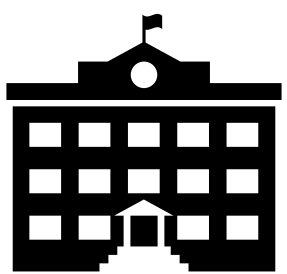
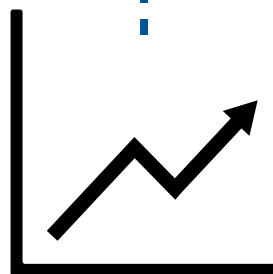


# UNE SCIENCE DES POLITIQUES PUBLIQUES



- Doctorant en microéconométrie appliquée à l'Université Copenhague où j'évalue des politiques de santé... grâce à des méthodes statistiques qu'on appelle des méthodes d'**inférence causale**.
- Je suis ici pour vous présenter ces méthodes, car elles ont pour vous ...
  - ... un intérêt **citoyen** (savoir si les politiques que vous prônez sont des solutions aux problèmes qui vous préoccupent), **académique** (concours), et **professionnel** (recherche académique, économiste/statisticien dans des instituts comme l'IPP, *causal data-scientist* dans le privé, ex. *L'Oréal*).
- Pourquoi besoin de méthodes sophistiquées ? Le 'paradoxe de Kundera'.

## Le paradoxe de Kundera

- Dans *l'insoutenable légèreté de l'être*, Tomas (le personnage principal), qui avait fui à l'Ouest repasse le rideau de fer pour retrouver la femme qu'il aime, Tereza.
- Est-ce que cette décision l'a rendu plus heureux ?
- Il n'en a aucune idée. Comment pourrait-il savoir ?

## Etes-vous sûr de vos solutions ?

*« En travaux pratique de physique, n'importe quel collégien peut faire des expériences pour vérifier l'exactitude d'une hypothèse scientifique. Mais l'homme, parce qu'il n'a qu'une seule vie, n'a aucune possibilité de vérifier l'hypothèse par l'expérience de sorte qu'il ne saura jamais s'il a eu tort ou raison d'obéir à son sentiment » (Kundera, 1984).*

- Nous sommes face aux politiques, face aux traitements médicaux, comme Tomas face à Tereza : incapables de savoir ce que nous serions sans eux.

## Etes-vous sûr de vos solutions ?

- Pour estimer l'effet d'une politique publique ou d'une de nos décisions, nous ne pouvons pas observer le contrefactuel, c'est-à-dire ce qui se serait produit dans un monde parallèle qui ne différerait du nôtre que dans la mesure où on n'aurait pas implémenté cette politique ou pris cette décision.
- C'est en tout cas vrai dans une des approches mathématiques courantes de la causalité en économie/science politique US/ épidémiologie ...

## Le modèle causal de Rubin (Rubin, 1971)

Effet de A = ce qu'il s'est produit avec A - ce qu'il se serait produit dans des conditions exactement similaires sauf que sans A.

- Pas si trivial : dans les rapports du Parlement et les discours de politiques, on mesure général l'efficacité d'une politique au nombre de personnes qui en ont bénéficié.

Exemple : Prime pour aller dans les déserts médicaux. Macron : « *nous avons rénové 1 million de logements avec notre prime* ».

Le but de cette conférence est de présenter des méthodes quantitatives (d'économistes notamment) qui permettent de connaître l'effet d'une politique/d'un traitement en estimant le **contrefactuel** \*

\* trouver un bon groupe de contrôle



## Disclaimer

1. Je présente **une seule** science des politiques publiques (celle fondée sur l'inférence causale, l'usage d'expériences et de quasi-expériences).
  - Il y a d'autres approches de l'évaluation des politiques publiques (entretiens qualitatifs, sondages sur les perceptions des bénéficiaires, ...).
  - Ces méthodes sont complémentaires aux méthodes que je présente...
  - Ces méthodes ont toute leur valeur *en soi*, notamment pour documenter la mise en place d'une politique publique, et les ressentis des personnes qui en ont bénéficié (≠/≠ affirmation causale).



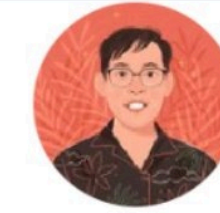
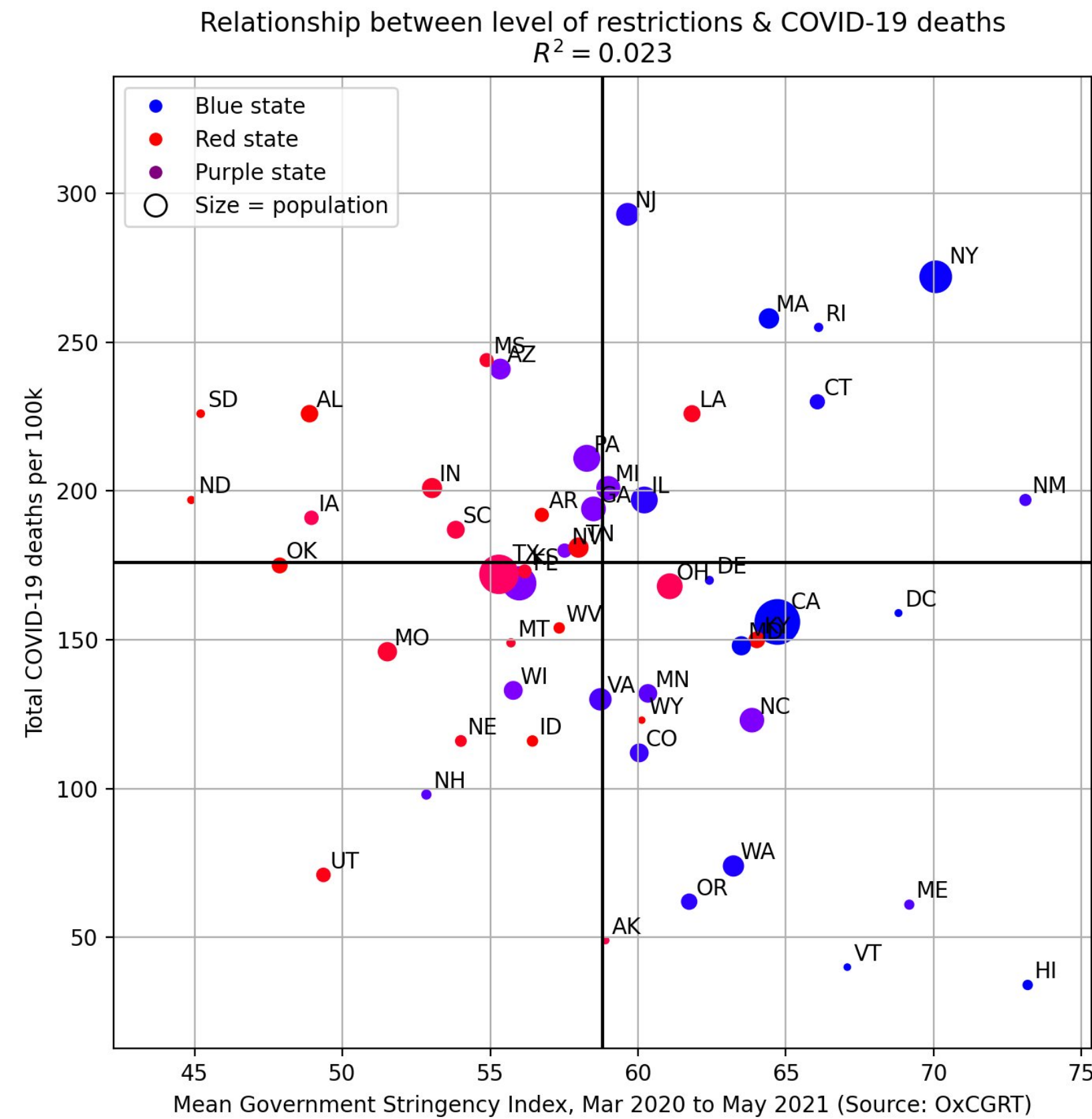
## **Au programme**

1. La causalité avec et sans expériences contrôlés randomisées (expériences et quasi-expériences)
2. Comment s'assurer que ce qui a fonctionné à petite échelle fonctionnera à plus grande échelle ?
3. Un temps d'échange sur la conférence, mon parcours et sur la recherche

### **Souvent dans les médias**

- Pour estimer le contrefactuel, comparer ceux qui ont reçu la politique à ceux qui n'ont pas reçu la politique pour avoir une idée ?
- Par exemple, on peut vérifier si les pays qui ont mis en place une politique plus intensément (SaM plus haut, confinement plus dur) affichent de meilleurs résultats ?

- Fausse bonne idée : sauf exceptions, ce type de corrélation/ce type de comparaisons entre pays ne permettent pas de faire d'affirmation causale/de savoir si une politique fonctionne ou non.
- Plus précisément, deux biais courants :
  - a. Le biais de simultanéité (la poule ou l'oeuf)
  - b. Le biais de la variable omise



Youyang Gu

@youyanggu

We can dive deeper by looking at levels of restrictions in each state, using the Oxford Covid-19 Government Response Tracker (OxCGRT).

More restrictions in a state is NOT correlated with fewer COVID-19 deaths.

However, more restrictions IS correlated with higher unemployment.

o Exemple : (i) pas de corrélation entre la dureté des mesures de restriction des mouvements (des confinements) et le nombre de morts du Covid => (ii) LE CONFINEMENT NE MARCHE PAS.

(i) est vrai, mais n'implique pas (ii).



