

# 効果測定の方法

伊藤成朗

アジア経済研究所

2022 年秋学期

国際交流学科, 聖心女子大学

# 祝: ノーベル経済学賞受賞3人(2019年)



# 祝: ノーベル経済学賞受賞3人(2019年)

Abhijit Banerjee インド人。理論経済学者。MIT 教授。

Esther Duflo フランス人。開発経済学者。MIT 教授。

Michael Kremer アメリカ人。理論経済学者。ハーバード大教授。

受賞理由:

「貧困緩和に実験的手法を導入した功績。貧困(という大きな)問題を小さな扱いや  
すい問題に分解し、実験を使って対策を示した。」

# 祝: ノーベル経済学賞受賞3人(2019年)

長所

# 祝: ノーベル経済学賞受賞3人(2019年)

長所

- ① 歪みなく効果を計測できる (**internal validity**)

# 祝: ノーベル経済学賞受賞3人(2019年)

長所

- ① 歪みなく効果を計測できる (**internal validity**)
- ② (結果に疑問を挟む余地は少ないので無駄な議論を節約できる)

# 祝: ノーベル経済学賞受賞3人(2019年)

長所

- ① 歪みなく効果を計測できる (*internal validity*)
- ② (結果に疑問を挟む余地は少ないので無駄な議論を節約できる)

- △ 被験者をランダムに治療群 *the treated* と統御群 *the control* に割り振り、前者にのみ介入
- △ Randomisation of treatment: 治療群と統御群は相似、異なるのは介入の有無だけ
- △ 結果指標の違いは介入が原因と解釈可能

# 祝: ノーベル経済学賞受賞3人(2019年)

## 長所

- ① 歪みなく効果を計測できる (**internal validity**)
- ② (結果に疑問を挟む余地は少ないので無駄な議論を節約できる)

△ 被験者をランダムに治療群 the treated と統御群 the control に割り振り、前者にのみ介入

△ Randomisation of treatment: 治療群と統御群は相似、異なるのは介入の有無だけ

△ 結果指標の違いは介入が原因と解釈可能

## 貢献

- ① 政策の根拠を推測から科学的証拠に変えた: evidence based policy making

# 祝: ノーベル経済学賞受賞3人(2019年)

## 長所

- ① 歪みなく効果を計測できる (**internal validity**)
- ② (結果に疑問を挟む余地は少ないので無駄な議論を節約できる)

△ 被験者をランダムに治療群 the treated と統御群 the control に割り振り、前者にのみ介入

△ Randomisation of treatment: 治療群と統御群は相似、異なるのは介入の有無だけ

△ 結果指標の違いは介入が原因と解釈可能

## 貢献

- ① 政策の根拠を推測から科学的証拠に変えた: evidence based policy making
- ② 研究者も証拠の質を議論するようになった

# 祝: ノーベル経済学賞受賞3人(2019年)

## 長所

- ① 歪みなく効果を計測できる (**internal validity**)
- ② (結果に疑問を挟む余地は少ないので無駄な議論を節約できる)

△ 被験者をランダムに治療群 the treated と統御群 the control に割り振り、前者にのみ介入

△ Randomisation of treatment: 治療群と統御群は相似、異なるのは介入の有無だけ

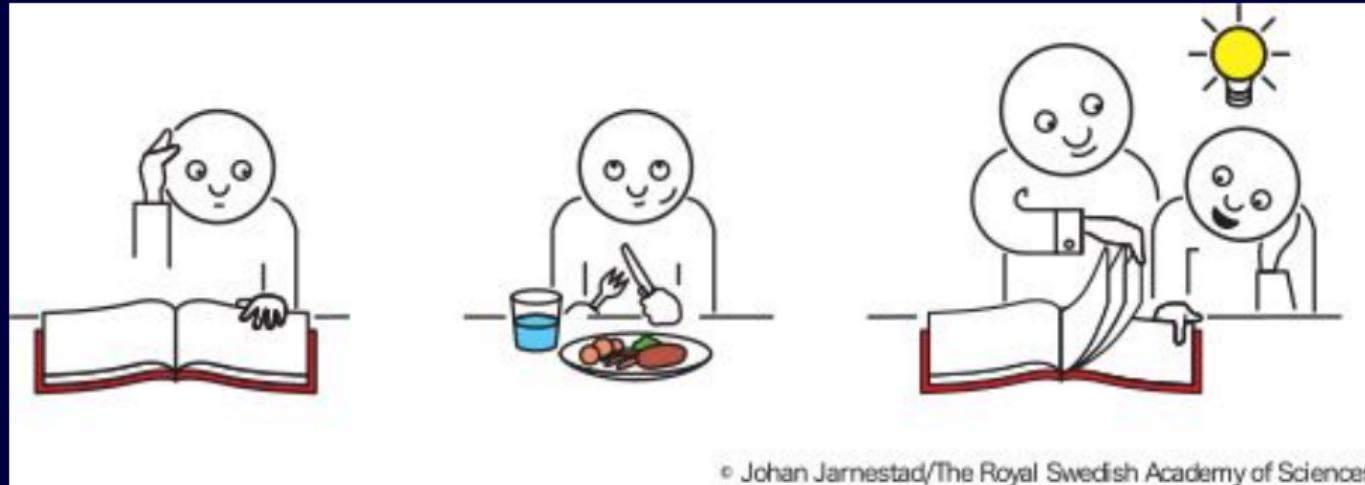
△ 結果指標の違いは介入が原因と解釈可能

## 貢献

- ① 政策の根拠を推測から科学的証拠に変えた: evidence based policy making
- ② 研究者も証拠の質を議論するようになった
- ③ 実験可能な事象・研究が提示すべき証拠の質の基準を上げた → 転じて観察データを使う研究の対象を明確化

# 祝: ノーベル経済学賞受賞3人(2019年)

例



© Johan Jarnestad/The Royal Swedish Academy of Sciences

教科書無料配布よりも(ケニア西部)、無料学校給食よりも(ケニア西部)、習熟の遅い生徒に補習させる方が試験点数を上げることが分かった(インド、ムンバイ近郊)

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?

# 祝: ノーベル経済学賞受賞3人(2019年)

## 短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?
  - 実施主体: NGOは能力もモラルも高く、政策担当者と比較にならない

# 祝: ノーベル経済学賞受賞3人(2019年)

## 短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?
  - 実施主体: NGOは能力もモラルも高く、政策担当者と比較にならない
  - (該当する理論が存在しないときに先入観無しにできることが良いときもある)

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?
  - 実施主体: NGOは能力もモラルも高く、政策担当者と比較にならない
  - (該当する理論が存在しないときに先入観無しにできることが良いときもある)
- ② 大きな政策を扱えない。大規模実験(e.g., ジャムナ橋建設)は統御群をなくす。

# 祝: ノーベル経済学賞受賞3人(2019年)

## 短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?
  - 実施主体: NGOは能力もモラルも高く、政策担当者と比較にならない
  - (該当する理論が存在しないときに先入観無しにできることが良いときもある)
- ② 大きな政策を扱えない。大規模実験(e.g., ジャムナ橋建設)は統御群をなくす。
- ③ 標本サイズが小さい(←予算がかかるから)ので推計値の精度が低い。参加率が低いと分析に使える標本がさらに減る。マイクロファイナンス実験。

# 祝: ノーベル経済学賞受賞3人(2019年)

## 短所

- ① メカニズム・理由(なぜ効果があったか)と無関係に実施可能。このため、理論を意識しなくても実験が可能。メカニズムが不明なので、その他地域への適用可能性(**external validity**)が不明。
  - 地域: インドで効果1ならガボンではどのくらいの効果?
  - 実施主体: NGOは能力もモラルも高く、政策担当者と比較にならない
  - (該当する理論が存在しないときに先入観無しにできることが良いときもある)
- ② 大きな政策を扱えない。大規模実験(e.g., ジャムナ橋建設)は統御群をなくす。
- ③ 標本サイズが小さい(←予算がかかるから)ので推計値の精度が低い。参加率が低いと分析に使える標本がさらに減る。マイクロファイナンス実験。
- ④ 実験バイアス: Hawthorne effect (treated), John Henry effect (control; raced against machine)

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)?

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)
    - ☞ かと思われたが、やはり、批判されている(?, “Should scholars be allowed to start a riot to see how violence spreads?”)。350 香港ドル=45 ドルくらい。

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)
    - ☞ かと思われたが、やはり、批判されている(?, “Should scholars be allowed to start a riot to see how violence spreads?”)。350 香港ドル=45 ドルくらい。
    - ☞ 筆者たちの主張(オンライン付論):

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)?
    - ☞ かと思われたが、やはり、批判されている(?, “Should scholars be allowed to start a riot to see how violence spreads?”)。350 香港ドル=45 ドルくらい。
    - ☞ 筆者たちの主張(オンライン付論):
      - 4大学(Munich, Stanford, UC Berkeley, HKUST)のIRB承認を得ている

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)
    - ☞ かと思われたが、やはり、批判されている(?, “Should scholars be allowed to start a riot to see how violence spreads?”)。350 香港ドル=45 ドルくらい。
    - ☞ 筆者たちの主張(オンライン付論):
      - 4大学(Munich, Stanford, UC Berkeley, HKUST)のIRB承認を得ている
      - リスクは小さい(10/15回で逮捕者ゼロ、2003年からのべ135万人が参加なのでデモ参加は日常から乖離せず、実験当時の2017-2018に言論の自由は保障されていた、軍による鎮圧可能性は小さい)

# 祝: ノーベル経済学賞受賞3人(2019年)

短所

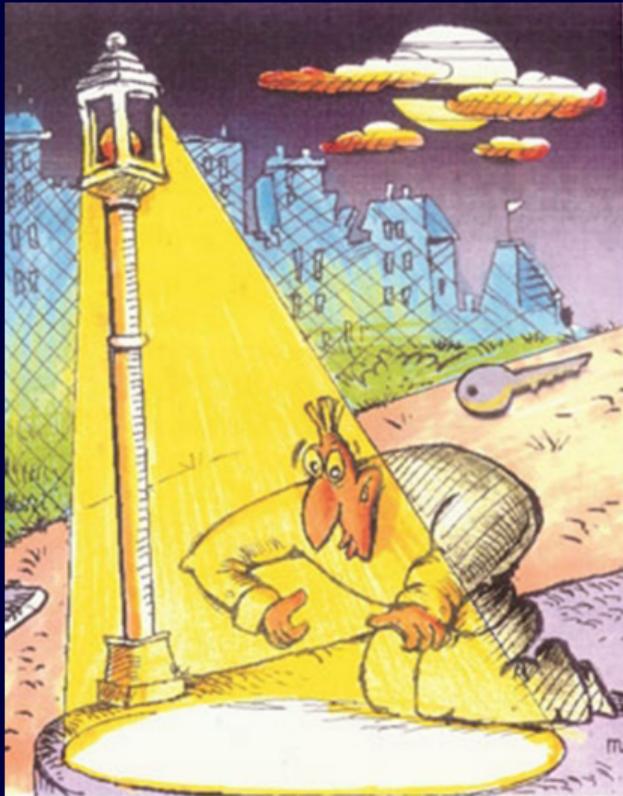
- ⑤ 検討手段が実験可能なものに集中: 薬、職業訓練、教材、補助教員、肥料、携帯
  - ランダム化しやすい: 親の学歴や年齢、家族構成は無理
  - 小さい(分割可能で被験者に割当可能): 橋や為替レートは無理
  - 倫理的に許要できる: 母乳育児、違法行為(贈賄: インド)推奨は駄目、政治デモ参加推奨(して参加人数計測)は文脈による(?)
    - ☞ かと思われたが、やはり、批判されている(?, “Should scholars be allowed to start a riot to see how violence spreads?”)。350 香港ドル=45 ドルくらい。
    - ☞ 筆者たちの主張(オンライン付論):
      - 4大学(Munich, Stanford, UC Berkeley, HKUST)のIRB承認を得ている
      - リスクは小さい(10/15回で逮捕者ゼロ、2003年からのべ135万人が参加なのでデモ参加は日常から乖離せず、実験当時の2017-2018に言論の自由は保障されていた、軍による鎮圧可能性は小さい)←先読み感ゼロ、想像力が... 当局が写真撮るかもよ?

# 祝: ノーベル経済学賞受賞3人(2019年)

鍵を捜す男

# 祝: ノーベル経済学賞受賞3人(2019年)

鍵を捜す男



出所: [https://diabetes.  
diabetesjournals.org/  
content/64/4/1105](https://diabetes.diabetesjournals.org/content/64/4/1105)

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

実験がうまくいき、効果があると分かっても、政策に採用されるかは別問題

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

実験がうまくいき、効果があると分かっても、政策に採用されるかは別問題

政治家および投票基盤が政策実施=得策と思わねばならないから

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

実験がうまくいき、効果があると分かっても、政策に採用されるかは別問題

政治家および投票基盤が政策実施=得策と思わねばならないから

- 個別補習の費用対効果が最も大きい: 学校教育の不十分さを示す結果。実験は学校教育の質の低さへの対症療法を示したが、根治療法を示していない。根治療法は教員・公務員組合等の反対で政治的に困難だろう。

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

実験がうまくいき、効果があると分かっても、政策に採用されるかは別問題

政治家および投票基盤が政策実施=得策と思わねばならないから

- 個別補習の費用対効果が最も大きい: 学校教育の不十分さを示す結果。実験は学校教育の質の低さへの対症療法を示したが、根治療法を示していない。根治療法は教員・公務員組合等の反対で政治的に困難だろう。

実験は思いついた政策に効果があるか気軽に試せるが、必要となる作業監理と予算は多いために、本当に検討する価値のある政策を選ばないと資源の無駄

倫理上、母乳育児や母乳育児に金銭的誘因を与える「推奨」を実験できないが、母乳育児の非金銭的「推奨」(内容伝達)を実験しても倫理的に問題ない

ただし、推奨内容が比較対象の女性に伝わらないか統御は難しい

実験がうまくいき、効果があると分かっても、政策に採用されるかは別問題

政治家および投票基盤が政策実施=得策と思わねばならないから

- 個別補習の費用対効果が最も大きい: 学校教育の不十分さを示す結果。実験は学校教育の質の低さへの対症療法を示したが、根治療法を示していない。根治療法は教員・公務員組合等の反対で政治的に困難だろう。

実験は思いついた政策に効果があるか気軽に試せるが、必要となる作業監理と予算は多いために、本当に検討する価値のある政策を選ばないと資源の無駄

- 投与と反応を見る疫学研究ではなく、人々の意志決定と行動選択を含む経済学研究なので計測に費用がかかる

# HA! HA! HA! HA!

日本経済新聞 2016年(平成28年)2月4日



普段ほとんど笑わない  
お年寄りが「自身の健康状態が良くない」と感じる割合は、毎日よく笑うお年寄りに比べて1・5倍以上高いことが東京大や千葉大などの研究グループの調査で分かった。

これまでの研究では健康状態の自己評価が低い人は喜びがたりになる割合や死亡率が高い傾向があるという。研究グループは「よく笑う人は健康である可能性を示す結果だ」と分析している。

お年寄り対象 東大など調査  
笑いの効用 強調

笑う頻度	男性	女性
週に1~5回	1.25	1.33
週に1~3回	1.37	1.55
ほとんどない	1.54	1.78

研究チームは、全国の65歳以上の男女約2万人を対象に、2013年秋、笑いの頻度と健康状態の関係をアンケートで尋ね、分析した。

普段の日常生活で声を出して笑う頻度を尋ねたところ、「ほぼ毎日」と答えたのは、男性が38%、女性が49%。一方、「ほとんどない」と答えたのは男性が10%、女性が5%だった。

その上で、笑いの頻度に応じて自分が評価する健康状態を分析した。ほとんど笑わない人で現

在の健康状態について、「あまり良くない」と答えた人は、ほぼ毎日笑った。男性が1・78倍、女性が1・78倍だった。

研究チームなどの調査によると、笑いの頻度が少ない人は、年齢とともに笑わないと感じている人が多い。

差し押さえロールスロイス 公売

トム 知られ 一ネツ 国税 国税

Ito (IDE, Sacred Heart)

- 高齢者: 笑う頻度が減ると健康不調を報告する比率が増える

今回の調査にあたつた東大大学院の近藤尚己准教授(公衆衛生学)は「研究成果は『よく笑う人は健康である』という可能性を示している。メカニズムは今後さらに詳細に調べる必要があるが、お年寄りは健康のために、日ごろからよく笑うことを心がけてほしい」

# HA! HA! HA! HA!

この記事には問題があります。何でしょうか。

# HA! HA! HA! HA!

この記事には問題があります。何でしょうか。問題を理解するためには、インパクト評価のキー・タームを知る必要があります

HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  (“A causes B.”)

相關 correlation  $\text{corr}[A, B] \neq 0$  (“A is correlated with B.”)

HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  (“A causes B.”)

相關 correlation  $\text{corr}[A, B] \neq 0$  (“A is correlated with B.”)

HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る



HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る



HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る



HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る

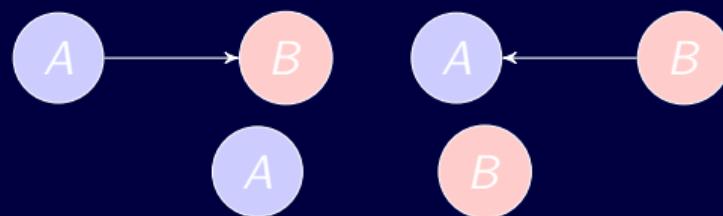


HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る

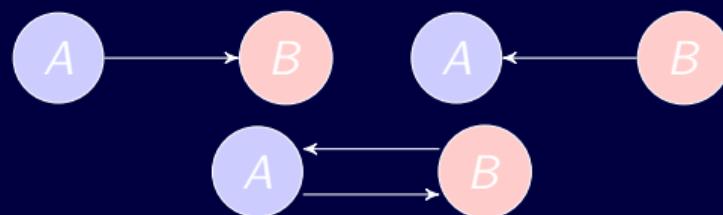


HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る

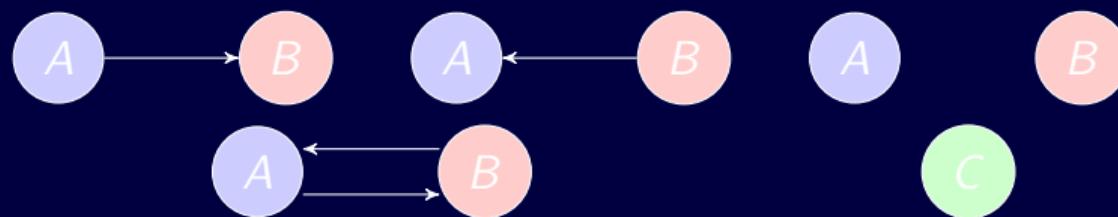


HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る

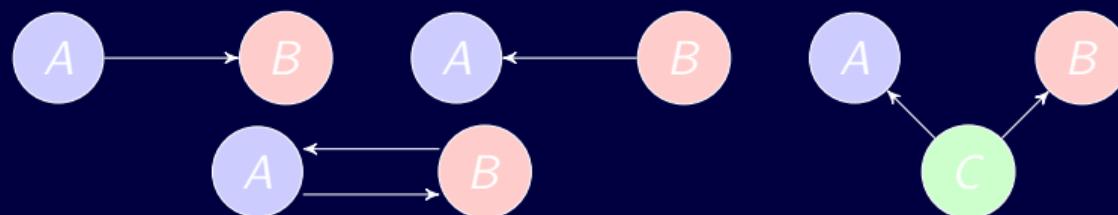


HA! HA! HA! HA!

因果 causality  $A \Rightarrow B$  ("A causes B.")

相関 correlation  $\text{corr}[A, B] \neq 0$  ("A is correlated with B.")

$A$  と  $B$  が相関: さまざまな因果関係があり得る



見せかけの相関

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

この研究のデザインは因果関係を明らかにできるか?

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

この研究のデザインは因果関係を明らかにできるか?

- できない。

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

この研究のデザインは因果関係を明らかにできるか?

- できない。(なぜかはすぐに分かります)

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

この研究のデザインは因果関係を明らかにできるか?

- できない。(なぜかはすぐに分かります)

では、何でもいいので何らかの因果関係を示すことはできるか。

# HA! HA! HA! HA!

笑いは健康(自己申告)を引き起こすか?

- 可能性はある
- もしくは、健康が笑いを引き起こす
- もしくは、その両方

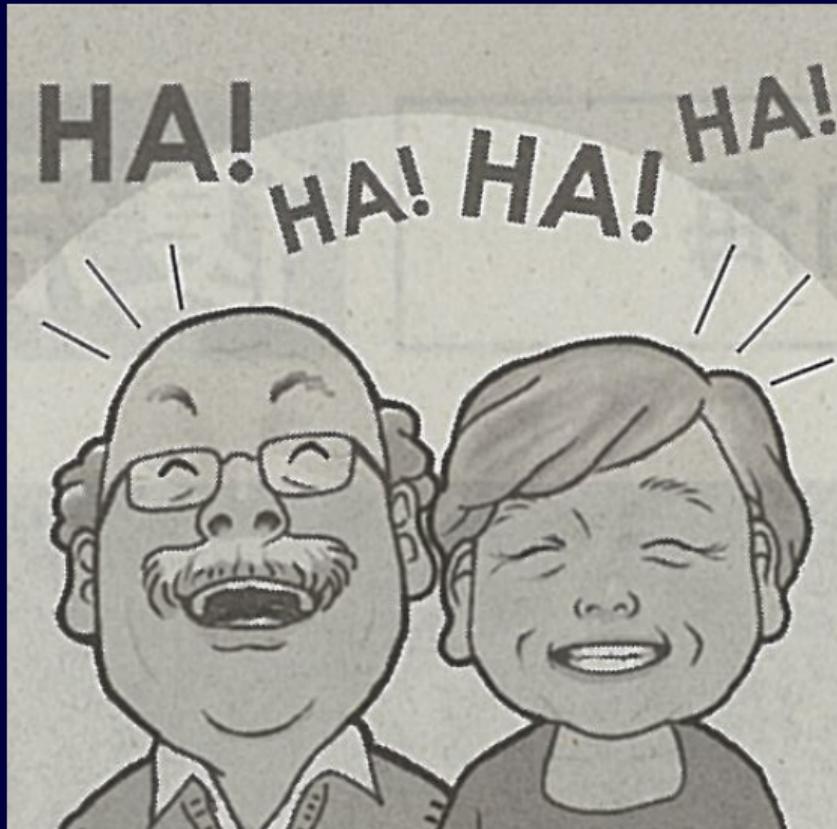
この研究のデザインは因果関係を明らかにできるか?

- できない。(なぜかはすぐに分かります)

では、何でもいいので何らかの因果関係を示すことはできるか。

- できない

HA! HA! HA! HA!



因果関係と相關関係は同じではない

因果関係と相關関係は同じではない

- 因果関係 ⇒ 相關関係

因果関係と相關関係は同じではない

- 因果関係 ⇒ 相關関係
- 相關関係 ⇒ 因果関係、ではない場合がある

## 因果関係と相關関係は同じではない

- 因果関係 ⇒ 相關関係
- 相關関係 ⇒ 因果関係、ではない場合がある

$$\text{試験点数} = 20 + 20 * D + e$$

$$D = \begin{cases} 0 & \text{if 塾に週 1 時間以上} \\ 1 & \begin{cases} \text{通わない} \\ \text{通う} \end{cases} \end{cases}$$

$D$  はダミー変数 dummy variable と呼ばれる 0 と 1 の 2 つの値をとる離散変数。2 つの値しかとらないので 2 項変数 binary variable とも呼ばれる。ここでは塾に週 1 時間以上通うと 1、そうではない場合は 0 という値をとる変数。塾に通う人と通わない人にグループ分けできる。 $e$  は誤差を表す確率変数で誤差項 error term という。

## 因果関係と相關関係は同じではない

- 因果関係 ⇒ 相關関係
- 相關関係 ⇒ 因果関係、ではない場合がある

$$\text{試験点数} = 20 + 20 * D + e$$

$$D = \begin{cases} 0 & \text{if 塾に週 1 時間以上} \\ 1 & \begin{cases} \text{通わない} \\ \text{通う} \end{cases} \end{cases}$$

$D$  はダミー変数 dummy variable と呼ばれる 0 と 1 の 2 つの値をとる離散変数。2 つの値しかとらないので 2 項変数 binary variable とも呼ばれる。ここでは塾に週 1 時間以上通うと 1、そうではない場合は 0 という値をとる変数。塾に通う人と通わない人にグループ分けできる。 $e$  は誤差を表す確率変数で誤差項 error term という。

ダミー変数は質的情報を表現できる:

## 因果関係と相関関係は同じではない

- 因果関係 ⇒ 相関関係
- 相関関係 ⇒ 因果関係、ではない場合がある

$$\text{試験点数} = 20 + 20 * D + e$$

$$D = \begin{cases} 0 & \text{if 塾に週 1 時間以上} \\ 1 & \begin{cases} \text{通わない} \\ \text{通う} \end{cases} \end{cases}$$

$D$  はダミー変数 dummy variable と呼ばれる 0 と 1 の 2 つの値をとる離散変数。2 つの値しかとらないので 2 項変数 binary variable とも呼ばれる。ここでは塾に週 1 時間以上通うと 1、そうではない場合は 0 という値をとる変数。塾に通う人と通わない人にグループ分けできる。 $e$  は誤差を表す確率変数で誤差項 error term という。

ダミー変数は質的情報を表現できる:

背景に連続変数がある 程度によるグループ分け。明るいと明るくない、早いと早くない、貧しいと貧しくない。

## 因果関係と相關関係は同じではない

- 因果関係 ⇒ 相關関係
- 相關関係 ⇒ 因果関係、ではない場合がある

$$\text{試験点数} = 20 + 20 * D + e$$

$$D = \begin{cases} 0 & \text{if 塾に週 1 時間以上} \\ 1 & \begin{cases} \text{通わない} \\ \text{通う} \end{cases} \end{cases}$$

$D$  はダミー変数 dummy variable と呼ばれる 0 と 1 の 2 つの値をとる離散変数。2 つの値しかとらないので 2 項変数 binary variable とも呼ばれる。ここでは塾に週 1 時間以上通うと 1、そうではない場合は 0 という値をとる変数。塾に通う人と通わない人にグループ分けできる。 $e$  は誤差を表す確率変数で誤差項 error term という。

ダミー変数は質的情報を表現できる:

背景に連続変数がある 程度によるグループ分け。明るいと明るくない、早いと早くない、貧しいと貧しくない。

背景に連続変数がない 分類によるグループ分け。右利きと左利き、テレビとそれ以外の家電、男とそれ以外のジェンダー、日本人と外国人。

$$\text{試験点数} = 20 + 20 * D + e$$

$$\text{試験点数} = 20 + 20 * D + e$$

これが因果関係の場合: 塾に通う  $\Rightarrow$  点数が 40 点になる、という解釈になる。

$$\text{試験点数} = 20 + 20 * D + e$$

これが因果関係の場合: 塾に通う  $\Rightarrow$  点数が 40 点になる、という解釈になる。



でも、方程式は相関関係を表す場合もある。

- 仮に、塾に試験点数を上げる効果は全く無いときに、勉強好きで試験点数がもともと 20 点程度良い人が勉強の機会を増やすために塾に行っている場合にも、この方程式は成り立つ。この場合、勉強好き ⇒ 点数、勉強好き ⇒ 塾通いという因果関係はあっても、塾通い ⇒ 試験点数という因果関係はない。

でも、方程式は相関関係を表す場合もある。

- 仮に、塾に試験点数を上げる効果は全く無いときに、勉強好きで試験点数がもともと 20 点程度良い人が勉強の機会を増やすために塾に行っている場合にも、この方程式は成り立つ。この場合、勉強好き ⇒ 点数、勉強好き ⇒ 塾通いという因果関係はあっても、塾通い ⇒ 試験点数という因果関係はない。「勉強好き」という欠落変数が試験点数と塾通いに同時に影響を与えていて、試験点数と塾通いの間に因果的な関係はない。

でも、方程式は相関関係を表す場合もある。

- 仮に、塾に試験点数を上げる効果は全く無いときに、勉強好きで試験点数がもともと 20 点程度良い人が勉強の機会を増やすために塾に行っている場合にも、この方程式は成り立つ。この場合、勉強好き ⇒ 点数、勉強好き ⇒ 塾通いという因果関係はあっても、塾通い ⇒ 試験点数という因果関係はない。「勉強好き」という欠落変数が試験点数と塾通いに同時に影響を与えていて、試験点数と塾通いの間に因果的な関係はない。



でも、方程式は相関関係を表す場合もある。

- 仮に、塾に試験点数を上げる効果は全く無いときに、勉強好きで試験点数がもともと 20 点程度良い人が勉強の機会を増やすために塾に行っている場合にも、この方程式は成り立つ。この場合、勉強好き  $\Rightarrow$  点数、勉強好き  $\Rightarrow$  塾通りという因果関係はあっても、塾通り  $\Rightarrow$  試験点数という因果関係はない。「勉強好き」という欠落変数が試験点数と塾通りに同時に影響を与えていて、試験点数と塾通りの間に因果的な関係はない。



方程式は逆の因果関係を表す場合もある。

- 仮に、試験点数が 20 点ほど良い人だけ選んで塾に行くことを強制しても、この方程式は成り立つ。

方程式は逆の因果関係を表す場合もある。

- 仮に、試験点数が 20 点ほど良い人だけ選んで塾に行くことを強制しても、この方程式は成り立つ。この場合、点数 ⇒ 塾通いという逆の因果関係が成り立っている。

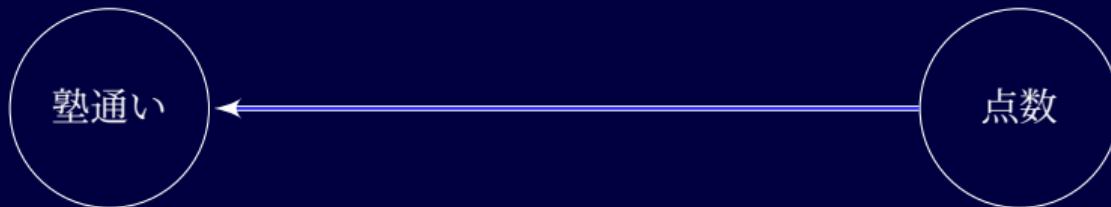
方程式は逆の因果関係を表す場合もある。

- 仮に、試験点数が 20 点ほど良い人だけ選んで塾に行くことを強制しても、この方程式は成り立つ。この場合、 $\text{点数} \Rightarrow \text{塾通い}$  という逆の因果関係が成り立っている。



方程式は逆の因果関係を表す場合もある。

- 仮に、試験点数が 20 点ほど良い人だけ選んで塾に行くことを強制しても、この方程式は成り立つ。この場合、点数 ⇒ 塾通りという逆の因果関係が成り立っている。



方程式は必ずしも右辺 ⇒ 左辺の因果関係ばかりではなく、逆方向の因果関係や欠落変数を通じた相関関係も含む。

因果関係を示すためには特定の条件が必要。その条件がない通常の回帰式の場合、  
相関関係までしか読み取ることができない。

因果関係を示すためには特定の条件が必要。その条件がない通常の回帰式の場合、  
相関関係までしか読み取ることができない。

予測だけなら相関関係で十分。塾に通う人は(なぜか分からないけど)試験点数が高い傾向がある、という関係だけで予測はできる。これは便利。

因果関係を示すためには特定の条件が必要。その条件がない**通常の回帰式の場合、相関関係までしか読み取ることができない。**

**予測だけなら相関関係で十分。**塾に通う人は(なぜか分からぬけど)試験点数が高い傾向がある、という関係だけで予測はできる。これは便利。

でも、**相関関係からは理由やメカニズム(因果関係の組み合わせ)は分からない。**よって、何かの事情でメカニズムが変わった場合、相関関係の強さも変わって、それまで通りの予測はできなくなる。

因果関係を示すためには特定の条件が必要。その条件がない**通常の回帰式の場合、相関関係までしか読み取ることができない。**

**予測だけなら相関関係で十分。**塾に通う人は(なぜか分からないけど)試験点数が高い傾向がある、という関係だけで予測はできる。これは便利。

でも、**相関関係からは理由やメカニズム(因果関係の組み合わせ)は分からない。**よって、何かの事情でメカニズムが変わった場合、相関関係の強さも変わって、それまで通りの予測はできなくなる。

相関関係に頼った予測はメカニズムを検討しないため、理論なき計測 measurement without a theory と揶揄されることもある

# 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

## 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る

## 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず

## 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく

## 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく
- 治療群にのみ塾に通わせる

# 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく
- 治療群にのみ塾に通わせる
  - ⇒ 治療の spillover を防ぐ: 統御群被験者が塾に通わないように、治療群被験者が塾に通うように、かつ、治療群被験者が統御群被験者に塾で学んだことを教えないように、被験者の行動を統御しなければいけない

# 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく
- 治療群にのみ塾に通わせる
  - ⇒ 治療の spillover を防ぐ: 統御群被験者が塾に通わないように、治療群被験者が塾に通うように、かつ、治療群被験者が統御群被験者に塾で学んだことを教えないように、被験者の行動を統御しなければいけない
  - ⇒ でも、被験者はやりたいことをやるので、そうした不完全な制御の政策の効果を測定していると解釈

# 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく
- 治療群にのみ塾に通わせる
  - ⇒ 治療の spillover を防ぐ: 統御群被験者が塾に通わないように、治療群被験者が塾に通うように、かつ、治療群被験者が統御群被験者に塾で学んだことを教えないように、被験者の行動を統御しなければいけない
  - ⇒ でも、被験者はやりたいことをやるので、そうした不完全な制御の政策の効果を測定していると解釈
- 後日、試験をして採点する

## 因果関係を示す方法: ランダム化統御試験 randomised controlled trial (RCT)

- 被験者・被験対象をランダムに「治療群 (the treated)」「統御群 (the control)」に割り振る
  - ⇒ ランダムに割り振ったので、両群の試験点数 (とその他変数) の分布 (の特徴である平均値) はほぼ同じのはず
  - ⇒ 標本が大きいほど誤差が減って平均値差はゼロに近づく
- 治療群にのみ塾に通わせる
  - ⇒ 治療の spillover を防ぐ: 統御群被験者が塾に通わないように、治療群被験者が塾に通うように、かつ、治療群被験者が統御群被験者に塾で学んだことを教えないように、被験者の行動を統御しなければいけない
  - ⇒ でも、被験者はやりたいことをやるので、そうした不完全な制御の政策の効果を測定していると解釈
- 後日、試験をして採点する
- 治療群の方が成績が高くなったら、塾に通う ⇒ 試験点数が高い、という因果関係を示すことができる

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

問い合わせるべきこと: 比較している対象は何か?

