?**
    - ▶ Plus on en met plus on peut être précis...
    - ▶ Mais plus on va reposer sur des couples mal appariés
  - ▶ Se dote-t-on d'exigences supplémentaires sur la qualité de l'appariement ou les valeurs du score de propension ?
    - ▶ Question qu'il faut en toute rigueur toujours se poser

## Appariement sur le score de propension : en pratique

- ▶ Comme pour la stratification on peut ajouter à tout ça des **corrections par régression linéaire**
  - ▶ Permet de prendre en compte le fait que l'on apparie à chaque individu un ou des individus qui sont en fait un peu différents de lui au regard de leurs caractéristiques observables

## Appariement sur le score de propension : préparer l'appariement

```
#On commence par apparier chaque individu observé à ses 5 plus proches voisins  
# de l'autre groupe, avec remplacement (on n'élimine donc pas au fur et à  
# mesure les individus que l'on a réussi à apparier)  
nb_match<-5
```

#Première étape : la petite table avec les individus de départ, la variable  
# qui dit s'ils ont reçu ou non l'appel téléphonique, et la valeur estimée  
# du score de propension

```
GerberGreenImai[,  
    indiv_id:=as.character(.I)]  
individus_de_depart<-GerberGreenImai[,  
    c("indiv_id",  
    "PHN.C1",  
    "pscore_estimates")]
```

Appariement sur le score de propension : l'appariement en tant que tel

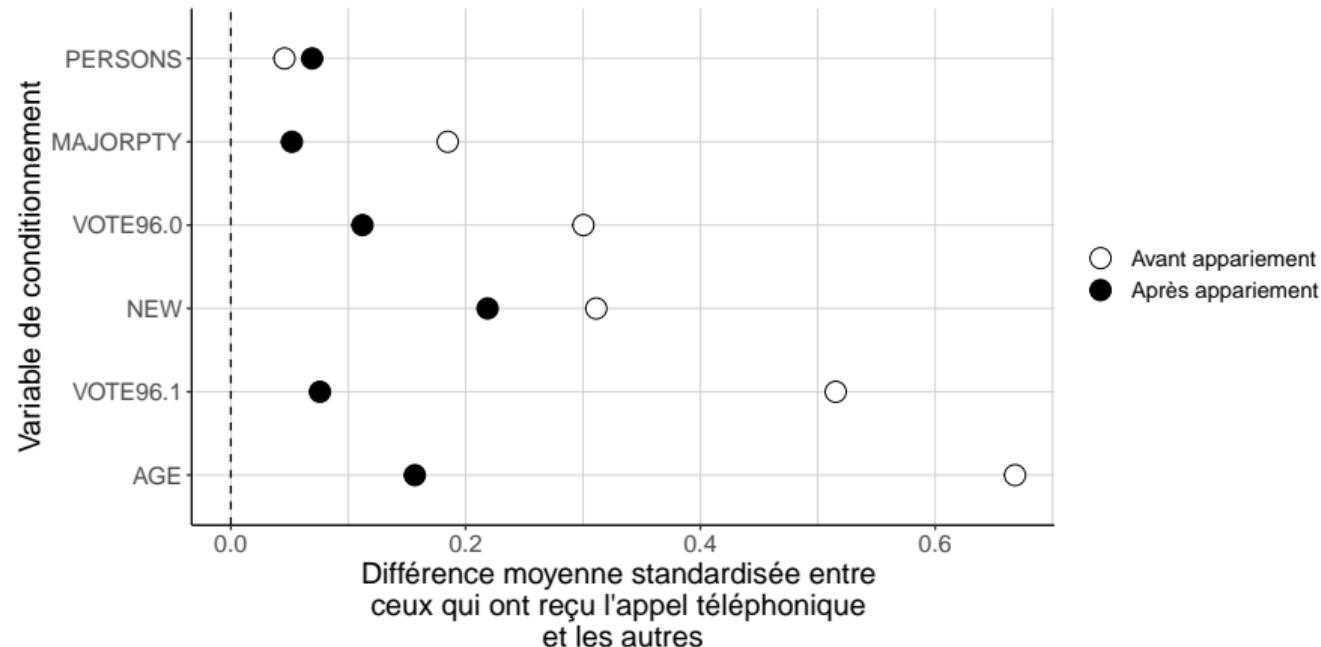
## Appariement sur le score de propension : l'appariement en tant que tel

```
#On classe pour chaque individu de départ les couples potentiels dans l'ordre
# croissant de la valeur absolue de la différence des valeurs estimées du
# score de propension
appariement[,  
            abs_diff_pscore:=abs(pscore_estimates.x-
                                pscore_estimates.y)]  
setorder(appariement,  
        indiv_id,  
        abs_diff_pscore)
```

## Appariement sur le score de propension : l'appariement en tant que tel

```
#On garde seulement pour chaque individu à apparter les 5 premières
# propositions : ce sont les individus avec les plus petites différentes de
# score de propension estimé
appariement[,
  rang_couple:=cumsum(PHN.C1+1-PHN.C1),
  by=indiv_id]
appariement<-appariement[rang_couple<=nb_match,
  c("indiv_id",
    "PHN.C1",
    "indiv_id_a_apparter")]
```