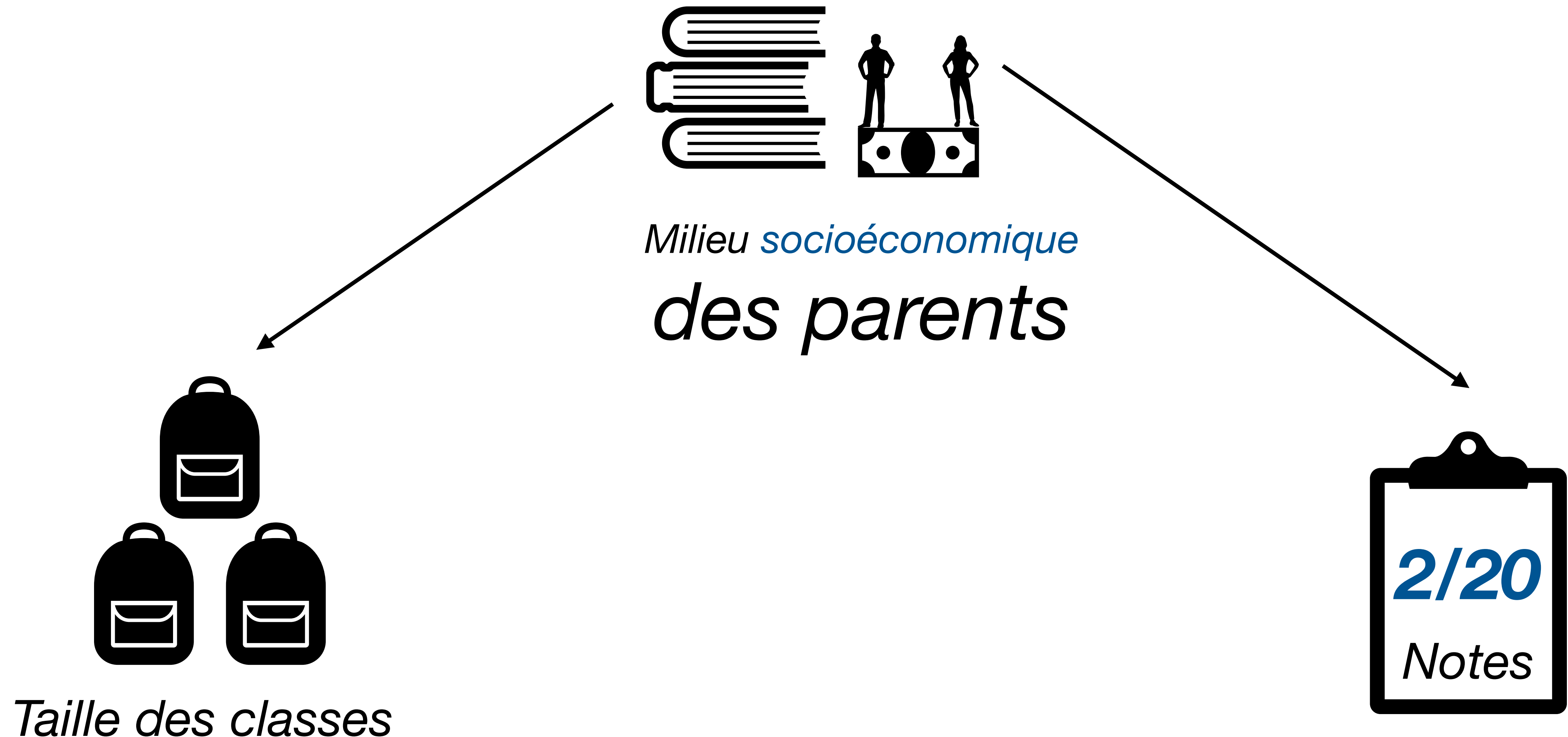
*Qui de la poule ou de l'oeuf ?*



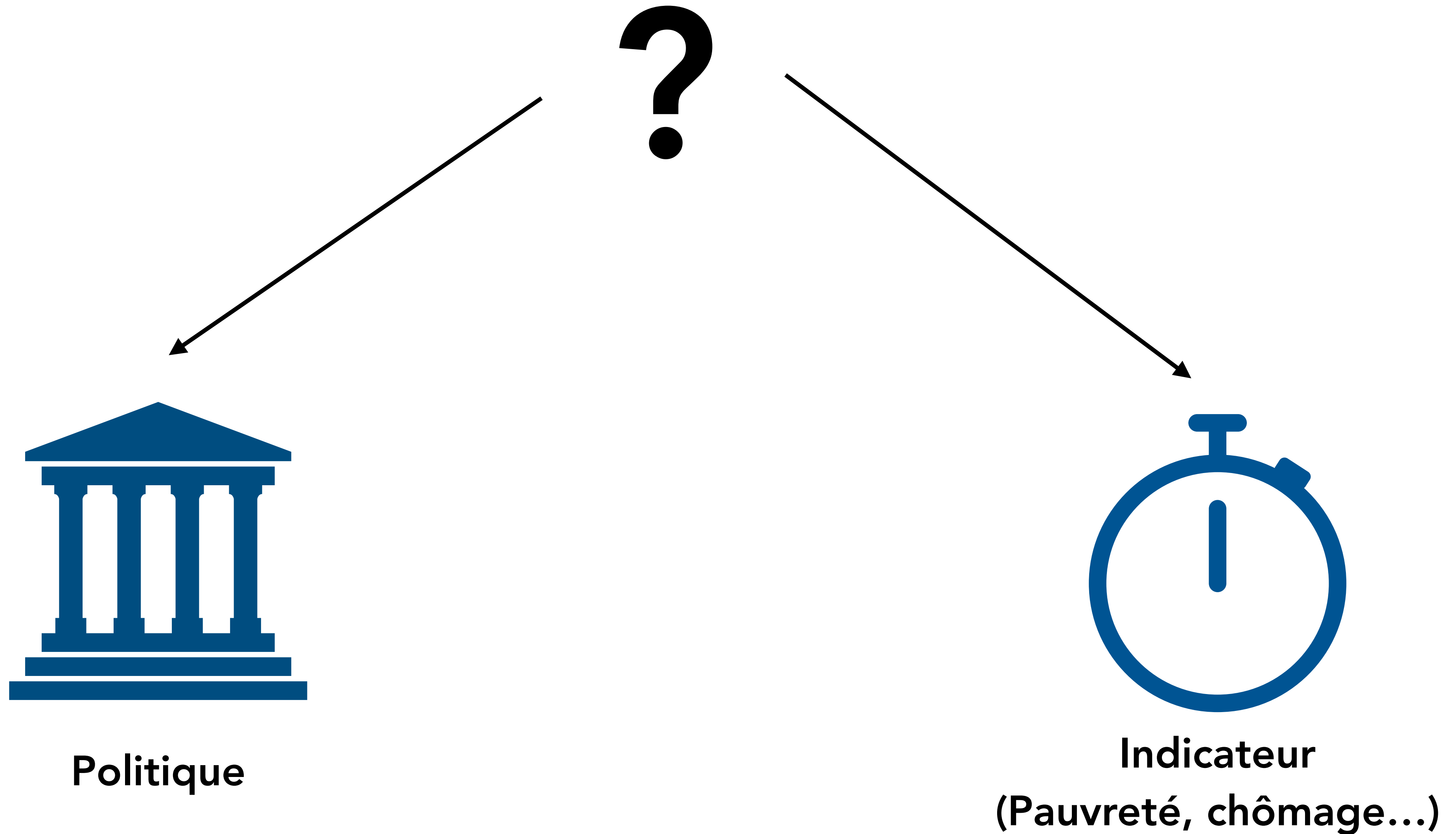








***Biais de la variable omise***

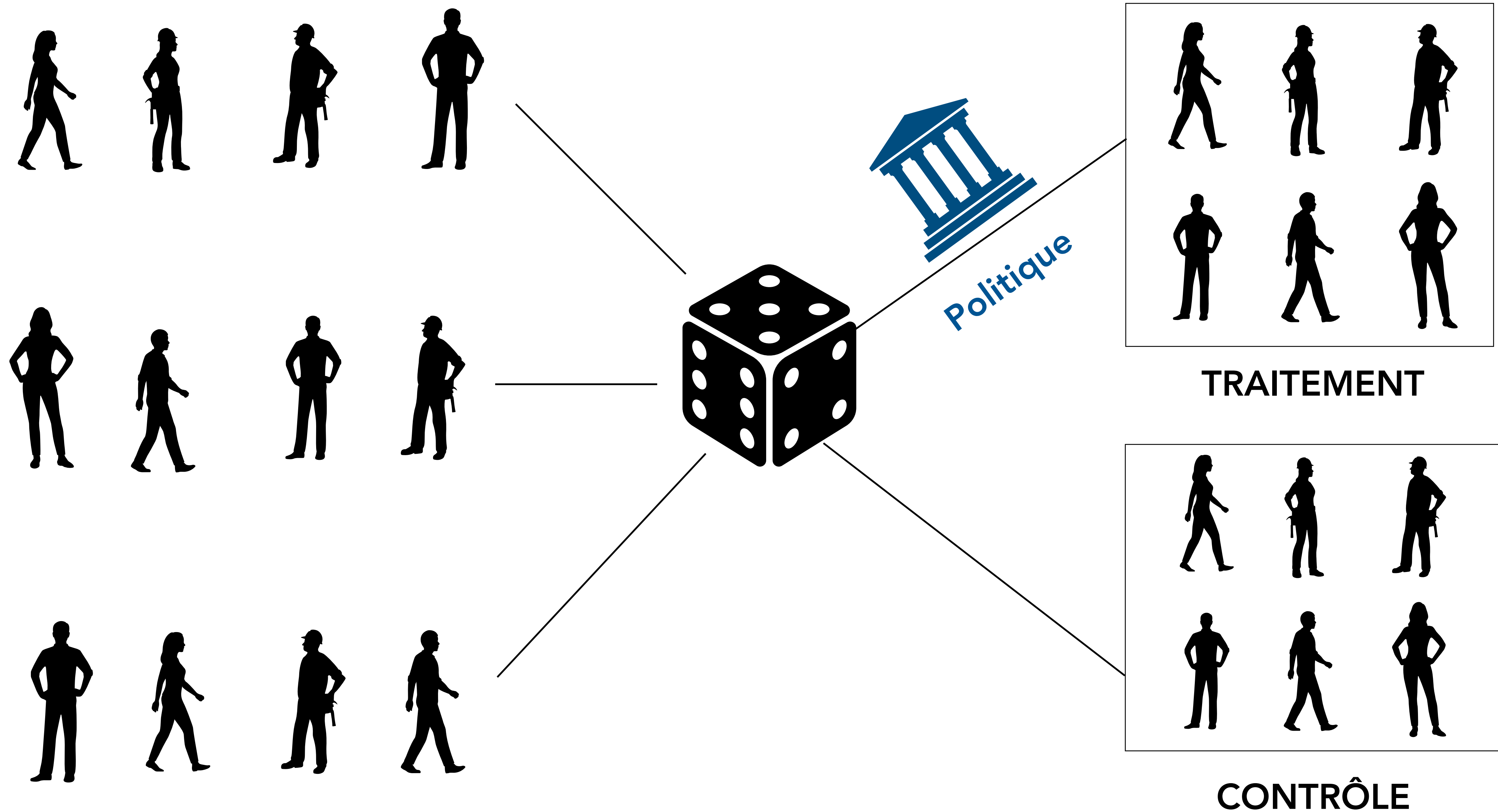


***Biais de la variable omise***

- Des papiers de recherche sur le crime et la police, et la taille de classe et les notes ont été entachés par ces biais.
- Est-ce que les auteurs de ces études étaient si naïfs que ça ?
- Non : dans les études, les chercheurs *contrôlaient* par certaines variables. *Comme on peut vérifier si l'écart de salaire entre les femmes et les hommes est significatif à éducation, expérience égales, ces chercheurs vérifiaient par exemple, si une hausse des policiers était associé à une hausse du crime à taux de chômage, de pauvreté égaux...*
- Ce n'était pas suffisant !

- Le problème c'est qu'on ne peut pas contrôler par tout (*sélection par les inobservables*), et que contrôler ne permet pas nécessairement de corriger le biais de simultanéité.
- Seules les méthodes que je vais vous présenter peuvent complètement corriger ces biais.

**Pourquoi ?**



- Le tirage au sort rend les deux groupes parfaitement comparables *a priori* (*modulo* quelques complications dont je vais parler plus tard).
- Si on tire au sort, pas de raison *systématique* qui puisse expliquer que certaines personnes avec un revenu plus haut, une santé meilleure... se retrouvent dans le groupe contrôle plutôt que dans le groupe traitement. Toutes les variables (*observables* ou *inobservables*) sont les mêmes dans les deux groupes.
- Dans ce cas, le seul facteur qui distingue le groupe contrôle du groupe traitement, c'est ... le traitement. Toute différence entre les groupes ne peut alors être due qu'au traitement.

- De fait, beaucoup d'ECRs ont été menés depuis les années 1930 pour évaluer des politiques publiques :
- Beaucoup dans les pays en développement (j'y reviendrai), pour tester l'impact du microcrédit, des moustiquaires... (J-Pal).

### **Mais aussi dans les pays développés**

- En éducation, Perry Preschool Project, dans les années 1960. => soutien préscolaire de 2h30 chaque matin pendant 2 ans. + visites au domicile familial de psychologues pour des jeunes enfants de 3,4 ans aux US => hausse spectaculaire de l'emploi, des salaires, baisse du crime;
- Sur l'ex de la baisse des classes, des ECRs comme le projet STAR dans les années 1990 sur des jeunes américains en maternelles.
- En France, l'expérience des internats d'excellence de Sourdun, Energie Jeunes...



- Il est aussi possible de faire des ECRs pour savoir comment augmenter le soutien aux politiques environnementales :
- Dechezleprêtre, Fabre, Santcheva et al. (2021) ont mené des ECRs avec 40,000 participants issus de 20 pays qui représentent 72% des émissions de la planète pour savoir comment augmenter le soutien pour des politiques environnementales. Ils montrent qu'informer sur les dangers du changement climatique a peu d'effet. Ce qui compte est de montrer qu'une politique donnée est (a) efficace pour réduire les émissions ; (b) n'affectera pas disproportionnellement les plus pauvres ; (c) qu'individuellement, le foyer de l'étude ne risque pas de trop y perdre/de gagner.

- Il est aussi possible de faire des ECRs pour évaluer des politiques environnementales, en particulier pour tenir compte de biais humains non inclus dans des modèles d'ingénierie :
  - Fowlie *et al.* (2015) rapportent les résultats d'un ECR mené auprès de plus de 30 000 foyers évaluant un grand programme d'efficacité énergétique résidentielle aux É-U.
  - Les économies réelles d'énergie sont environ **2,5 fois** inférieures à celles projetées par les modèles d'ingénierie.

Pourquoi ? Christensen *et al.* (2018) : 43% du gap qualité des travaux de rénovation (autre facteur humain), 6% effet rebond.

- Toutefois, parfois, il n'est pas possible de mener un ECR.
- Première solution : expérience naturelle.
  - => Ici, trouver des cas où ce ne sont pas les chercheurs qui ont fait le tirage au sort.
- Beaman et al. (2012) en Inde exploitent une loi de 1993 qui a réservé aux femmes des postes de direction dans des conseils de village sélectionnés au hasard.
- Ils constatent que dans les villages où les postes ont été réservés,...
  - L'écart entre les sexes dans le niveau d'éducation des adolescents a été effacé.
  - Les femmes consacrent moins de temps aux tâches ménagères.

- Angrist (2021) (« *Quand j'étais au lycée, comme beaucoup de ma génération, je fumais énormément de weed. (...)* ») Son frère - dans un autre lycée, meilleur- en fumait beaucoup moins. Est-ce les fréquentations de l'autre lycée qui ont empêché son frère de fumer ?

*Est-ce qu'on a plus de chance d'avoir des proches qui se comportent comme nous surtout parce que nos proches nous influencent (effet de pairs ), ou surtout parce que se ressemble s'assemble, et que nos proches sont du même milieu social (biais de sélection) ?*

- Est-ce que fréquenter des camarades meilleurs scolairement améliore ses résultats, ou est-ce qui se rassemble s'assemble ?
  - => Expérience naturelle: *Darmouth college* aux Etats-Unis, où l'allocation dans les chambres de dortoirs est aléatoire. Avoir un colocataire qui est meilleur que la moyenne augmente ses propres résultats.
- Est-ce que ceux qui ont des amis/de la famille médecins sont en meilleure santé à cause des conseils de leurs amis médecins ?
  - => Expérience naturelle: Persson et al. (2015) utilisent une loterie d'admission à la faculté de médecine en Suède. Avoir un membre de sa famille médecin augmente son espérance de vie.

### Macroéconomie (Broda and Parker (2014))

- Expérience naturelle pour résoudre le problème de simultanéité dans l'effet des dépenses publiques sur le PIB :

*Les dépenses publiques affectent (probablement) le PIB, mais le PIB affecte aussi les dépenses publiques (si baisse du PIB, hausse du chômage, hausse des transferts sociaux...).*

- Difficile à contourner, mais parfois on a de la chance, comme avec la relance 2008 aux US : Impossible, d'un point de vue logistique, d'envoyer tous les chèques le même jour, l'administration fiscale américaine a déterminé la date exacte à laquelle chaque ménage recevrait le chèque grâce à l'avant-dernier chiffre du numéro de sécurité sociale, qui est attribué de manière **aléatoire**.

=> Comme un ECR, avec comme groupe traitement ceux qui ont reçu le chèque en premier contre ceux qui n'ont pas encore reçu.