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

問い合わせるべきこと: 比較している対象は何か?

Before-after: 以前の自分

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

問い合わせるべきこと: 比較している対象は何か?

Before-after: 以前の自分

With-without: アクセスのない誰か

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

問い合わせるべきこと: 比較している対象は何か?

Before-after: 以前の自分

With-without: アクセスのない誰か

インパクトを知る上で適切な比較か?

プロジェクト評価の報告書にはさらっとこんな結論が散見される  
「プロジェクトによって健康状態が改善された」

問うべきこと: Counterfactual(比較すべき対象)は何か?

問い合わせるべきこと: 比較している対象は何か?

Before-after: 以前の自分

With-without: アクセスのない誰か

インパクトを知る上で適切な比較か?

殆どの場合、不適切

下記のようにインパクトを計測できれば、政策が結果指標  $y_i$  を変えたといえる

( $i$  が政策に影響されたときの  $y_i$ )

- ( $i$  が政策に影響されなかったときの  $y_i$ ).

下記のようにインパクトを計測できれば、政策が結果指標  $y_i$  を変えたといえる

$(i$  が政策に影響されたときの  $y_i$ )  
–( $i$  が政策に影響されなかったときの  $y_i$ ).

数学表記すると

$$(y_i|D_i = 1) - (y_i|D_i = 0), \quad \text{or},$$

$$y_{i1} - y_{i0}.$$

下記のようにインパクトを計測できれば、政策が結果指標  $y_i$  を変えたといえる

( $i$  が政策に影響されたときの  $y_i$ )  
– ( $i$  が政策に影響されなかったときの  $y_i$ ).

数学表記すると

$$(y_i|D_i = 1) - (y_i|D_i = 0), \quad \text{or},$$

$$y_{i1} - y_{i0}.$$

$D_i = 0, 1$   $i$  が治療群 (政策に影響されるグループ) に属するとき  $D_i = 1$ 、 $i$  が統御群 (政策に影響されないグループ) に属するとき  $D_i = 0$

“|” Reads “given” or “when”. 「次が所与のとき」「次が成り立つとき」と読む

$y_i|D_i = 1$   $i$  が治療群に属しているときの  $y_i$ 、 $y_{i1}$  とも書く

$y_i|D_i = 0$   $i$  が統御群に属しているときの  $y_i$ 、 $y_{i0}$  とも書く

個人  $i$  の治療効果 treatment effect:

$$(y_i|D_i = 1) - (y_i|D_i = 0) = y_{1i} - y_{0i}.$$

プログラム評価での根源的問題

個人  $i$  の治療効果 treatment effect:

$$(y_i|D_i = 1) - (y_i|D_i = 0) = y_{1i} - y_{0i}.$$

プログラム評価での根源的問題

治療群に属するときの  $y_i$  と統御群に属するときの  $y_i$  を同時に観察できない

個人  $i$  の治療効果 treatment effect:

$$(y_i|D_i = 1) - (y_i|D_i = 0) = y_{1i} - y_{0i}.$$

プログラム評価での根源的問題

治療群に属するときの  $y_i$  と統御群に属するときの  $y_i$  を同時に観察できない

各個人  $i$  の結果指標  $y_i$  の *counterfactual (CF)* を観測することはできない

個人  $i$  の治療効果 treatment effect:

$$(y_i|D_i = 1) - (y_i|D_i = 0) = y_{1i} - y_{0i}.$$

プログラム評価での根源的問題

治療群に属するときの  $y_i$  と統御群に属するときの  $y_i$  を同時に観察できない

各個人  $i$  の結果指標  $y_i$  の *counterfactual (CF)* を観測することはできない

言い換えれば、仮定なしには政策の効果を計算することはできない

個人  $i$  の治療効果 treatment effect:

$$(y_i|D_i = 1) - (y_i|D_i = 0) = y_{1i} - y_{0i}.$$

プログラム評価での根源的問題

治療群に属するときの  $y_i$  と統御群に属するときの  $y_i$  を同時に観察できない

各個人  $i$  の結果指標  $y_i$  の *counterfactual (CF)* を観測することはできない

言い換えれば、仮定なしには政策の効果を計算することはできない

$y_i|D_i = 1$  の CF 現実には治療群 ( $D_i = 1$ ) に属する  $i$  が統御群に属したときの結果指標  $y_i$  (" $y_{0i}$ ")。  $y_{0i}|D_i = 1$  と表記。

$y_i|D_i = 0$  の CF 現実には統御群 ( $D_i = 0$ ) に属する  $i$  が治療群に属したときの結果指標  $y_i$  (" $y_{1i}$ ")。  $y_{1i}|D_i = 0$  と表記。

一定の仮定の下、政策の平均治療効果 average treatment effect (ATE) を推計できる

$$ATE = \mathcal{E}[y_i | D_i = 1] - \mathcal{E}[y_i | D_i = 0].$$

一定の仮定の下、政策の平均治療効果 average treatment effect (ATE) を推計できる

$$ATE = \mathbb{E}[y_i | D_i = 1] - \mathbb{E}[y_i | D_i = 0].$$

以下の思考実験を考える

多数  $n$  の個人に対し、ランダムに治療状態  $D_i$  を割り当てる。ランダム化がうまくできたとする（公平なコインを使うなど）。

両グループの  $y_i$  の分布は似通うはず。分布が近似していれば平均値も近似する。極限を取って  $n \rightarrow \infty$  ( $n$  が無限大の場合)、両グループの分布は同一、平均値も同じになる。

ATE を (一致推計量 consistent estimator [標本が無限大になると真の値になる推計量]として) 得る条件

- ① 政策前に、統御群と治療群の ( $y_i$ ) 分布が近似していること。
- ② ( $n$  が十分に大きいこと。)
- ③ インパクトはすべての  $i$  で同じ。

ATE を (一致推計量 consistent estimator [標本が無限大になると真の値になる推計量]として) 得る条件

- ① 政策前に、統御群と治療群の ( $y_i$ ) 分布が近似していること。
- ② ( $n$  が十分に大きいこと。)
- ③ インパクトはすべての  $i$  で同じ。

2 は大数の法則。現実に  $n$  は無限大よりも小さいが、十分に大きければ標本平均が母集団平均に十分近くなる。

ATE を (一致推計量 consistent estimator [標本が無限大になると真の値になる推計量]として) 得る条件

- ① 政策前に、統御群と治療群の ( $y_i$ ) 分布が近似していること。
- ② ( $n$  が十分に大きいこと。)
- ③ インパクトはすべての  $i$  で同じ。

2 は大数の法則。現実に  $n$  は無限大よりも小さいが、十分に大きければ標本平均が母集団平均に十分近くなる。

3 は単純化のために利用。グループごとにインパクトが違うなら、グループをもっと細かく分ければいい。

ATE を (一致推計量 consistent estimator [標本が無限大になると真の値になる推計量]として) 得る条件

- ① 政策前に、統御群と治療群の ( $y_i$ ) 分布が近似していること。
- ② ( $n$  が十分に大きいこと。)
- ③ インパクトはすべての  $i$  で同じ。

2 は大数の法則。現実に  $n$  は無限大よりも小さいが、十分に大きければ標本平均が母集団平均に十分近くなる。

3 は単純化のために利用。グループごとにインパクトが違うなら、グループをもっと細かく分ければいい。

1 が最も重要。ランダムに割り振ることによって、各グループの特徴の分布が近似。

ATE を (一致推計量 consistent estimator [標本が無限大になると真の値になる推計量]として) 得る条件

- ① 政策前に、統御群と治療群の ( $y_i$ ) 分布が近似していること。
- ② ( $n$  が十分に大きいこと。)
- ③ インパクトはすべての  $i$  で同じ。

2 は大数の法則。現実に  $n$  は無限大よりも小さいが、十分に大きければ標本平均が母集団平均に十分近くなる。

3 は単純化のために利用。グループごとにインパクトが違うなら、グループをもっと細かく分ければいい。

1 が最も重要。ランダムに割り振ることによって、各グループの特徴の分布が近似。

グループの割り振りがランダムだと、ATE の一致推計量が得られる。

☞ ATE の一致推計量を得るために、治験は患者をプラセボ (統御群) と治療群にランダムに割り振る。

# 実験をしてからの手順

- ① ATE を推計 (estimation)

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

## ☞ $p$ 値 ( $p$ value) は帰無仮説が成り立つ確率と考えていい

- 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

## △ $p$ 値 ( $p$ value) は帰無仮説が成り立つ確率と考えていい

- 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

## ☞ $p$ 値 ( $p$ value) は帰無仮説が成り立つ確率と考えていい

- 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき → 帰無仮説が正しい確率は  $p$  値ほど小さい

# 実験をしてからの手順

## ① ATE を推計 (estimation)

- 両群の平均値の差を計算

## ② 推計値を検定する統計学的推論 (inference)

- 帰無仮説「両群の平均値の差はゼロ」が正しい場合の分布を使って平均値の差が極端かを計算、計算する極端さの指標は  $p$  値
- 極端であれば帰無仮説に疑問を呈し、 $p$  の確率で ATE はゼロではない、と推論

### ☞ $p$ 値 ( $p$ value) は帰無仮説が成り立つ確率と考えていい

- 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき → 帰無仮説が正しい確率は  $p$  値ほど小さい

### ☞ $p$ 値が 5% をカットオフにして、5%未満だったら「統計的に有意」、以上だったら「統計的に有意ではない」と表現されることが多いが、**推奨できない** ← 4.99% と 5.01% の差は無視可能なのに表現が違いすぎる

# 実験をしてからの手順

- ③ 推論の頑健性をチェック (robustness checks)

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

**Randomisation checks** 実験前の特徴が両群で似ていること=ランダム化が成功していることを検定、帰無仮説「実験前の(全)変数において、両群の平均値の差はゼロ」

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

**Randomisation checks** 実験前の特徴が両群で似ていること=ランダム化が成功していることを検定、帰無仮説「実験前の(全)変数において、両群の平均値の差はゼロ」

- 他要因排除の確認: 他要因による効果を検定。効果が見出されたら、メインの結論もその要因が引き起こしたかも。

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

**Randomisation checks** 実験前の特徴が両群で似ていること=ランダム化が成功していることを検定、帰無仮説「実験前の(全)変数において、両群の平均値の差はゼロ」

- 他要因排除の確認: 他要因による効果を検定。効果が見出されたら、メインの結論もその要因が引き起こしたかも。

**Placebo tests** 介入変数を編集して、効果が無いことを検定 (e.g., 一部 control に介入=1 と割当てる、介入量 dose を変える)

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

**Randomisation checks** 実験前の特徴が両群で似ていること=ランダム化が成功していることを検定、帰無仮説「実験前の(全)変数において、両群の平均値の差はゼロ」

- 他要因排除の確認: 他要因による効果を検定。効果が見出されたら、メインの結論もその要因が引き起こしたかも。

**Placebo tests** 介入変数を編集して、効果が無いことを検定 (e.g., 一部 control に介入=1 と割当てる、介入量 dose を変える)

**Falsification tests (negative controls)** 効果が発生し得ない標本や結果に介入=1 と割当て、効果がないことを検定 (e.g., 介入前の標本、ラマダン介入に影響され得ないラマダン下のキリスト教徒標本)

# 実験をしてからの手順

## ③ 推論の頑健性をチェック (robustness checks)

- ランダム化の確認

**Randomisation checks** 実験前の特徴が両群で似ていること=ランダム化が成功していることを検定、帰無仮説「実験前の(全)変数において、両群の平均値の差はゼロ」

- 他要因排除の確認: 他要因による効果を検定。効果が見出されたら、メインの結論もその要因が引き起こしたかも。

**Placebo tests** 介入変数を編集して、効果が無いことを検定 (e.g., 一部 control に介入=1 と割当てる、介入量 dose を変える)

**Falsification tests (negative controls)** 効果が発生し得ない標本や結果に介入=1 と割当て、効果がないことを検定 (e.g., 介入前の標本、ラマダン介入に影響され得ないラマダン下のキリスト教徒標本)

☞ 両方ともほぼ同じ内容。敢えて言うならば、placebo tests はどんなデータでも実施可能、falsification tests は特定の文脈を使って実施。後者だけ実施可能な対象を狭める限定があるので、後者は前者に含まれる概念。

# ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a  
平均値はグループ “a” 平均値の分布の中  
央付近に位置するはず。

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞ p 値 (p value) は帰無仮説が成り立つ確率と考えていい

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

- ☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- ☞ この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

- ☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- ☞ この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

- ☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- ☞ この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき → 帰無仮説が正しい確率は  $p$  値ほど小さい

ランダム化確認: permutation test (並べ替え検定), randomisation test (確率化検定)

帰無仮説 null hypothesis: グループ a の分布=グループ b の分布

両グループのデータ全てを並べ、グループ名 (a, b) をランダムに並べ替えてグループ a 平均値を計算。これを多数回繰り返すと、グループ “a” 平均値の分布が描ける。

帰無仮説が正しければ、本当のグループ a 左右どちらかの端にあれば、グループ a と平均値はグループ “a” 平均値の分布の中 央付近に位置するはず。  
グループ b の分布は異なっていると判断できる。

☞  $p$  値 ( $p$  value) は帰無仮説が成り立つ確率と考えていい

- ☞ 正確には、帰無仮説が正しい下で推計値 (この場合は平均値の差) が発生する確率
- ☞ この確率 ( $p$  値) が小さいとき、得られた推計値は帰無仮説から見て極端な事象 → 帰無仮説の正しさを疑うべき → 帰無仮説が正しい確率は  $p$  値ほど小さい

平均値以外の (25%) 分位値などにも適用可。グループ “a” の (25%) 分位値を計算すれば同じ手順。グループ “a” (25%) 分位値分布の端に位置すると、両グループの

Sacred Heart, IDE

# バングラデシュ最貧困層への貸付実験: 大規模貸付グループと小規模貸付グループの比較

## PERMUTATION TEST RESULTS OF ARM ASSIGNMENT, TRADITIONAL VS. NON-TRADITIONAL ARMS

variables	NonTradArm	TradArm	p-value.lower	p-value.mid	p-value.upper
HeadLiteracy	0.116	0.100	0.498	0.545	0.591
HeadAge	38.023	38.536	0.549	0.550	0.551
HHsize	4.188	4.128	0.618	0.629	0.640
FloodInRd1	0.508	0.464	0.265	0.285	0.305
HAssetAmount	782	690	0.085	0.085	0.085
PAssetAmount	1114	1083	0.823	0.823	0.823
LivestockValue	5951	5184	0.328	0.328	0.328
NumCows	0.266	0.237	0.458	0.472	0.487
NetValue	7675	6604	0.334	0.334	0.334
n	580	180	(rate: 0.237)		

Source: Estimated with GUK administrative and survey data.

Notes: 1. R's package coin is used for baseline group mean covariates to conduct approximate permutation tests.

2. Number of repetition is set to 100000. Step-down method is used to adjust for multiple testing of a multi-factor grouping variable. 40 are lost to flood before arm assignment.

治療対象選定をランダム化すると ATE の一致推計量が得られる

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも...

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が  
断ったらどうする？

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が  
断ったらどうする？

さらに...

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が  
断ったらどうする？

さらに... 人々は時に「する」をする。統御群に割り振られても何とかして治療群として  
参加するかもしれない。被験者が同意事項に違反するときどうする？

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が断ったらどうする？

さらに... 人々は時にするをする。統御群に割り振られても何とかして治療群として参加するかもしれない。被験者が同意事項に違反するときどうする？

北朝鮮のような独裁国家以外では、人々には選ぶ権利がある。被験者にグループ割り振りを強制することはできない。

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が断ったらどうする？

さらに... 人々は時にするをする。統御群に割り振られても何とかして治療群として参加するかもしれない。被験者が同意事項に違反するときどうする？

北朝鮮のような独裁国家以外では、人々には選ぶ権利がある。被験者にグループ割り振りを強制することはできない。

実験者も完璧ではないので非同意者を必ず出してしまう

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が断ったらどうする？

さらに... 人々は時にするをする。統御群に割り振られても何とかして治療群として参加するかもしれない。被験者が同意事項に違反するときどうする？

北朝鮮のような独裁国家以外では、人々には選ぶ権利がある。被験者にグループ割り振りを強制することはできない。

実験者も完璧ではないので非同意者を必ず出してしまって

われわれが計測できるのは非同意者を含むグループ平均値の差。非同意者がいるとインパクトが小さくなる。

治療対象選定をランダム化すると ATE の一致推計量が得られる

でも... 人々には同意するか決める権利がある。治療を断るかもしれない。全員が断ったらどうする？

さらに... 人々は時にするをする。統御群に割り振られても何とかして治療群として参加するかもしれない。被験者が同意事項に違反するときどうする？

北朝鮮のような独裁国家以外では、人々には選ぶ権利がある。被験者にグループ割り振りを強制することはできない。

実験者も完璧ではないので非同意者を必ず出してしまって

われわれが計測できるのは非同意者を含むグループ平均値の差。非同意者がいるとインパクトが小さくなる。

非同意者を含む効果推計値を治療意図に基づく効果 **intention-to-treat (ITT) effect** という。実験室での効力 *efficacy* ではなくフィールドでの有効性 *effectiveness*。ATE を推計できる研究は少ない。

さまざまな効果推計量 (実証研究の大半が ITT か LATE)

ATE Average treatment effects: 全個人の平均効果

$$ATE = \mathbb{E}[y_i | D_i = 1] - \mathbb{E}[y_i | D_i = 0].$$

さまざまな効果推計量 (実証研究の大半が ITT か LATE)

**ATE** Average treatment effects: 全個人の平均効果

$$ATE = \mathcal{E}[y_i|D_i = 1] - \mathcal{E}[y_i|D_i = 0].$$

**ITT** Intention-to-treat effects: 実施群非同意者 (比率  $1 - \alpha \in [0, 1]$ ) と統御群非同意者 (比率  $1 - \beta \in [0, 1]$ ) を含む全個人の平均効果, 実際の割当て  $A_i = 0, 1$

$$\begin{aligned} ITT &= \alpha \mathcal{E}[y_i|D_i = 1, A_i = 1] + (1 - \alpha) \mathcal{E}[y_i|D_i = 1, A_i = 0] && \mathcal{E}[y_i|D_i = 1] \\ &\quad - \beta \mathcal{E}[y_i|D_i = 0, A_i = 0] - (1 - \beta) \mathcal{E}[y_i|D_i = 0, A_i = 1]. && -\mathcal{E}[y_i|D_i = 0] \end{aligned}$$

さまざまな効果推計量 (実証研究の大半が ITT か LATE)

ATE Average treatment effects: 全個人の平均効果

$$ATE = \mathbb{E}[y_i|D_i = 1] - \mathbb{E}[y_i|D_i = 0].$$

ITT Intention-to-treat effects: 実施群非同意者 (比率  $1 - \alpha \in [0, 1]$ ) と統御群非同意者 (比率  $1 - \beta \in [0, 1]$ ) を含む全個人の平均効果, 実際の割当て  $A_i = 0, 1$

$$ITT = \alpha \mathbb{E}[y_i|D_i = 1, A_i = 1] + (1 - \alpha) \mathbb{E}[y_i|D_i = 1, A_i = 0] - \mathbb{E}[y_i|D_i = 1]$$

$$- \beta \mathbb{E}[y_i|D_i = 0, A_i = 0] - (1 - \beta) \mathbb{E}[y_i|D_i = 0, A_i = 1]. - \mathbb{E}[y_i|D_i = 0]$$

ATT Average treatment effects on the treated: 実施群における平均効果

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

さまざまな効果推計量 (実証研究の大半が ITT か LATE)

ATE Average treatment effects: 全個人の平均効果

$$ATE = \mathbb{E}[y_i|D_i = 1] - \mathbb{E}[y_i|D_i = 0].$$

ITT Intention-to-treat effects: 実施群非同意者 (比率  $1 - \alpha \in [0, 1]$ ) と統御群非同意者 (比率  $1 - \beta \in [0, 1]$ ) を含む全個人の平均効果, 実際の割当て  $A_i = 0, 1$

$$ITT = \alpha \mathbb{E}[y_i|D_i = 1, A_i = 1] + (1 - \alpha) \mathbb{E}[y_i|D_i = 1, A_i = 0] - \beta \mathbb{E}[y_i|D_i = 0, A_i = 0] - (1 - \beta) \mathbb{E}[y_i|D_i = 0, A_i = 1].$$

ATT Average treatment effects on the treated: 実施群における平均効果

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

LATE Local average treatment effects: 割当てられて初めて介入を受ける同意者 compliers ( $\alpha - \beta$  だけいる) の平均効果、均一効果

$$\mathbb{E}[\Delta y_i|D_i = 1, A_i = 1] = \mathbb{E}[\Delta y_i|D_i = 0, A_i = 0] = \mu \text{ を仮定}$$

$$LATE = \frac{\alpha \mathbb{E}[\Delta y_i|D_i = 1, A_i = 1] - \beta \mathbb{E}[\Delta y_i|D_i = 0, A_i = 0]}{\alpha - \beta} = \frac{(\alpha - \beta)\mu}{\alpha - \beta} = \mu.$$

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

- $y_{0i}|D_i = 1$  治療群の CF

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

- $y_{0i}|D_i = 1$  治療群の CF
- ATE と ATT の違い: 全員 vs. 治療群

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

- $y_{0i}|D_i = 1$  治療群の CF
- ATE と ATT の違い: 全員 vs. 治療群
- ITT と ATT の違い: 非同意者を含む全員 vs. 治療群

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

- $y_{0i}|D_i = 1$  治療群の CF
- ATE と ATT の違い: 全員 vs. 治療群
- ITT と ATT の違い: 非同意者を含む全員 vs. 治療群

統御群の平均治療効果 average treatment effects on the control (ATC) も以下のように定義できる

$$ATC = \mathbb{E}[y_{1i}|D_i = 0] - \mathbb{E}[y_{0i}|D_i = 0].$$

治療群の平均治療効果 Average treatment effect on the treated, (ATT, 治療群における平均治療効果)

$$ATT = \mathbb{E}[y_{1i}|D_i = 1] - \mathbb{E}[y_{0i}|D_i = 1].$$

- $y_{0i}|D_i = 1$  治療群の CF
- ATE と ATT の違い: 全員 vs. 治療群
- ITT と ATT の違い: 非同意者を含む全員 vs. 治療群

統御群の平均治療効果 average treatment effects on the control (ATC) も以下のように定義できる

$$ATC = \mathbb{E}[y_{1i}|D_i = 0] - \mathbb{E}[y_{0i}|D_i = 0].$$

ATE は ATT と ATC の加重平均値

$$ATE = bATT + (1 - b)ATC, \quad b = \frac{n_1}{n_0 + n_1}.$$

# 裨益

ターゲット対象

		No	Yes
No	No	compliers	
	Yes		compliers

- ターゲティング(対象設定)が正しくても、実現するかは別問題
- LATE: compliers (Yes, Yes) - compliers (No, No)

# 裨益

ターゲット対象

		No	Yes
No	No	compliers	
	Yes		compliers

- ターゲティング(対象設定)が正しくても、実現するかは別問題
- LATE: compliers (Yes, Yes) - compliers (No, No)

# 裨益

		No	Yes
ターゲット対象	No	compliers	leakage
	Yes		compliers

- ターゲティング(対象設定)が正しくても、実現するかは別問題
- LATE: compliers (Yes, Yes) - compliers (No, No)

# 裨益

		No	Yes
ターゲット対象	No	compliers	leakage
	Yes	exclusion	compliers

- ターゲティング(対象設定)が正しくても、実現するかは別問題
- LATE: compliers (Yes, Yes) - compliers (No, No)

## A Numerical Example

		TREATMENT		$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$D = 0$	$D = 1$		
	$n_0$	80	20	0.20	
	$y_0$	1000	2000		1200
	$z = 1$	$n_1$	50	200	0.80
	$y_1$	1000	2000		1800
$\bar{y}_D$		1000	2000	$ITT = 600$	

$$ATE = 1000 = 2000 - 1000 = (y|D=1) - (y|D=0).$$

## A Numerical Example

		TREATMENT		$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$D = 0$	$D = 1$		
	$n_0$	80	20	0.20	
	$y_0$	1000	2000		1200
	$z = 1$	$n_1$	50	200	0.80
	$y_1$	1000	2000		1800
	$\bar{y}_D$	1000	2000	$ITT = 600$	

## A Numerical Example

		TREATMENT		$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$D = 0$	$D = 1$		
	$n_0$	80	20	0.20	
	$y_0$	1000	2000		1200
	$z = 1$	$n_1$	50	200	0.80
	$y_1$	1000	2000		1800
	$\bar{y}_D$	1000	2000	$ITT = 600$	

$$ATE = 1000 = 2000 - 1000 = (y|D=1) - (y|D=0).$$

$$ITT = \bar{y}_1 - \bar{y}_0 = 1800 - 1200 = 600.$$

## A Numerical Example

		TREATMENT			
		$D = 0$	$D = 1$	$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$n_0$	80	20	0.20
		$y_0$	1000	2000	1200
	$z = 1$	$n_1$	50	200	0.80
		$y_1$	1000	2000	1800
	$\bar{y}_D$	1000	2000	$ITT = 600$	

$$ATE = 1000 = 2000 - 1000 = (y|D=1) - (y|D=0).$$

$$ITT = \bar{y}_1 - \bar{y}_0 = 1800 - 1200 = 600.$$

Difference in participation rates  
 $\Delta \bar{D}_z = \bar{D}_1 - \bar{D}_0 = .8 - .2 = .6.$

# LATE

## A Numerical Example

		TREATMENT		$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$D = 0$	$D = 1$		
	$n_0$	80	20	0.20	
	$y_0$	1000	2000		1200
	$z = 1$	$n_1$	50	200	0.80
		$y_1$	1000	2000	1800
	$\bar{y}_D$	1000	2000	$ITT = 600$	

$$ATE = 1000 = 2000 - 1000 = (y|D=1) - (y|D=0).$$

$$ITT = \bar{y}_1 - \bar{y}_0 = 1800 - 1200 = 600.$$

Difference in participation rates  
 $\Delta \bar{D}_z = \bar{D}_1 - \bar{D}_0 = .8 - .2 = .6.$

ITT per participant=ATE (only holds when impacts are homogenous).

## LATE

## A Numerical Example

		TREATMENT		$\bar{D}_z$	$\bar{y}_z$
ELIGIBILITY	$z = 0$	$D = 0$	$D = 1$		
	$n_0$	80	20	0.20	
	$y_0$	1000	2000		1200
	$z = 1$	$n_1$	50	200	0.80
		$y_1$	1000	2000	1800
	$\bar{y}_D$	1000	2000	$ITT = 600$	

$$ATE = 1000 = 2000 - 1000 = (y|D=1) - (y|D=0).$$

$$ITT = \bar{y}_1 - \bar{y}_0 = 1800 - 1200 = 600.$$

Difference in participation rates  
 $\Delta \bar{D}_z = \bar{D}_1 - \bar{D}_0 = .8 - .2 = .6.$

ITT per participant = ATE (only holds when impacts are homogenous).

$$\frac{\bar{y}_1 - \bar{y}_0}{\bar{D}_1 - \bar{D}_0} = \frac{ITT}{\Delta \bar{D}_z} = \frac{600}{.6} = 1000.$$

This is not ATE but called LATE.