# Appariement sur le score de propension : mise à l'épreuve de la propriété équilibrante du score de propension



## Appariement sur le score de propension : l'estimation proprement dite

```
#On commence par estimer la différence de taux de participation électorale
# à l'intérieur de chaque groupe défini par l'intervention, entre les individus
# de départ et ceux auxquels on les a appariés
effet_par_groupe<-
  GerberGreenImai_matching[,  

    lapply(X=.SD,  

      FUN=function(x){  

        diff_moy(  

          groupe_intervention = appel_ego,  

          variable = x,  

          poids=poids_matching)  

        }),  

      .SDcols="VOTED98",  

      by=c("appel_indiv_dep")]
```

## Appariement sur le score de propension : l'estimation proprement dite

```
#On n'a plus qu'à agréger ces effets spécifiques à chaque groupe
effet_par_groupe[,
  poids_groupe:=
    fcase(
      appel_indiv_dep==1,
      mean(GerberGreenImai$PHN.C1),
      appel_indiv_dep==0,
      1-mean(GerberGreenImai$PHN.C1)
    )]
ATE_matching<-effet_par_groupe[,  
  list(ATE=sum(VOTED98*poids_groupe)/  
        sum(poids_groupe))]
```

```
ATE_matching
```

```
##          ATE
## 1: 0.07143781
```

Repondération

## Repondération à partir du score de propension : en pratique

- ▶ **Ce que l'on voudrait faire idéalement :**
  - ▶ répondérer avec l'inverse du (complément du) vrai score de propension
  - ▶ c'est impossible
- ▶ **Ce que l'on fait à la place :**
  - ▶ répondérer avec l'inverse du (complément du) score de propension estimé
  - ▶ la différence :
    - ▶ Score de propension estimé vs. vrai score de propension

## Repondération à partir du score de propension : en pratique

- ▶ Quelques questions à se poser :
  - ▶ En théorie la moyenne des poids est 1
    - ▶ Ce serait notamment le cas si on calculait les poids à partir du vrai score de propension
    - ▶ Cela suggère qu'on n'a pas forcément besoin de mettre la somme des poids au dénominateur quand on calcule des moyennes mais seulement le nombre d'individus de chaque groupe
  - ▶ En pratique cette moyenne peut être (légèrement) différente de 1
    - ▶ On peut normaliser les poids pour que leur moyenne soit effectivement 1
    - ▶ Ou ce qui revient au même mettre la somme des poids au dénominateur
    - ▶ Cela conduit parfois à des estimations plus stables

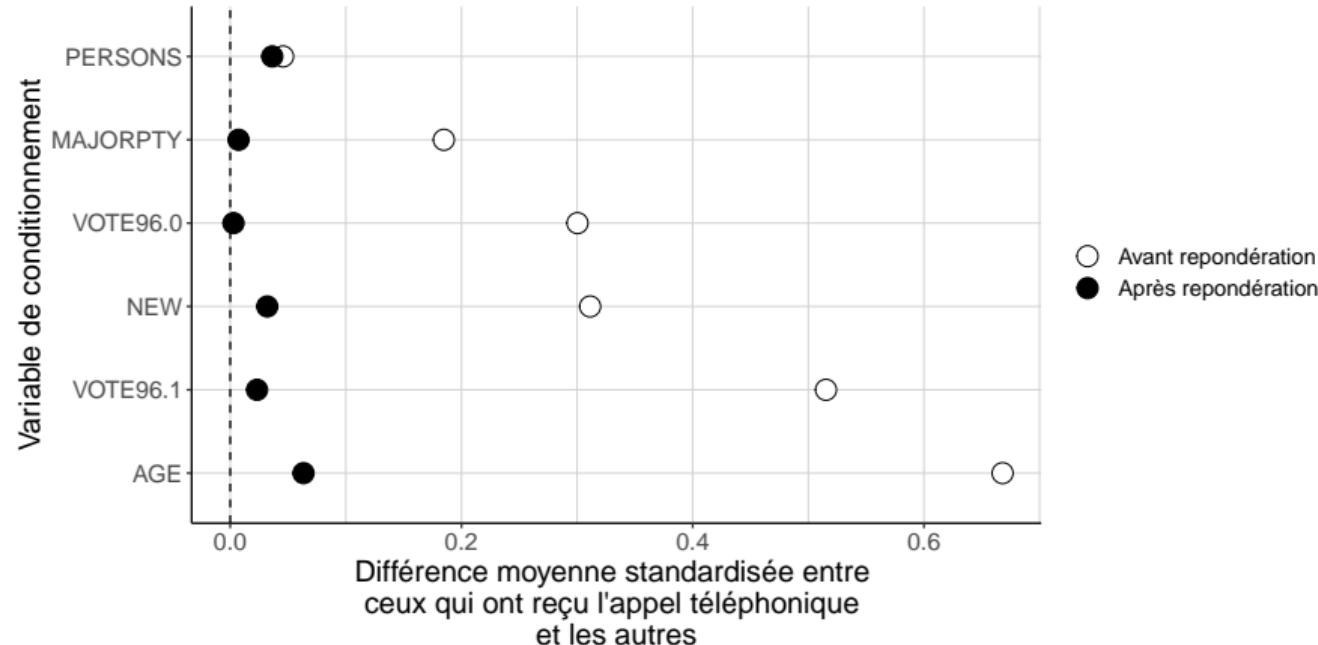
## Repondération à partir du score de propension : en pratique