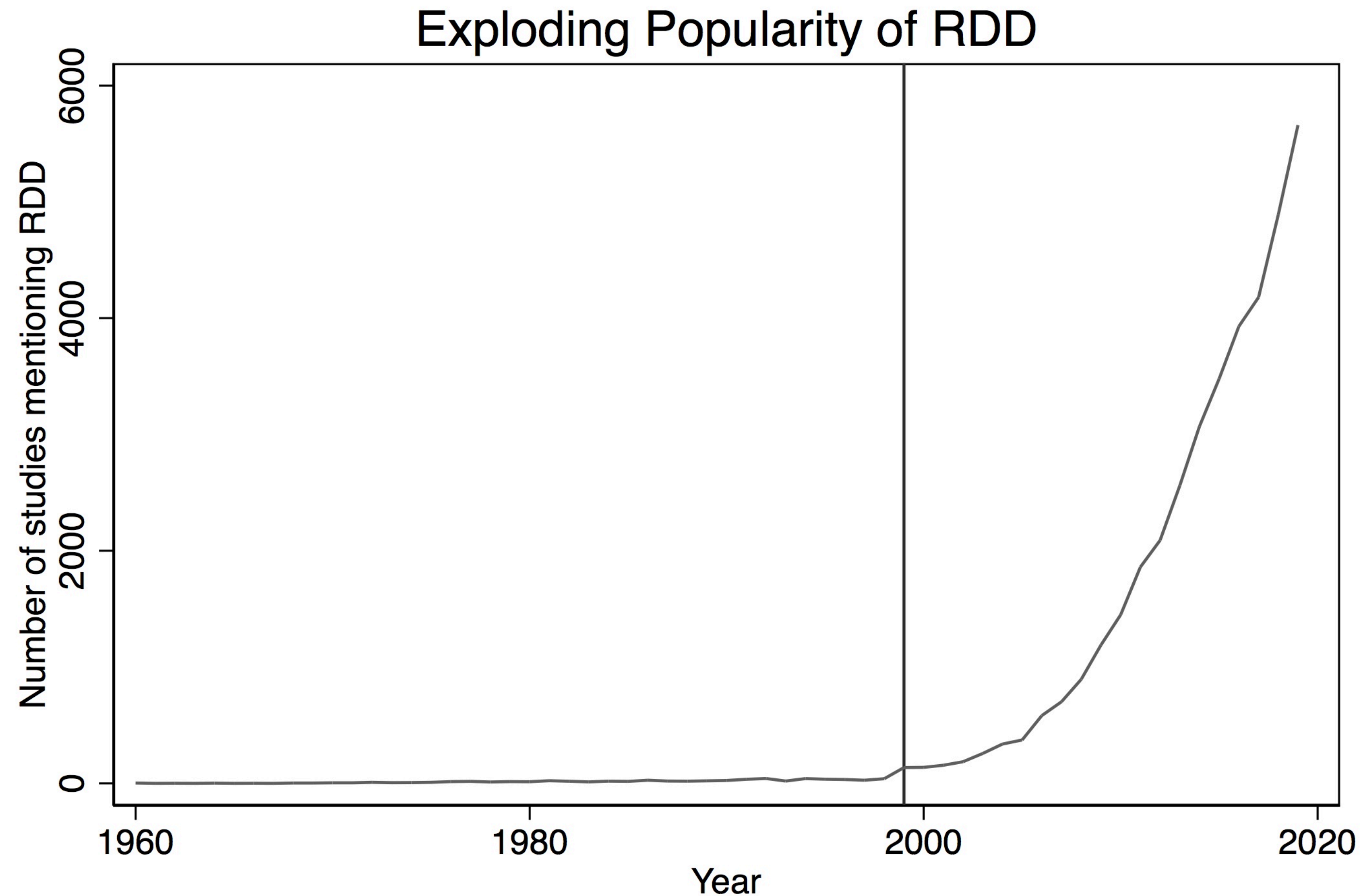
## **Macroéconomie (Broda and Parker (2014))**

- Les ménages ont bien en majorité dépensé cette argent, et ne l'ont pas épargné
- Résultat important (première étape) pour estimer l'effet multiplicateur (De combien augmente le PIB quand l'Etat augmente ses dépenses publiques d'1 milliard).
  - => On pense souvent que la révolution causale n'a pas bénéficié à la macroéconomie. C'est très discutable. La dernière génération de modèles (théories mathématiques) en macroéconomie est celle des HANKs, qui tentent d'être cohérents avec les résultats des expériences naturelles.

- Parfois, pas de phénomènes aléatoire aussi explicites, solution : quasi-expériences.
- Quasi-expériences : ensemble de méthodes qui permettent de tenter de distinguer corrélation et causalité quand on ne peut faire d'expériences (et en corrigeant la sélection par les inobservables).
- Comment on sait qu'elles permettent de distinguer corrélation et causalité ?  
Annexe 3.



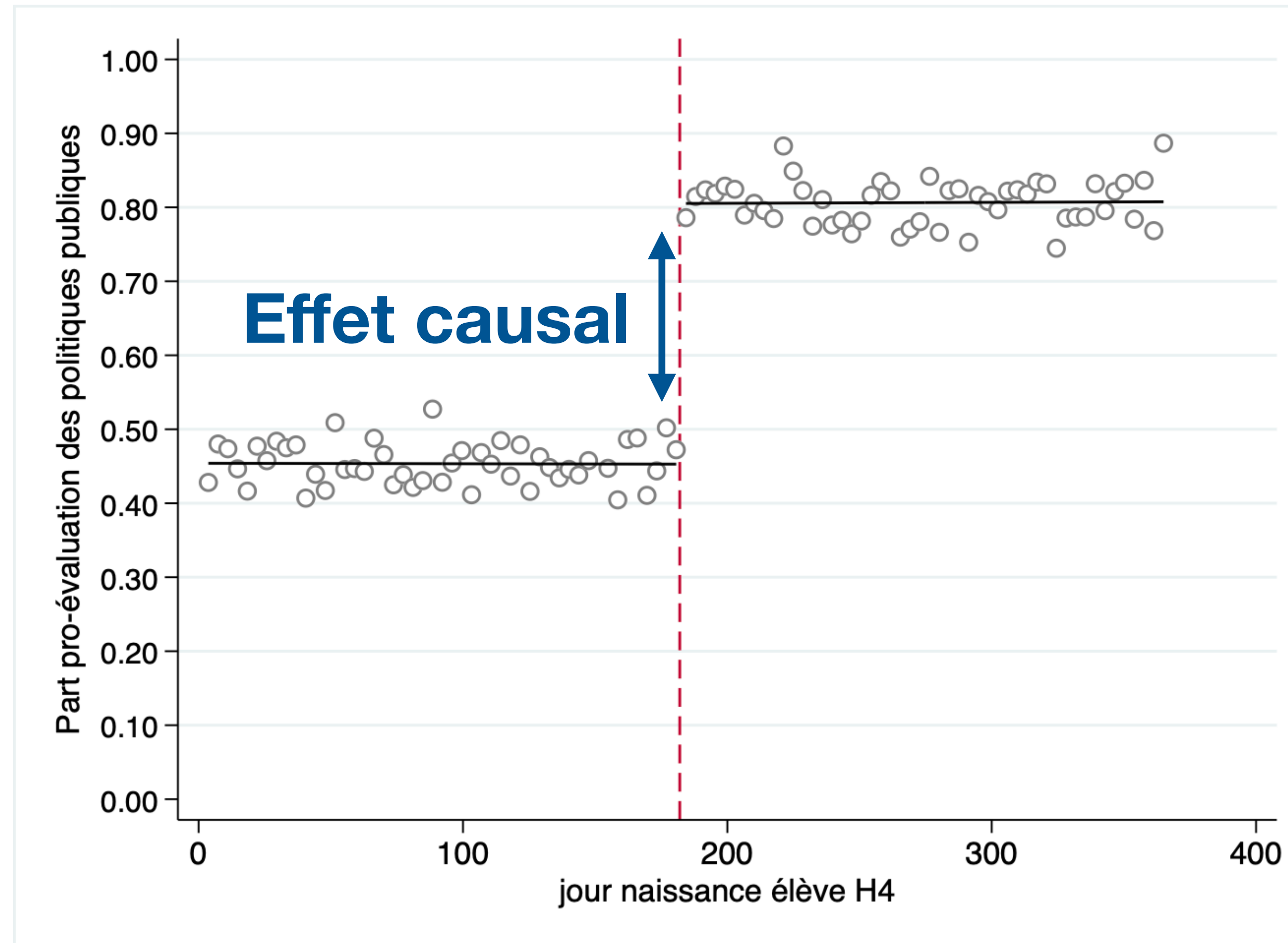
- Je ne présente pas toutes les méthodes (impossible), mais deux exemples basiques :
  - La reine des quasi-expériences : la Régression en Discontinuité.
  - La plus populaire : le Différence-en-différence.
- L'implémentation de ces méthodes peut être (très technique), en particulier celle de leurs extensions, mais elles sont toutes fondées sur des principes intuitifs.



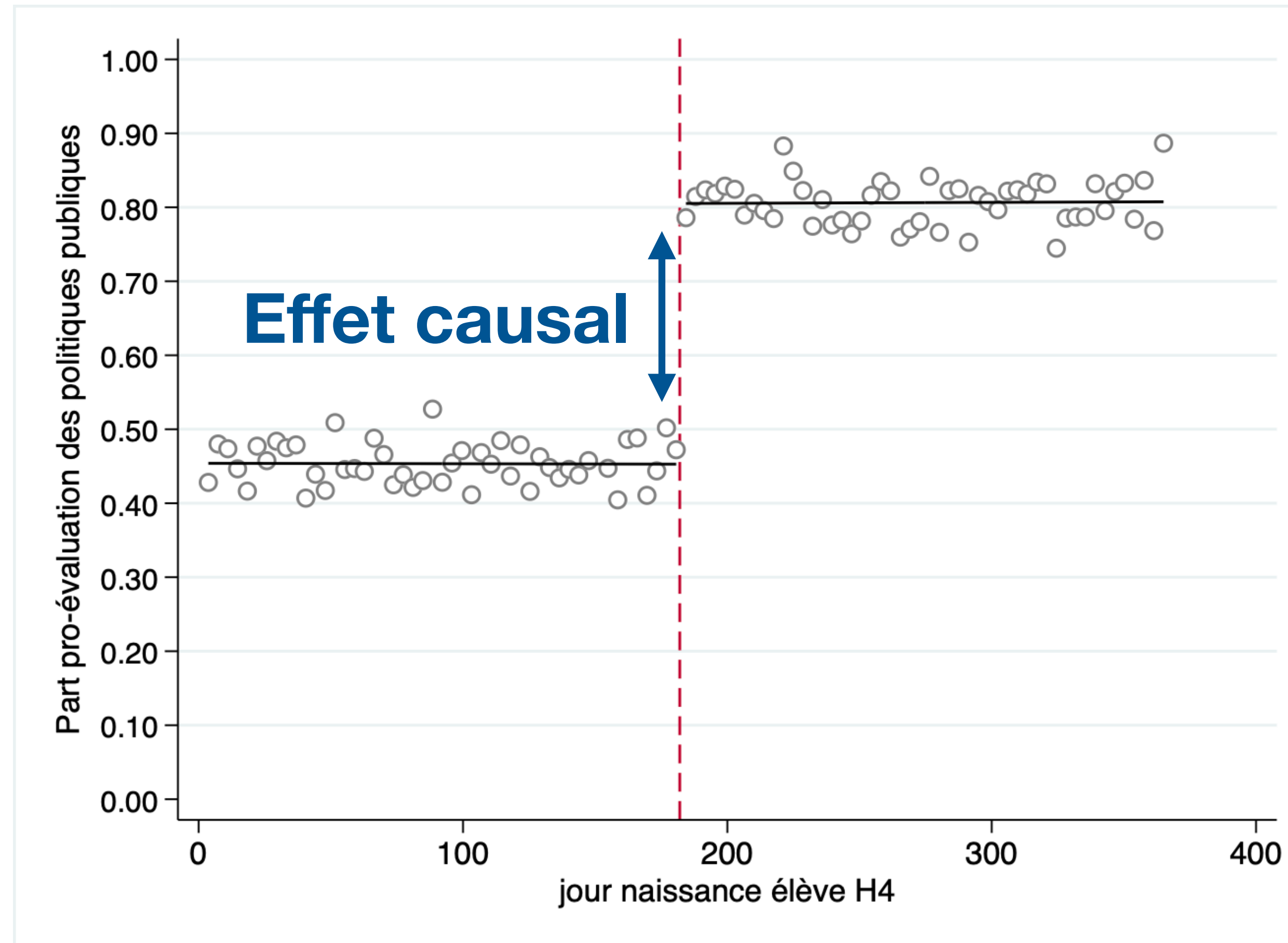
Vertical bar is Angrist and Pischke (1999) and Deaton (1999)

From *Causal Inference* (2021), Scott Cunningham.

- o Méthode créée en psychologie à la popularité explosive depuis que les économistes s'en sont emparés en 1999.



- Si une intervention est déclenchée à partir d'un critère arbitraire, ... ceux juste au dessus, ou juste en dessous de ce seuil devraient être statistiquement similaires.
- Ex sur données simulées : imaginez que cette conférence soit réservée à ceux nés à partir du 182ème jour de l'année.



- Cette méthode n'isole un effet que sous deux hypothèses : (a) non-manipulation de la variable du seuil (ici, la date de naissance) ; et (b) exogénéité (Ex: sans l'intervention, ceux qui sont nés le 181ème vs 182ème jour de l'année auraient la même probabilité d'être pro-évaluation des politiques publiques).