Figure 1: A graphical representation of ITT and LATE (Wald estimator)

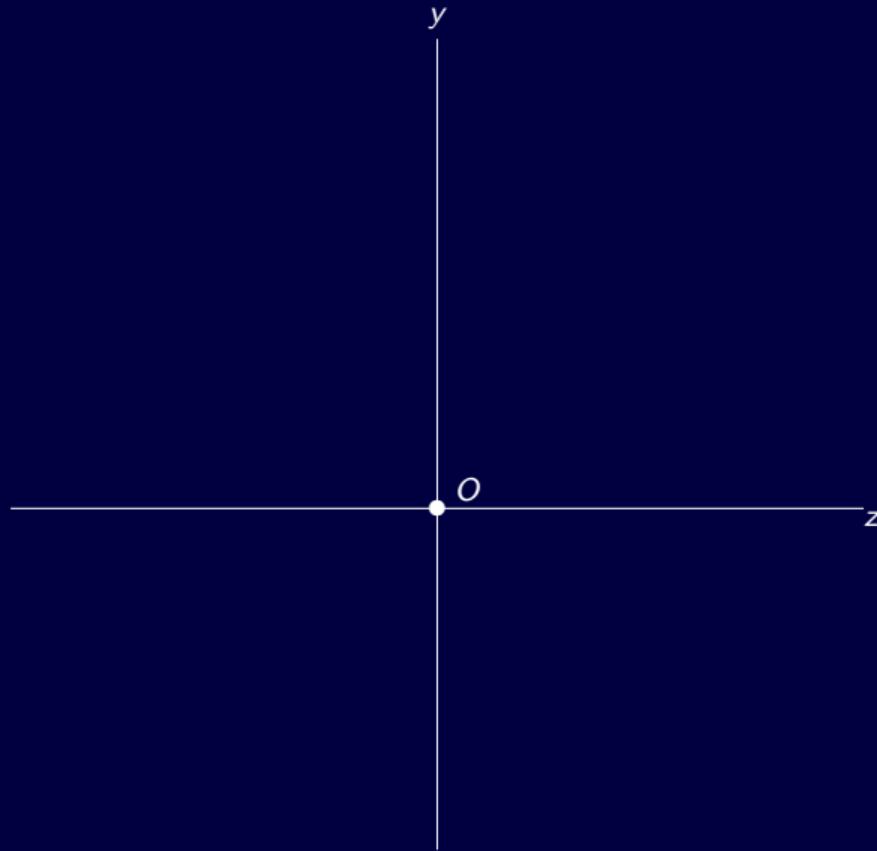


Figure 1: A graphical representation of ITT and LATE (Wald estimator)

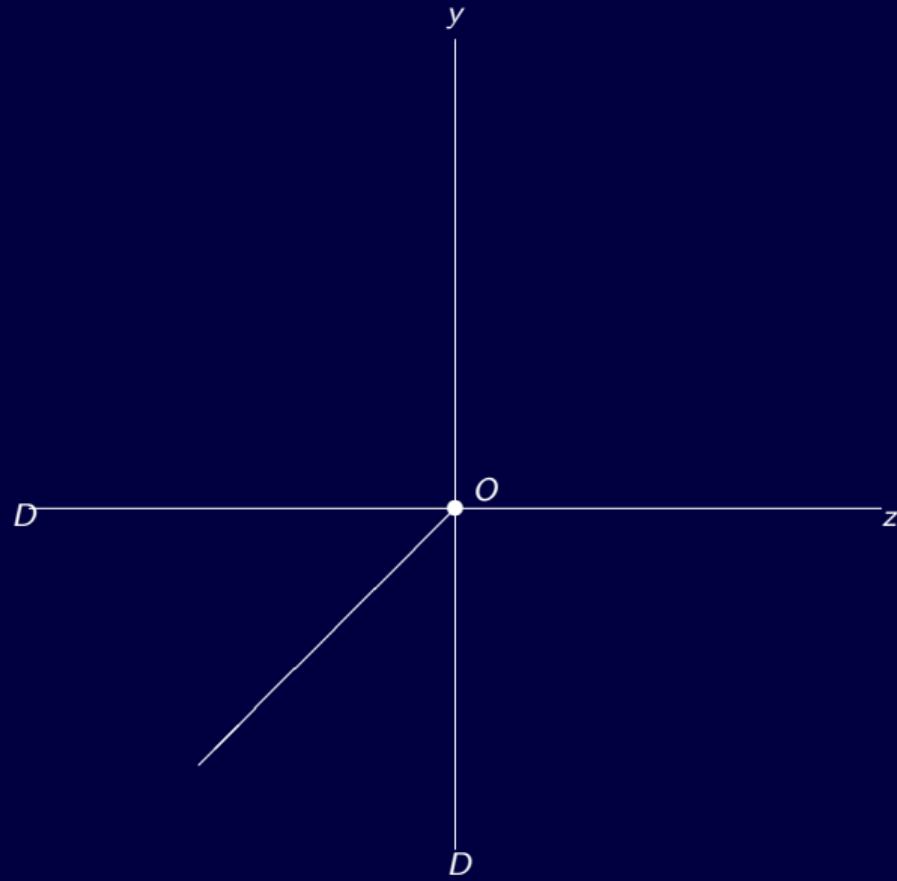


Figure 1: A graphical representation of ITT and LATE (Wald estimator)

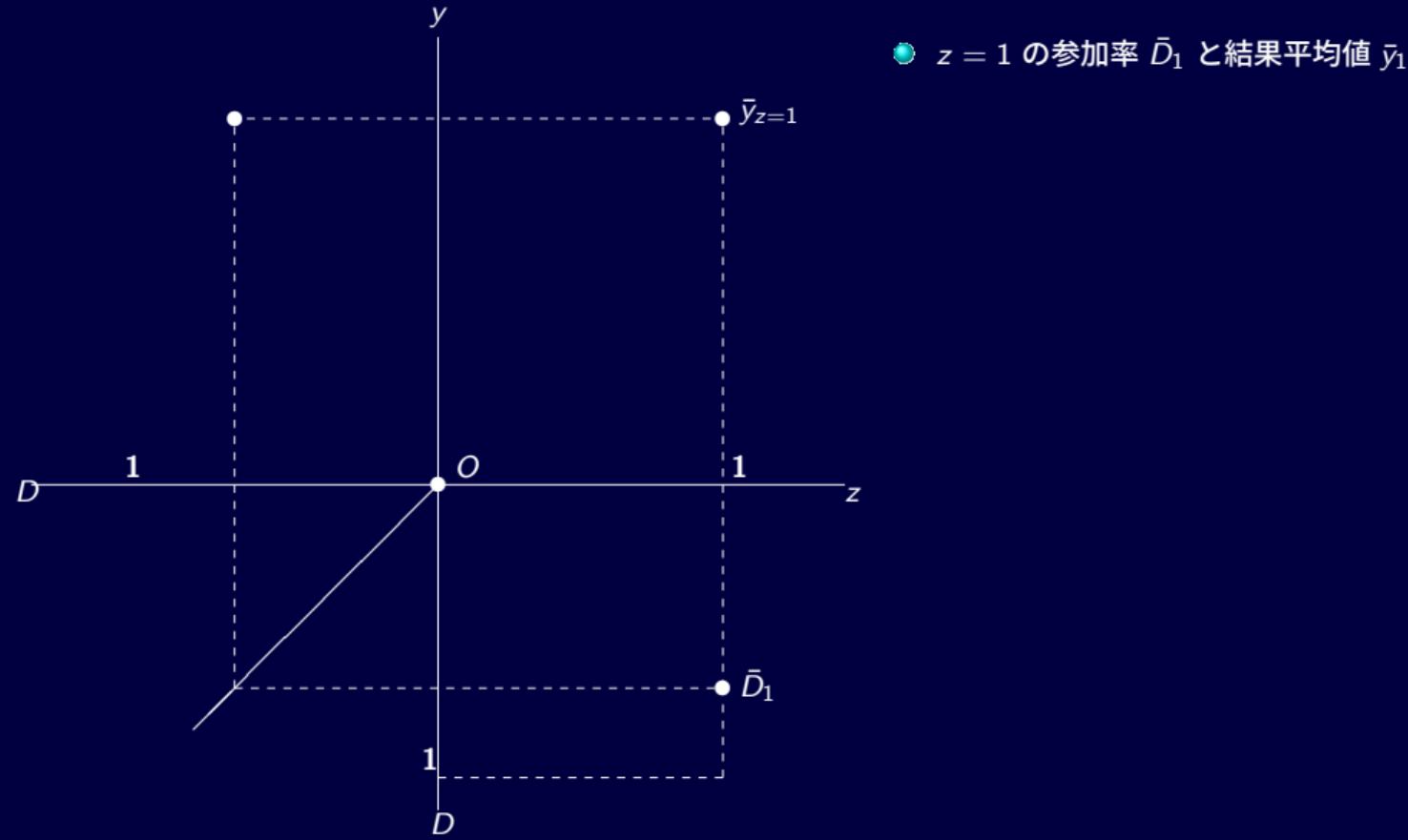


Figure 1: A graphical representation of ITT and LATE (Wald estimator)

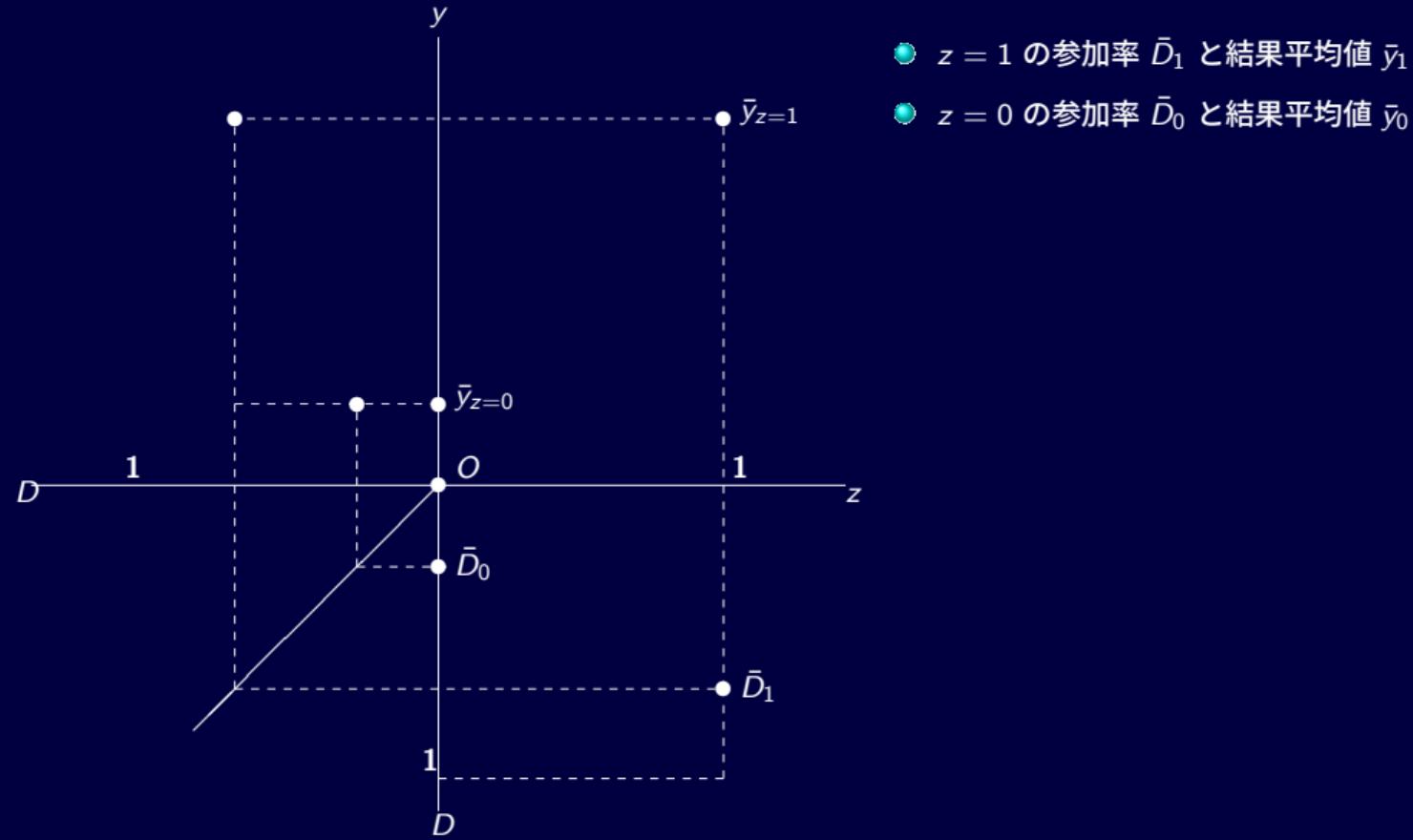


Figure 1: A graphical representation of ITT and LATE (Wald estimator)

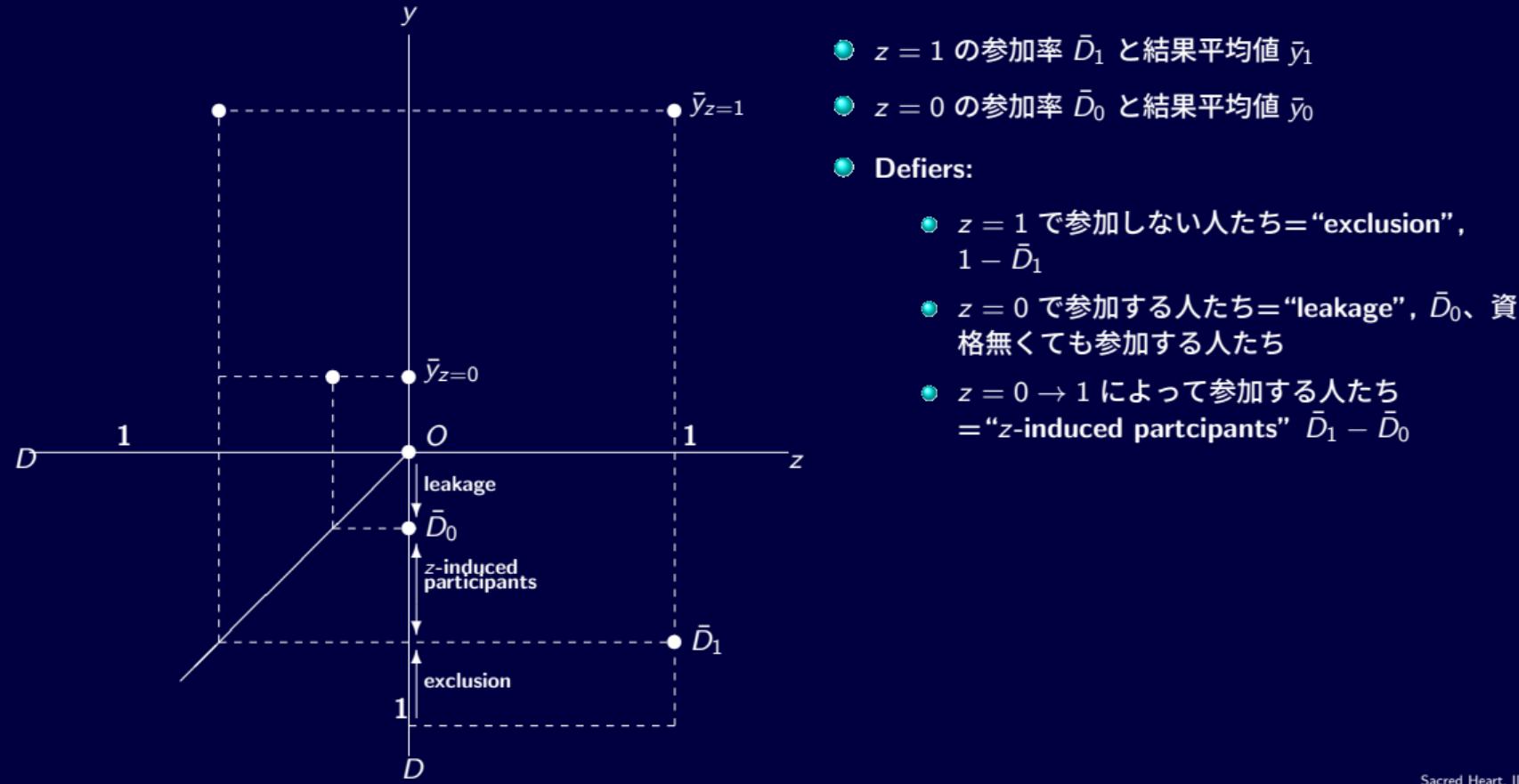


Figure 1: A graphical representation of ITT and LATE (Wald estimator)

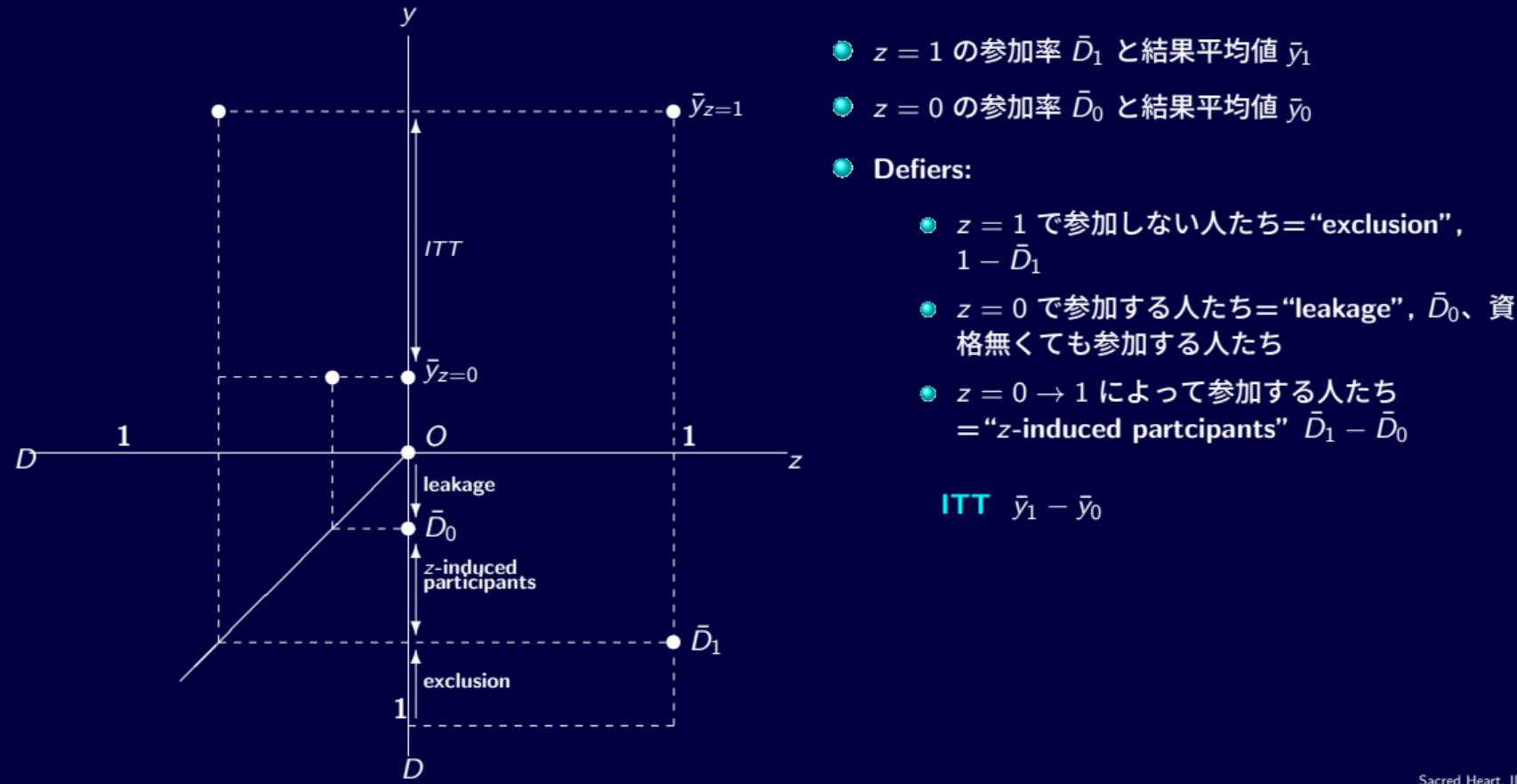
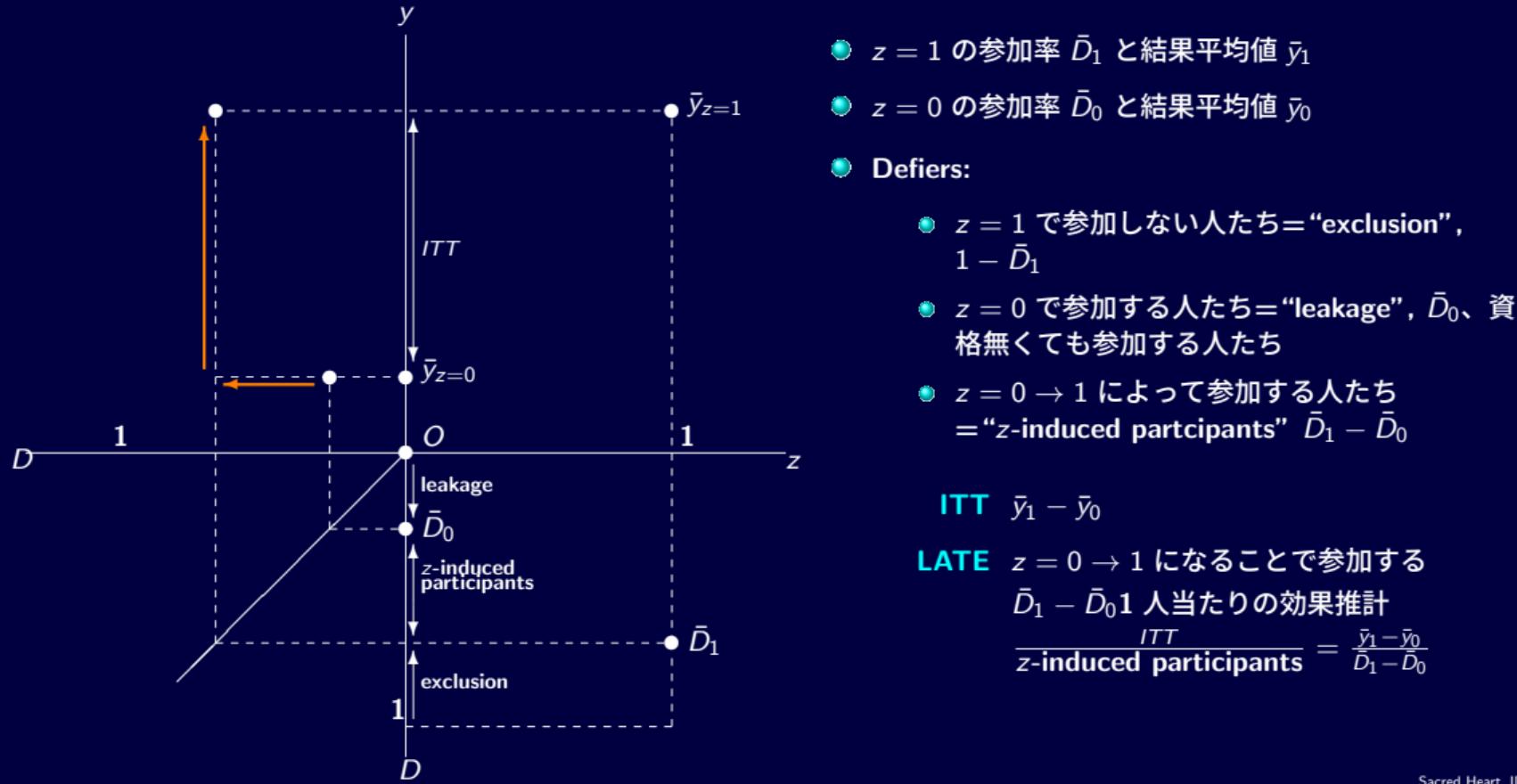


Figure 1: A graphical representation of ITT and LATE (Wald estimator)



実験には人が意図して実施する科学実験と偶然発生する自然実験 natural experiments  
がある

実験には人が意図して実施する科学実験と偶然発生する自然実験 natural experiments  
がある

科学実験は非倫理的なものは実施しない

実験には人が意図して実施する科学実験と偶然発生する自然実験 natural experiments がある

科学実験は非倫理的なものは実施しない

自然実験は意図せず発生するので非倫理的であっても実施されてしまう

実験には人が意図して実施する科学実験と偶然発生する自然実験 natural experiments がある

科学実験は非倫理的なものは実施しない

自然実験は意図せず発生するので非倫理的であっても実施されてしまう

☞ チャンス!

実験には人が意図して実施する科学実験と偶然発生する自然実験 natural experiments がある

科学実験は非倫理的なものは実施しない

自然実験は意図せず発生するので非倫理的であっても実施されてしまう

☞ チャンス!

非倫理的な自然実験の例:

- ☞ 妊婦のアルコール摂取 (?)
- ☞ 胎内での経済的・心的ショック (?)
- ☞ 親の変更 (?)

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

## 親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

## 親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

☞ 子にとっては親がランダムに変更される

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

? : 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

- ☞ 子にとっては親がランダムに変更される
- ☞ 親のこれら特徴+その他をランダムに変える実験

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

?: 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

- ☞ 子にとっては親がランダムに変更される
- ☞ 親のこれら特徴+その他をランダムに変える実験

ランダムであることの確認:

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

?: 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

- ☞ 子にとっては親がランダムに変更される
- ☞ 親のこれら特徴+その他をランダムに変える実験

ランダムであることの確認: 養子全員が treated なので permutation test は使えない

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

?: 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

- ☞ 子にとっては親がランダムに変更される
- ☞ 親のこれら特徴+その他をランダムに変える実験

ランダムであることの確認: 養子全員が treated なので permutation test は使えない

- NGO の手続き説明書類

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

?: 韓国の生活苦の乳児が養子縁組でノルウェイに行く

NGO: ノルウェイで養親候補を書類審査+面接、合格者の書類を韓国に送付、先着順で子どもと縁組

養親は子どもに関する希望を出せず、到着順はランダム

養親は年齢、学歴、所得、純資産額などで異なる

- ☞ 子にとっては親がランダムに変更される
- ☞ 親のこれら特徴+その他をランダムに変える実験

ランダムであることの確認: 養子全員が treated なので permutation test は使えない

- NGO の手続き説明書類
- 養子の特徴(月齢と性別)が養親の特徴と無相関: 養子の特徴を養親の特徴に回帰、推計された係数が統計学的にゼロ(帰無仮説: 係数はゼロ、が棄却できません)

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

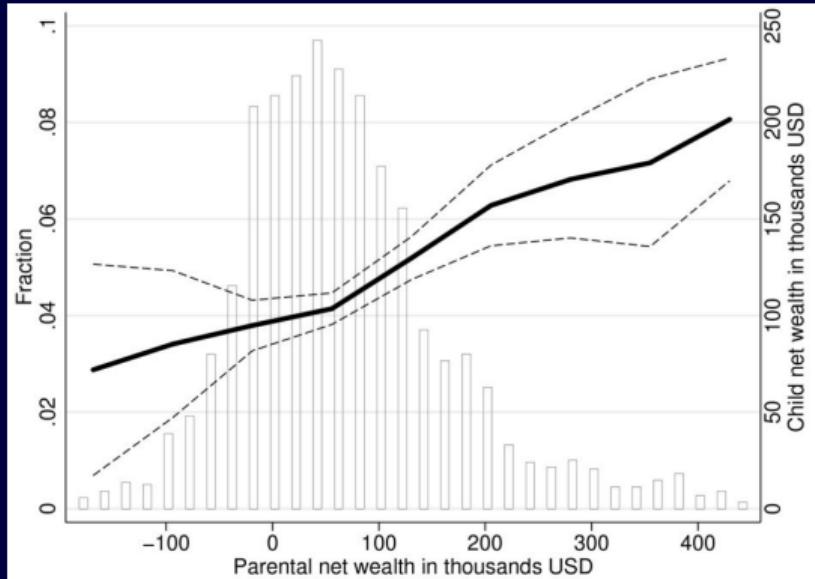


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

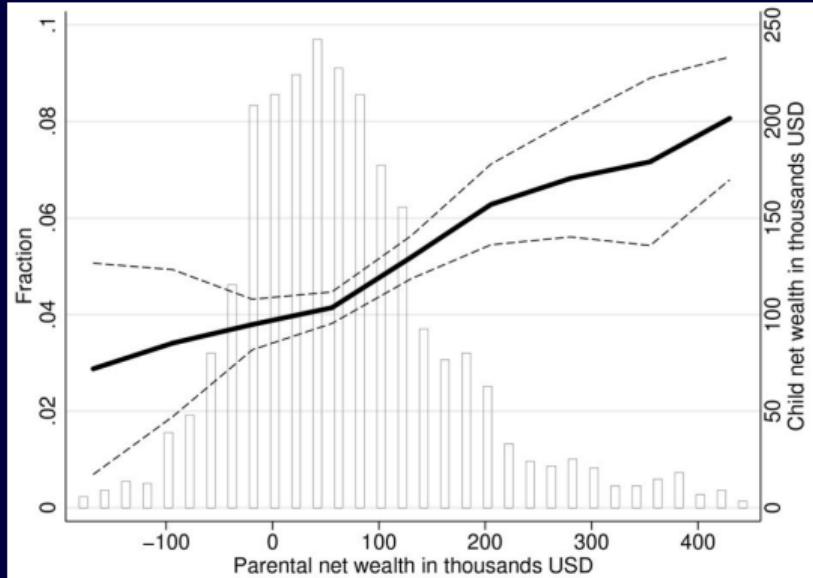


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

- 右上がり：養親の純資産額が増える  
⇒ 養子の純資産額が増える（右縦軸）

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

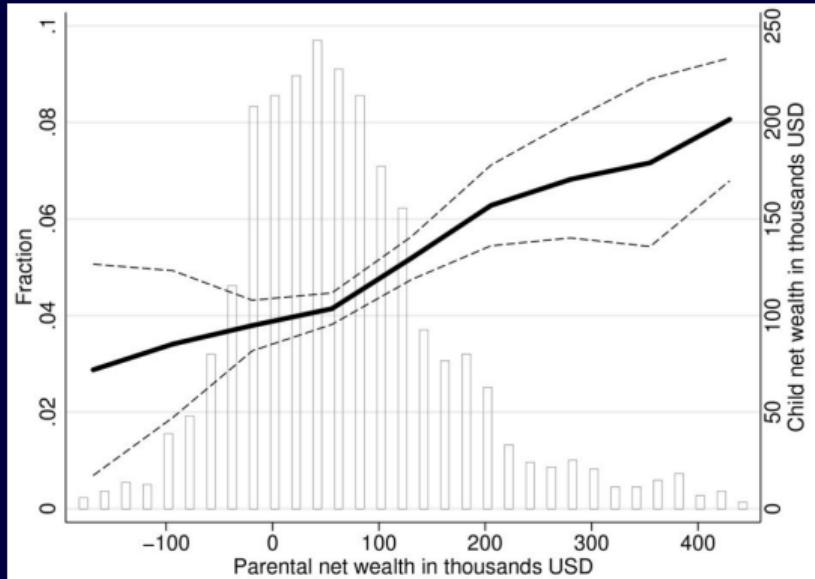


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

- 右上がり: 養親の純資産額が増える  
⇒ 養子の純資産額が増える (右縦軸)
- 実線: 推計値

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

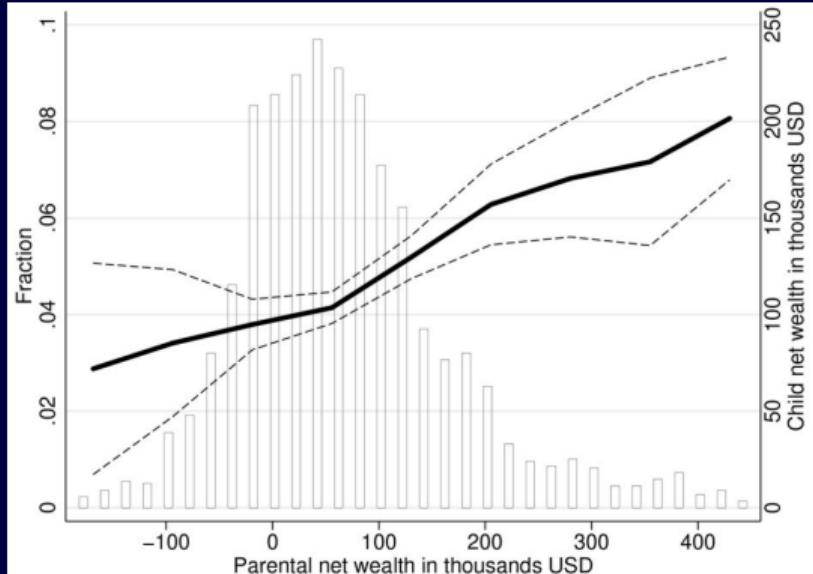


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

- 右上がり: 養親の純資産額が増える  
⇒ 養子の純資産額が増える (右縦軸)
  - 実線: 推計値
  - 破線: 95%信頼区間 (95% confidence interval)

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

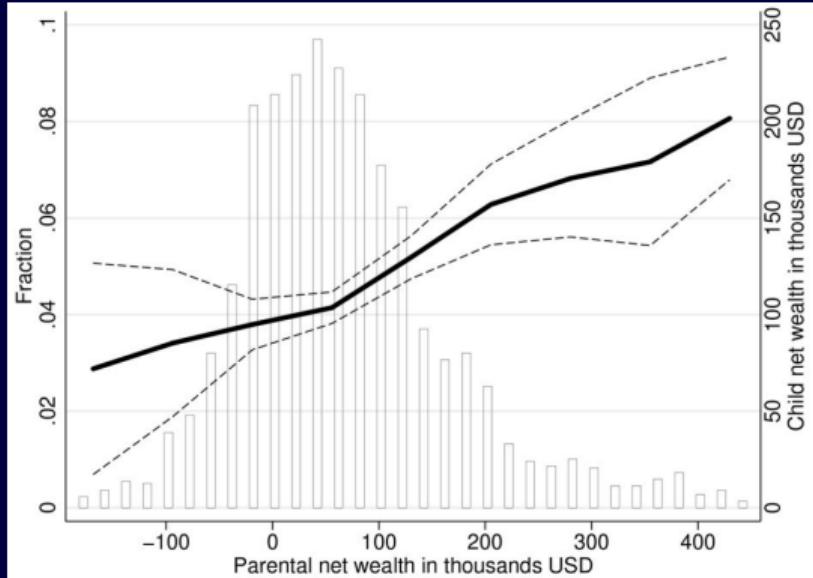


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

- 右上がり: 養親の純資産額が増える  
⇒ 養子の純資産額が増える (右縦軸)
  - 実線: 推計値
  - 破線: 95%信頼区間 (95% confidence interval)
- 養親の純資産額の度数分布 (左縦軸)

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

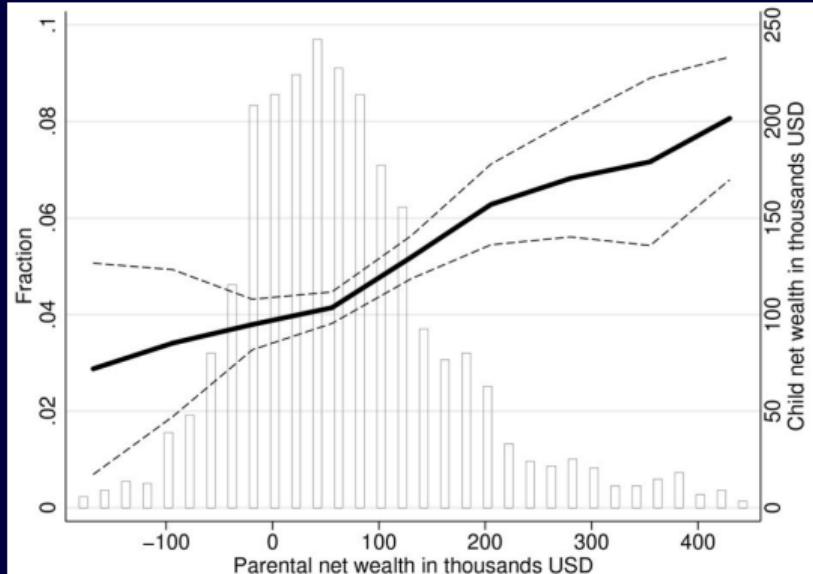


FIG. 4.—Association between adoptee's net wealth and adoptive parents' net wealth. This figure is based on the baseline sample consisting of 2,254 Korean adoptees adopted at infancy and their adoptive parents. The histogram shows the density of parental wealth (the left y-axis). The solid line shows estimates from a local linear regression of net wealth of the adoptee as an adult (measured as an average of 2012–14) on the net wealth of her adoptive parents (measured as an average of 1994–96), conditional on a full set of indicators for year of adoption and birth years of child and parents. Dashed lines show 90% confidence intervals.