- ▶ On donne beaucoup d'importance à certains individus pour lesquels le score de propension est très faible ou très élevé
  - ▶ Comme pour l'appariement, pose la question de conserver ou pas de telles observations
  - ▶ Une partie du problème est qu'en enlevant ces observations on change aussi la quantité estimée

## Repondération à partir du score de propension : calcul des poids

```
#On va calculer les poids à partir du score de propension
GerberGreenImai[,
  poids_pscore:=
    fcase(
      PHN.C1==1,
      mean(GerberGreenImai$PHN.C1)/
        pscore_estimates,
      PHN.C1==0,
      mean(1-GerberGreenImai$PHN.C1)/
        (1-pscore_estimates)
    )]
```

## Repondération à partir du score de propension : mise à l'épreuve de la propriété équilibrante du score de propension



## Repondération à partir du score de propension : l'estimation à proprement parler

```
#Estimation de l'ATE avec les poids non-normalisés
ATE_reponderation_nonnorm<-
  GerberGreenImai[,  

    list(ATE=
      sum(poids_pscore*PHN.C1*VOTED98)/
      sum(PHN.C1)-
      sum(poids_pscore*(1-PHN.C1)*VOTED98)/
      sum(1-PHN.C1))]  

ATE_reponderation_nonnorm  
  

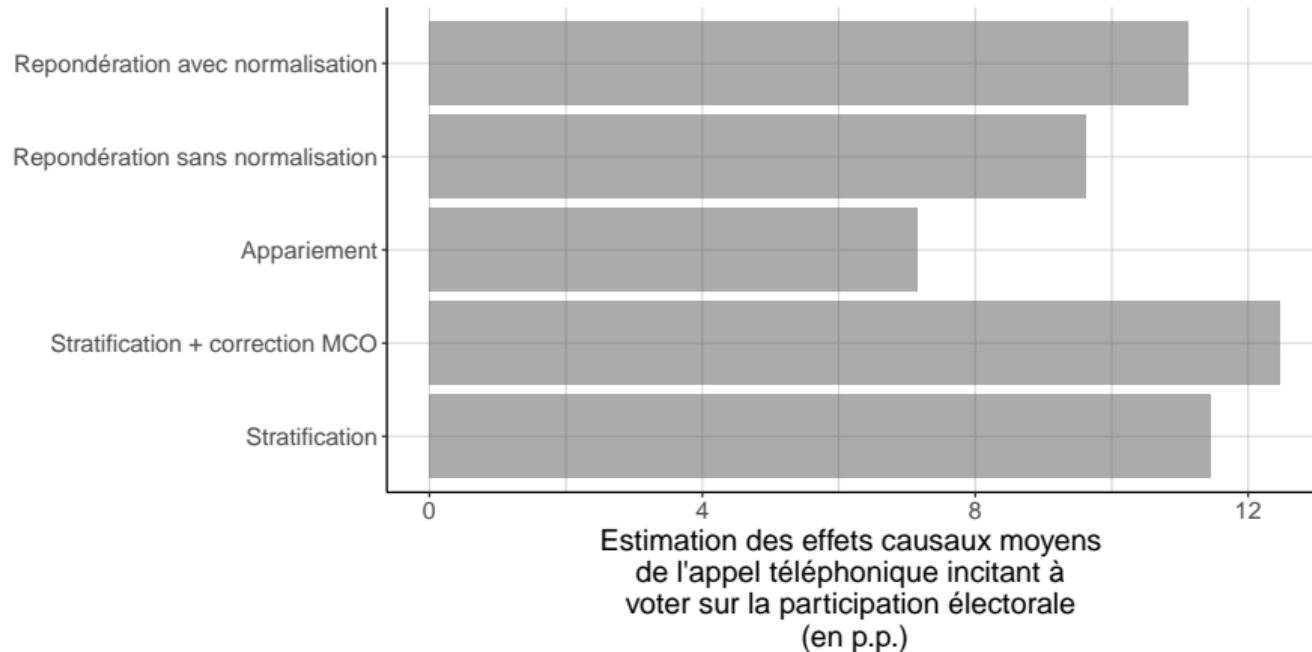
##           ATE
## 1: 0.09615629
```

## Repondération à partir du score de propension : l'estimation à proprement parler

```
#Estimation de l'ATE avec les poids normalisés
ATE_reponderation_norm<-  
  GerberGreenImai[,  
    list(ATE=  
        sum(poids_pscore*PHN.C1*VOTED98)/  
        sum(poids_pscore*PHN.C1)-  
        sum(poids_pscore*(1-PHN.C1)*VOTED98)/  
        sum(poids_pscore*(1-PHN.C1)))]  
  
ATE_reponderation_norm
```

```
##          ATE  
## 1: 0.1113282
```

## Les résultats sont-ils robustes ?



Et les régressions linéaires dans tout ça ?