*D.S. Lee / Journal of Econometrics 142 (2008) 675–697*

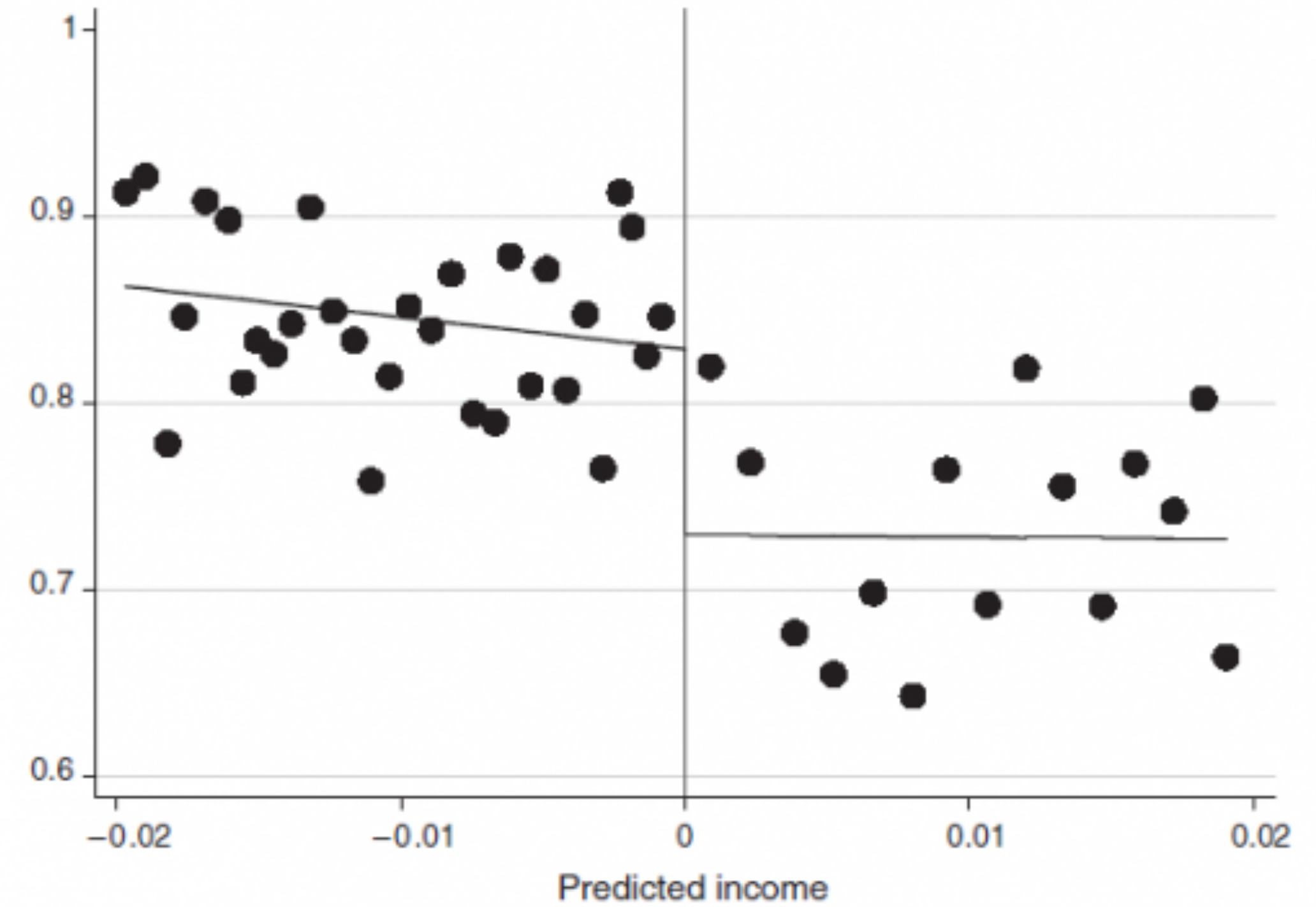
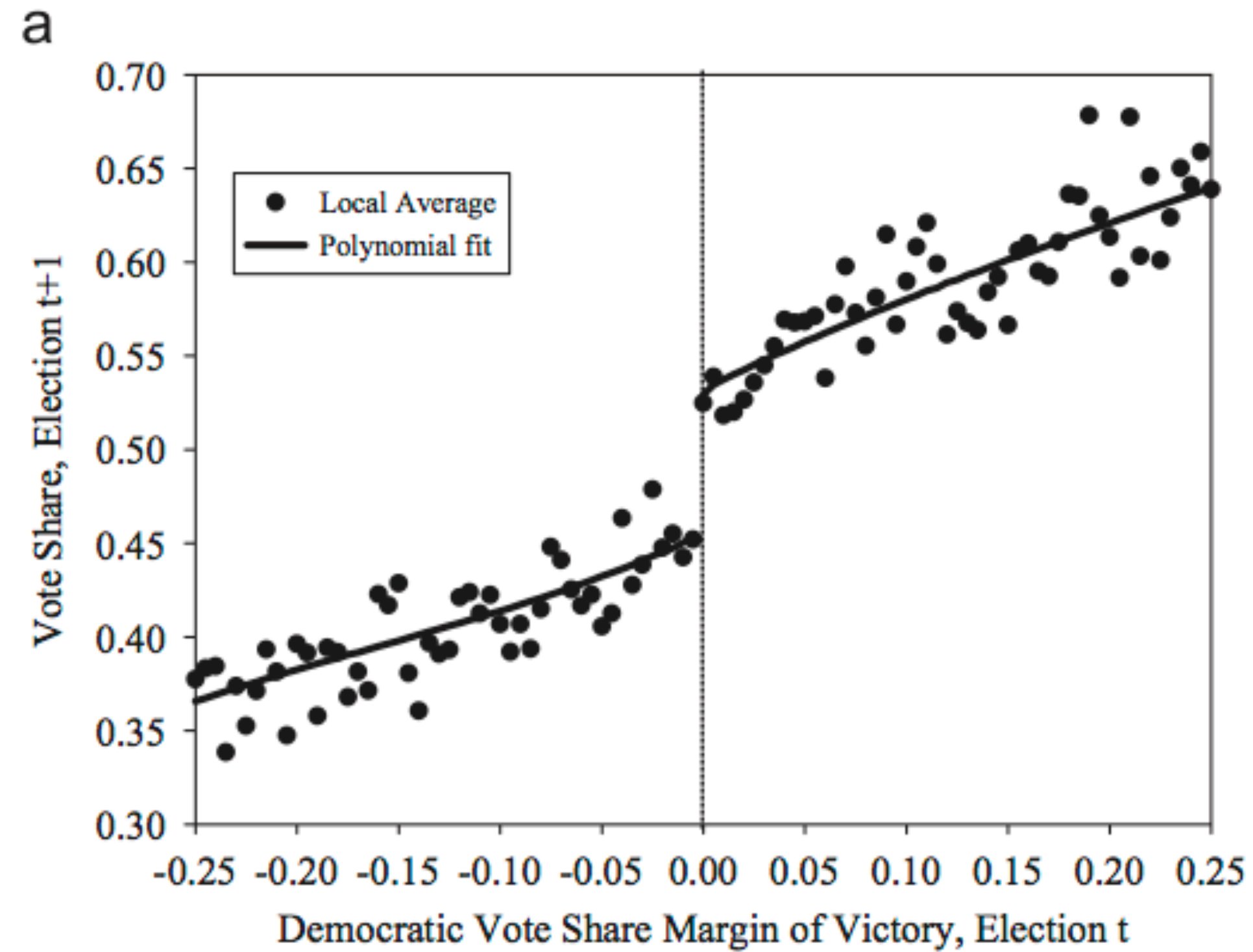


FIGURE 4. PANES PROGRAM ELIGIBILITY AND POLITICAL SUPPORT FOR THE GOVERNMENT, 2008 FOLLOW-UP SURVEY ROUND



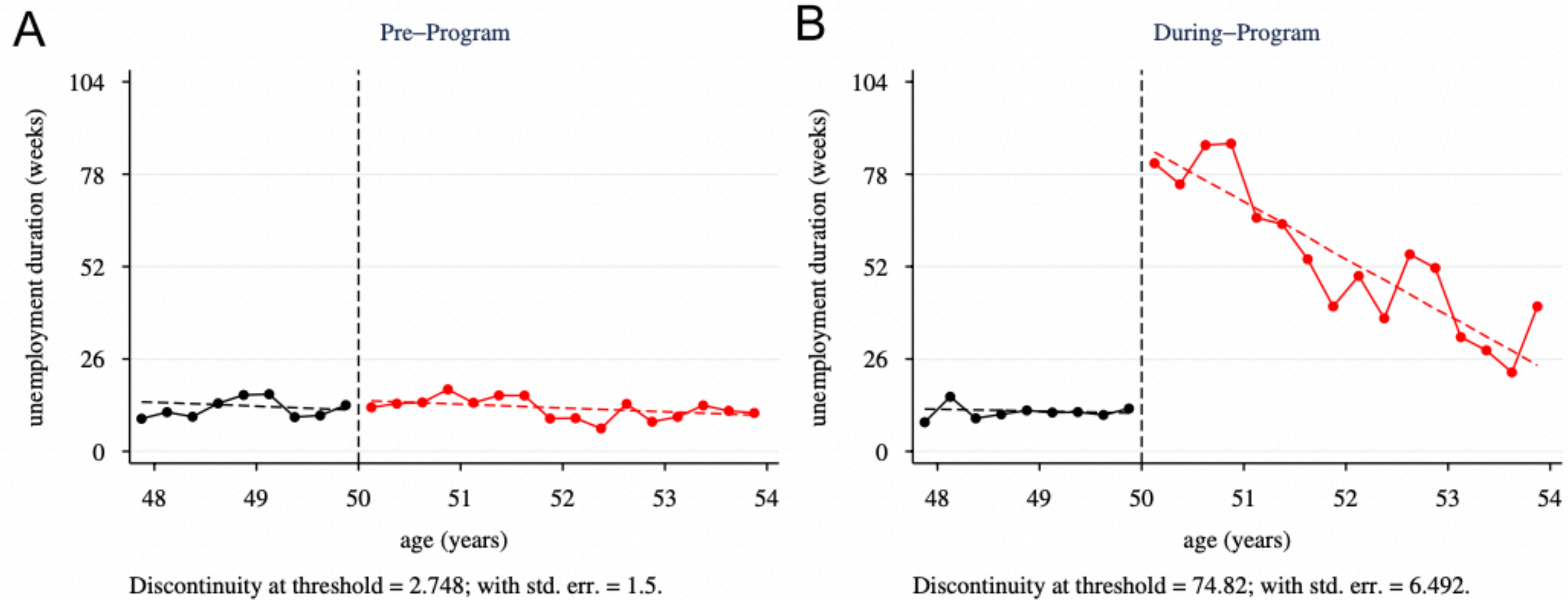
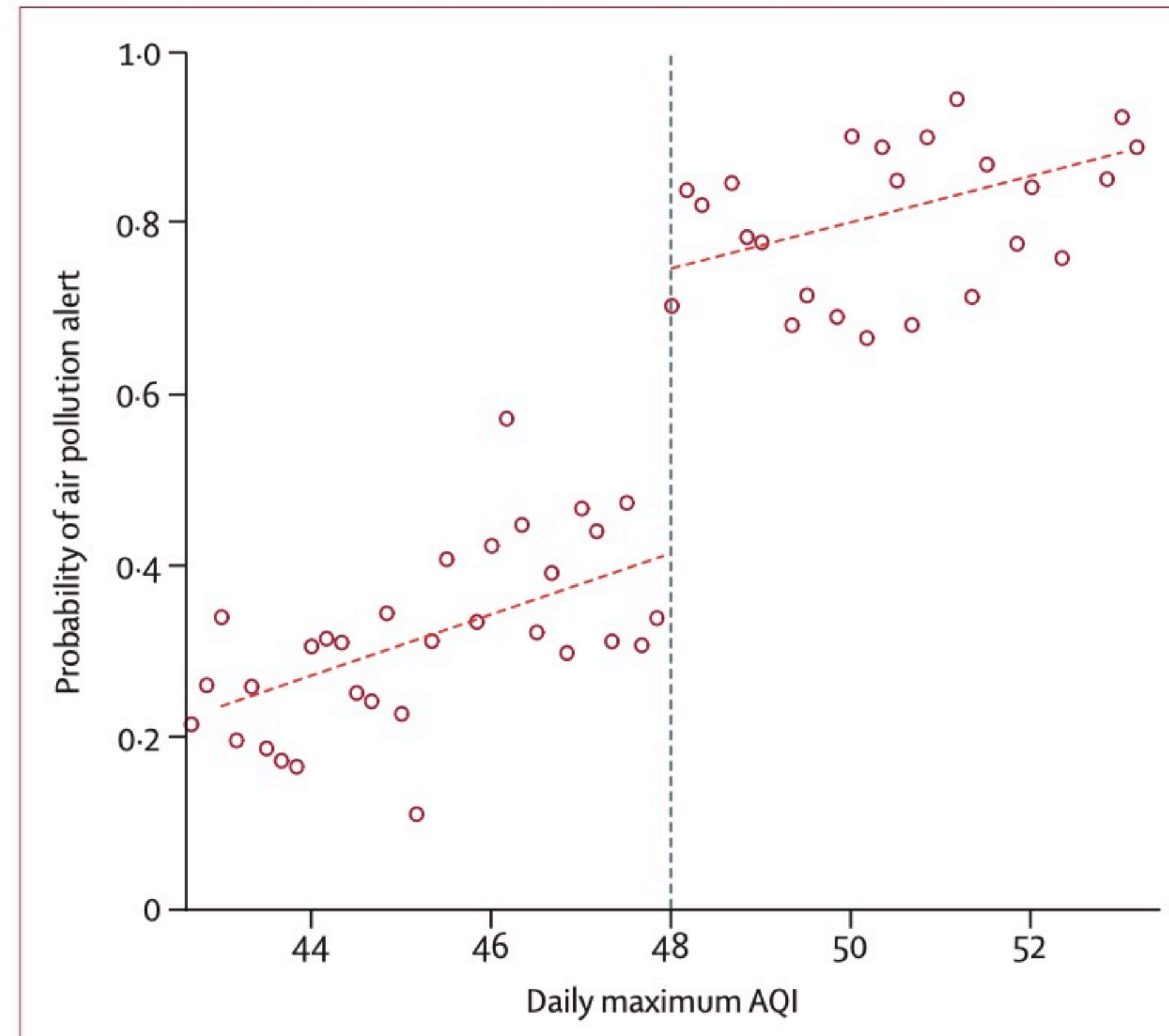


Fig. 10. Results for repeatedly unemployed women. Pre-program refers to women with pre-program unemployment duration information. During-program refers to the same women when REBP in force.



**Figure 1: Probability of an air quality alert in Toronto, Canada, from 2003 to 2012, by daily maximum AQI**

Plot is centred at an AQI of 48 (blue vertical dotted line). Eligible days were days with daily maximum AQI  $\geq 48$  (dots to the right) and non-eligible days were days with daily maximum AQI  $< 48$  (dots to the left). Red dashed lines depict linear regression lines. AQI=air quality index.