- 右上がり: 養親の純資産額が増える  
⇒ 養子の純資産額が増える (右縦軸)
  - 実線: 推計値
  - 破線: 95%信頼区間 (95% confidence interval)
- 養親の純資産額の度数分布 (左縦軸)
  - 標本が多い (度数が多い) 資産額では信頼区間が狭い=推計値の精度が高い

## 親の違いによる子の純資産額、所得、学歴、金融投資への影響

養子  $i$  の特徴  $Y_i$  と養親  $j$  の  $k + 1$  個の特徴  $W_j, x_{1j}, \dots, x_{kj}$  がどのように関係しているかを見たい (養親純資産額は  $W_j$ , 養子の特徴も  $m$  個:  $x_{1i}, \dots, x_{mi}$ )

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

養子  $i$  の特徴  $Y_i$  と養親  $j$  の  $k+1$  個の特徴  $W_j, x_{1j}, \dots, x_{kj}$  がどのように関係しているかを見たい (養親純資産額は  $W_j$ , 養子の特徴も  $m$  個:  $x_{1i}, \dots, x_{mi}$ )

- ◆ 特徴: 学歴、年齢、性別、所得、リスク資産投資比率など

## 親の違いによる子の純資産額、所得、学歴、金融投資への影響

養子  $i$  の特徴  $Y_i$  と養親  $j$  の  $k+1$  個の特徴  $W_j, x_{1j}, \dots, x_{kj}$  がどのように関係しているかを見たい (養親純資産額は  $W_j$ , 養子の特徴も  $m$  個:  $x_{1i}, \dots, x_{mi}$ )

☞ 特徴: 学歴、年齢、性別、所得、リスク資産投資比率など

実験 → 養子にとって養親の特徴はランダムに与えられている

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

養子  $i$  の特徴  $Y_i$  と養親  $j$  の  $k+1$  個の特徴  $W_j, x_{1j}, \dots, x_{kj}$  がどのように関係しているかを見たい (養親純資産額は  $W_j$ , 養子の特徴も  $m$  個:  $x_{1i}, \dots, x_{mi}$ )

▣ 特徴: 学歴、年齢、性別、所得、リスク資産投資比率など

実験 → 養子にとって養親の特徴はランダムに与えられている

▣ 実験の場合に限り、OLS(普通の回帰式)で歪みのない効果が推計できる

$$Y_i = \underbrace{\alpha_{1965} Z_{1965} + \dots + \alpha_{1986} Z_{1986}}_{\text{adoption year effects}} + \beta W_j + \underbrace{\eta_1 x_{1j} + \dots + \eta_k x_{kj}}_{\text{養親の特徴の効果}} + \underbrace{\lambda_1 x_{1i} + \dots + \lambda_m x_{mi}}_{\text{養子の特徴を制御}} + \underbrace{\gamma \kappa_j + \delta \chi_i}_{\text{養親と養子の固定効果}} + u_i.$$

1965 年の効果 + ⋯ + 1986 年の効果  
就学年数の効果 + 養親の X 歳効果 + 養子の X 歳効果 + 縁組み日齢の効果 + 兄弟人數の効果 + 所得 1 単位増えることの効果 + 居住地域所得中央値の効果

$\beta$ : 養親から養子への純資産額の伝播係数

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

養親純資産額 ⇒ 養子純資産額

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD						Nonadoptees (8)	
	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES				
	(1)	(2)	(3)	(4)	(5)	(6)		
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	
Gender		Yes	Yes	Yes	Yes	Yes	Yes	
Adoption age (in days)		Yes	Yes				Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650		515	1,105	

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
							Adoptees	Nonadoptees
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)	Yes	Yes					Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1)	(2)	(3)	(4)	(5)	(6)	Adoptees (7)	Nonadoptees (8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)	Yes	Yes					Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1)	(2)	(3)	(4)	(5)	(6)	Adoptees (7)	Nonadoptees (8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes		Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes		Yes	Yes	Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1)	(2)	(3)	(4)	(5)	(6)	Adoptees (7)	Nonadoptees (8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes		Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes		Yes	Yes	Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1)	(2)	(3)	(4)	(5)	(6)	Adoptees (7)	Nonadoptees (8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes				Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD						Nonadoptees (8)	
	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES				
	(1)	(2)	(3)	(4)	(5)	(6)		
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	
Gender		Yes	Yes	Yes	Yes	Yes	Yes	
Adoption age (in days)		Yes	Yes				Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650		515	1,105	

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
							Adoptees	Nonadoptees
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes		Yes	Yes		Yes
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)						Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差  
その他共変数

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
							Adoptees	Nonadoptees
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)	.468*** (.122)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes		Yes	Yes		
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)					Yes	Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差  
その他共変数  
標本サイズ

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1) (2) (3)			(4) (5) (6)			Adoptees	Nonadoptees
	Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes				Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)					Yes	Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ , \*\*\*  $p < .01$

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差  
その他共変数  
標本サイズ

\* \*  $p < .05$ , \* \* \*  $p < .01$  は  $p$  value を示す

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1) (2) (3)			(4) (5) (6)			Adoptees	Nonadoptees
	Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adoption age (in days)		Yes	Yes				Yes	
Family characteristics			Yes		Yes	Yes		
Matched sample (propensity score)					Yes	Yes		
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ , \*\*\*  $p < .01$

(1)-(3) の係数: .204 - .225

養親の純資産額が5クローネ増えると養子の純資産額を(少なくとも)1クローネ増やす

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差  
その他共変数  
標本サイズ

\* \*  $p < .05$ , \* \* \*  $p < .01$  は  $p$  value を示す

# 親の違いによる子の純資産額、所得、学歴、金融投資への影響

TABLE 3  
INTERGENERATIONAL LINKS IN WEALTH

	KOREAN-NORWEGIAN ADOPTEES			NONADOPTEES			FAMILIES WITH BOTH ADOPTED CHILD AND NONADOPTED CHILD	
	(1) (2) (3)			(4) (5) (6)			Adoptees	Nonadoptees
	Child-parent net wealth relation	.225*** (.041)	.225*** (.041)	.204*** (.042)	.575*** (.011)	.547*** (.011)	.548*** (.018)	.276** (.139)
Adoption year indicators	Yes	Yes	Yes				Yes	
Birth year indicator of child and parents	Ycs	Ycs	Ycs	Yes	Yes	Yes	Yes	Yes
Gender				Yes	Yes	Yes	Yes	Yes
Adoption age (in days)				Yes	Yes	Yes	Yes	Yes
Family characteristics				Yes	Yes	Yes	Yes	Yes
Matched sample (propensity score)								
Observations	2,254			1,206,650			515	1,105

NOTE.—In this table, Korean-Norwegian adoptees are born in South Korea between 1965 and 1986 and adopted at infancy by Norwegian parents. Nonadoptees are born in Norway between 1965 and 1986 and raised by their biological parents. Family characteristics include education (in years) of the mother and father, number of siblings, (log of) parents' income, and (log of) the median income in parents' municipality of residence, all measured at the time of birth of the child. In col. 6, the observations in the sample of nonadoptees are weighted by the propensity score for being an adoptee (based on predetermined characteristics; see sec. VB and table B9). In cols. 7 and 8, we restrict the sample to families with both a Korean-Norwegian adopted child and a nonadopted child. Using this restricted sample, we then estimate the intergenerational wealth transmission separately for the 515 adopted children (col. 7) and for the 1,105 nonadopted children (col. 8). The sample restriction ensures that we are comparing adoptees and nonadoptees with exactly the same set of parents. Standard errors (in parentheses) are clustered at the level of the mother.

\*\*  $p < .05$ , \*\*\*  $p < .01$

(1)-(3) の係数: .204 - .225

養親の純資産額が5クローネ増えると養子の純資産額を(少なくとも)1クローネ増やす

(4)-(6) の係数: .547 - .575

養親の純資産額が5クローネ増えると実子の純資産額を(少なくとも)2.5クローネ増やす

養親純資産額 ⇒ 養子純資産額

(1)-(3) 韓国出自の養子  
(4)-(6) 実子

韓国出自の養子のいる家庭の  
(7) 養子と(8) 実子のサンプル

推計値と標準誤差  
その他共変数  
標本サイズ

\* \*  $p < .05$ , \* \* \*  $p < .01$  は  $p$  value を示す

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- 間接効果の80%が生前贈与

純資産額の効果は 60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ❖ 間接効果の 80%が生前贈与
- ❖ 逆に言えば、養子の純資産額は 68%くらい(直接効果の 60%+40%の 2割)が生前贈与以外のもの

純資産額の効果は 60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の 80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は 68%くらい(直接効果の 60%+40%の 2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

純資産額の効果は 60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の 80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は 68%くらい(直接効果の 60%+40%の 2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

- ☛ ノルウェイが高学歴でより平等な社会だから差が出なかったのか?

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は68%くらい(直接効果の60%+40%の2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

☛ ノルウェイが高学歴でより平等な社会だから差が出なかったのか?

ただし、リスク資産投資比率は養親から養子に直接効果あり

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は68%くらい(直接効果の60%+40%の2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

- ☛ ノルウェイが高学歴でより平等な社会だから差が出なかったのか?

ただし、リスク資産投資比率は養親から養子に直接効果あり

所得や学歴以外の家庭内の何かが子の純資産額を高める

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は68%くらい(直接効果の60%+40%の2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

- ☛ ノルウェイが高学歴でより平等な社会だから差が出なかったのか?

ただし、リスク資産投資比率は養親から養子に直接効果あり

所得や学歴以外の家庭内の何かが子の純資産額を高める

- ☛ 学校ではない...学校教育は平等化装置ではない(ノルウェイでは)

純資産額の効果は60%程度が直接効果、40%程度が間接効果(養子の所得、学歴、養子への生前贈与、経済系の学位)を経由

- ☛ 間接効果の80%が生前贈与
- ☛ 逆に言えば、養子の純資産額は68%くらい(直接効果の60%+40%の2割)が生前贈与以外のもの

養親の純資産額は養子の学歴は僅少効果、所得には統計学的にはゼロの効果

- ☛ ノルウェイが高学歴でより平等な社会だから差が出なかったのか?

ただし、リスク資産投資比率は養親から養子に直接効果あり

所得や学歴以外の家庭内の何かが子の純資産額を高める

- ☛ 学校ではない...学校教育は平等化装置ではない(ノルウェイでは)
- ☛ 何かが分からぬ...何をすれば貧しい家庭の子どもが同じリソースにアクセスできるのか

# selection

グループの割り振り ( $D_i$ ) がランダム化していないと、ごく稀なケースを除き、政策がない場合の結果指標の分布はグループ間で異なる。

# selection

グループの割り振り ( $D_i$ ) がランダム化していないと、ごく稀なケースを除き、政策がない場合の結果指標の分布はグループ間で異なる。

被験者=目的意識を持って参加する人間なので、参加者と不参加者は特徴が異なる

# selection

グループの割り振り ( $D_i$ ) がランダム化していないと、ごく稀なケースを除き、政策がない場合の結果指標の分布はグループ間で異なる。

被験者=目的意識を持って参加する人間なので、参加者と不参加者は特徴が異なる

自己選抜 self-selection 対象者自身による選抜。参加利益のある人は参加。

# selection

グループの割り振り ( $D_i$ ) がランダム化していないと、ごく稀なケースを除き、政策がない場合の結果指標の分布はグループ間で異なる。

被験者=目的意識を持って参加する人間なので、参加者と不参加者は特徴が異なる

自己選抜 self-selection 対象者自身による選抜。参加利益のある人は参加。

実施対象選抜 placement selection 政策担当者による選抜。政策担当者が特定の集団を選ぶように指示もしくは誘因を得ているとき、政策がないときに対象者(治療群)と非対象者(統御群)の分布が近似する保証はない。

# self-selection

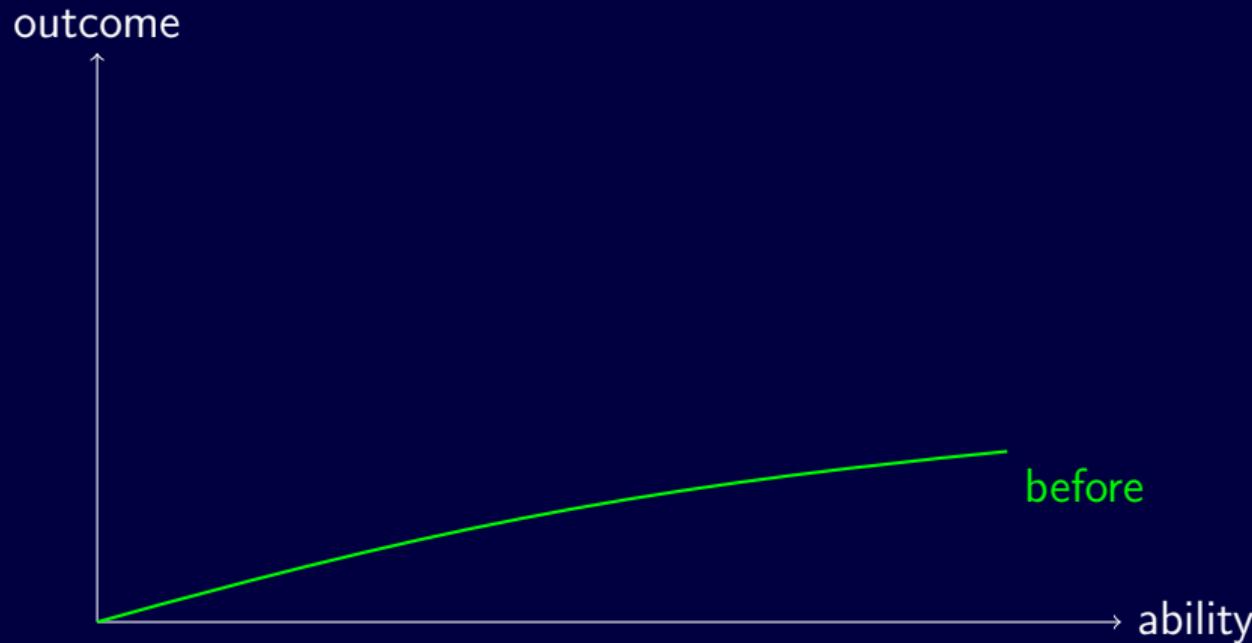
What we will learn:

- ① Mechanism of self-selection
- ② Bias of the naïve estimator (simple comparison among treated and control)
- ③ Difference-in-differences (DID) estimator and how before-after data of both treated and control can give a consistent estimate of ATT under a mild condition

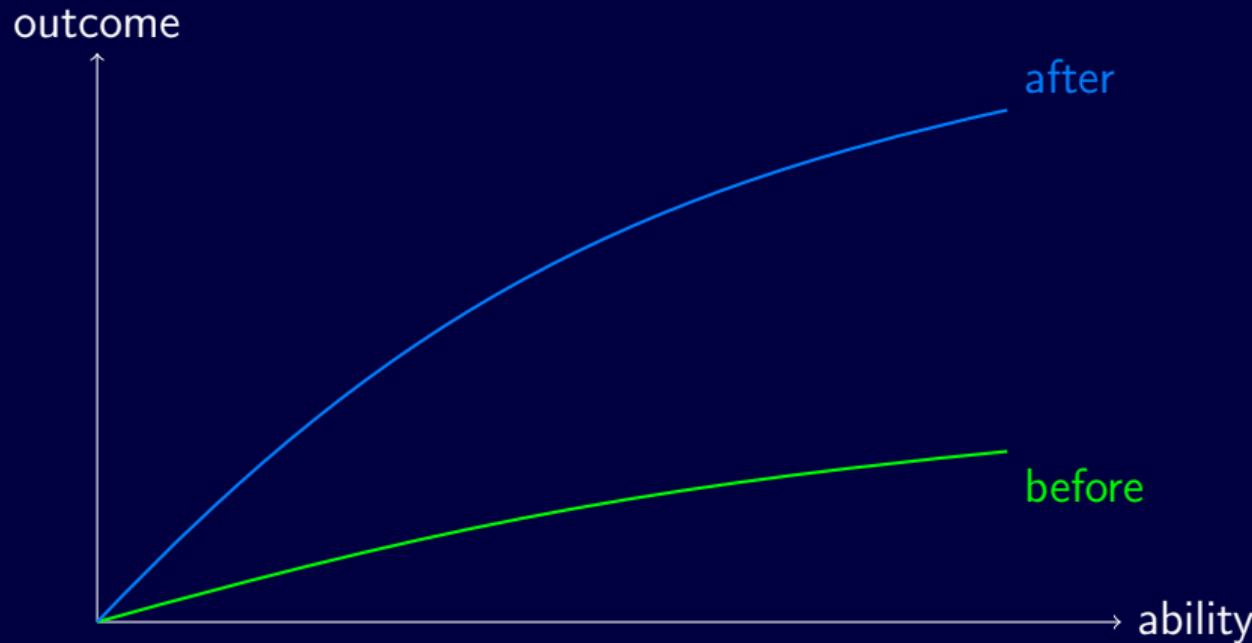
# self-selection (benefits)



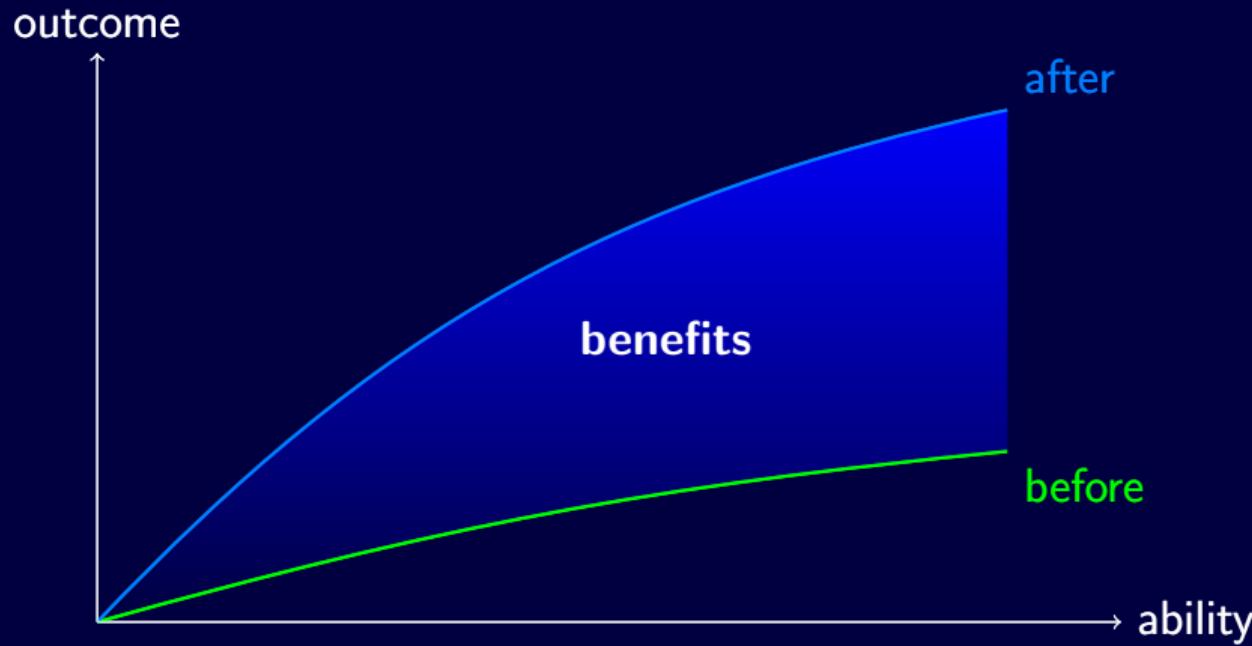
# self-selection (benefits)



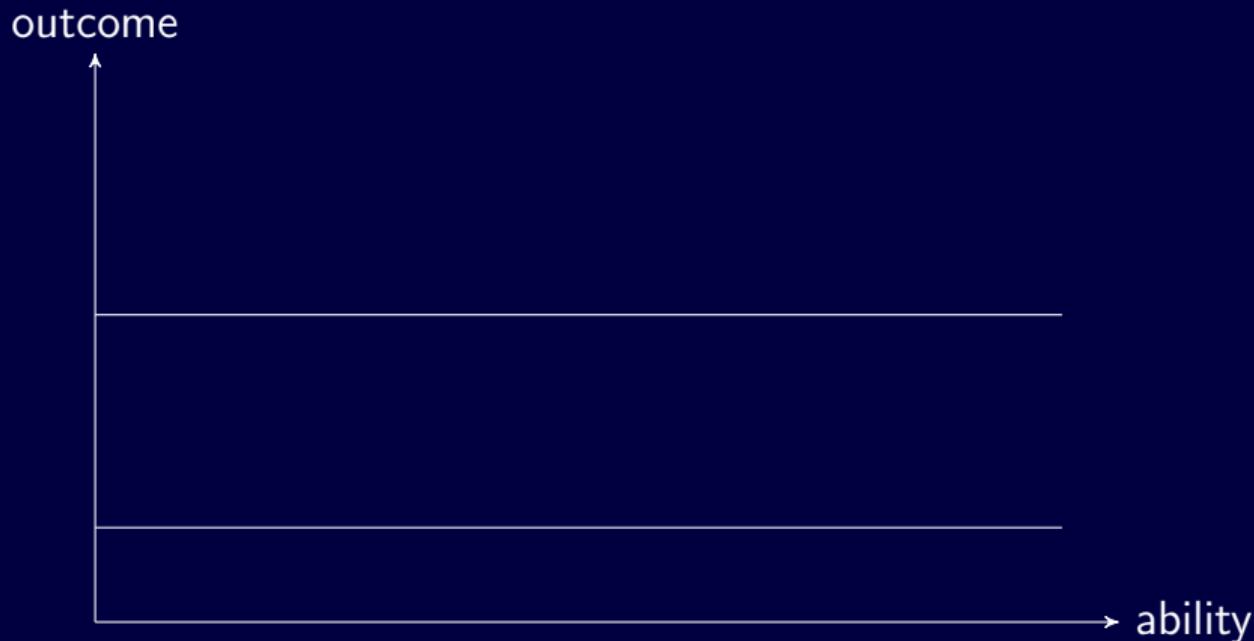
# self-selection (benefits)



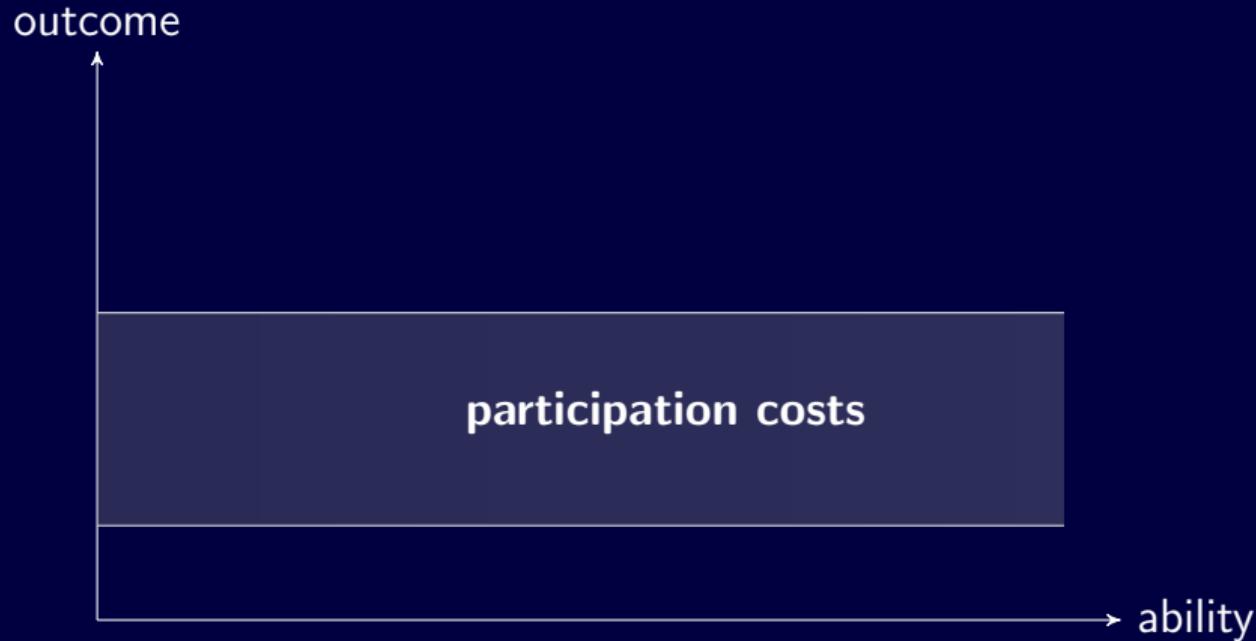
# self-selection (benefits)



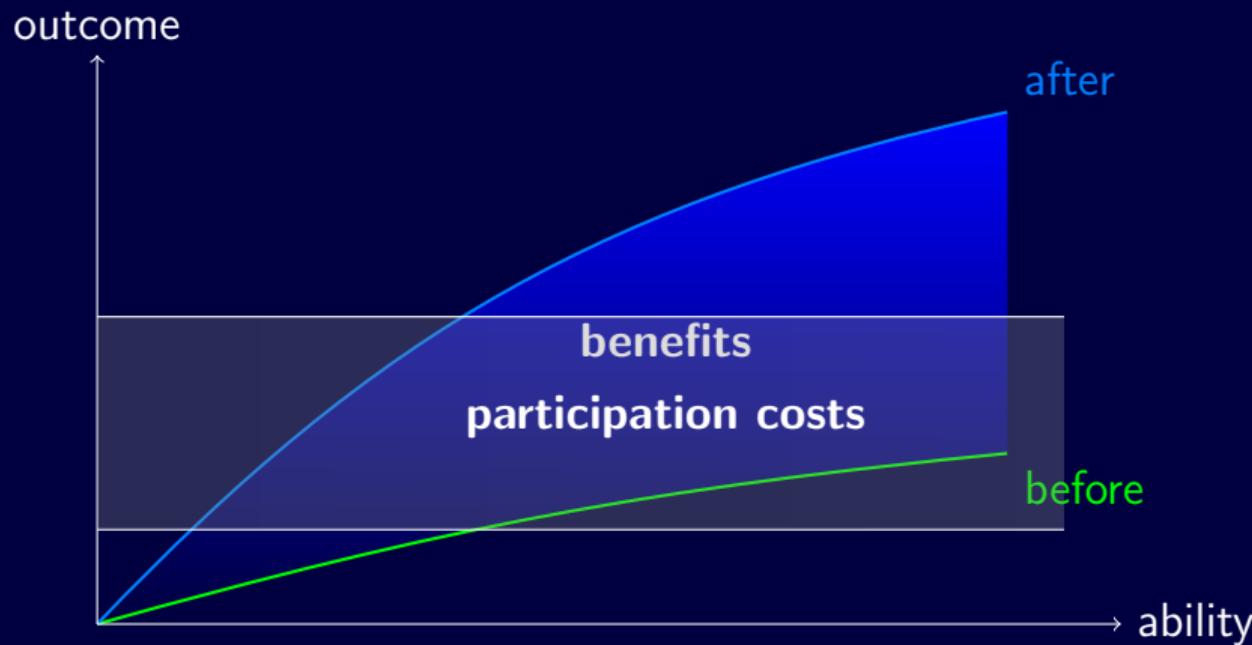
# self-selection (costs)



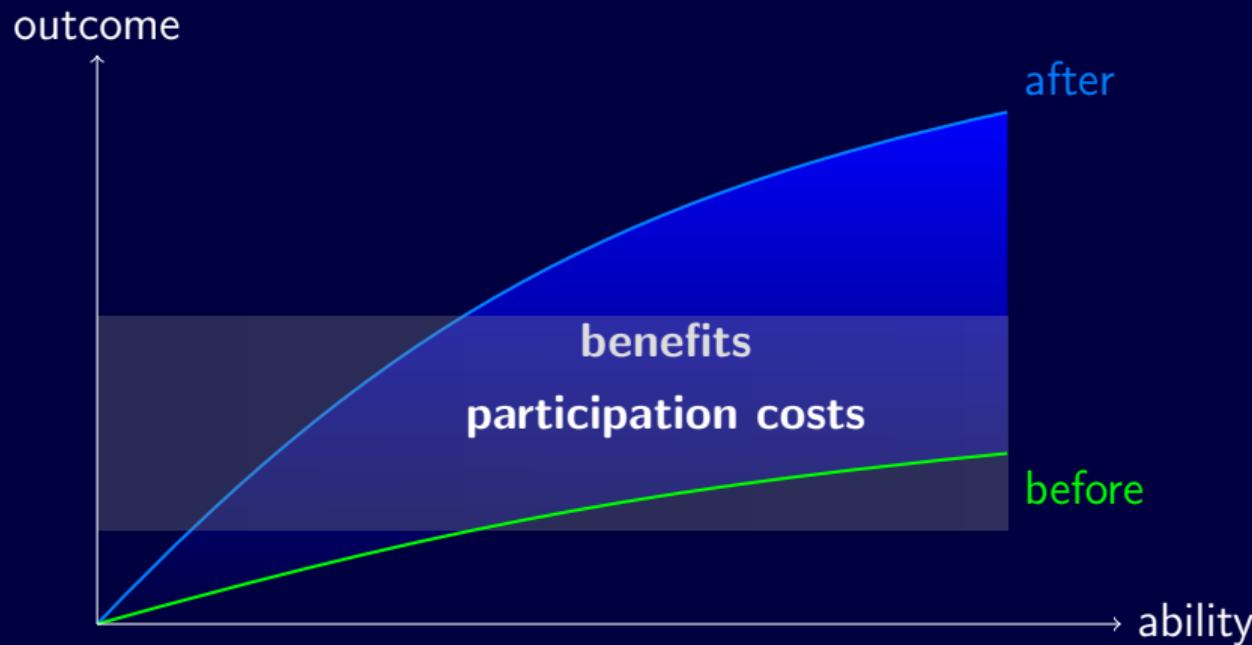
# self-selection (costs)