## Pourquoi les régressions linéaires ?

- ▶ Les régressions linéaires par les moindres carrés ordinaires sont souvent utilisées pour **contrôler de certaines variables**
- ▶ Cela signifie-t-il la même chose que ce que l'on fait ici ?
  - ▶ **Oui, à certaines conditions**
- ▶ Y a-t-il un intérêt à passer par une régression linéaire quand ces conditions sont remplies ?
  - ▶ **Oui !**

Usage habituel des moindres carrés ordinaires

# Interprétation causale des régressions linéaires

- ▶ On se concentre sur l'**usage habituel**
  - ▶ Dans le cas de Gerber and Green (2000) révisé par Imai (2005) il reviendrait à régresser la participation individuelle  $Y_i$  sur le fait d'avoir reçu ou non un appel téléphonique  $D_i$  et toutes les variables de conditionnement  $X_i$  :

$$Y_i = \beta D_i + X'_i \gamma + \epsilon_i \text{ avec } \mathbb{E}[X_i \epsilon_i] = 0$$

- ▶ La question que l'on se pose : le coefficient  $\beta$  a-t-il une interprétation causale ?

## Interprétation causale des régressions linéaires : le cas des régressions saturées

- ▶ Pour simplifier on va supposer que l'hypothèse d'indépendance conditionnelle tient en conditionnant seulement sur le quartier (variable WARD)
- ▶ Estimez la régression dans ce cas
- ▶ Peut-on calculer  $\beta$  à partir de statistiques descriptives simples estimées sur les données ? (suggestion : se souvenir d'un résultat vu sur la régression du salaire sur le sexe et la profession dans la séance sur les régressions linéaires)
- ▶ Le coefficient a-t-il une interprétation causale ?

# Interprétation causale des régressions linéaires : estimation de la régression

```
#On récupère les données et on convertit en data.table  
data(GerberGreenImai)  
GerberGreenImai<-data.table(GerberGreenImai)  
  
#On estime le modèle de régression linéaire et on récupère le coefficient sur  
# la variable représentant l'appel téléphonique  
estimation_OLS<-lm(VOTED98 ~ PHN.C1 + WARD,  
                      data=GerberGreenImai)$coefficients["PHN.C1"]
```

Le coefficient sur l'appel téléphonique est une moyenne de contrastes évalués quartier par quartier, avec des poids que l'on sait calculer

```
#Pour chaque quartier, on estime : la différence de taux de participation
# électorale entre ceux qui ont et ceux qui n'ont pas reçu l'appel téléphonique
# la taille du quartier et la probabilité d'y recevoir l'appel téléphonique
estimation_par_quartier<-GerberGreenImai[,
                                list(contraste_participation=
                                      sum(PHN.C1*VOTED98) /
                                      sum(PHN.C1) -
                                      sum((1-PHN.C1)*VOTED98) /
                                      sum(1-PHN.C1) ,
                                         taille=.N,
                                         proba_appel=sum(PHN.C1)/.N),
                                         by=c("WARD")]
```

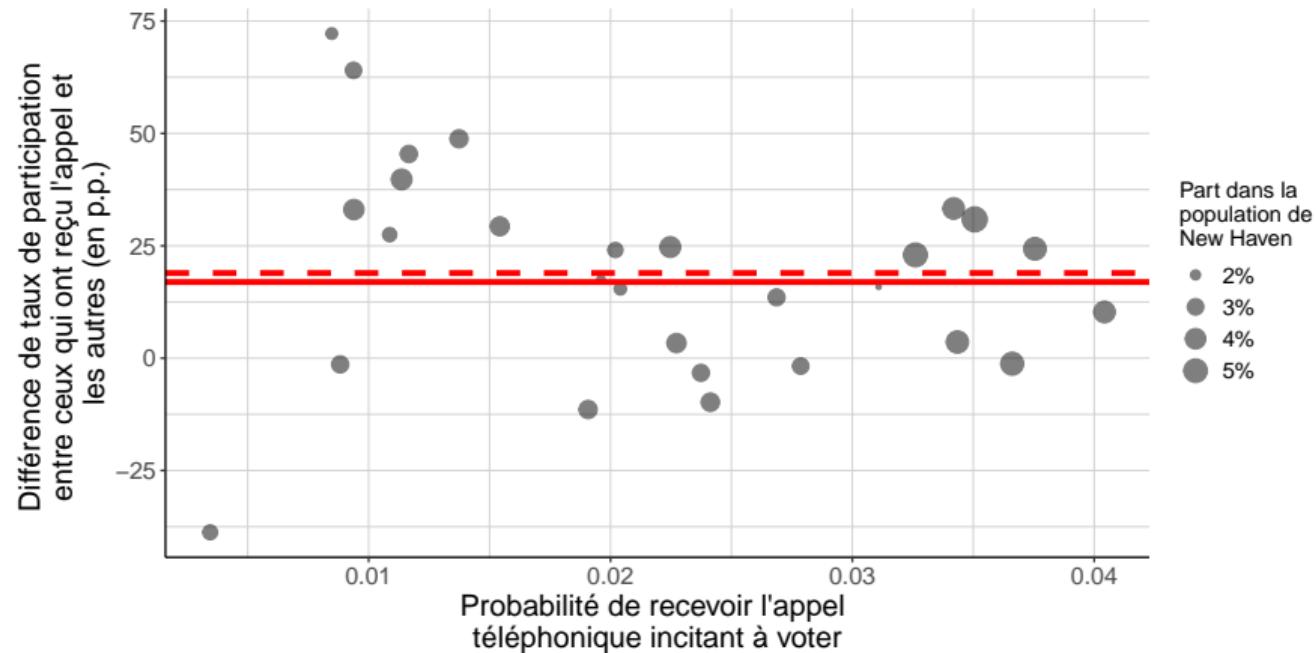
Le coefficient sur l'appel téléphonique est une moyenne de contrastes évalués quartier par quartier, avec des poids que l'on sait calculer

```
#On peut agréger ces différences de taux de participation avec des poids
# proportionnels à la taille du quartier et à la variance conditionnelle
#  $(1 - \text{proba\_appel}) * \text{proba\_appel}$ 
contraste_participation_agrege<-
  estimation_par_quartier[!is.na(contraste_participation),
    list(moyenne_effet_causal=
      sum(taille*
        (1-proba_appel)*proba_appel*
        contraste_participation)/
      sum(taille*
        (1-proba_appel)*proba_appel))]

#On peut vérifier que cela permet bien de retrouver le coefficient
all.equal(as.numeric(estimation_OLS),
          as.numeric(contraste_participation_agrege))

## [1] TRUE
```