Journal of Clinical Epidemiology

Volume 68, Issue 2, February 2015, Pages 132-143



Review Article

## Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice

Ellen Moscoe<sup>a</sup>  , Jacob Bor<sup>b c</sup>, Till Bärnighausen<sup>a c</sup>

JOURNAL ARTICLE

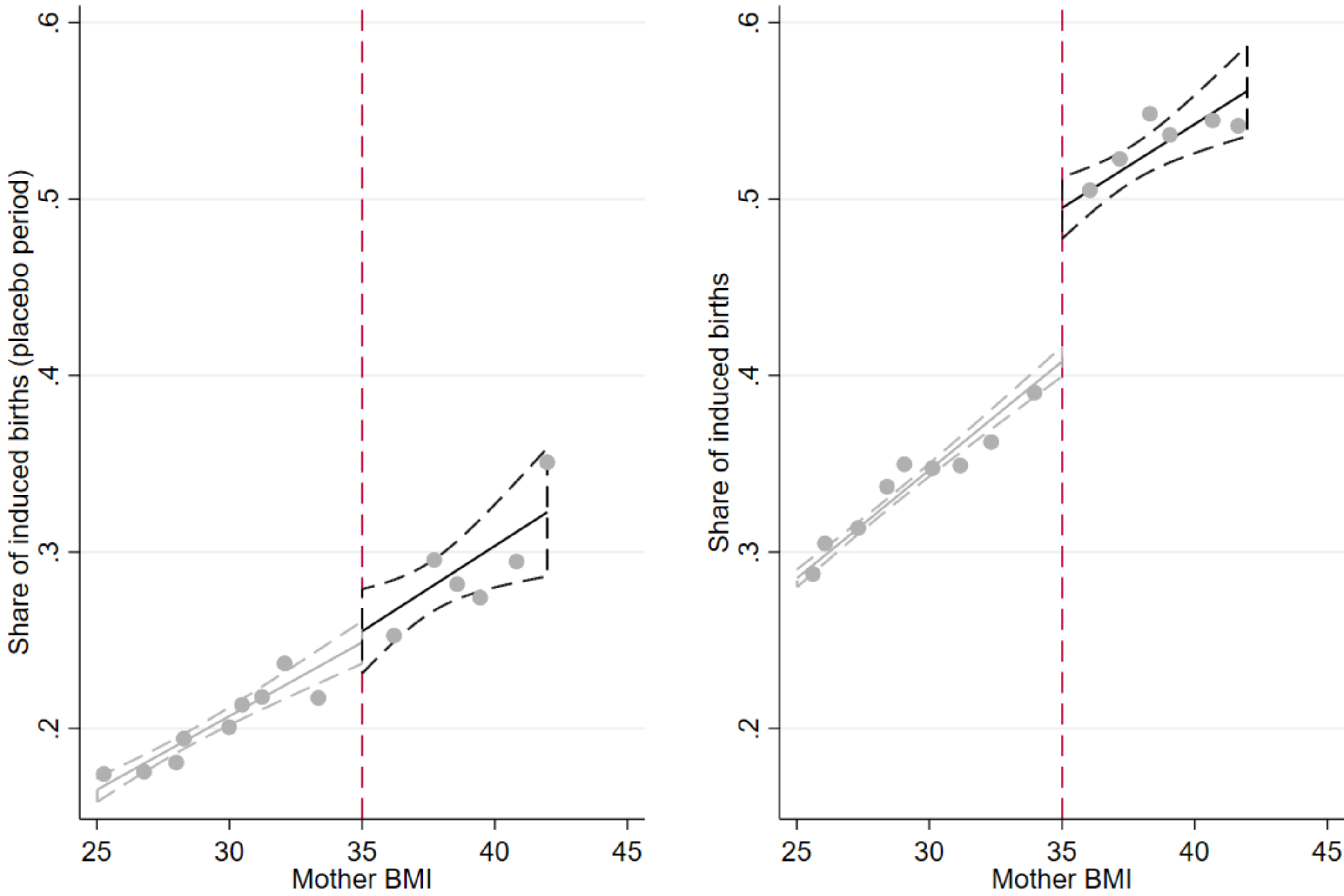
## Evaluating the Effectiveness of Vaccines Using a Regression Discontinuity Design

Nicole E Basta , M Elizabeth Halloran

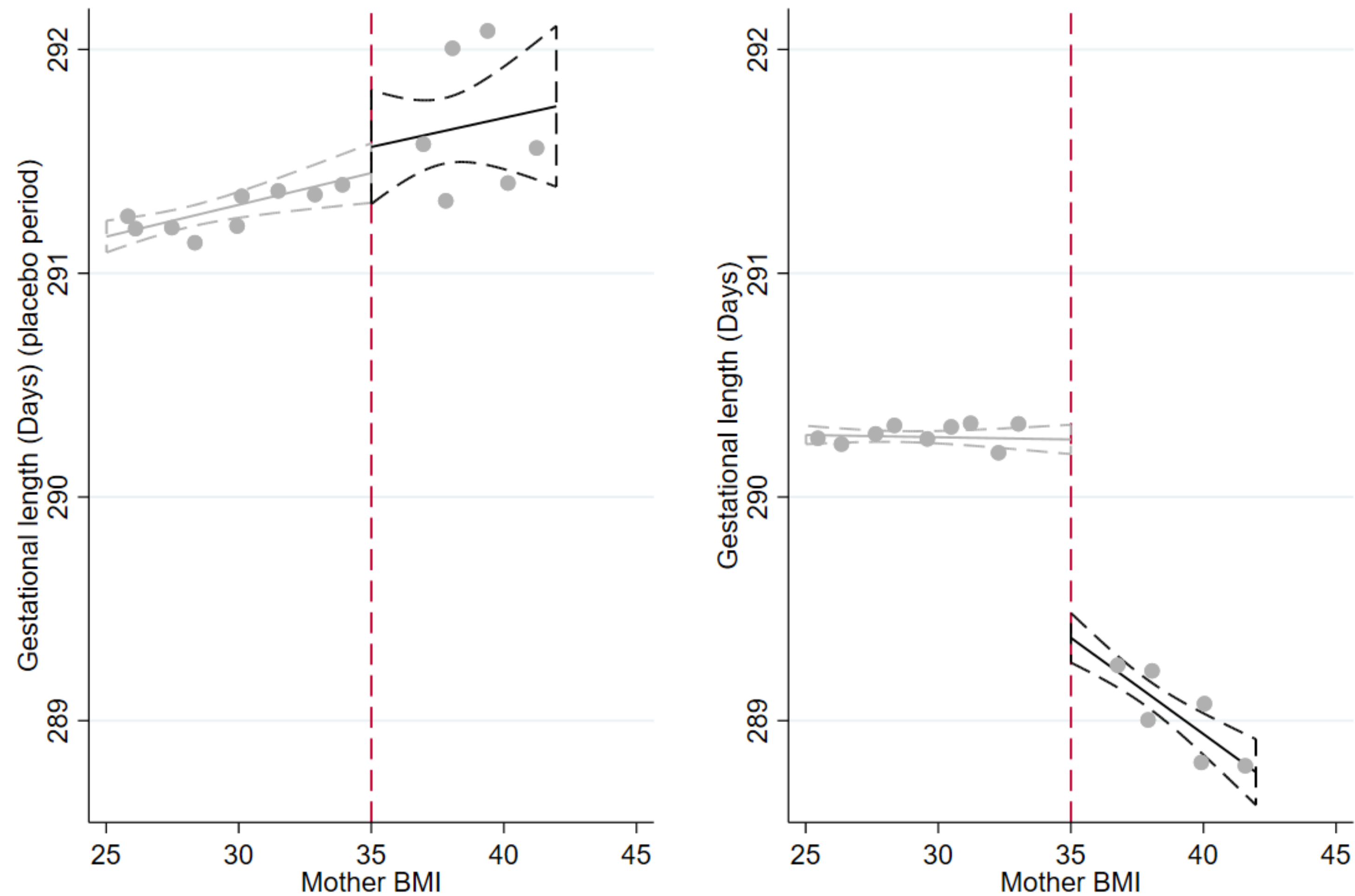
*American Journal of Epidemiology*, Volume 188, Issue 6, June 2019, Pages 987–990,

<https://doi.org/10.1093/aje/kwz043>

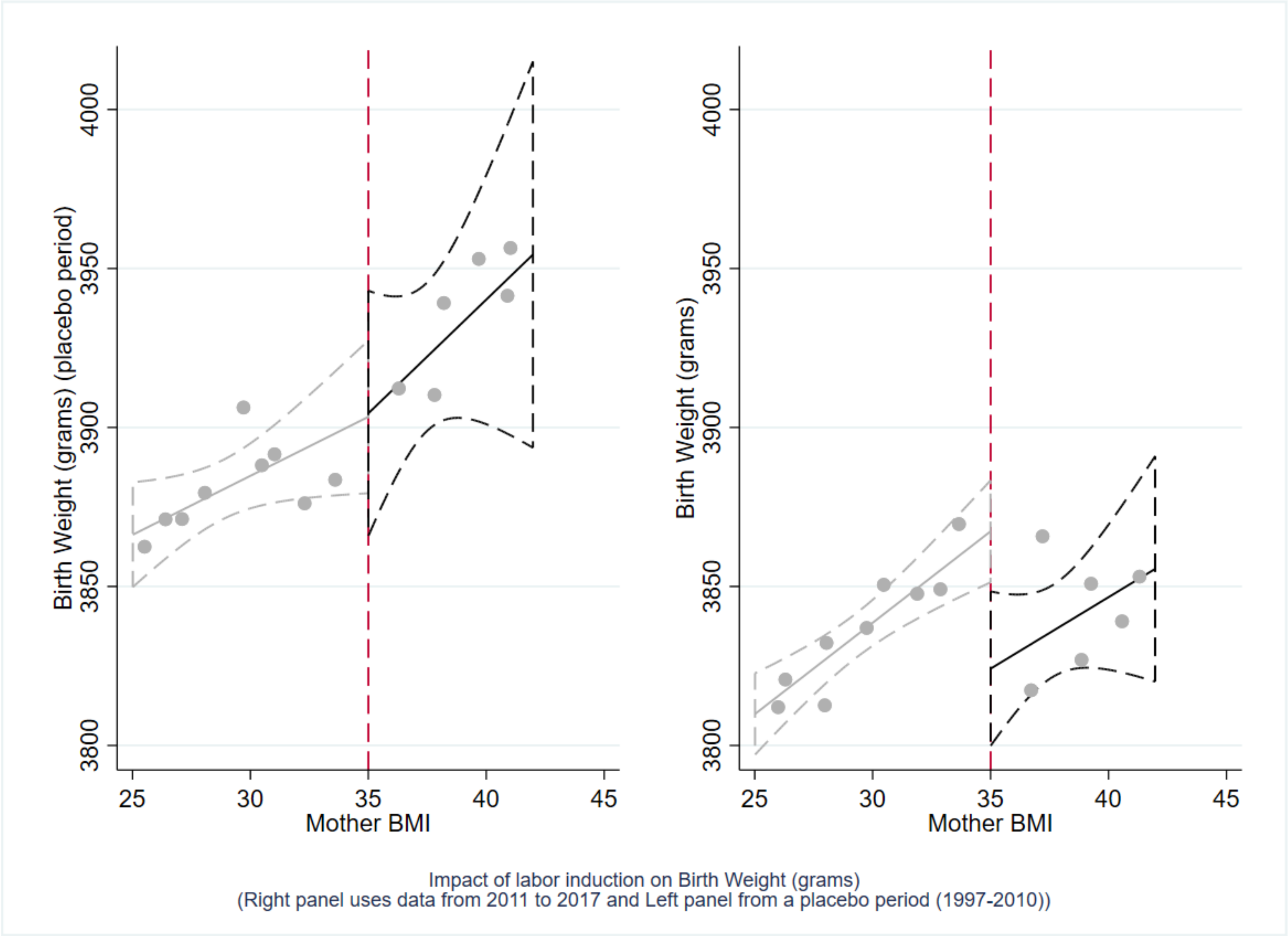
**Published:** 19 February 2019    **Article history** ▼



Impact of labor induction on Share of induced births  
(Right panel uses data from 2011 to 2017 and Left panel from a placebo period (1997-2010))



Impact of labor induction on Gestational length (Days)  
(Right panel uses data from 2011 to 2017 and Left panel from a placebo period (1997-2010))





- Limitations de la méthode :

- a. Effet causal local : sur ceux autour du seuil pas sur tous les bénéficiaires

- Solution : Être prudent sur l'interprétation.

- b. Beaucoup de paramètres libres (plus de détails dans l'annexe 2):

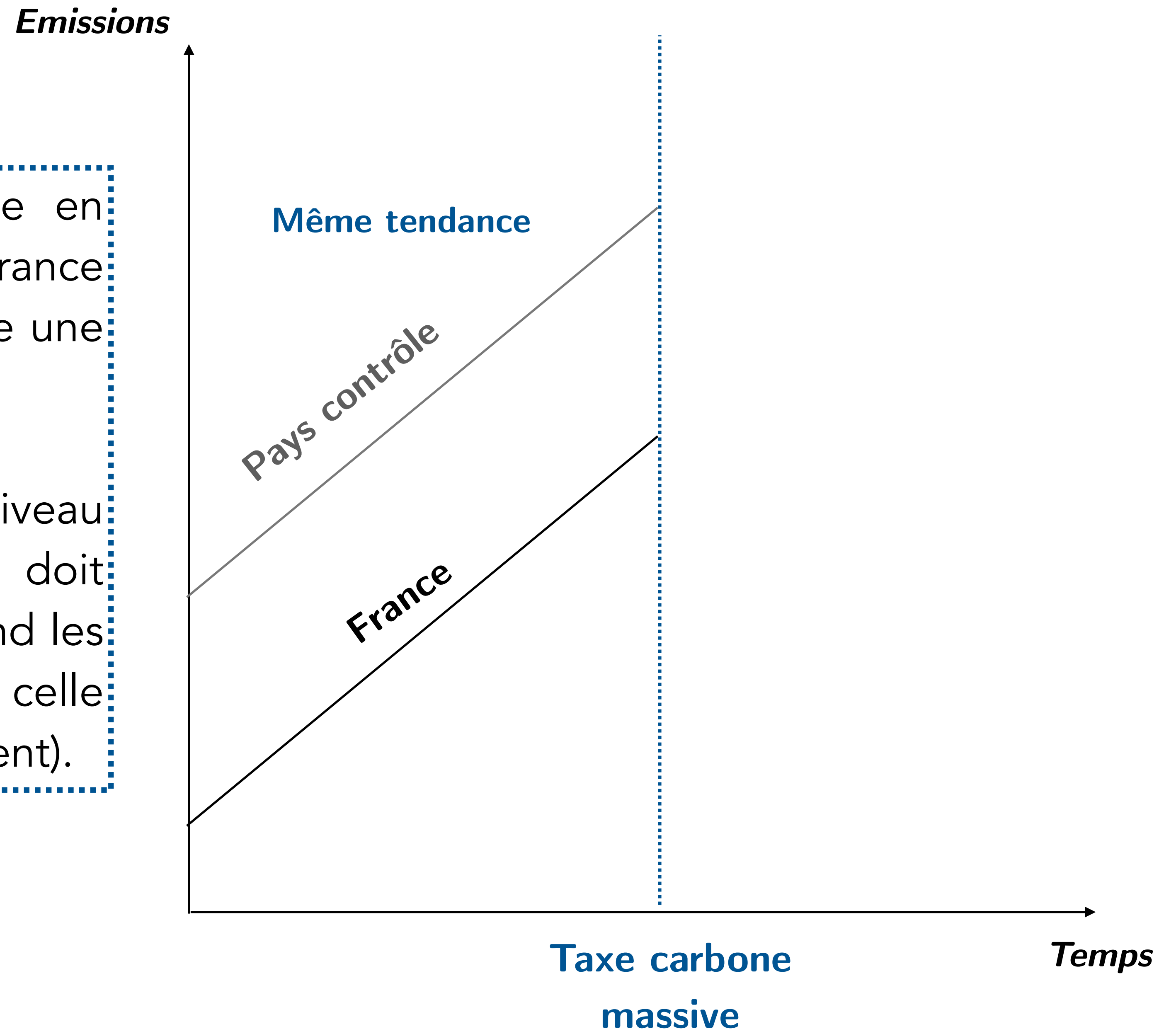
- Solution : Algorithmes qui calculent les choix optimaux. Ceci permet surtout d'empêcher de changer des paramètres jusqu'à tant qu'il trouve quelque chose qu'il lui plaît.