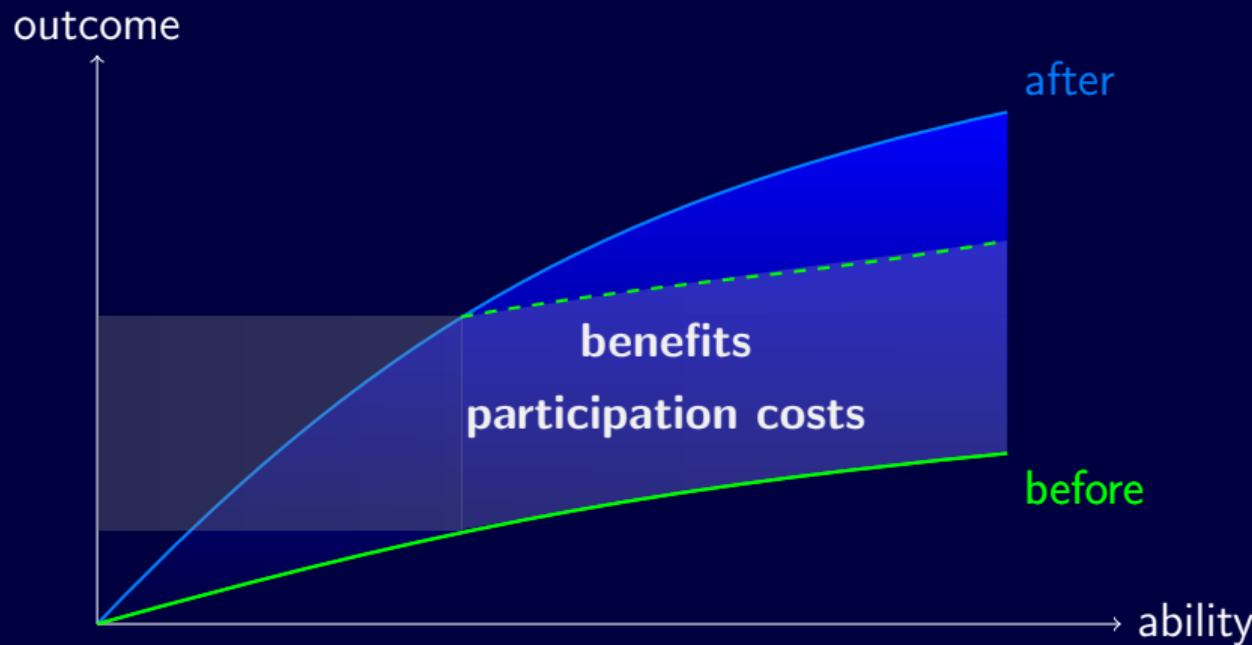
# self-selection (benefits and costs)



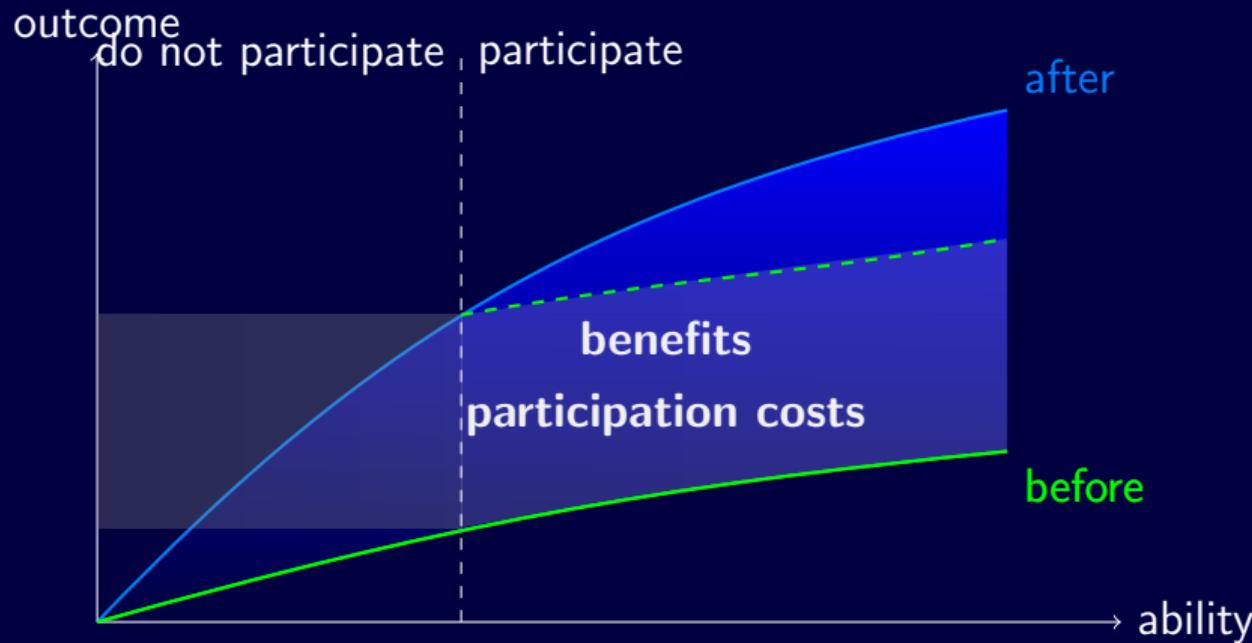
# self-selection (participation decisions)



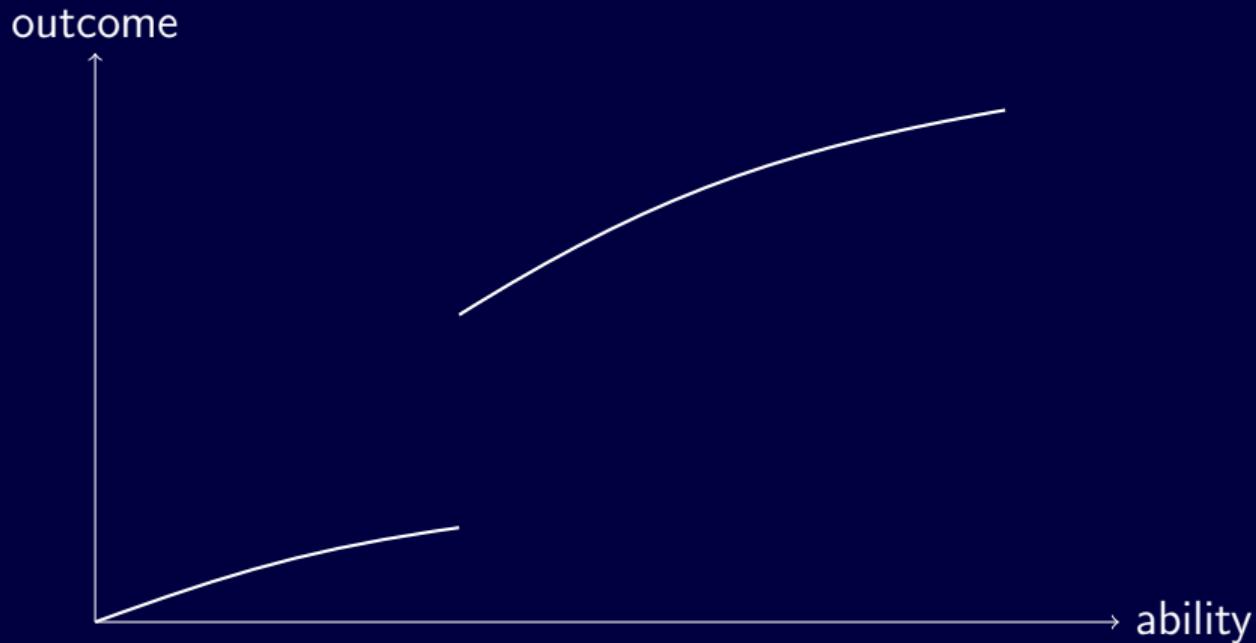
# self-selection (participation decisions)



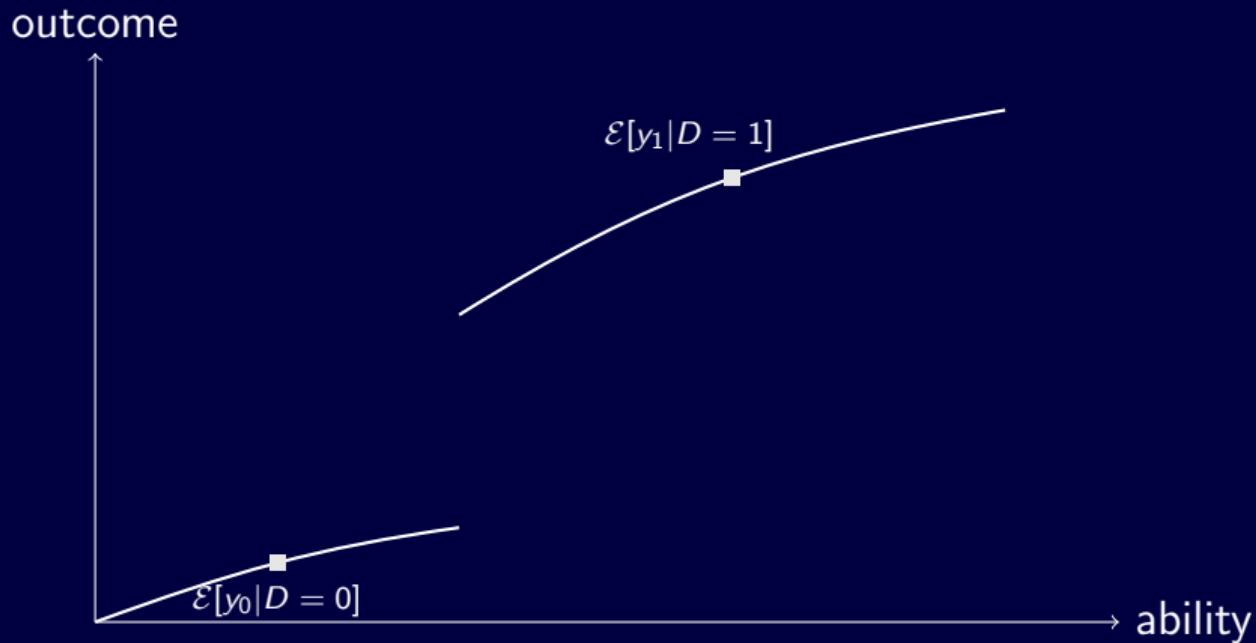
# self-selection (participation decisions)



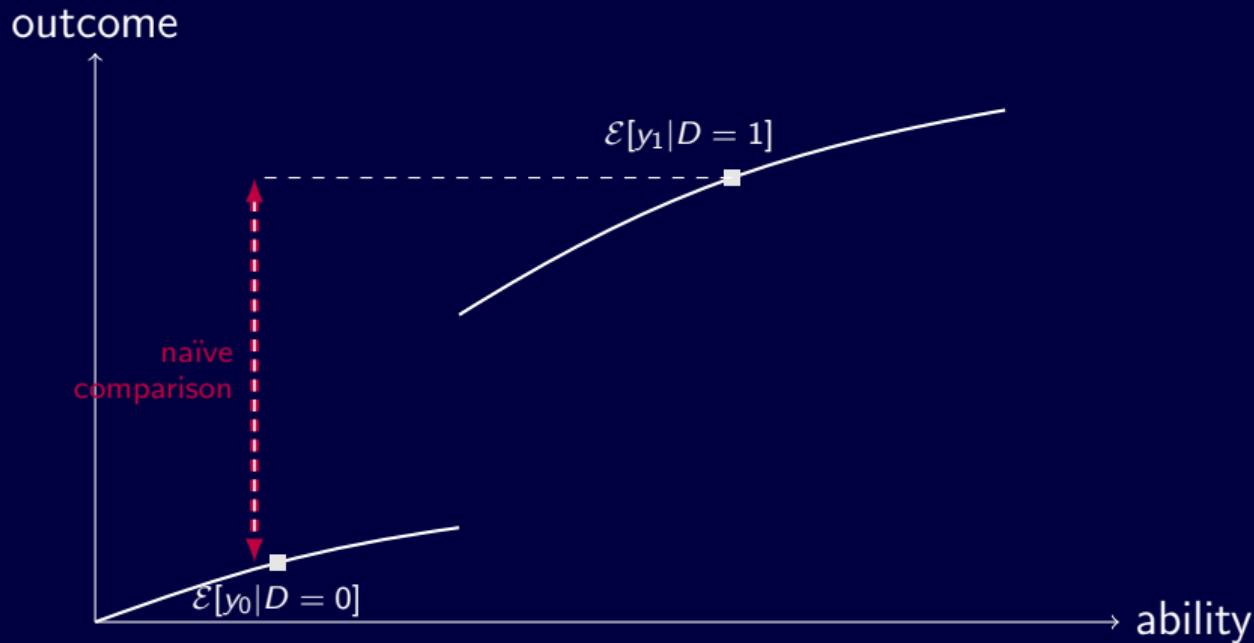
# self-selection (results)



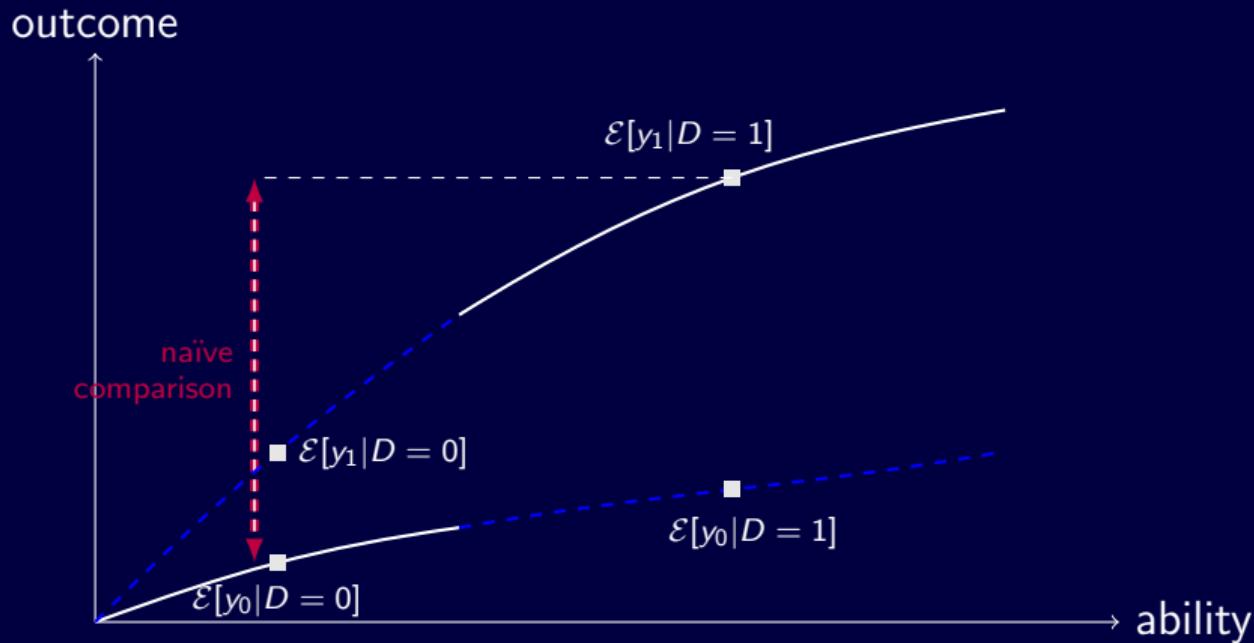
# self-selection (results)



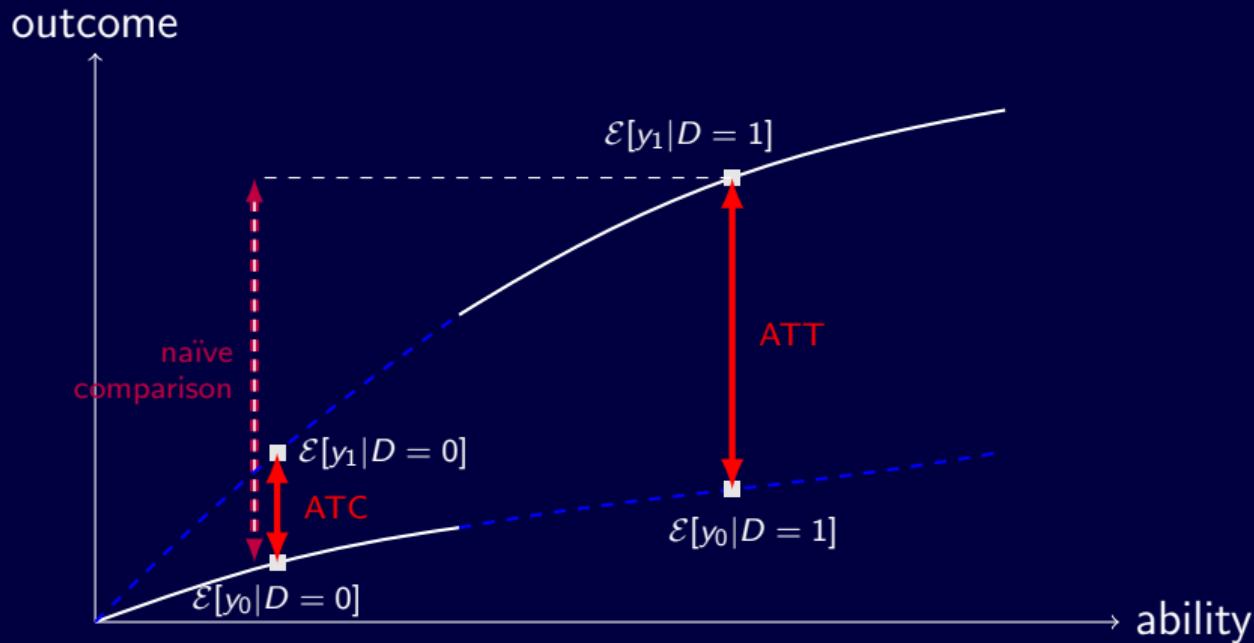
# self-selection (results)



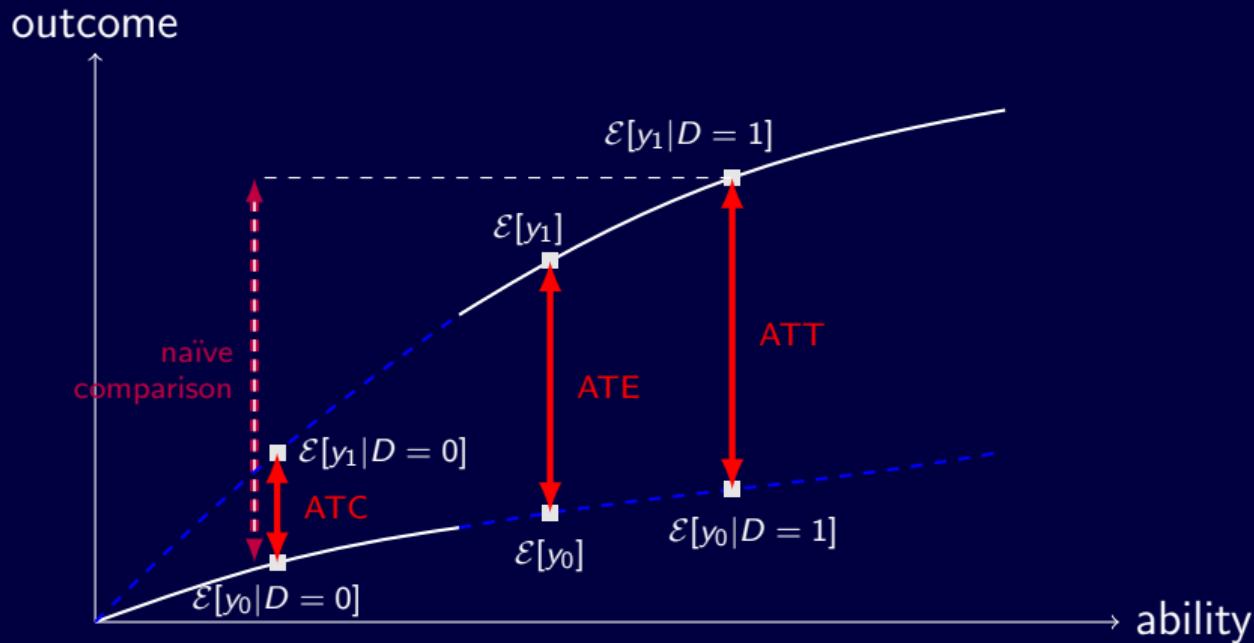
# self-selection (results)



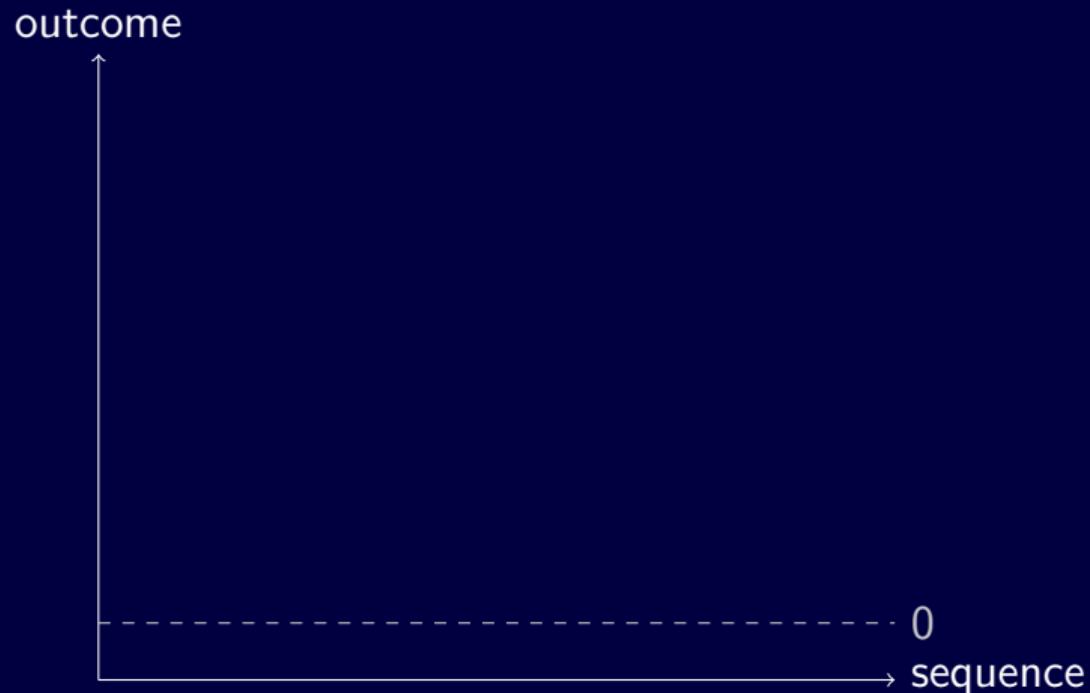
# self-selection (results)



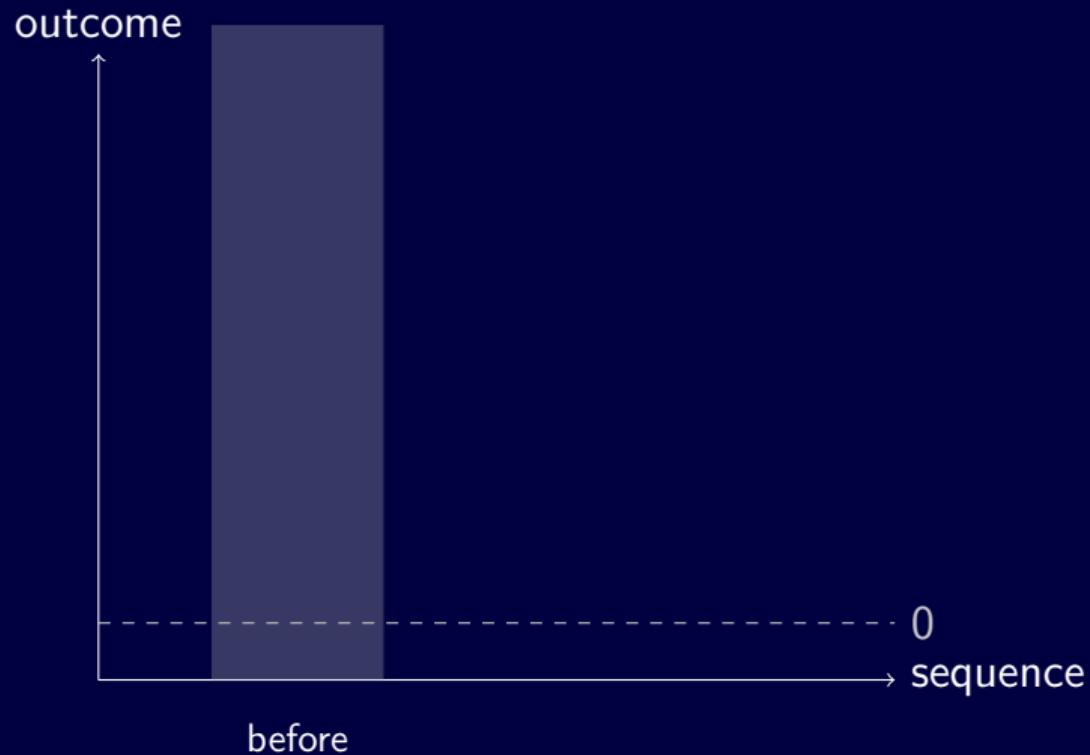
# self-selection (results)



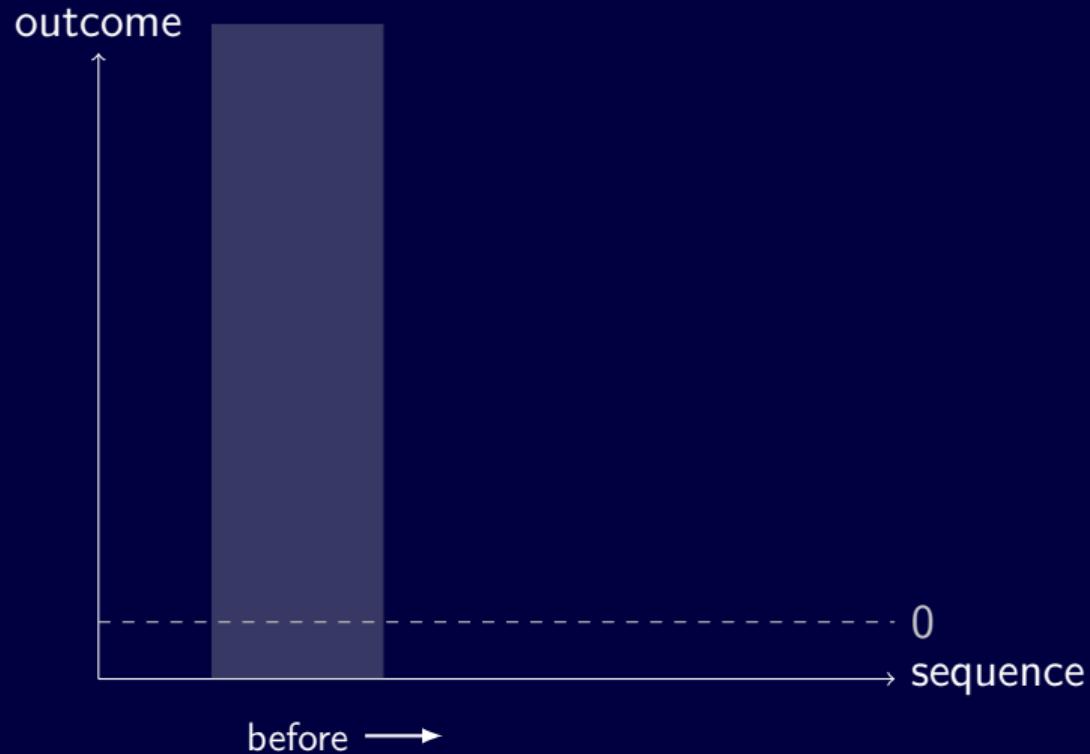
# self-selection (what we observe)



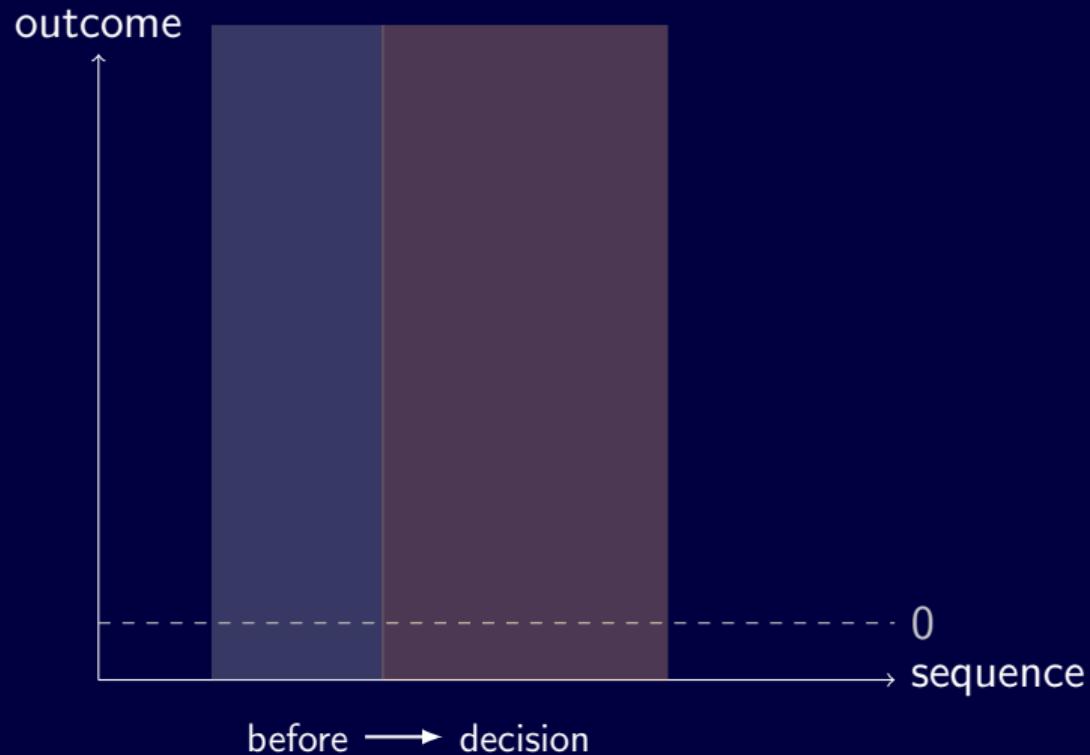
# self-selection (what we observe)