Le coefficient sur l'appel téléphonique a une interprétation causale ; il met davantage de poids sur les quartiers les plus peuplés et ceux où la probabilité de recevoir un appel est plus proche de 50%



## Interprétation causale des régressions linéaires : cas des régressions saturées

- ▶ Ce résultat tient parce que la variable de quartier décrit des strates mutuellement exclusives et qui recouvrent tous les cas possibles
- ▶ Dans ce cas le coefficient sur l'appel téléphonique est une moyenne d'effets causaux moyens spécifiques à chaque quartier
  - ▶ avec des poids proportionnels à
    - ▶ la taille du quartier
    - ▶ la variance de la variable d'appel téléphonique
- ▶ Moyenne qui est donc **different**e celle que l'on a envie de faire pour estimer les effets causaux moyens sur toute la population de New Haven !

## Interprétation causale des régressions linéaires : et dans le cas général ?

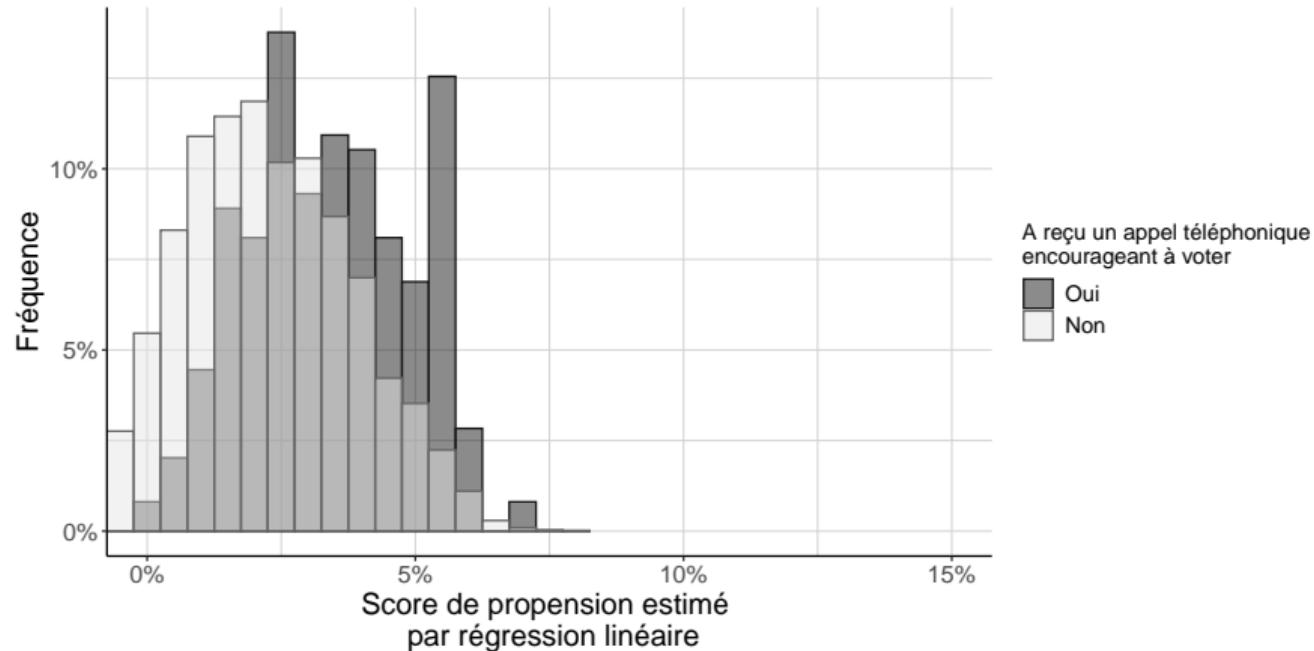
- ▶ Ce résultat s'étend au cas général si la **spécification des variables de conditionnement est assez riche** pour que :
  - ▶ le score de propension estimé par régression linéaire ne dépasse jamais 1 et ;
  - ▶ le vrai score de propension s'identifie à sa valeur estimée par régression linéaire ou
  - ▶ la moyenne de la variable d'intérêt en l'absence de traitement est linéaire en les variables de conditionnement
- ▶ **La bonne nouvelle** : on sait regarder à quoi ressemble le score de propension estimé par régression linéaire, et on sait regarder que c'est une estimation de bonne qualité !