## Différence-en-différence (DiD)

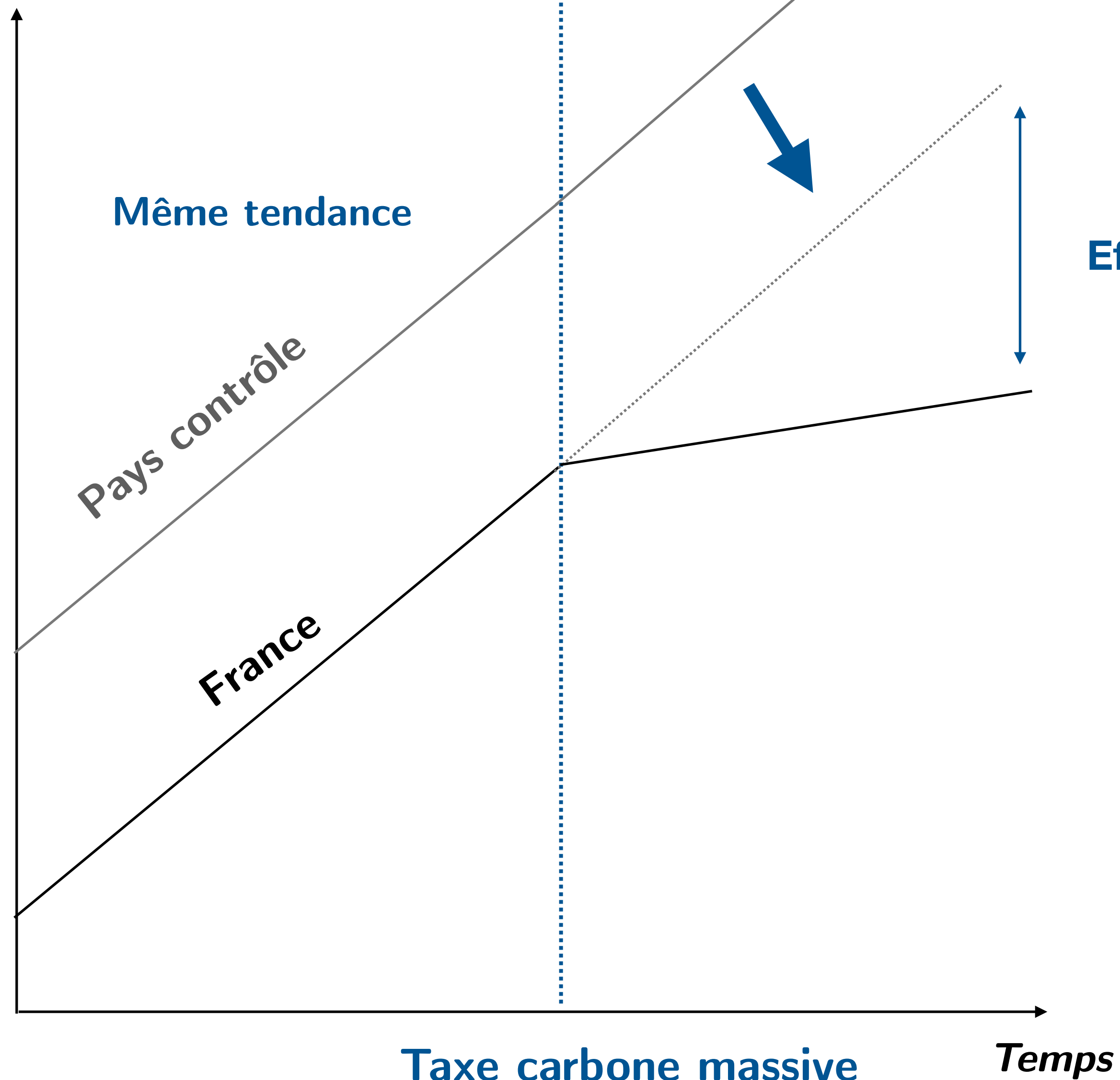
- Méthode qui vient de santé publique (employée par John Snow pour montrer que le choléra se transmet par l'eau sale) et qui a été modernisée par des économistes, initialement pour étudier l'effet du salaire minimum sur l'emploi.
- Elle a connu beaucoup de développement plus sophistiquées mais l'idée de base est très simple.
  - Imaginez que l'Etat français souhaite mettre en place une taxe carbone.



- Trouver un pays qui a une tendance en termes d'émissions comparable à la France *avant la taxe*, et qui n'a pas mis en place une telle taxe.
- Ce pays n'a pas à avoir le même niveau d'émission de la France, mais l'écart doit rester le même au cours du temps : quand les émissions de la France baisse une année, celle de ce pays aussi, ( $\leq$  quand elles montent).



- Quand la réforme intervient, l'effet de la taxe carbone sur les émissions est alors ...
- La différence entre la variation des émissions en France après la taxe et la variation des émissions dans le pays contrôle.

*Emissions*

## Quatre hypothèses (principales) :

- A. Tendances parallèles : si la politique n'avait pas eu lieu, les trajectoires d'émissions des pays auraient continué à évoluer en même temps, de manière synchronisée.  
=> Invérifiable directement, mais ... on peut vérifier si les trajectoires d'émissions de CO2 étaient les mêmes avant.
- B. Pas de politique mise en place en même temps.
- C. Pas de débordement (je vais en reparler)
- D. Pas d'effet d'anticipation :  
Solution: vérifier si les tendances se séparent non pas à partir de l'implémentation de la mesure, mais à partir du moment où elle est annoncée.

## Comment faire qu'on n'a pas de pays avec des tendances parallèles ?

1. Contrôle synthétique : on peut comparer les émissions en France à une moyenne pondérée d'autres pays - un algorithme fait en sorte que certains pays comptent moins que d'autres; dans la moyenne, de telle sorte que ce pays synthétique aura les mêmes émissions que la France avant la réforme.
  - C'est ce qu'Andersson (2019) a fait avec la Suède et sa taxe carbone en 1992.

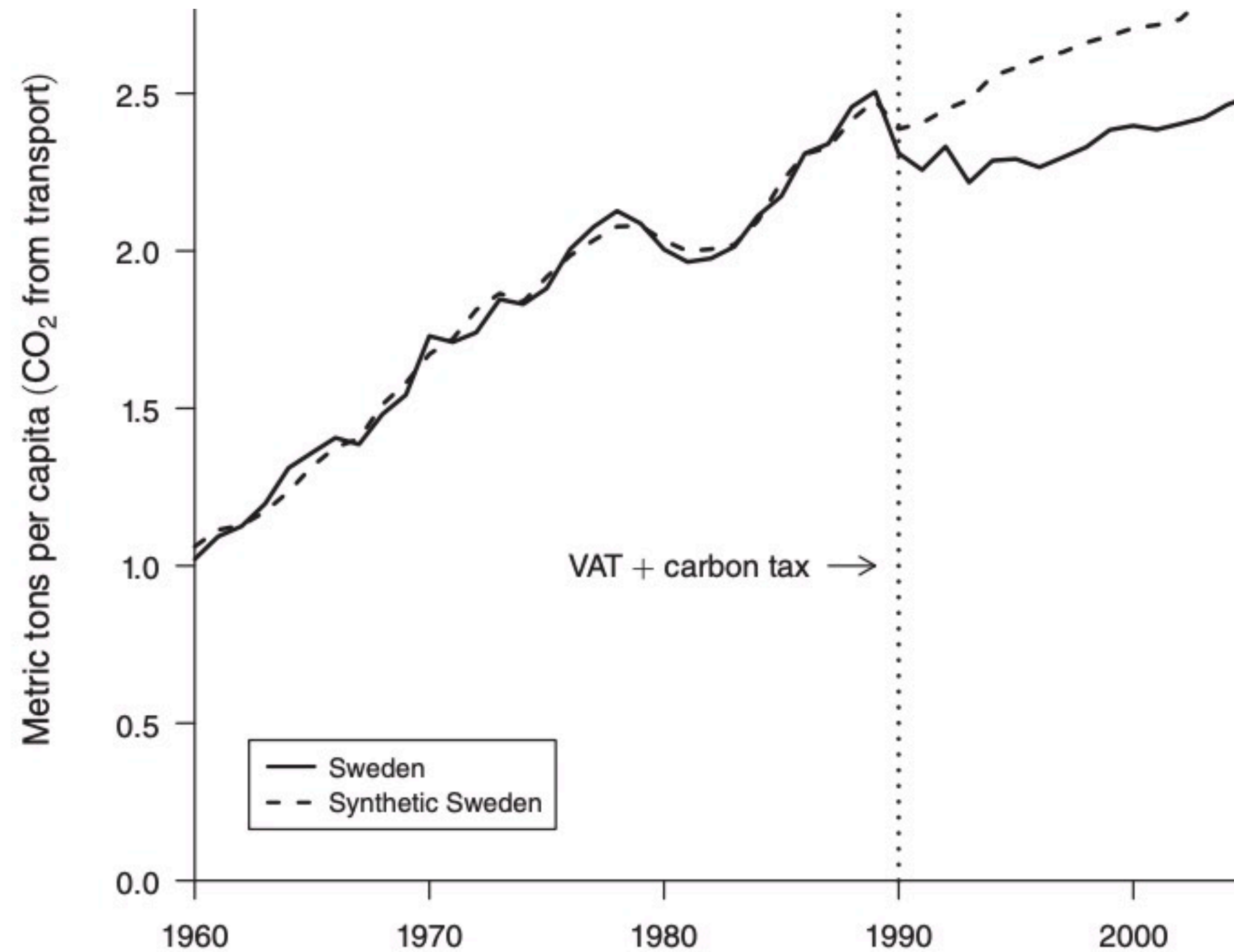
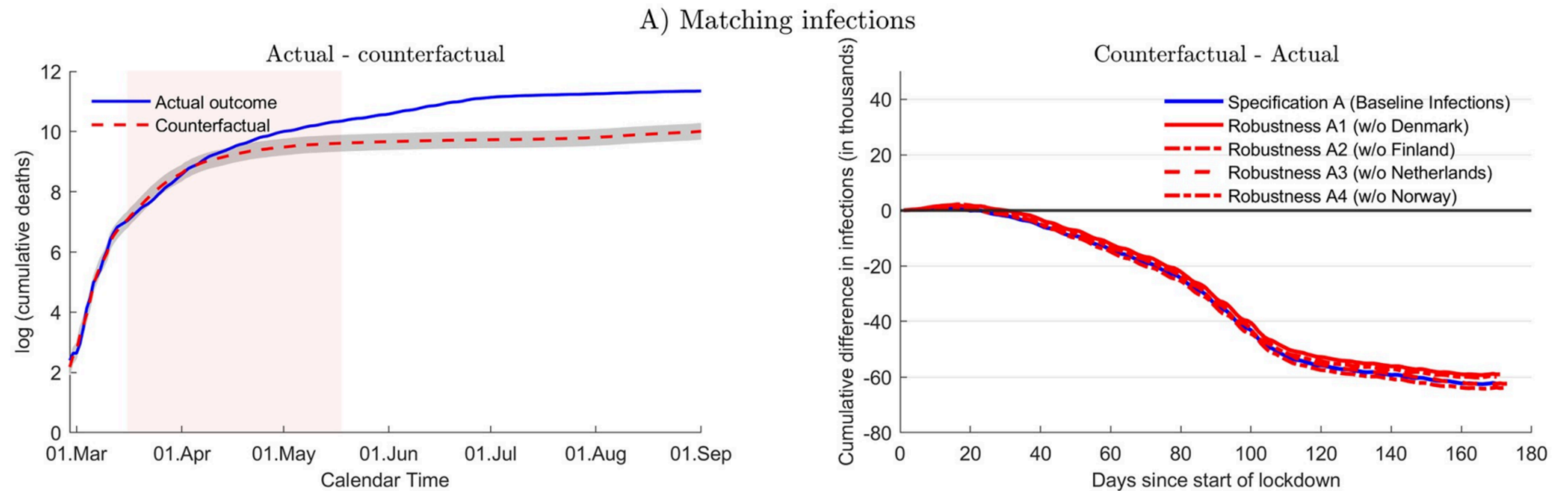


FIGURE 4. PATH PLOT OF PER CAPITA CO<sub>2</sub> EMISSIONS FROM TRANSPORT DURING 1960–2005:  
SWEDEN VERSUS SYNTHETIC SWEDEN

o Répliqué en Finlande.