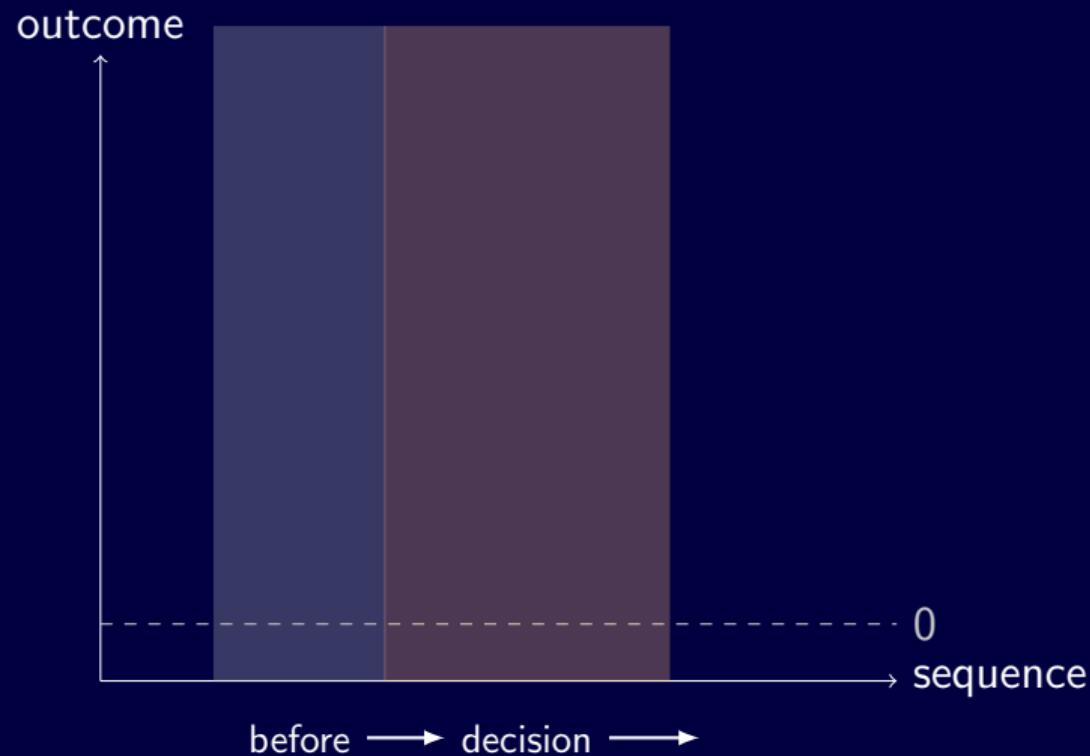
# self-selection (what we observe)



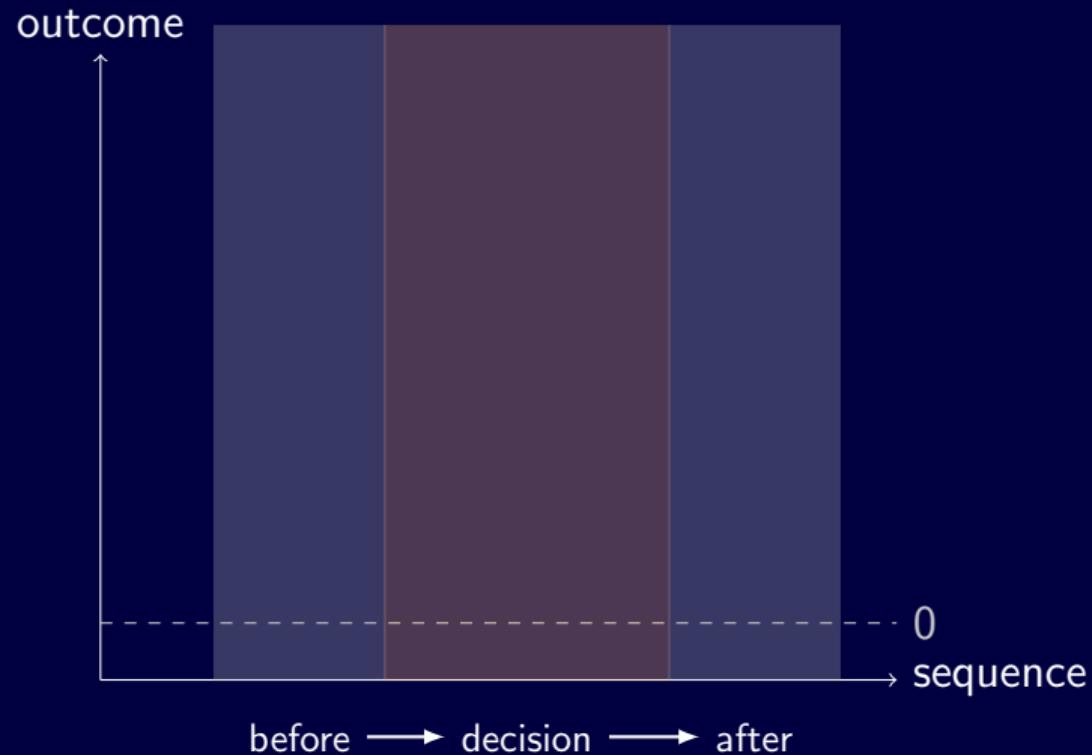
# self-selection (what we observe)



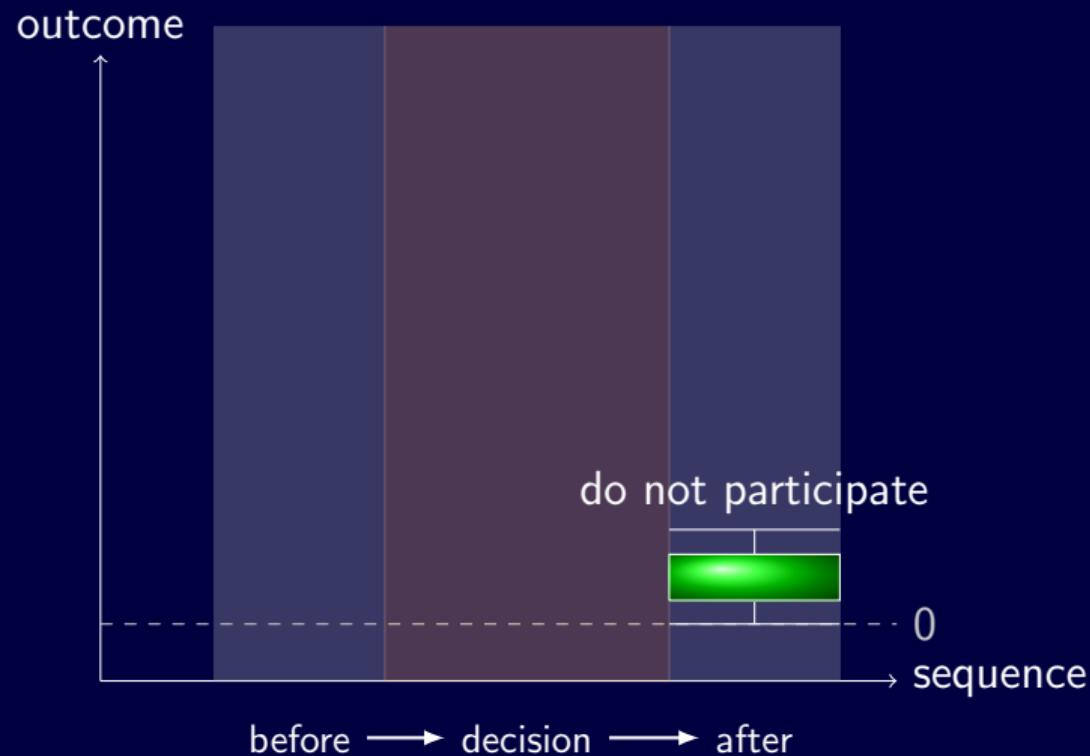
# self-selection (what we observe)



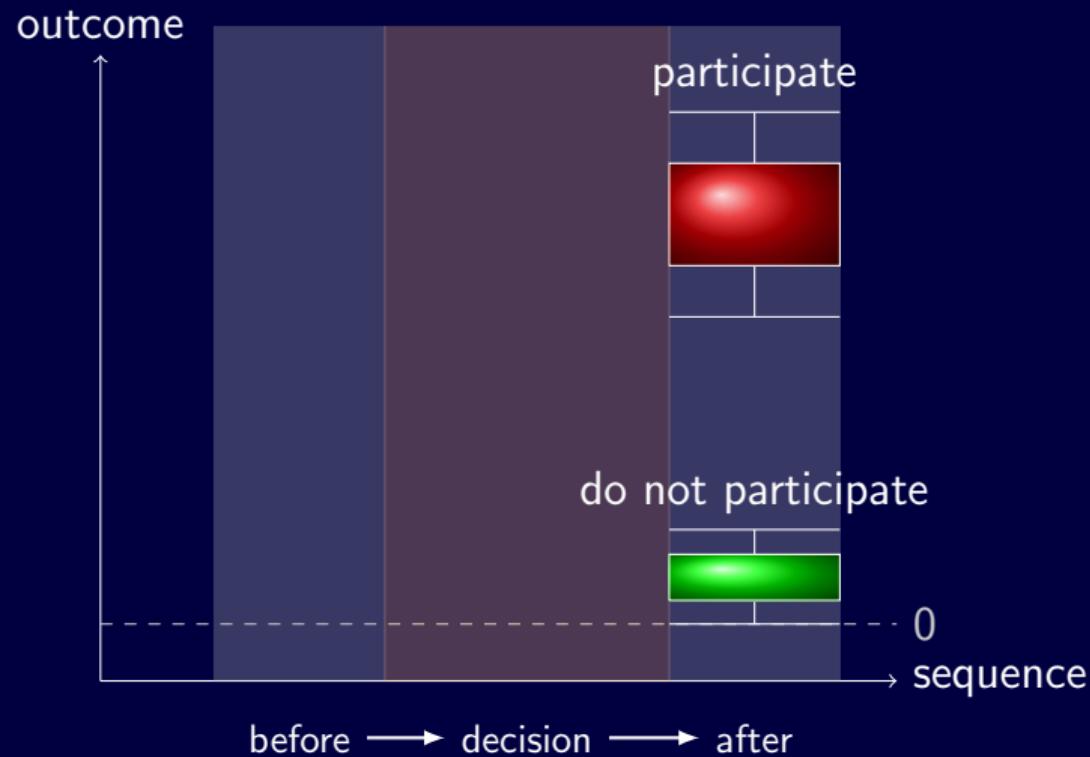
# self-selection (what we observe)



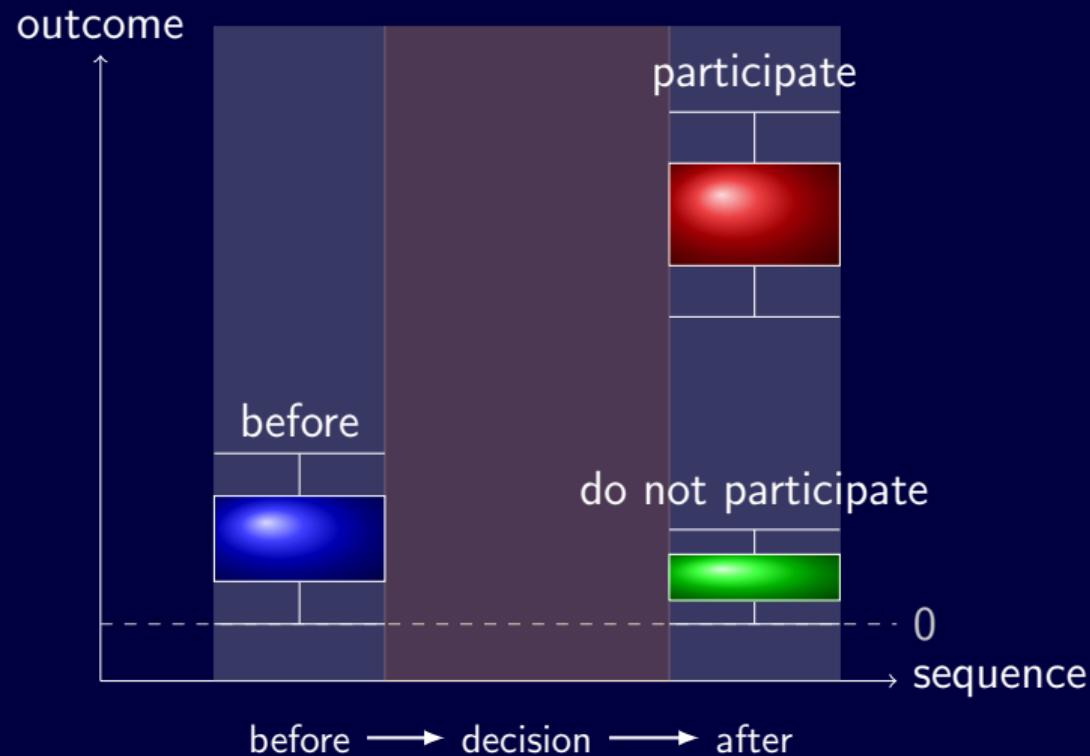
# self-selection (what we observe)



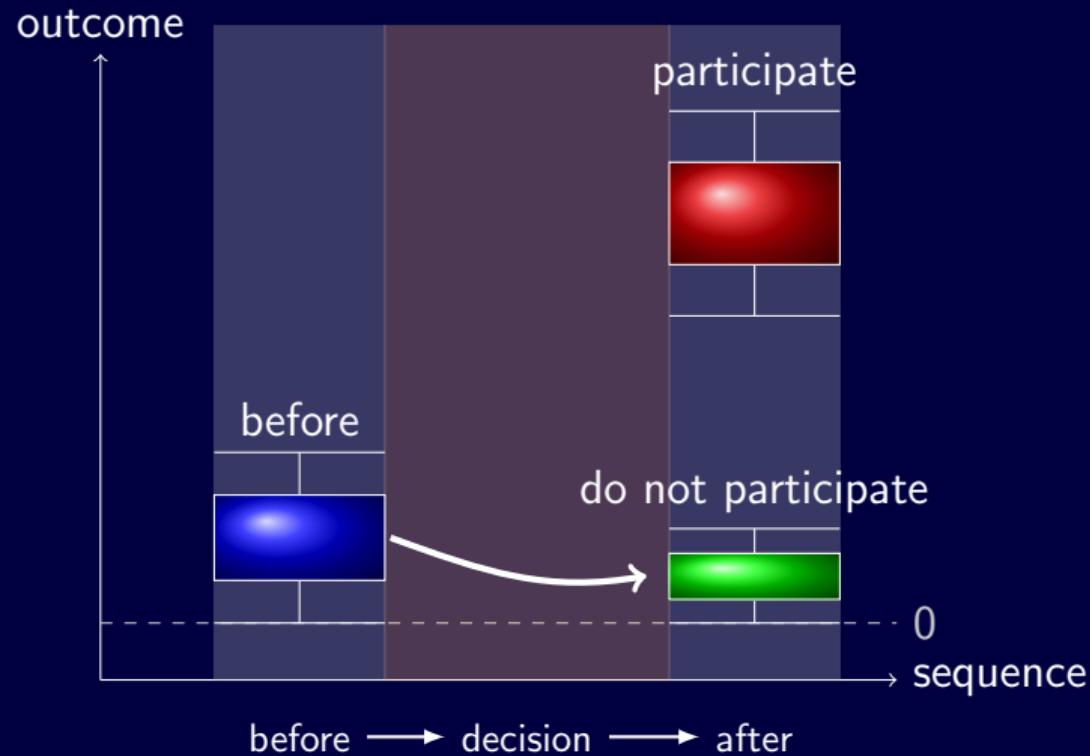
# self-selection (what we observe)



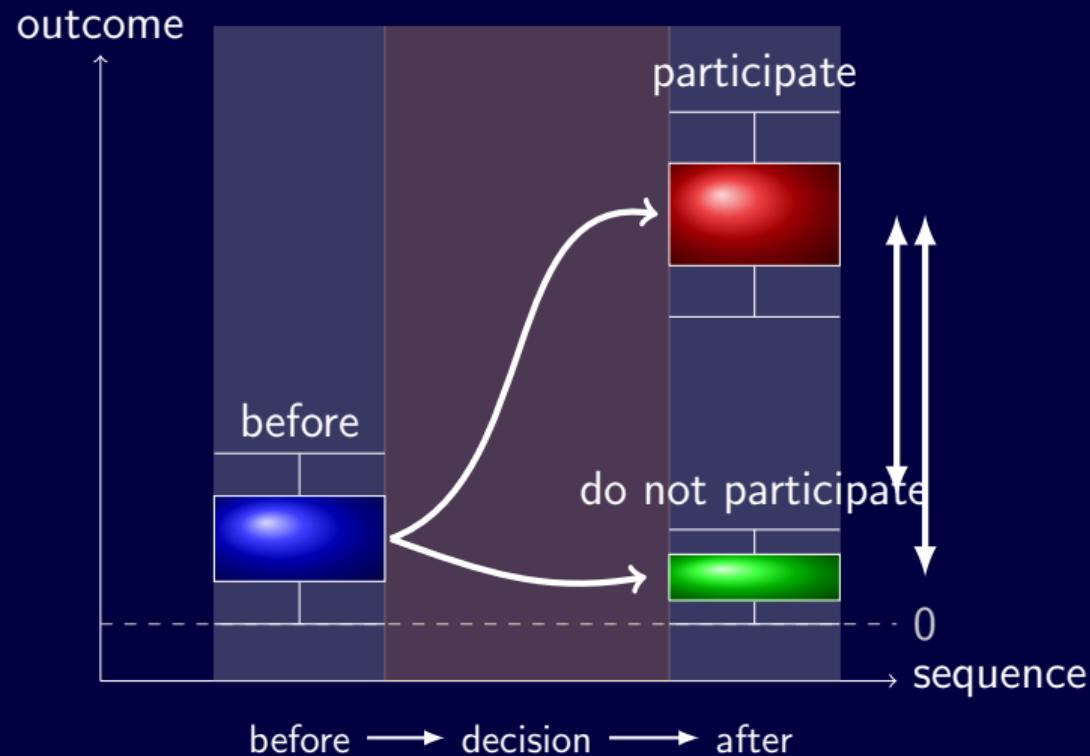
# self-selection (what we observe)



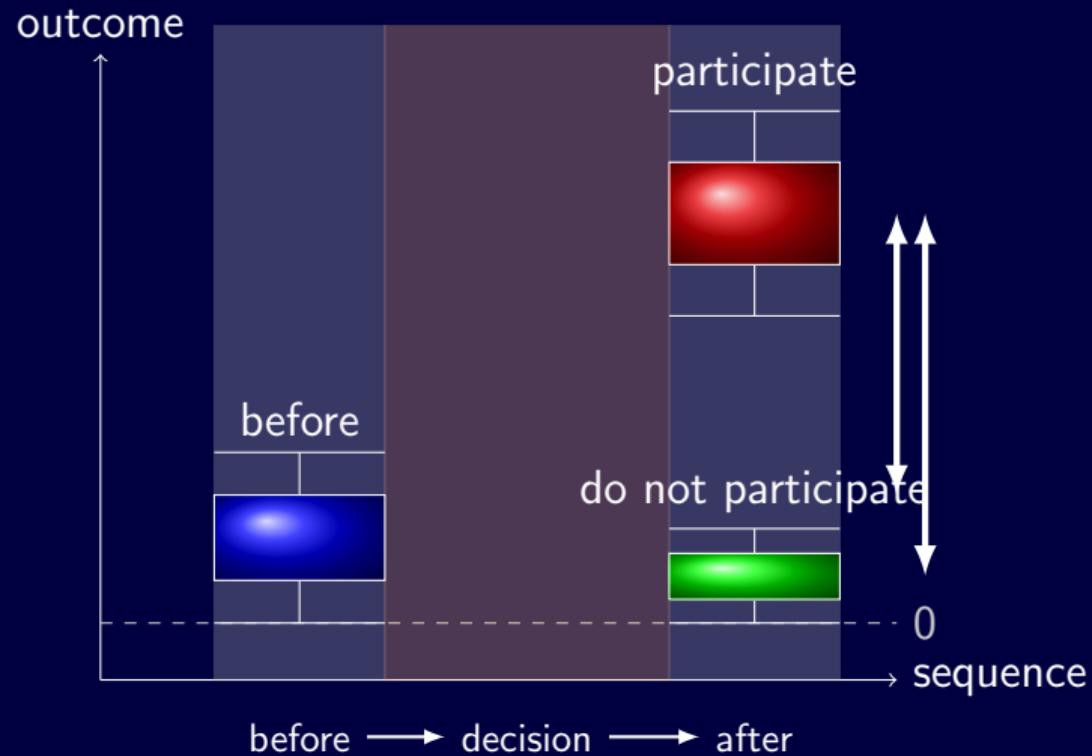
# self-selection (what we observe)



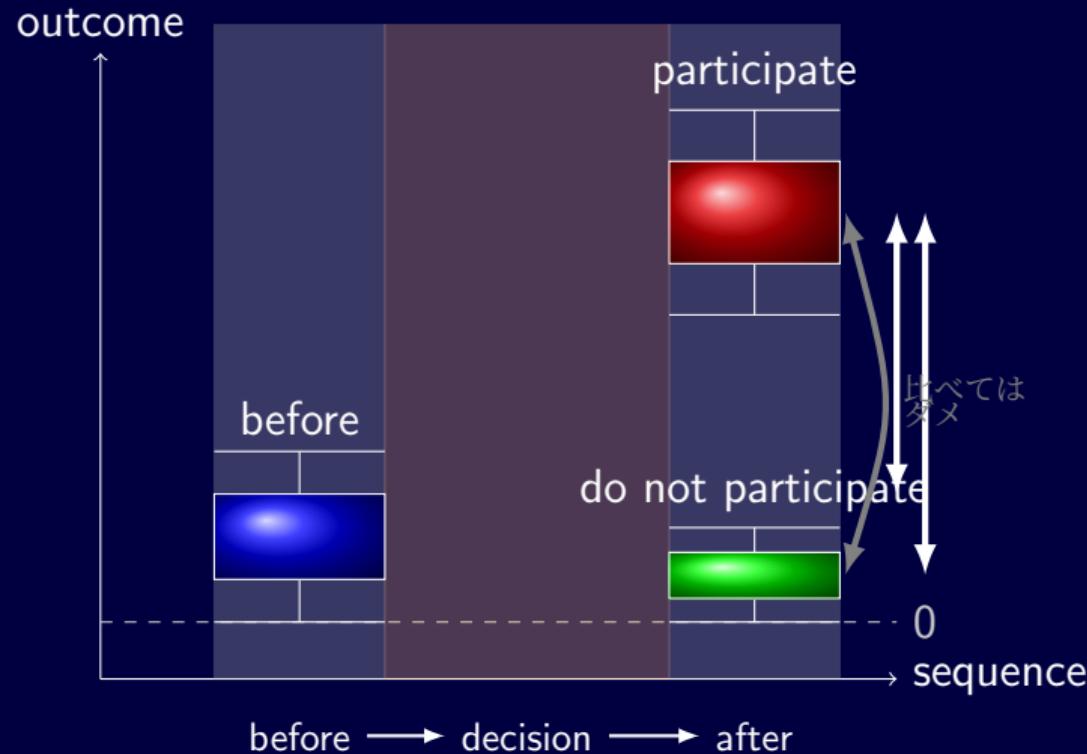
# self-selection (what we observe)