## Interprétation causale des régressions linéaires : application empirique

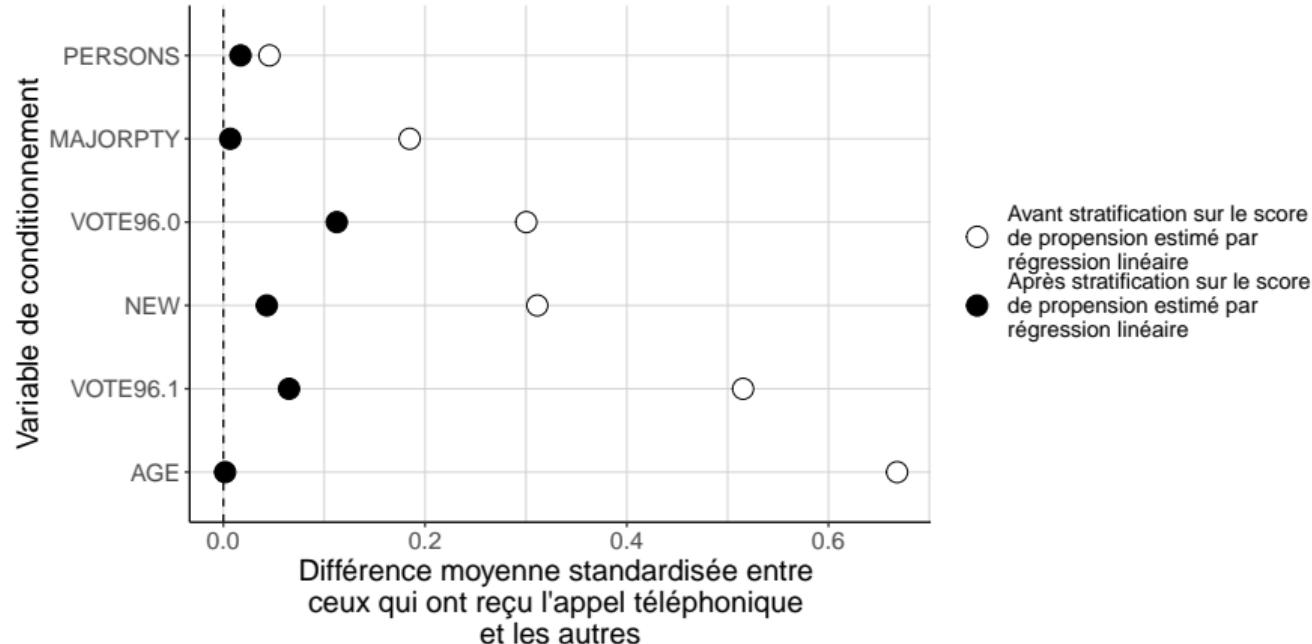
- ▶ Chargez les données sur l'expérience aléatoire défectueuse de Gerber and Green (2000) revisitée par Imai (2005)
- ▶ Estimez par régression linéaire le score de propension en suivant cette fois la spécification retenue par Imai (2005) :
  - ▶  $\text{PHN.C1} \sim \text{PERSONS} + \text{WARD} + \text{AGE} + \text{MAJORPTY} + \text{VOTE96.0} + \text{VOTE96.1} + \text{NEW} + \text{AGE2} + \text{PERSONS*VOTE96.0} + \text{PERSONS*NEW}$
- ▶ Représentez la distribution du score de propension estimé pour les deux groupes. L'hypothèse de support commun est-elle raisonnable ?
- ▶ Mettez à l'épreuve la propriété équilibrante du score de propension, par exemple par stratification. Que peut-on en déduire ?
- ▶ Estimez la régression linéaire de la participation électorale sur l'appel téléphonique et les variables de conditionnement. Que peut-on en conclure ?

## Interprétation causale des régressions linéaires : estimation du score de propension

# Interprétation causale des régressions linéaires : examen de l'hypothèse de support commun



# Interprétation causale des régressions linéaires : mise à l'épreuve de la propriété équilibrante du score de propension