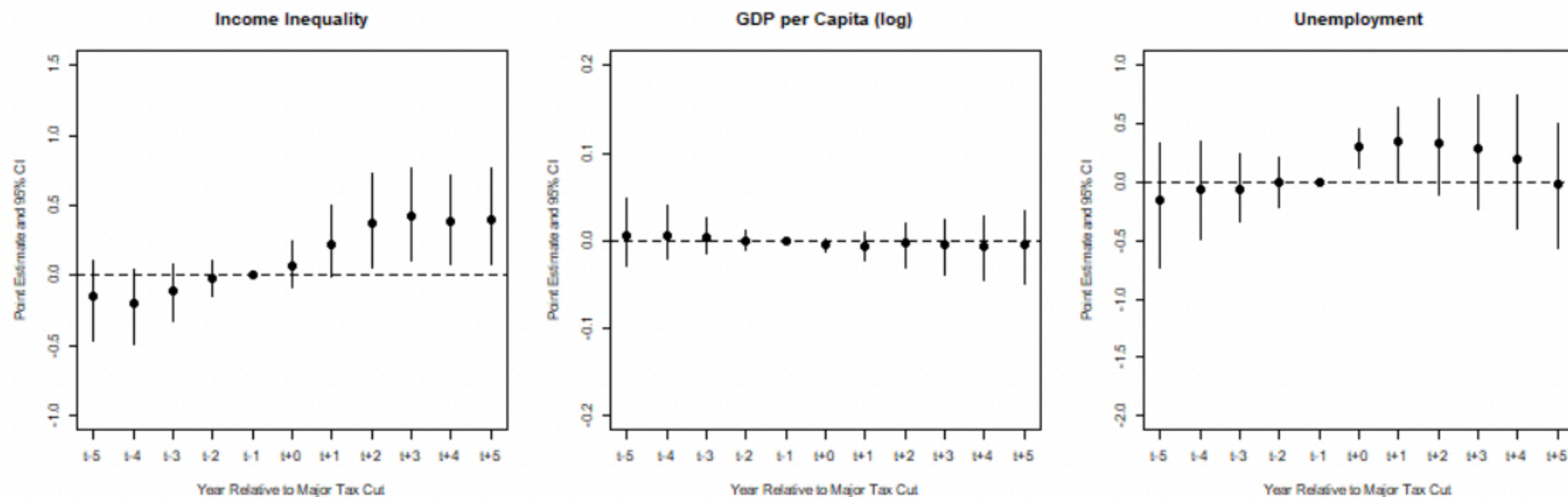




Born et al. (2021)



- DiD avec matching** : étude sur la théorie du ruissellement. On suit plusieurs pays qui ont implémenté des baisses de taxe sur les plus riches à des périodes différentes.
- A chaque fois qu'un pays baisse une taxe, un algorithme fouille la base de données pour trouver un autre pays mais avec une tendance comparable avant la baisse de taxe.



Source: Authors' calculations.

## 1. Néocolonialisme : les pays en développement, cobayes des ECRs ?

- Très courant en économie du développement. JPAL a mené beaucoup d'expériences dans les pays en développement.
- Une des études les plus influentes sur l'impact des masques (Abaluck et al., 2021) a été menée ... au Bangladesh.

Solution: rendre plus fréquent l'évaluation des politiques publiques par dans les pays développés.

## 2. Plus facile pour des politiques moins ambitieuses.

Pas de RCT pour sortir de l'Euro.

Nuance :

- Surtout l'évaluation par RCT qui est parfois difficile, ...
- Une fois que la politique est en place expériences naturelles et méthodes quasi-expérimentales (contrôle synthétique pour l'entrée de l'euro).

Pas toujours => monnaies locales.

## On ne peut pas savoir si une politique dans un contexte A va produire les mêmes effets dans un contexte B

- Ce n'est pas une question métaphysique. On peut répliquer dans un nouveau contexte qui nous intéresse et comparer les résultats avec celui d'avant.
  - Parfois, cela donne des résultats cohérents. Par exemple, programmes type internats d'excellence VS *charter schools* aux É-U.
  - Parfois non. Par exemple, le dédoublement des classes en REP+ : effet en mathématiques était proche de ce qu'on attendait, pas les résultats en français.
- Il faut toutefois suivre un certain nombre de précautions pour s'assurer que ce qui a marché à petite échelle va marcher à plus grande échelle.





## 1. Répliquer 3-4 fois les essais !

- Même si un traitement n'a pas d'effet, on pourrait trouver par hasard un écart entre le groupe traitement, et le groupe contrôle.
- Pour départager hasard et effet, on considère alors comme significatif un écart entre le groupe traitement et le groupe contrôle qu'on aurait eu qu'une probabilité de 5% d'observer si le traitement n'avait pas d'effet.
- Cela signifie que sur 100 expériences de méthodes dont on sait qu'elles n'ont aucun effet, 5 vont montrer un résultat significatif (faux positifs) !
- Solution : répliquer, par exemple 3-4 fois l'essai (John List). C'est cher, mais moins que de se tromper.

### Exemple :

- DARE : programme de lutte contre la drogue aux Etats-Unis, 10 millions de dollar par an à son apogée en 2002, déployé sur 75% du territoire Etats-unien. Faux positif de l'expérience pilote à Honulu.



## 2. Assurez-vous que la *population* de l'essai est représentative

- L'essai peut avoir des effets différents selon les populations.

Ex: Duflo (2014) et le sel enrichi en fer pour les carences en fer qui ne fonctionne que sur les jeunes adolescentes.

- Plus largement, il faut être très vigilant avec ce qu'on appelle la sélection dans l'essai : ceux qui anticipent que le traitement vont marcher s'inscrivent aux essais (pas généralisable au reste de la population).

### 3. Assurez-vous que la *situation* est représentative

- Lorsque vous appliquez une intervention testée dans un contexte A dans un nouveau contexte B, il n'est pas nécessaire qu'elle soit exactement similaire, mais il faut préserver des ingrédients non-négociables.
- Ex: JL a créé une école maternelle dans la banlieue de Chicago pour tester les théories en économie sur le rôle des compétences non-cognitives, le fait de payer les parents... Superbe résultats. Mais pas en Angleterre (même en intervention).
- Le fait de payer les parents était un ingrédient non-négociable !

## 4. On n'échappe jamais à la théorie : prêtez attention aux externalités et aux effets d'équilibre général.

- **Externalités** : même quand les groupes sont tirés au sort, le traitement peut affecter le groupe contrôle ce qui biaise les résultats de l'expérience.

Exemple: Dans les pays en développement, les bénéfices du déparasitage (contre des vers intestinaux comme les ankylostomes) ont été sous-estimés car ceux qui menaient des essais ne tenaient pas en compte l'effet de l'**immunité collective**.

- **Solution aux externalités** : Prêtez attention au design de l'expérience. Par exemple, ne pas allouer un traitement à l'échelle d'une école, d'une ville quand risque de débordement (d'externalités) ... Rarement une bonne idée à l'échelle individuelle. Ex. Hôpital Copenhague + Facebook.

## 4. On n'échappe jamais à la théorie : prêtez attention aux externalités et aux effets d'équilibre général.

- o **Effet d'équilibre général** : l'effet estimé à petite échelle peut ne pas être le même que ce que celui qu'on mesure à plus large échelle.

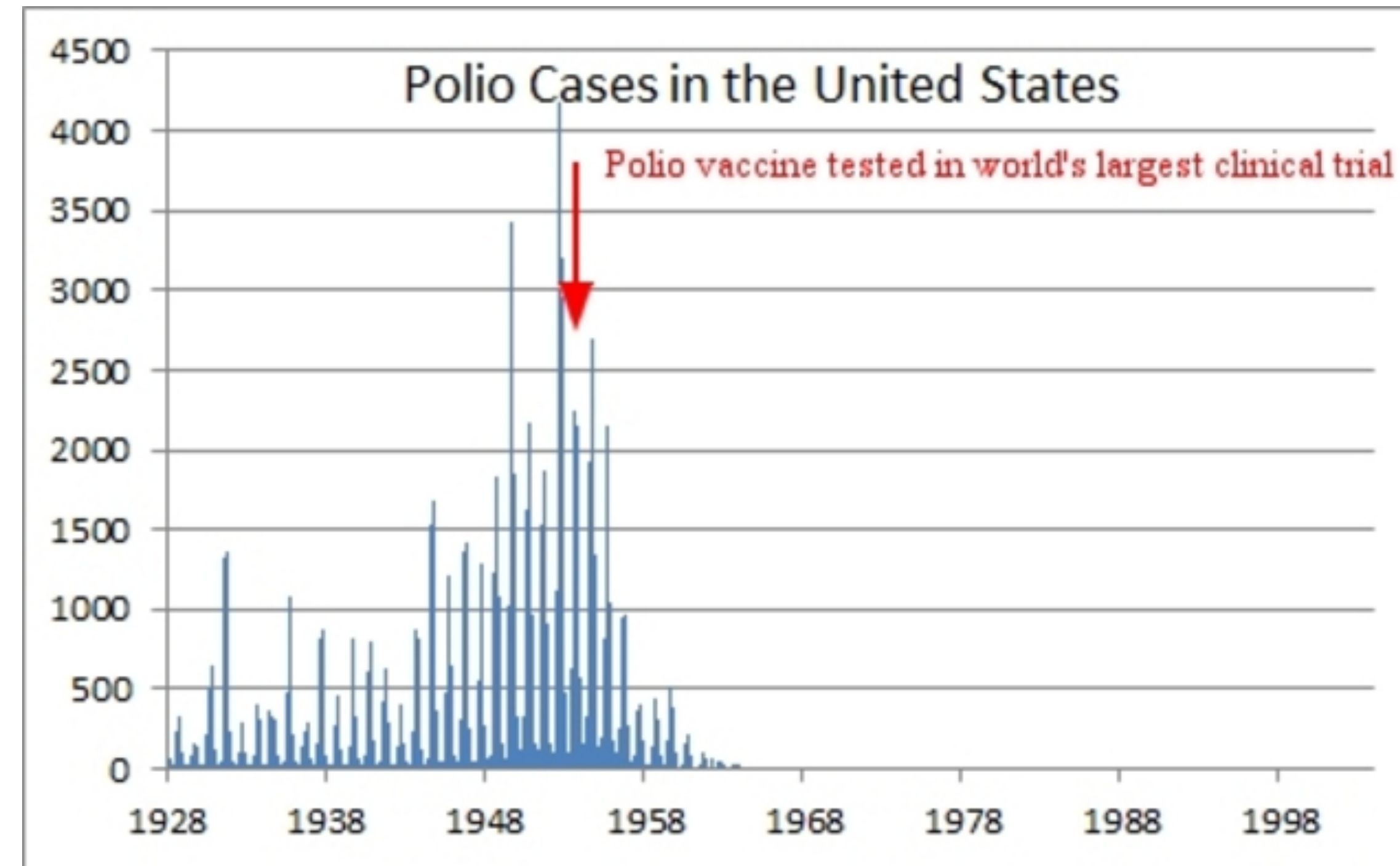
Exemple: Chez Uber, JL a introduit le système des pourboires pour faire revenir les chauffeurs qui quittaient l'application à petite échelle. A petite échelle, les pourboires augmentaient les revenus. A grande échelle, les nouveaux chauffeurs se sont faits concurrence, ce qui a ramené la rémunération à son niveau initial.

- o **Solution à l'effet d'EG** : Utiliser l'aléa pour quantifier les effets d'équilibre général : si le traitement est alloué au hasard dans plusieurs zones, le nombre de personnes traitées dans une zone donnée est aussi aléatoire. On peut ainsi savoir comment le nombre de personnes traitées autour de soi affecte l'effet du traitement. On peut aussi faire plusieurs bras avec plusieurs intensités.

## 5. Gare à la structure des coûts !

- La structure des coûts : comment le coût évolue à mesure qu'on produit plus/qu'on déploie plus.
- Important aussi dans le design des politiques publiques. Ex: John List (2022) et baisse de taille de classes en Californie.
- Débat entre l'usage de l'évaluation de politiques publiques pour tester une théorie (Levitt), ou pour tester l'efficacité de la politique (List):
  - Test de théorie : tester dans les meilleures conditions, sinon, si échec on ne sait pas si la théorie est fausse, ou les conditions étaient trop mauvaises.
  - Test d'efficacité : expérience plus proche de la politique finale possible (Ex: JL professeurs au salaire moyen à Chicago dans son école).





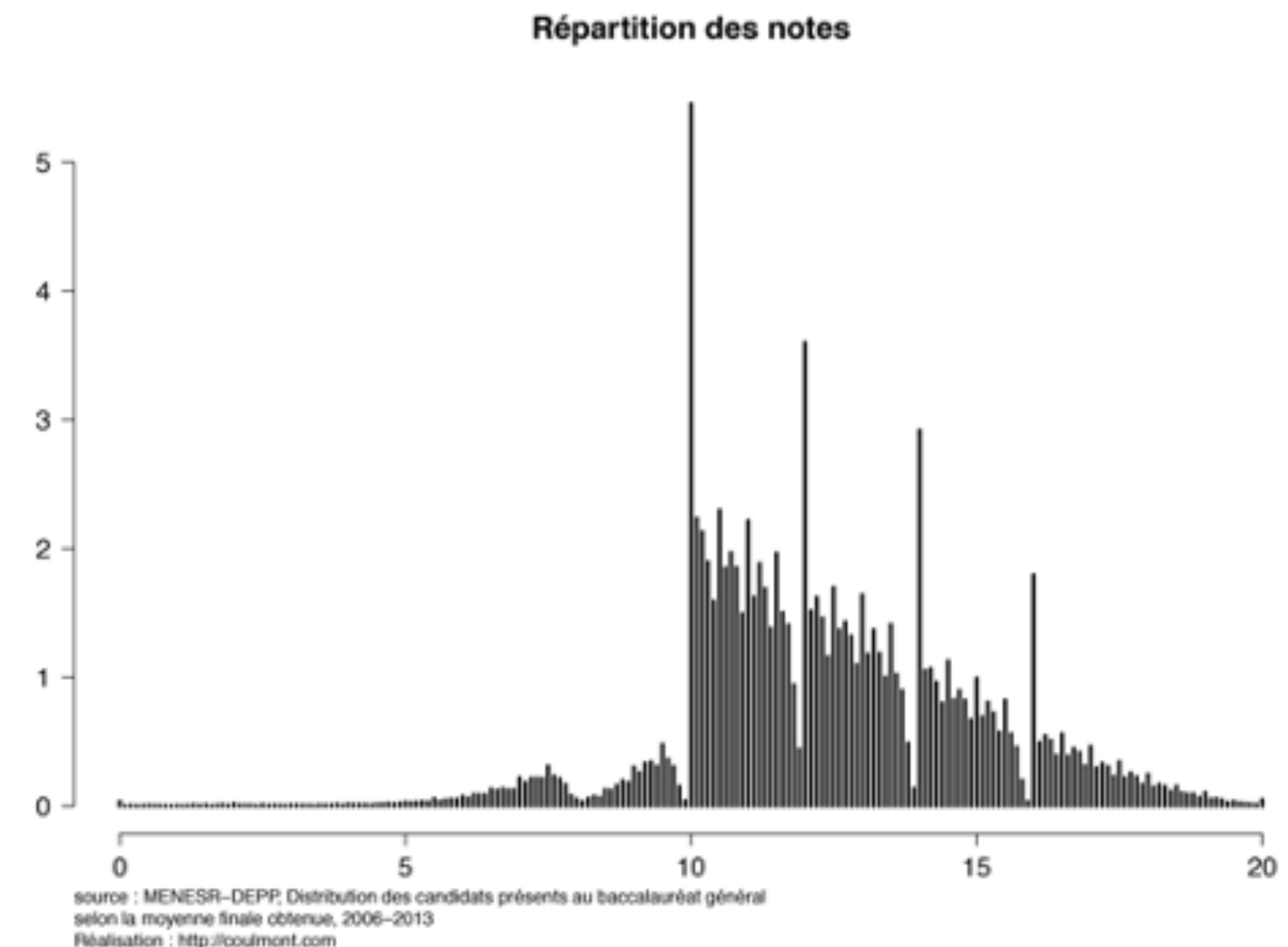
- o Pour John List, la vaccination contre la polio est l'exemple parfait d'une politique qui avait tout pour se déployer avec une grande efficacité : (1) les tests sur une population représentative d'enfants montraient qu'il était efficace sur tous ; (2) le vaccin était peu onéreux à produire, notamment grâce à des économies d'échelle ; (3) les seules externalités, celles générées par l'immunité collective, étaient positives.