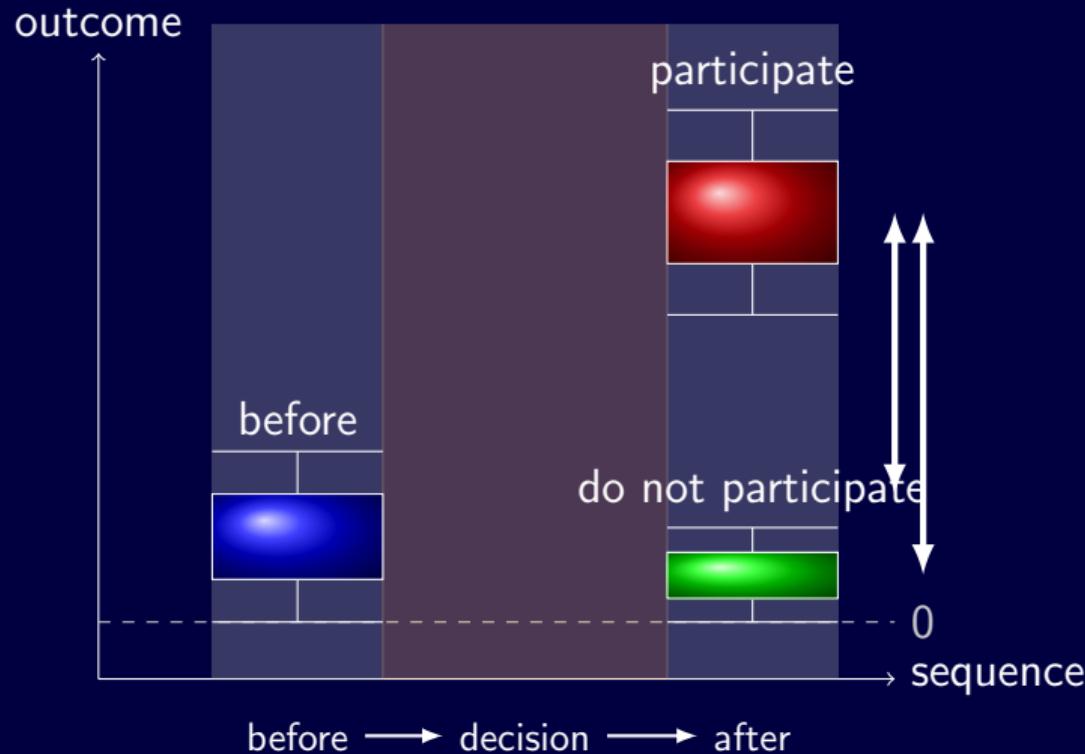
# evaluation: naïve comparison



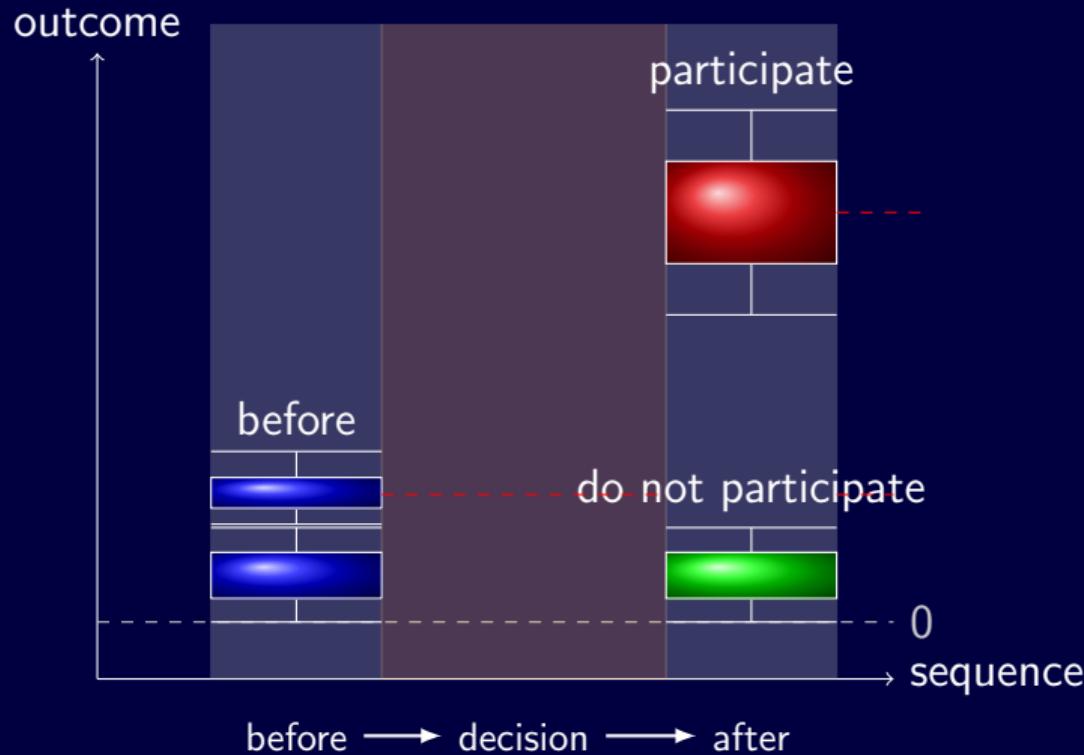
# evaluation: naïve comparison



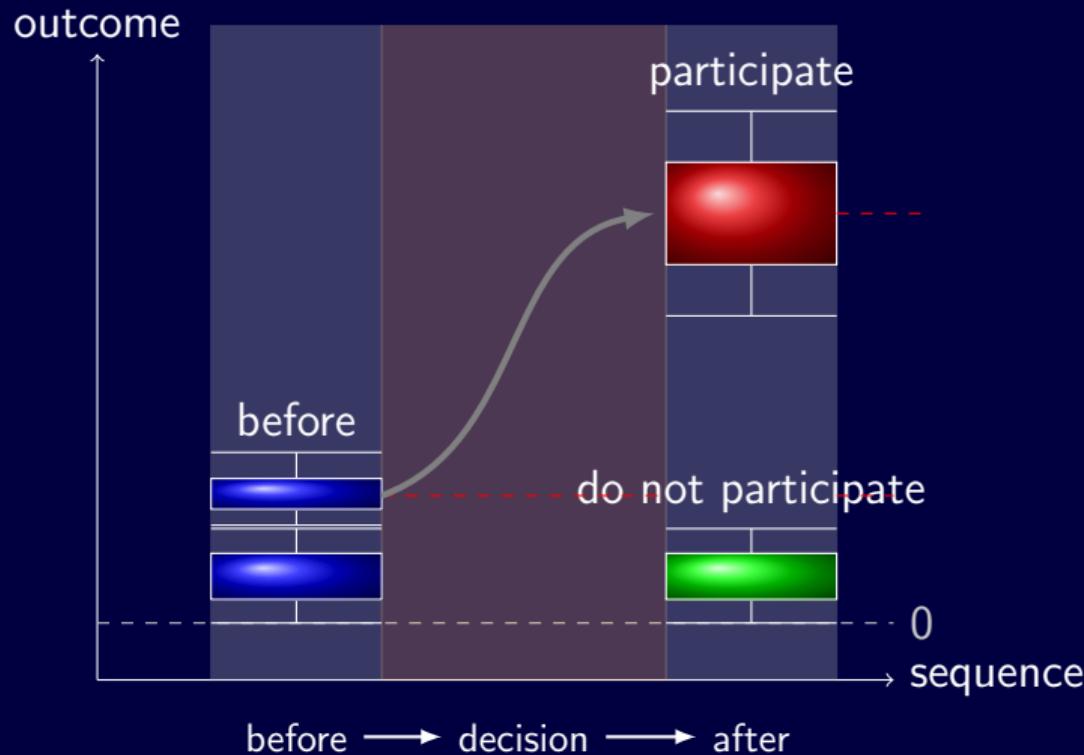
# evaluation: proper comparison



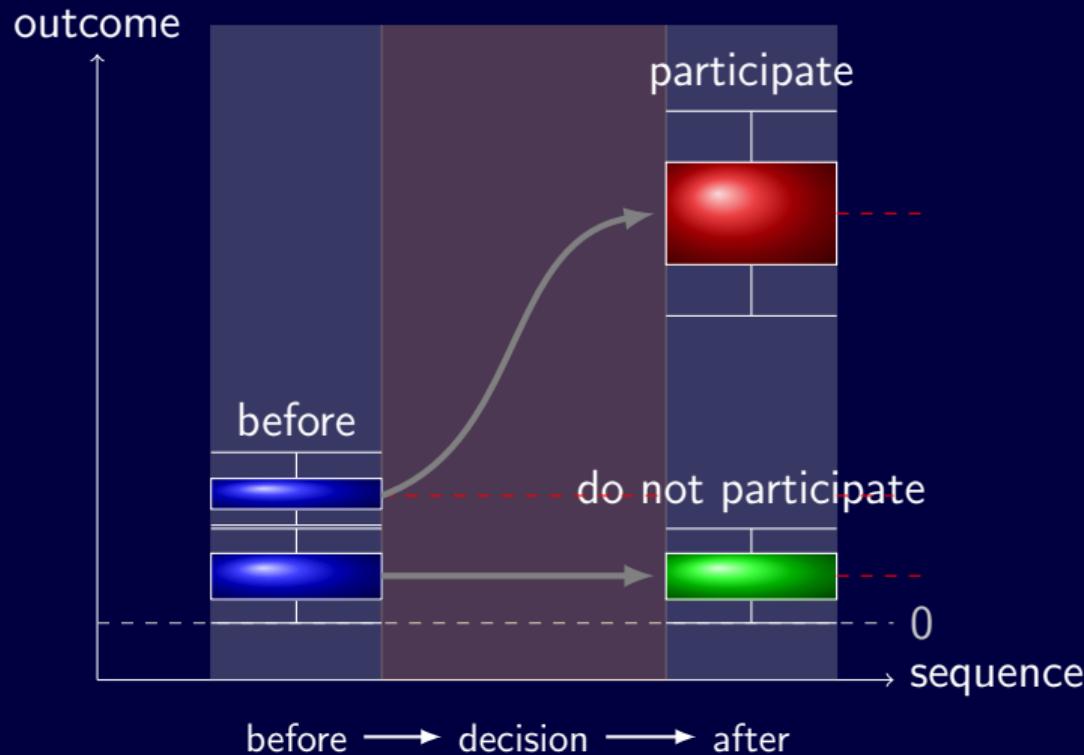
# evaluation: proper comparison



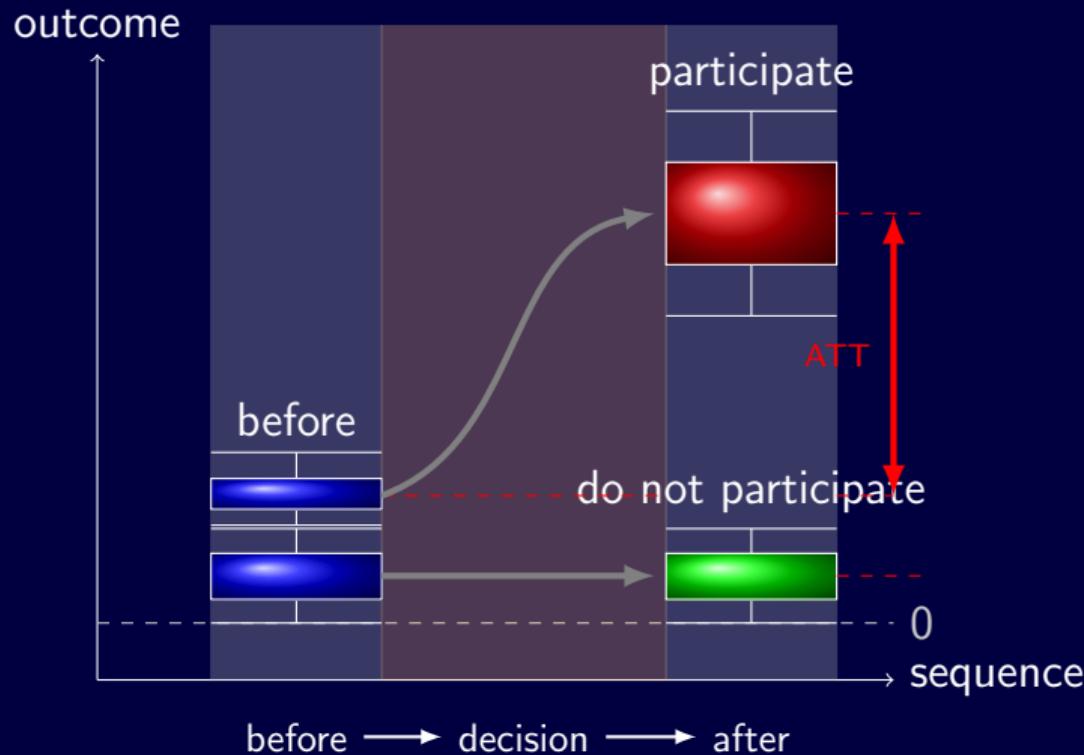
# evaluation: proper comparison



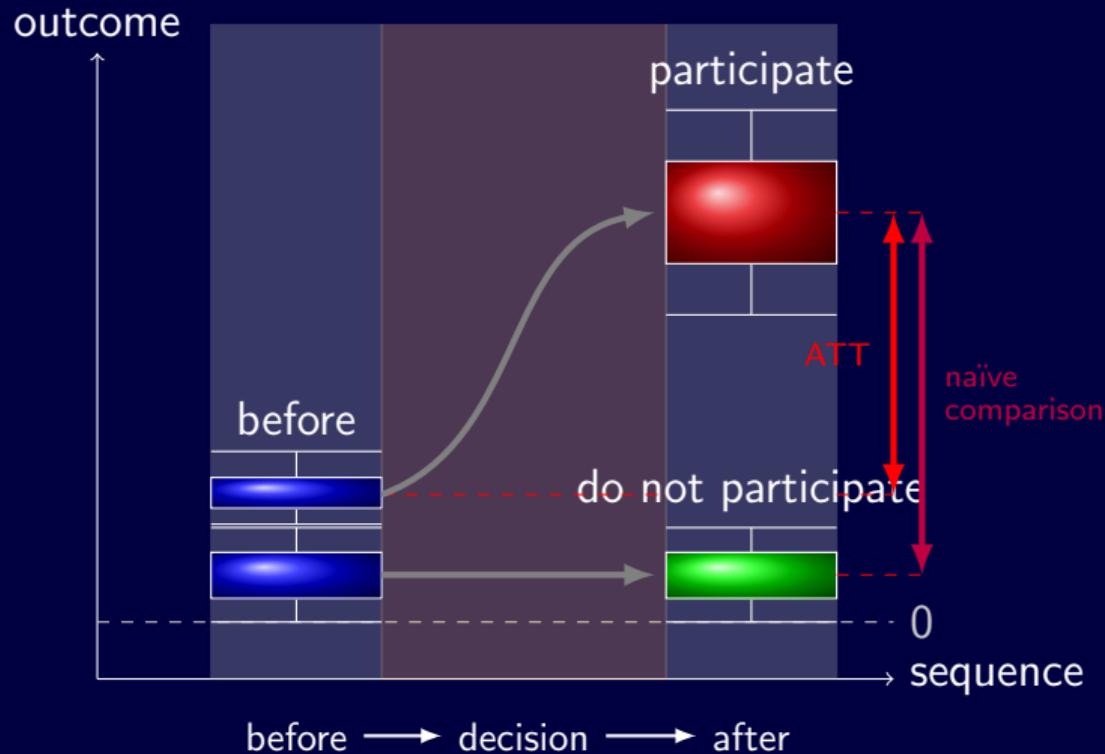
# evaluation: proper comparison



# evaluation: proper comparison



# evaluation: proper comparison



# ATC is usually not estimable

	factual	CF
ATT	treated	wants to be treated but not treated
ATC	control	wants to be not treated but treated

# ATC is usually not estimable

	factual	CF
ATT	treated	wants to be treated but not treated
ATC	control	wants to be not treated but treated

ATCは治療されたくないという意思が尊重されれば推計不可能。

# ATC is usually not estimable

	factual	CF
ATT	treated	wants to be treated but not treated
ATC	control	wants to be not treated but treated

ATCは治療されたくないという意思が尊重されれば推計不可能。

- ⇒ ATC推計には、治療を受けないことを選ぶ人たち  $D = 0$  と  $D = 0$  の人たちと似ているが治療を受けた人たちをマッチさせる必要あり。

# ATC is usually not estimable

	factual	CF
ATT	treated	wants to be treated but not treated
ATC	control	wants to be not treated but treated

ATCは治療されたくないという意思が尊重されれば推計不可能。

- ⇒ ATC推計には、治療を受けないことを選ぶ人たち  $D = 0$  と  $D = 0$  の人たちと似ているが治療を受けた人たちをマッチさせる必要あり。
- ⇒ これはあり得ない。「似た人たち」は治療を受けるはずがないから。しかし、なぜか治療を受けた。結局は似ていないのかもしれない。

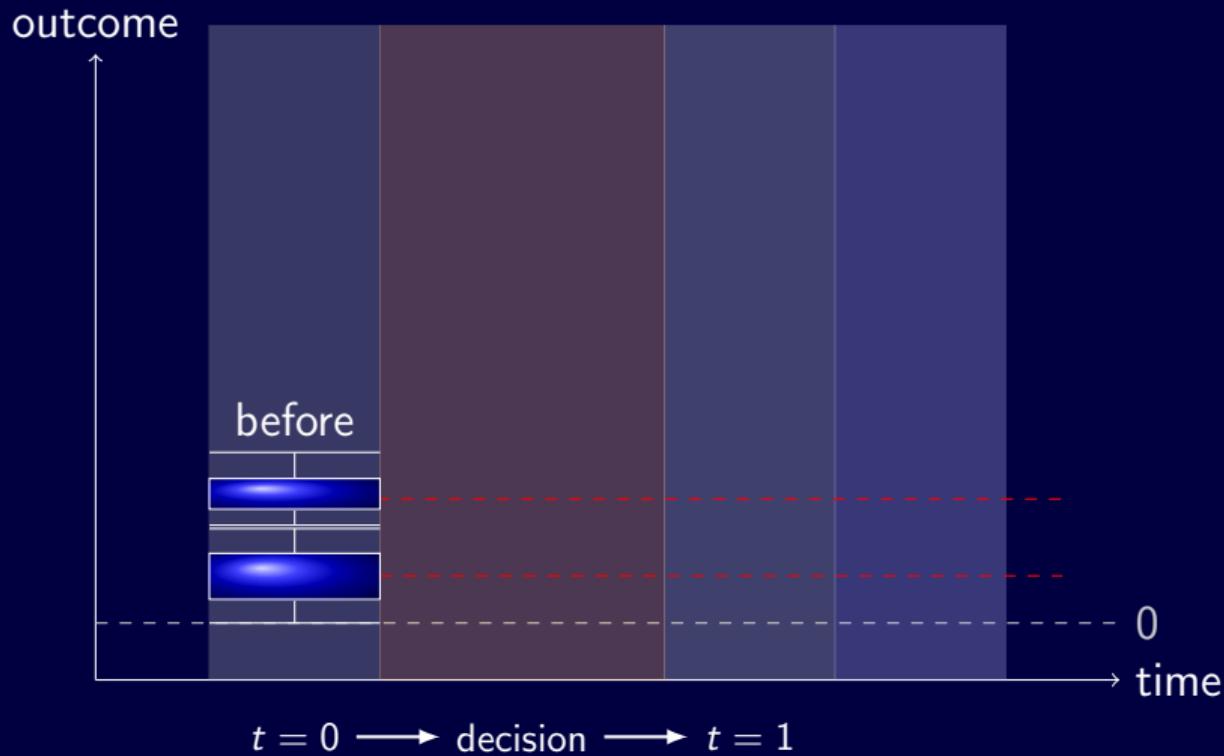
# ATC is usually not estimable

	factual	CF
ATT	treated	wants to be treated but not treated
ATC	control	wants to be not treated but treated

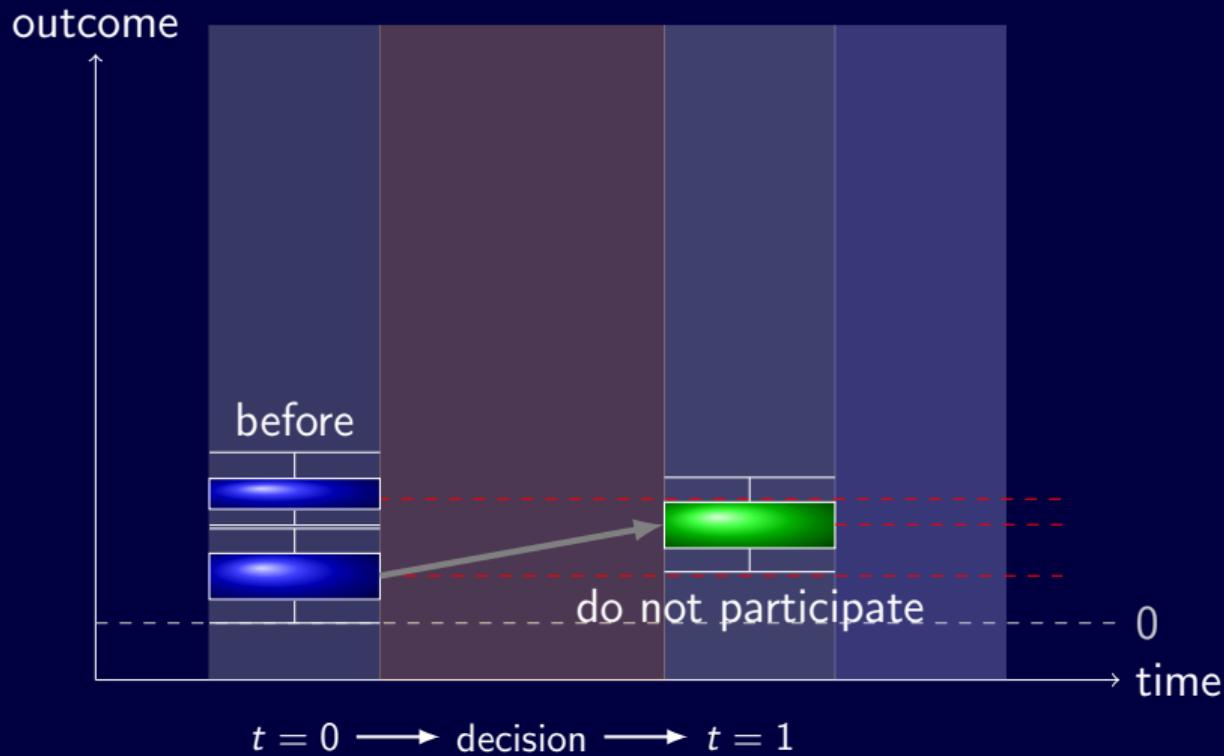
ATCは治療されたくないという意思が尊重されれば推計不可能。

- ⇒ ATC推計には、治療を受けないことを選ぶ人たち  $D = 0$  と  $D = 0$  の人たちと似ているが治療を受けた人たちをマッチさせる必要あり。
- ⇒ これはあり得ない。「似た人たち」は治療を受けるはずがないから。しかし、なぜか治療を受けた。結局は似ていないのかもしれない。
- ⇒ 例外: 天災やその他の個人には制御不能な事件・事故などによって「治療」が避けられないとき。資産損失、融資打ち切り、アルコール成分の高いビールの若年層(妊娠を知らない妊婦)向け販売解禁、など。

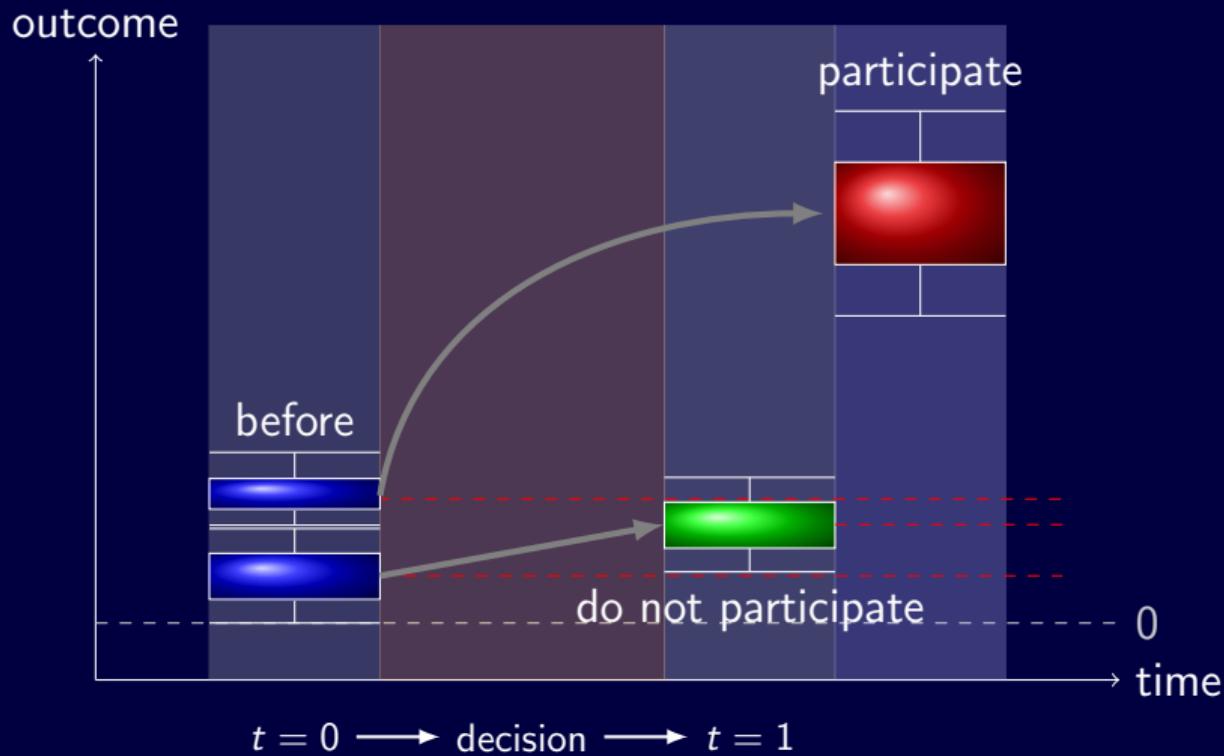
# difference-in-differences (idea)



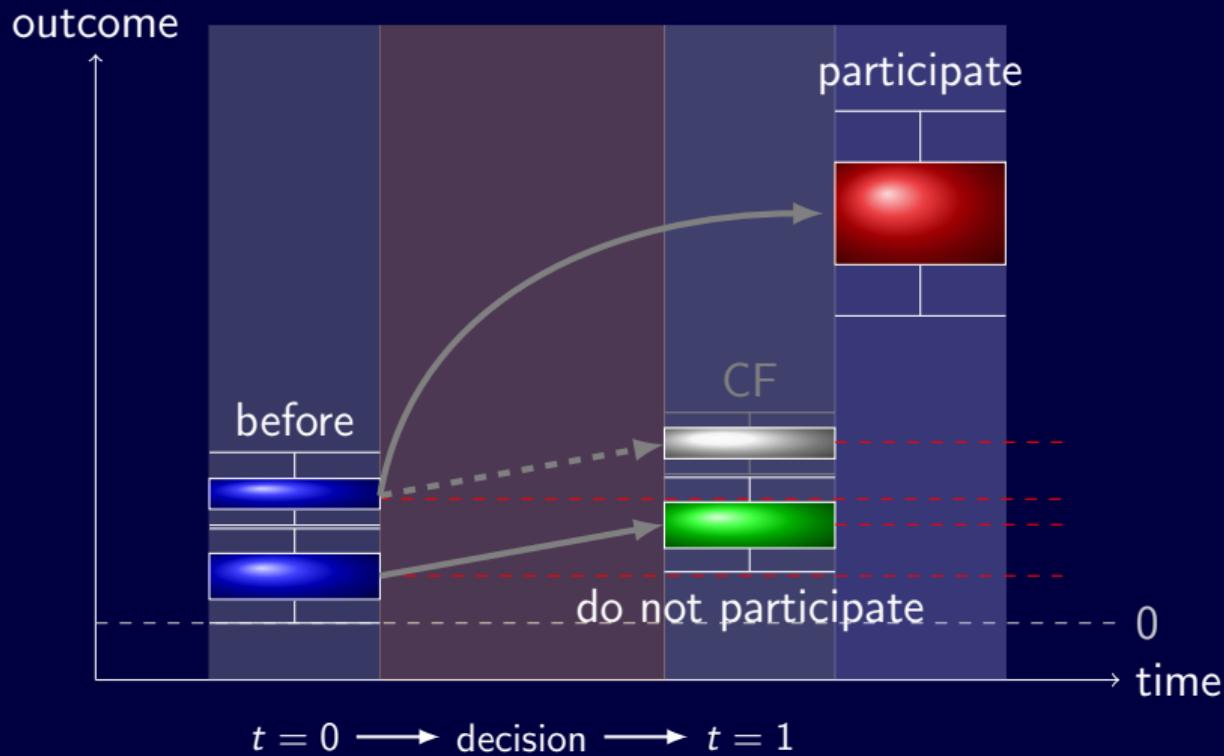
# difference-in-differences (idea)



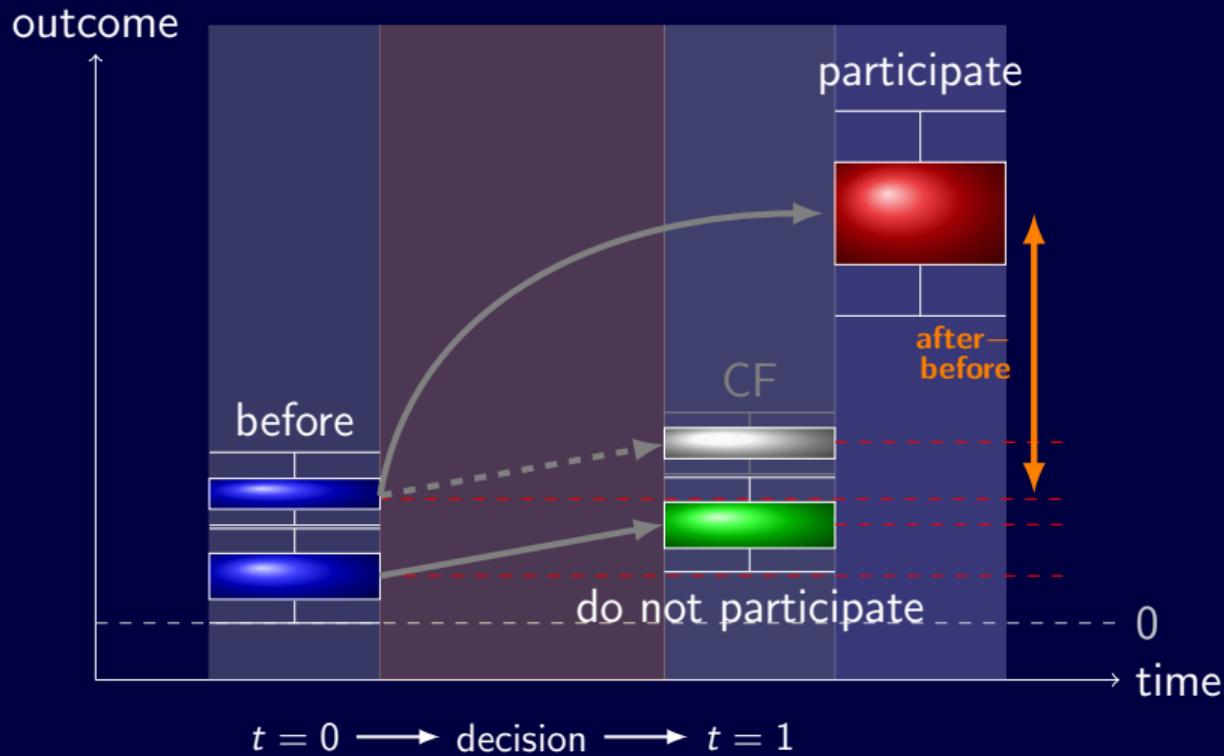
# difference-in-differences (idea)



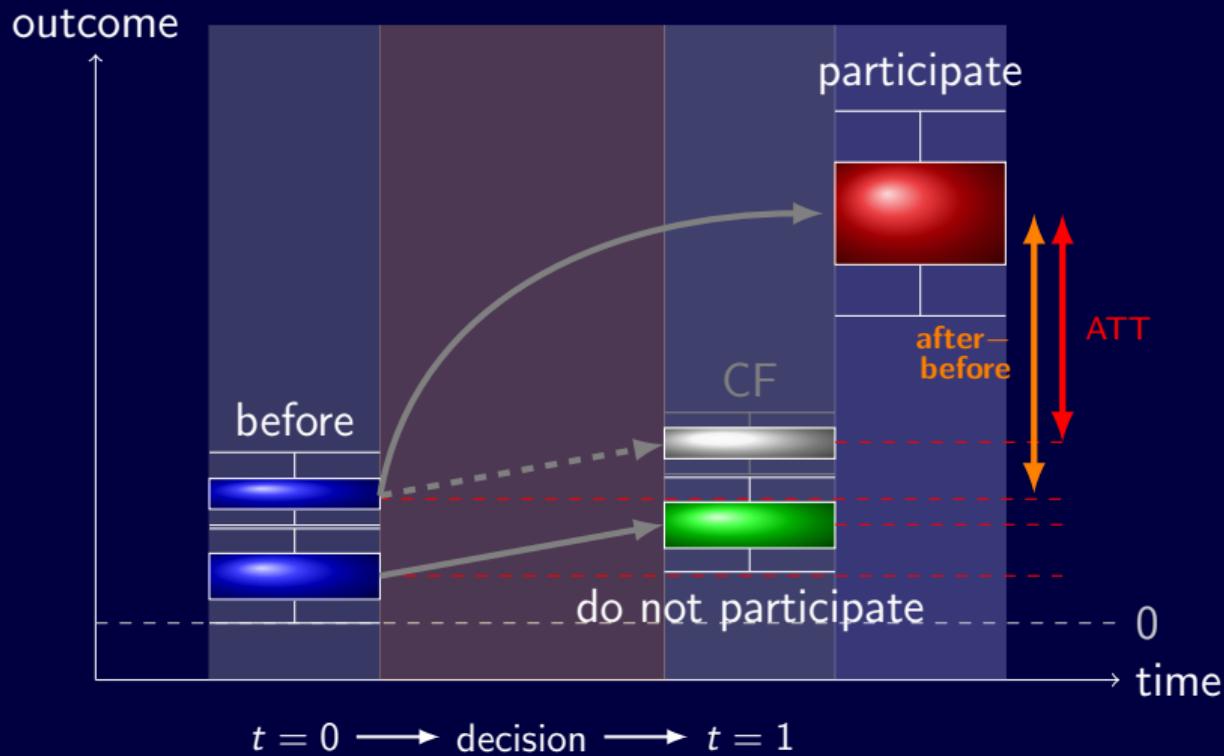
# difference-in-differences (idea)



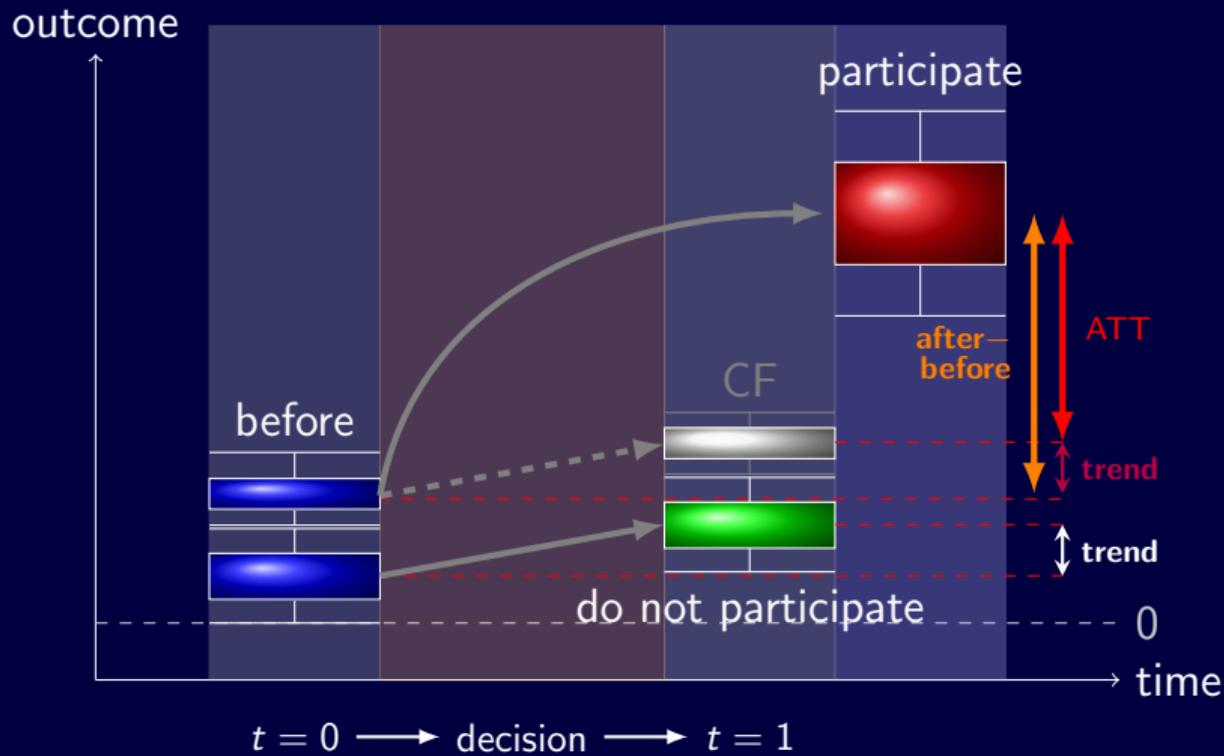
# difference-in-differences (idea)



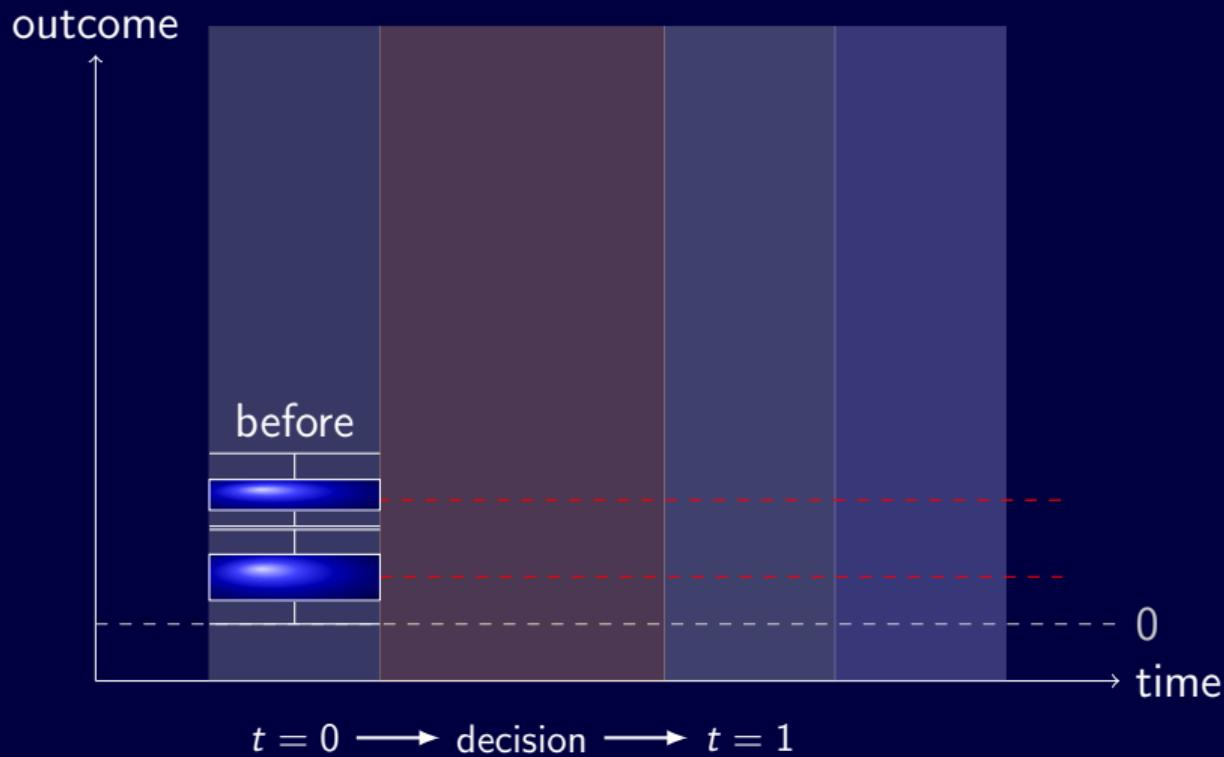
# difference-in-differences (idea)