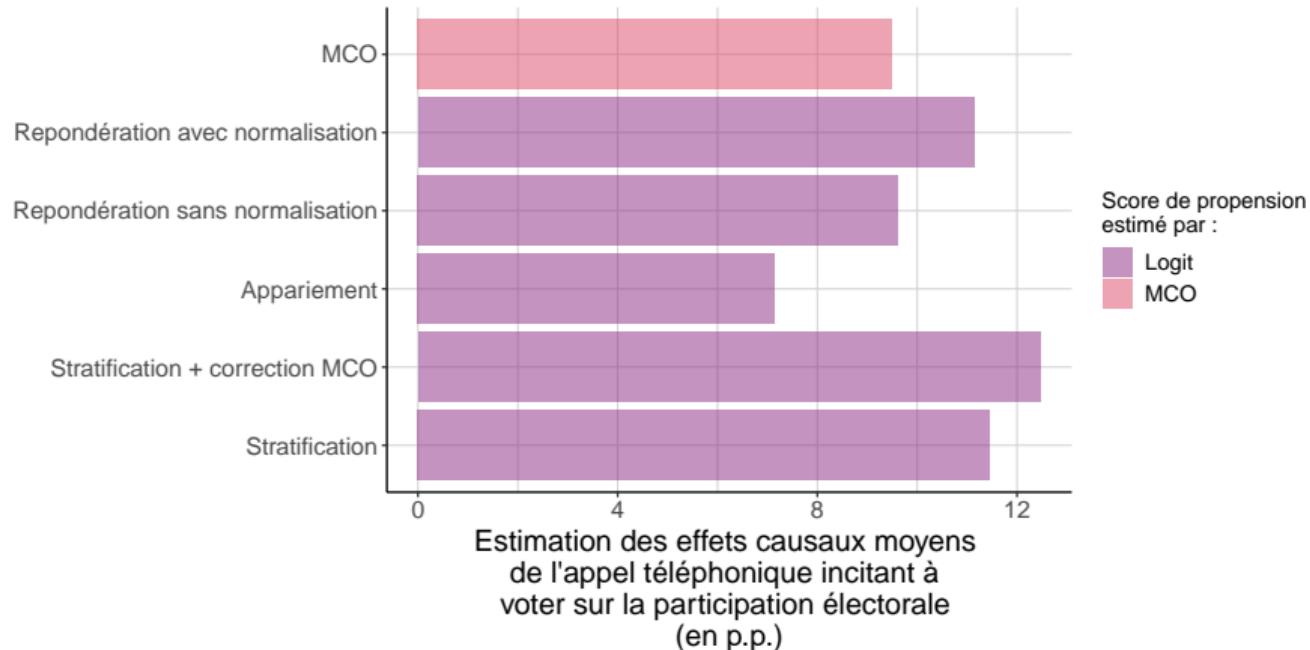
# Interprétation causale des régressions linéaires : l'estimation à proprement parler

```
#On estime la régression à proprement parler
effet_causal_OLS<-lm(VOTED98 ~PHN.C1 +
                      PERSONS + WARD + AGE + MAJORPTY + VOTE96.0 +
                      VOTE96.1 + NEW + AGE2 + PERSONS*VOTE96.0 +
                      PERSONS*NEW,
                      data=GerberGreenImai)$coefficients["PHN.C1"]

effet_causal_OLS

##      PHN.C1
## 0.09497093
```

# Comment le coefficient se compare-t-il à nos estimation des effets causaux moyens de l'appel téléphonique ?



De bonnes raisons de s'en tenir là

## Interprétation causale des régressions linéaires : un premier bilan

- ▶ **Si on a une spécification suffisamment riche, alors le coefficient sur la variable d'appel téléphonique a une interprétation causale**
  - ▶ C'est une moyenne d'effets causaux moyens spécifiques à chaque strate
- ▶ On peut vérifier que la spécification est suffisamment riche pour que cette interprétation soit raisonnable en s'intéressant au score de propension estimé par régression linéaire
- ▶ Mais la régression linéaire donne aux différentes strates un poid qui n'est pas le même que celui dont on a besoin pour récupérer les effets causaux moyens de l'intervention dans toute la population
- ▶ Cela condamne-t-il le passage par la régression linéaire ? Non !

## Quantifier l'incertitude dans les régressions linéaires n'est pas très difficile

- ▶ Résultats mentionnés dans la séance de rappel sur les régressions + presque tous les logiciels savent estimer l'incertitude sur les coefficients très facilement
- ▶ Au contraire **la quantification de l'incertitude pour les autres approches est l'objet de débats** dans la littérature
  - ▶ Certains auteurs considèrent que l'on travaille sur des données exhaustives et que l'incertitude vient de l'assignation aléatoire de l'intervention
  - ▶ D'autres considèrent qu'elle vient du travail sur un échantillon beaucoup plus petit que la population
  - ▶ Certaines approches usuelles ne sont pas adaptées à certaines techniques comme les appariements

## Quantifier l'incertitude : un exemple empirique

- ▶ Estimez l'incertitude qui porte sur le coefficient de l'appel téléphonique incitant à voter dans la régression précédente

## Quantifier l'incertitude : un exemple empirique

```
reg_OLS<-lm(VOTED98 ~PHN.C1 +
              PERSONS + WARD + AGE + MAJORPTY + VOTE96.0 +
              VOTE96.1 + NEW + AGE2 + PERSONS*VOTE96.0 +
              PERSONS*NEW,
              data=GerberGreenImai)

coeftest(reg_OLS,
         vcov=vcovHC(reg_OLS))[2,]

##      Estimate Std. Error     t value   Pr(>|t|)
## 0.0949709301 0.0271414717 3.4991076079 0.0004687003
```

## Une borne sur la précision

- ▶ **Résultat récent** : quand la spécification est assez riche, la quantité à laquelle renvoie le coefficient de l'appel téléphonique dans la régression est la moyenne d'effets causaux moyens spécifiques à chaque strate **la plus précisément estimée à partir des données** (Goldsmith-Pinkham, Hull, and Kolesár (2021))
  - ▶ Si on l'estime très imprécisément alors ça ne sert à rien de se fatiguer à chercher à estimer les effets causaux moyens dans toute la population

## Récupérer les effets causaux moyens avec des régressions linéaires

- ▶ **C'est possible** avec des techniques issues au départ de la quantification des discriminations sur le marché du travail
  - ▶ Décomposition de Oaxaca-Blinder (Blinder (1973) ; Oaxaca (1973))
- ▶ Il faut faire deux régressions plutôt qu'une mais ce n'est pas très coûteux
- ▶ Sujet traité en annexe du support

Comment sait-on que ça marche ?