- Lorsqu'on déploie une politique (un traitement), ce qui compte n'est pas ce qui s'est passé (combien de personnes ont reçu de l'argent, trouvé un emploi...) mais la différence entre ce qui s'est passé et le **contrefactuel** ... ce qui se serait passé dans un monde exactement similaire mais sans cette politique.
- On peut répondre à cette question grâce à des ECRs, ou même des quasi-expériences. Ces méthodes sont nécessaires du fait de la **sélection par les inobservables**.
- Il faut toutefois prendre certaines précautions pour s'assurer que ce qui a marché à petite échelle va marcher à plus grande échelle : (1) **répliquer** ; (2) tester sur une population **représentative** ; (3) identifier les **ingrédients essentiels** ; s'interroger sur ... (4) est-ce que le traitement peut affecter le contrôle ? ; (5) la structure des **coûts**.

**Merci !** Pour me contacter : [louis.freget@econ.ku.dk](mailto:louis.freget@econ.ku.dk)

- o Comment on sait-on que les quasi-expériences permettent de distinguer corrélation et causalité ?
    - a. Preuve mathématique que sous ses hypothèses, la méthode peut prouver la causalité.
    - b. Simulation : on simule des bases de données avec des effets qu'on connaît, on applique la méthode à cette base de données, et on vérifie que la méthode permet d'estimer correctement l'effet.
    - c. Comparer directement les résultats des ECRs aux quasi-expériences, et voir s'ils concordent.
- Ex: Card et al. (2018) méta-analysent des politiques actives d'emploi, ECRs et quasi-expérience ne diffèrent pas systématiquement dans leurs résultats.

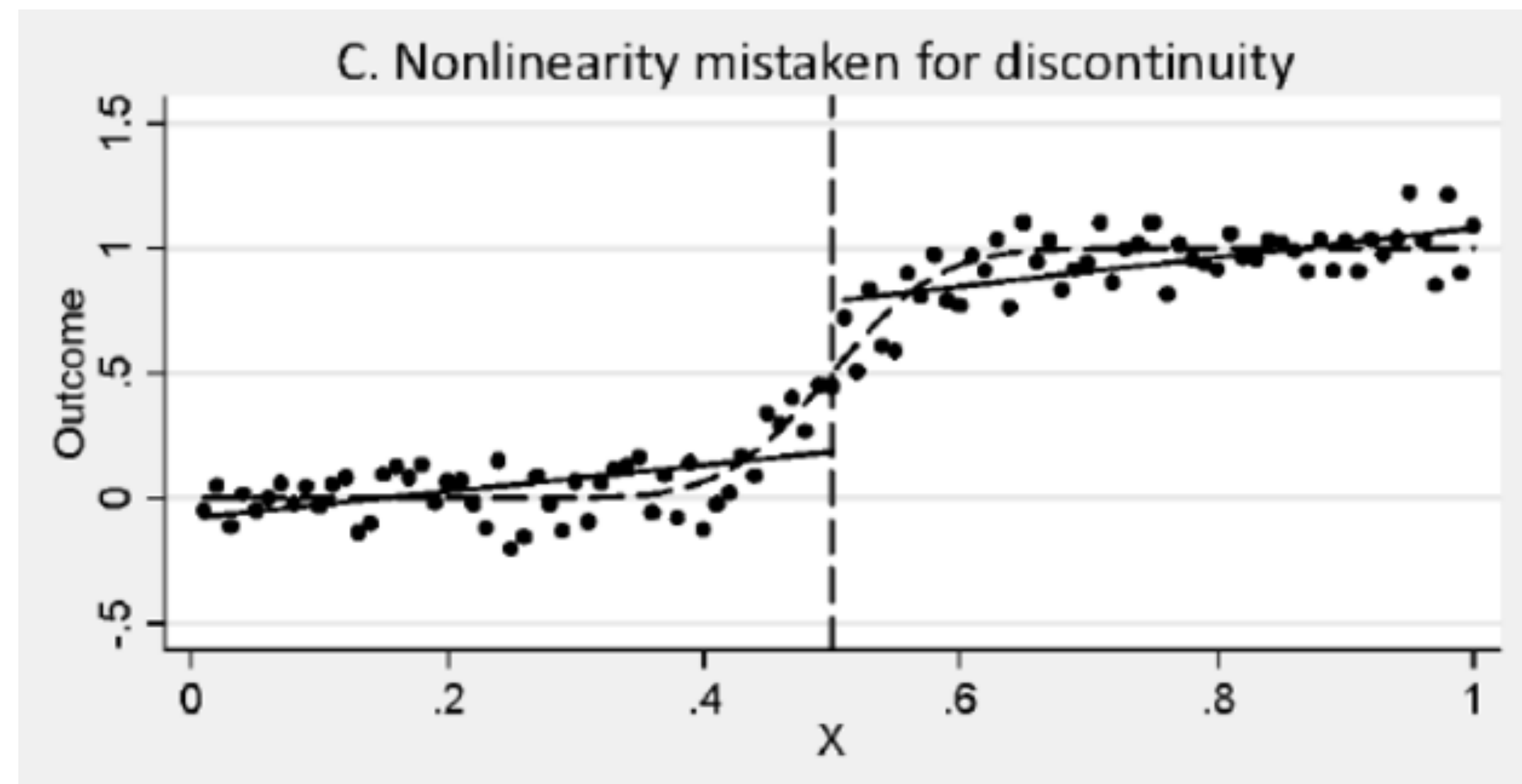
- Hypothèse (a) : si les bénéficiaires de l'intervention peuvent agir sur le fait d'être juste au dessus ou juste en dessous du seuil, ceux juste au dessus et juste en dessous de seuil risquent de ne plus être similaires.
- Ici (a) est naturellement plausible (importance de la connaissance du contexte). Mais on peut aussi tester l'absence de manipulation en vérifiant qu'il n'y a pas de *bunching* (beaucoup de gens au-dessus ou en dessous du seuil). Exemple de bunching :



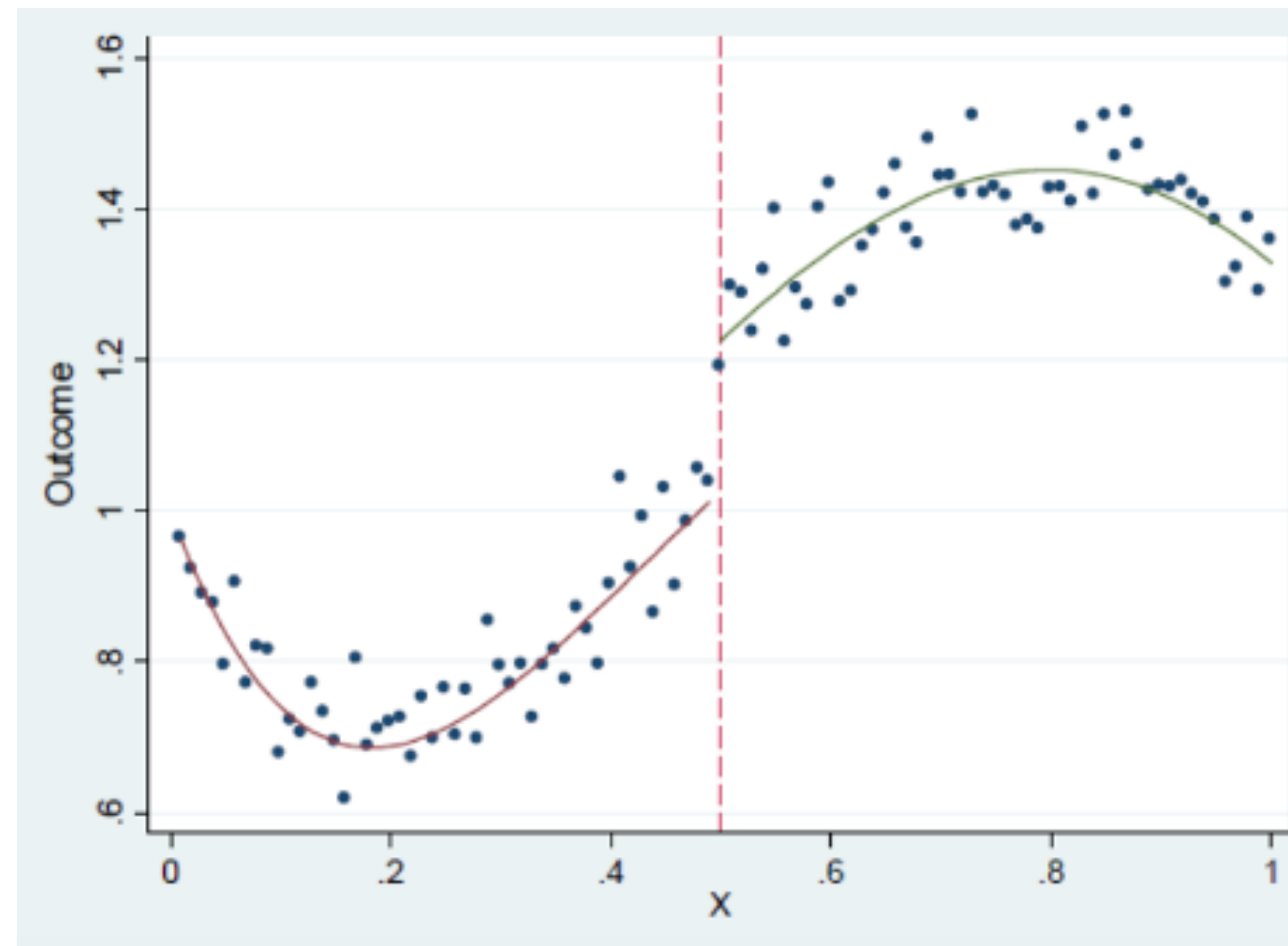
- Hypothèse (b) : Pas de différence systématique entre ceux juste en dessous du seuil, et juste au dessus du seuil.
  - Pour le vérifier, on peut faire la balance des observables. Vérifier que ceux juste en dessous et au dessus du seuil, ont bien le même niveau d'éducation, le même revenu, ... (comme dans un ECR).
  - (Annexe 3 pour d'autres tests).



- Hypothèse (b) : Pas de différence systématique entre ceux juste en dessous du seuil, et juste au dessus du seuil.
- Mais... est-ce que le saut autour du seuil n'est pas expliquée par une fonction mathématique plus compliquée ?



- Hypothèse (b) : Pas de différence systématique entre ceux juste en dessous du seuil, et juste au dessus du seuil.
- Pour le vérifier, on peut faire la balance des observables. Vérifier que ceux juste en dessous et au dessus du seuil, ont bien le même niveau d'éducation, le même revenu, ... (comme dans un ECR).
- Solution : estimer des fonctions mathématiques plus compliquées de chaque côté du seuil.



b. Beaucoup de paramètres libres :

- Bandwidth : est-ce que je compare des adhérents d'AF qui ont 34,9 à ceux qui 35,1 ans ? Ceux qui ont plutôt 33 à 36 ? Plus la bandwidth est grande, plus je peux détecter des petits effets.
- Forme fonctionnelle : quelle fonction mathématique exacte j'estime de chaque côté du seuil ?
- Kernel : est-ce que les observations plus proches du seuil comptent plus dans le calcul de la discontinuité ?

Solution : Algorithmes qui calculent les choix optimaux. Ceci permet surtout d'empêcher de changer des paramètres jusqu'à tant qu'il trouve quelque chose qu'il lui plaît. (Epistémologie du *shibari*).

b. Beaucoup de paramètres libres :

- Bandwidth : est-ce que je compare des adhérents d'AF qui ont 34,9 à ceux qui 35,1 ans ? Ceux qui ont plutôt 33 à 36 ? Plus la bandwidth est grande, plus je peux détecter des petits effets.
- Forme fonctionnelle : quelle fonction mathématique exacte j'estime de chaque côté du seuil ?
- Kernel : est-ce que les observations plus proches du seuil comptent plus dans le calcul de la discontinuité ?

Solution : Algorithmes qui calculent les choix optimaux. Ceci permet surtout d'empêcher de changer des paramètres jusqu'à tant qu'il trouve quelque chose qu'il lui plaît. (Epistémologie du *shibari*).