## Est-ce que ça marche ?

- ▶ Jusque là on a fait comme si l'hypothèse d'indépendance conditionnelle était assurée et que tous nos problèmes venaient du fait que la traduction de l'hypothèse de support commun dans des données empiriques était difficile
- ▶ Si on a travaillé correctement on a (au mieux !) une **justification théorique** pour la première
- ▶ Peut-on faire mieux ?
  - ▶ On ne peut jamais la tester en toute généralité parce qu'elle porte sur des valeurs potentielles inobservées
  - ▶ Mais on peut quand même faire mieux !

Mettre à l'épreuve l'hypothèse d'indépendance conditionnelle

## Faire un test placebo

- ▶ Ce que l'on vérifie quand on regarde la propriété équilibrante du score de propension :
  - ▶ après stratification / appariement / repondération, les deux groupes sont (presque) identiques du point de vue des caractéristiques observables que l'on a faites entrer dans le calcul du score de propension
- ▶ L'idée cette fois-ci : sont-ils identiques sur des dimensions observables que l'on a pas faites entrer dans ce calcul ?
  - ▶ **intéressant seulement si on a une bonne théorie sur les raisons pour lesquelles ils doivent l'être**
  - ▶ par exemple si ce sont des dimensions mesurées avant l'intervention et que son effet doit donc être nul

## Faire un test placebo

- ▶ Aucun de ces tests ne met à l'épreuve l'hypothèse d'indépendance conditionnelle en tant que telle
  - ▶ Mais chacun lui fait gagner un peu plus de crédibilité
  - ▶ Et ce d'autant plus qu'il est clair qu'ils ne sont pas corrélés entre eux ni avec les variables de conditionnement
  - ▶ Et d'autant plus qu'ils sont corrélés avec les valeurs potentielles de la variable d'intérêt

Comparer avec une expérience aléatoire contrôlée

## Retour sur LaLonde (1986), Dehejia and Wahba (1999), Dehejia and Wahba (2002)

- ▶ Exemple historiquement important
- ▶ Travaux consacrés à l'évaluation des effets du National Supported Work Demonstration
  - ▶ Un programme d'emploi aidé étastunien mis en place dans la seconde moitié des années 1970
  - ▶ Qui a fait l'objet d'une expérience aléatoire contrôlée
- ▶ Proposition de LaLonde (1986) : utiliser les résultats de l'expérience aléatoire contrôlée comme étalon pour évaluer la qualité des techniques économétriques usuelles à l'époque :
  - ▶ On garde seulement les données sur les participants au programme, en jetant les volontaires qui du fait de l'aléas n'ont pas été retenu
  - ▶ On utilise des données d'enquête pour constituer un groupe d'individus éligibles

## Retour sur LaLonde (1986), Dehejia and Wahba (1999), Dehejia and Wahba (2002)

- ▶ Parvient-ton avec des techniques de régression appliquées à ces données à retrouver les résultats de l'expérience ?
- ▶ La réponse est négative pour un très large panel de méthodes
  - ▶ Résultat qui contribue à la préoccupation de toute une génération d'économètres appliqués de proposer des approches plus crédibles et plus transparentes d'évaluation des politiques publiques
- ▶ Dehejia and Wahba (1999) et Dehejia and Wahba (2002) : nouvel examen des données de LaLonde (1986) en mobilisant les techniques d'appariement sur le score de propension
  - ▶ Résultats plutôt satisfaisants !
- ▶ Beaucoup de travaux consacrés aux techniques d'estimation disponibles sous l'hypothèse d'indépendance conditionnelle mobilisent ces données comme un terrain de jeu pour évaluer empiriquement la qualité de l'estimation

## Proposition d'application : retour sur LaLonde (1986)

- ▶ Reprenez les données non-expérimentales de LaLonde (1986)
- ▶ En utilisant les outils développés pour évaluer la qualité de l'estimation du score de propension, aurait-il été possible de repérer à l'avance les difficultés posées par l'usage de la régression linéaire ?

## Bibliographie

## Bibliographie I

- Blinder, Alan S. 1973. "Wage Discrimination : Reduced Form and Structural Estimates." *The Journal of Human Resources* 8 (4) : 436–55.
- Cochran, W. G. 1968. "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies." *Biometrics* 24 (2) : 295–313.  
<http://www.jstor.org/stable/2528036>.
- Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies : Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448) : 1053–62. <http://www.jstor.org/stable/2669919>.
- . 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics* 84 (1) : 151–61.
- Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout : A Field Experiment." *The American Political Science Review* 94 (3) : 653–63. <http://www.jstor.org/stable/2585837>.
- Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2021. "Contamination Bias in Linear Regressions." arXiv. <https://doi.org/10.48550/ARXIV.2106.05024>.

## Bibliographie II

- Imai, Kosuke. 2005. "Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review* 99 (2) : 283–300.  
[https://EconPapers.repec.org/RePEc:cup:apsrev:v:99:y:2005:i:02:p:283-300\\_05](https://EconPapers.repec.org/RePEc:cup:apsrev:v:99:y:2005:i:02:p:283-300_05).
- Imbens, Guido W. 2015. "Matching Methods in Practice: Three Examples." *The Journal of Human Resources* 50 (2) : 373–419. <http://www.jstor.org/stable/24735990>.
- LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 76 (4) : 604–20.  
<http://www.jstor.org/stable/1806062>.
- Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (3) : 693–709.