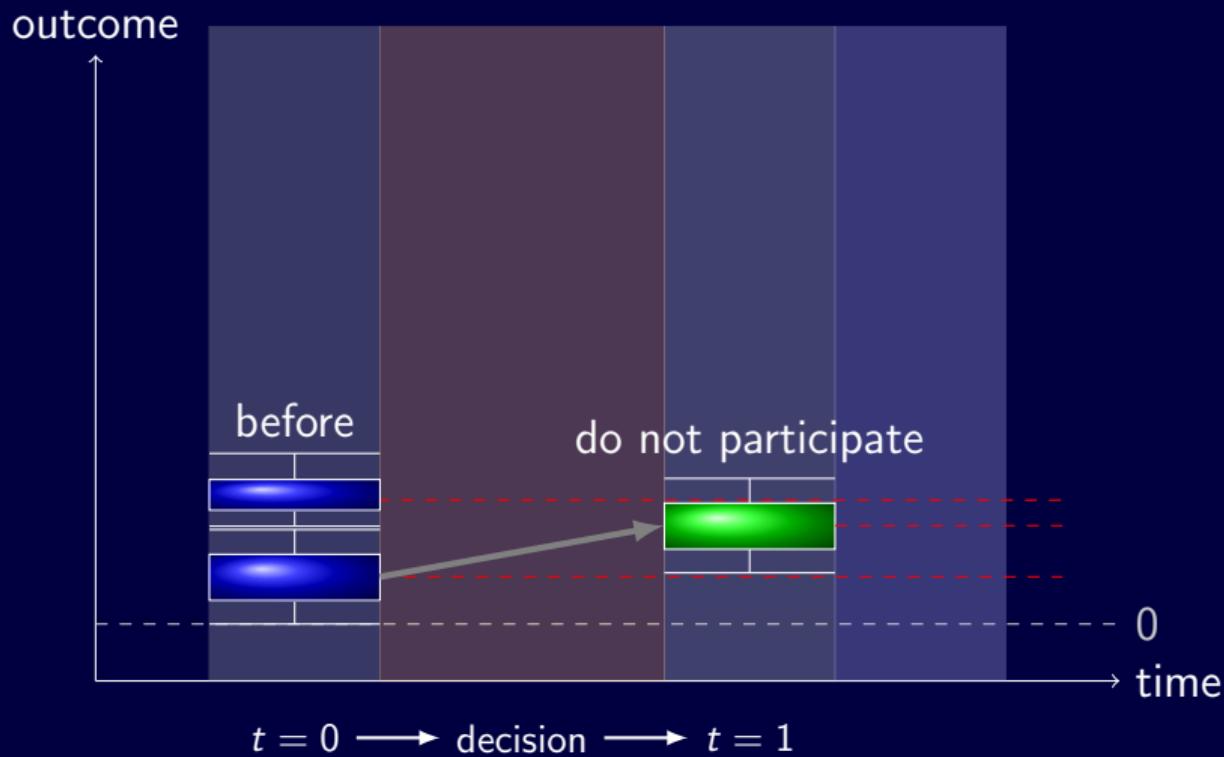
# difference-in-differences (idea)



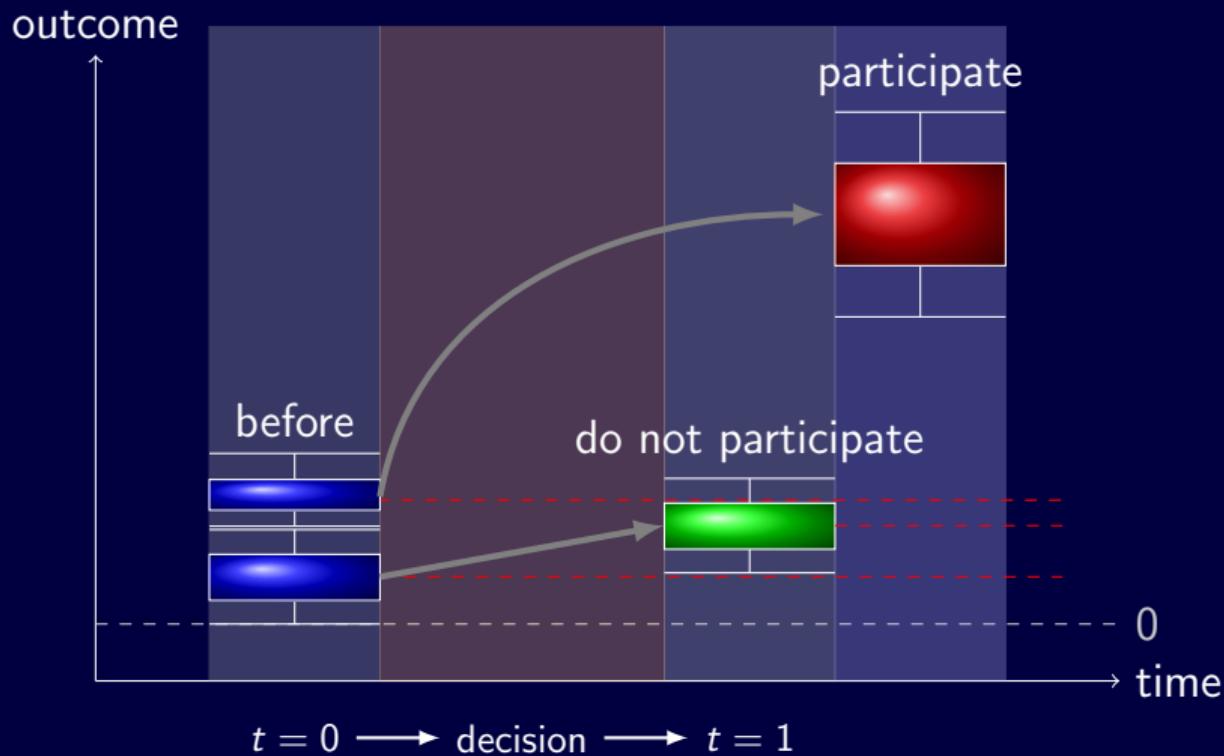
# difference-in-differences (implementation)



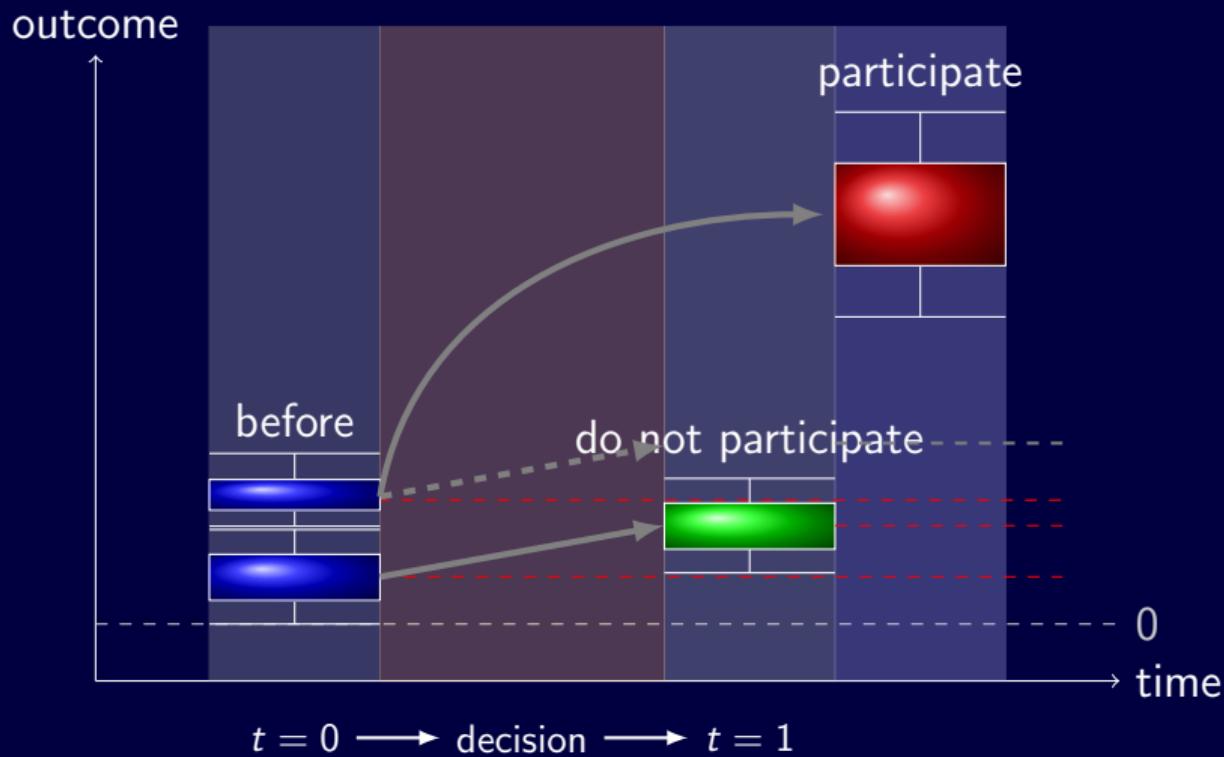
# difference-in-differences (implementation)



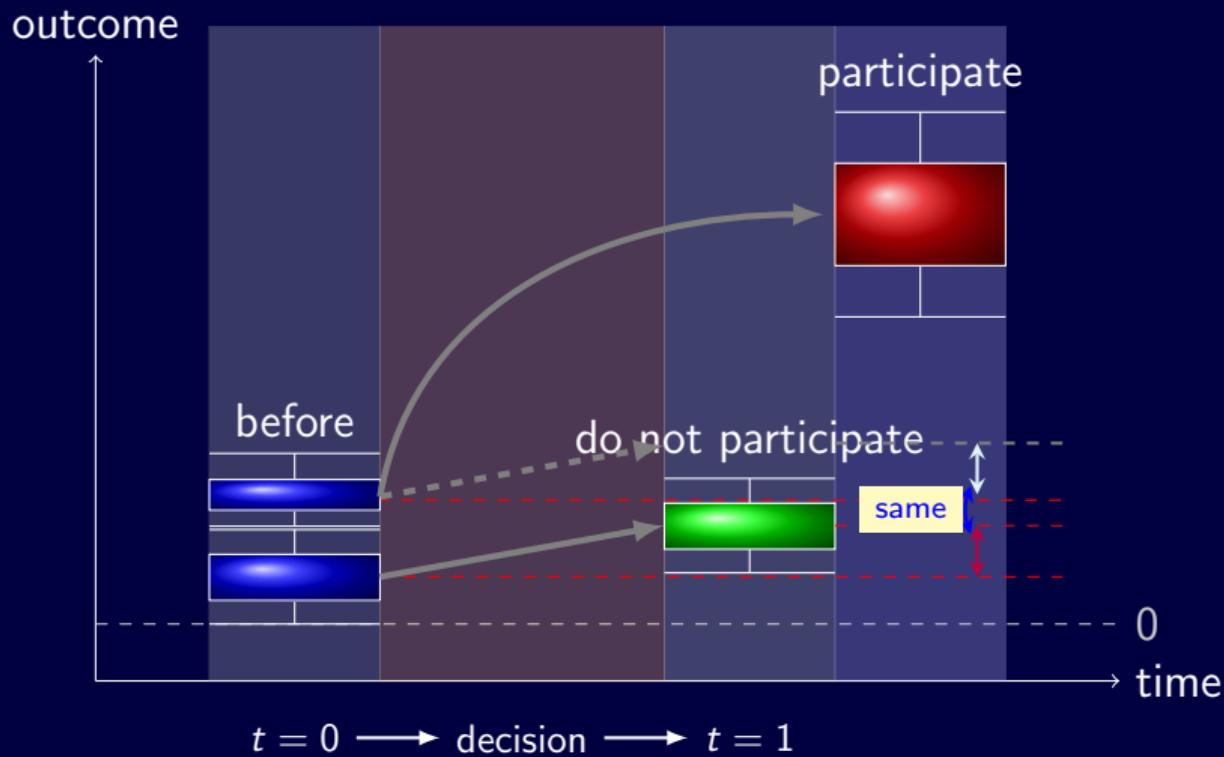
# difference-in-differences (implementation)



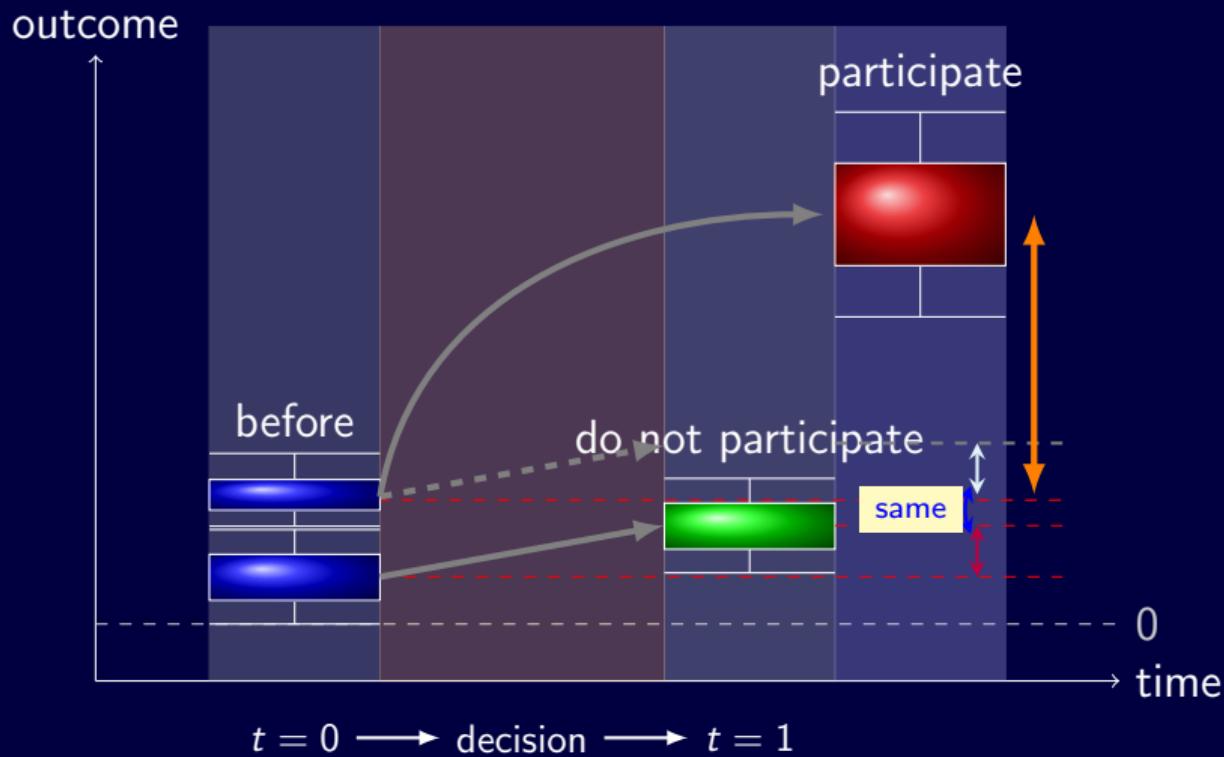
# difference-in-differences (implementation)



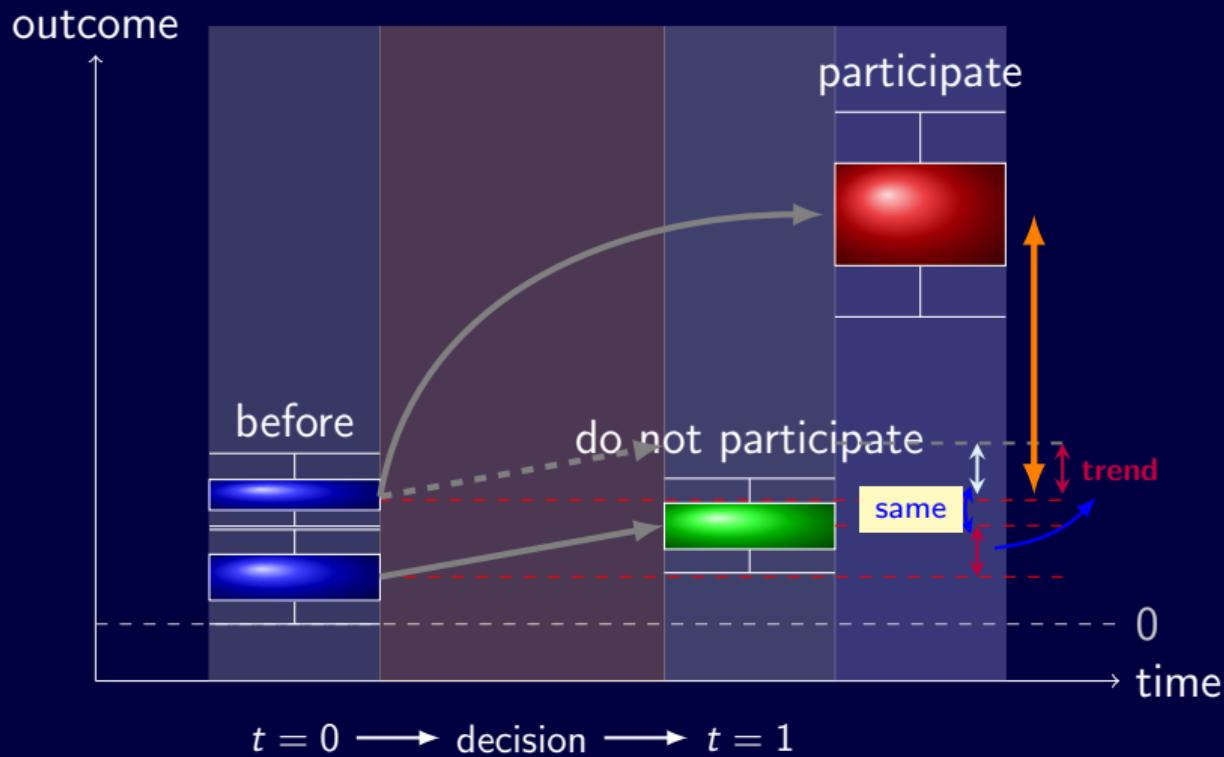
# difference-in-differences (implementation)



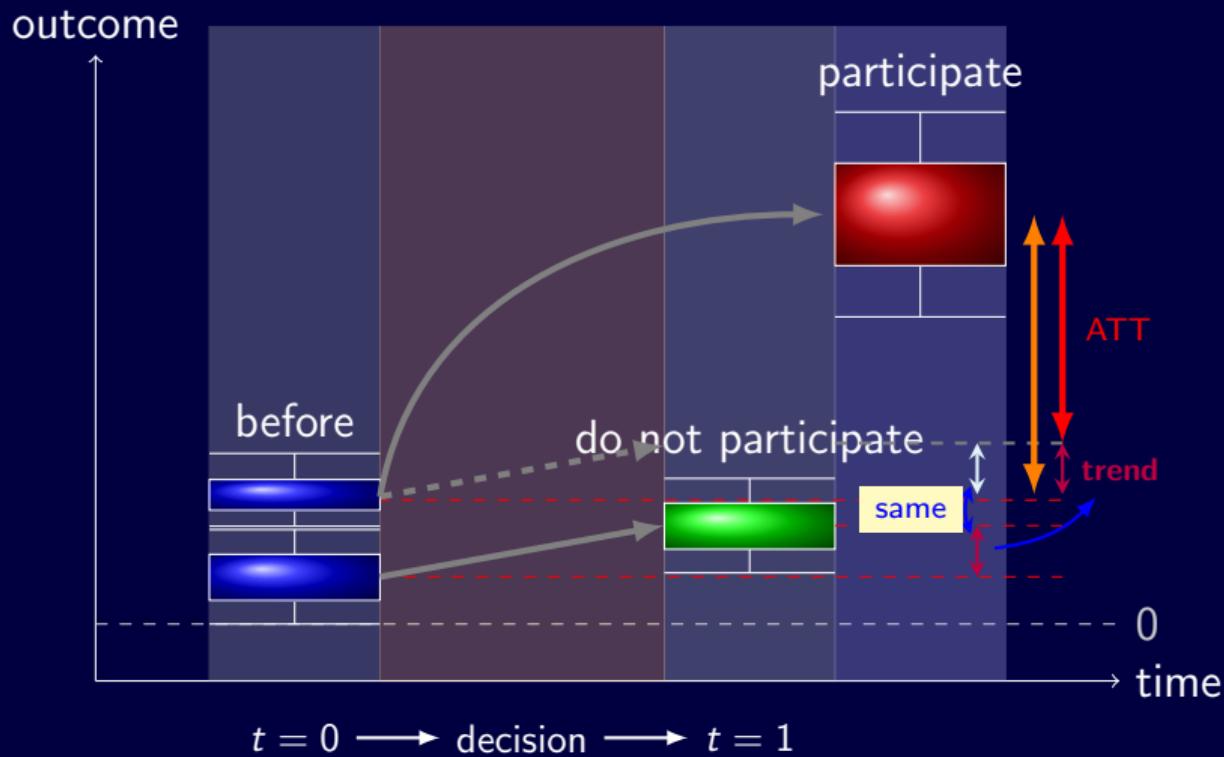
# difference-in-differences (implementation)



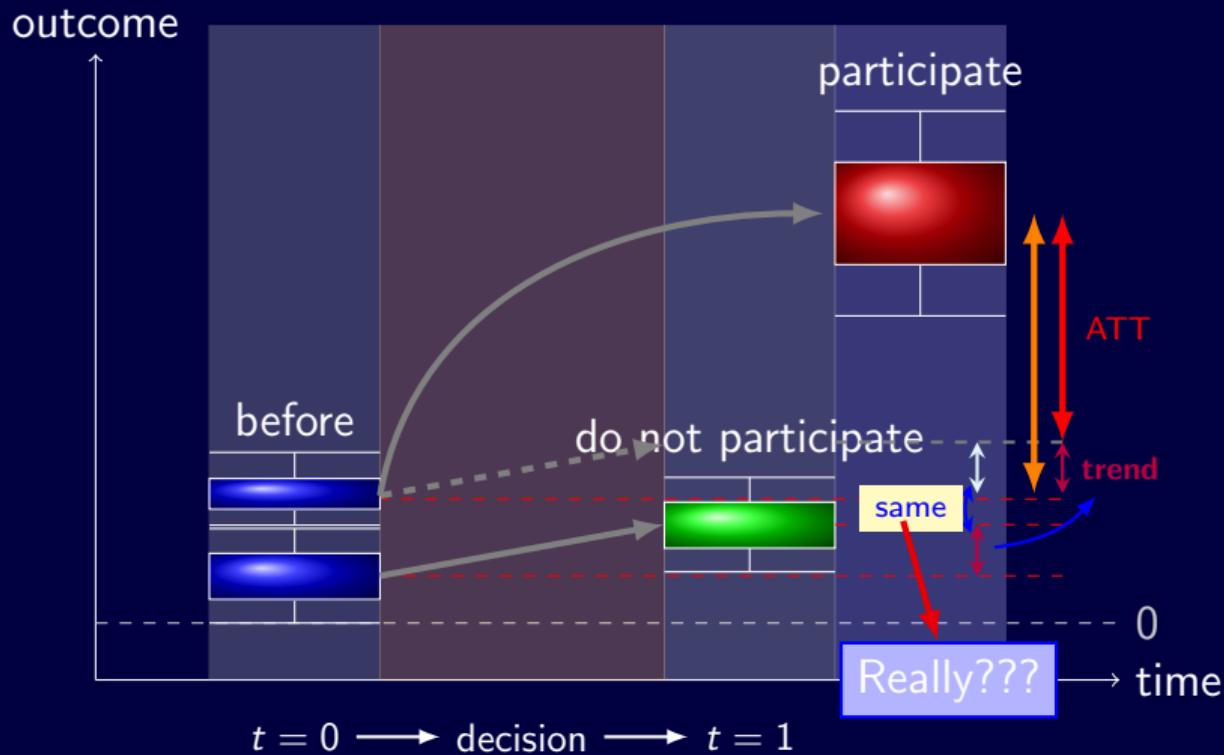
# difference-in-differences (implementation)



# difference-in-differences (implementation)



# difference-in-differences (implementation)



## Implementation:

Get data:

- Treated group outcomes  $y_{i,t}^1$  before and after the policy. If there are  $n^1$  individuals,  
 $\underbrace{y_{1,t}^1, \dots, y_{n^1,t}^1}_{\text{year } t}, \underbrace{y_{1,t+1}^1, \dots, y_{n^1,t+1}^1}_{\text{year } t+1}.$

## Implementation:

Get data:

- Treated group outcomes  $y_{i,t}^1$  before and after the policy. If there are  $n^1$  individuals,  
$$\underbrace{y_{1,t}^1, \dots, y_{n^1,t}^1}_{\text{year } t}, \underbrace{y_{1,t+1}^1, \dots, y_{n^1,t+1}^1}_{\text{year } t+1}.$$
- Control group outcomes  $y_{i,t}^0$  before and after the policy. If there are  $n^0$  individuals,  
$$\underbrace{y_{1,t}^0, \dots, y_{n^0,t}^0}_{\text{year } t}, \underbrace{y_{1,t+1}^0, \dots, y_{n^0,t+1}^0}_{\text{year } t+1}.$$

## Implementation:

Get data:

- Treated group outcomes  $y_{i,t}^1$  before and after the policy. If there are  $n^1$  individuals,  
$$\underbrace{y_{1,t}^1, \dots, y_{n^1,t}^1}_{\text{year } t}, \underbrace{y_{1,t+1}^1, \dots, y_{n^1,t+1}^1}_{\text{year } t+1}.$$
- Control group outcomes  $y_{i,t}^0$  before and after the policy. If there are  $n^0$  individuals,  
$$\underbrace{y_{1,t}^0, \dots, y_{n^0,t}^0}_{\text{year } t}, \underbrace{y_{1,t+1}^0, \dots, y_{n^0,t+1}^0}_{\text{year } t+1}.$$
- ☞ One needs individual level data, not just group level averages, to do inferences (=compute  $p$  values). For testing a null hypothesis, one needs standard errors of the estimates, which need variances and covariances, which can only be computed with individual level data.
- Let us denote the smaller of  $n^1, n^0$  as  $n^{min}$ .

Steps:

- ① Compute before and after means for both groups.

$$\bar{y}_t^1 = \frac{y_{1,t}^1 + \cdots + y_{n^1,t}^1}{n^1} = \frac{\sum_{i=1}^{n^1} y_{i,t}^1}{n^1}, \quad \bar{y}_t^0 = \frac{\sum_{i=1}^{n^0} y_{i,t}^0}{n^0}, \quad \bar{y}_{t+1}^1, \quad \bar{y}_{t+1}^0.$$

Steps:

- ① Compute before and after means for both groups.

$$\bar{y}_t^1 = \frac{y_{1,t}^1 + \cdots + y_{n^1,t}^1}{n^1} = \frac{\sum_{i=1}^{n^1} y_{i,t}^1}{n^1}, \quad \bar{y}_t^0 = \frac{\sum_{i=1}^{n^0} y_{i,t}^0}{n^0}, \quad \bar{y}_{t+1}^1, \quad \bar{y}_{t+1}^0.$$

- ② Compute difference-in-differences:  $y^{DID} = (\bar{y}_{t+1}^1 - \bar{y}_t^1) - (\bar{y}_{t+1}^0 - \bar{y}_t^0)$ .

Steps:

- ① Compute before and after means for both groups.

$$\bar{y}_t^1 = \frac{y_{1,t}^1 + \cdots + y_{n^1,t}^1}{n^1} = \frac{\sum_{i=1}^{n^1} y_{i,t}^1}{n^1}, \quad \bar{y}_t^0 = \frac{\sum_{i=1}^{n^0} y_{i,t}^0}{n^0}, \quad \bar{y}_{t+1}^1, \quad \bar{y}_{t+1}^0.$$

- ② Compute difference-in-differences:  $y^{DID} = (\bar{y}_{t+1}^1 - \bar{y}_t^1) - (\bar{y}_{t+1}^0 - \bar{y}_t^0)$ .
- ③ Compute the standard deviation of  $y^{DID}$  with:

$$\sigma^{DID} = \sqrt{\mathcal{V} [\bar{y}_{t+1}^1 - \bar{y}_t^1 - \bar{y}_{t+1}^0 + \bar{y}_t^0]}$$

where  $\mathcal{V} [\bar{y}_{t+1}^1 - \bar{y}_t^1 - \bar{y}_{t+1}^0 + \bar{y}_t^0]$  equals to

$$\begin{aligned} & \frac{\hat{\sigma}_{y_{t+1}^1}^2}{n^1} + \frac{\hat{\sigma}_{y_t^1}^2}{n^1} + \frac{\hat{\sigma}_{y_{t+1}^0}^2}{n^0} + \frac{\hat{\sigma}_{y_{t+1}^0}^2}{n^0} - 2\frac{\widehat{\text{cov}}[y_{t+1}^1, y_t^1]}{n^1} - 2\frac{\widehat{\text{cov}}[y_{t+1}^1, y_{t+1}^0]}{n^{\min}} + 2\frac{\widehat{\text{cov}}[y_{t+1}^1, y_t^0]}{n^{\min}} \\ & + 2\frac{\widehat{\text{cov}}[y_t^1, y_{t+1}^0]}{n^{\min}} - 2\frac{\widehat{\text{cov}}[y_t^1, y_t^0]}{n^{\min}} - 2\frac{\widehat{\text{cov}}[y_{t+1}^0, y_t^0]}{n^0}. \end{aligned}$$

## Steps:

- ④ Compute  $p$  value. Using R,  $\text{pt}(y^{DID}/\sigma^{DID}, n^{min}, \text{lower.tail} = \text{F})$  gives the  $p$  value of the null hypothesis of zero effect.

Steps:

- ④ Compute  $p$  value. Using R,  $\text{pt}(y^{DID}/\sigma^{DID}, n^{min}, \text{lower.tail} = F)$  gives the  $p$  value of the null hypothesis of zero effect.

A simpler way: It can be shown that  $y^{DID} = a_3$  in the following regression.

- ① Regress  $y_{i,t} = a_0 + a_1 D_i + a_2 \text{after}_{t+1} + a_3 D_i * \text{after}_{t+1} + e_{i,t}$  using  $t, t+1$  data.

- $D_i$ : Treatment dummy variable.  $D_i = \begin{cases} 1 & \text{if treated} \\ 0 & \text{control} \end{cases}$
- $\text{after}_{t+1}$ : 'after' dummy variable.  $\text{after}_{t+1} = \begin{cases} 1 & \text{if } t+1 \\ 0 & t \end{cases}$

- ② Test  $a_3 = 0$  and get its  $p$  value.

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

It:

- 水準は違ってもいい(だから、水準の分布は似ていなくてよい)

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

It:

- 水準は違ってもいい(だから、水準の分布は似ていなくてよい)

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

It:

- 水準は違ってもいい(だから、水準の分布は似ていなくてよい)

群同士が似ていると仮定するよりも現実的

- 低所得者: 消費の分布は低位ばかり

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

It:

- 水準は違ってもいい(だから、水準の分布は似ていなくてよい)

群同士が似ていると仮定するよりも現実的

- 低所得者: 消費の分布は低位ばかり

- 高所得者: 消費の分布は高位ばかり

DID の識別仮定: 介入がないとき、結果指標の変化が群間で似ていること。共通トレンドの仮定 **common trend assumption**.

It:

- 水準は違ってもいい(だから、水準の分布は似ていなくてよい)

群同士が似ていると仮定するよりも現実的

- 低所得者: 消費の分布は低位ばかり

- 高所得者: 消費の分布は高位ばかり

- 消費水準の分布は異なるかもしれないが、消費変化の分布は似ている可能性あり

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

For DID we need:

- Before-after data.

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

For DID we need:

- Before-after data.
- For both the treated and the control.

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

For DID we need:

- Before-after data.
- For both the treated and the control.

## List of identifying assumptions.

In the absence of the policy:  $\delta(A)$  reads “distribution of  $A$ ”

With-without  $\delta(y_{i1}) \simeq \delta(y_{i0})$ .

Before-after  $\delta(y_{i,\text{after}}) \simeq \delta(y_{i,\text{before}})$ .

DID  $\delta(\Delta y_{i1}) \simeq \delta(\Delta y_{i0})$ .

For DID we need:

- Before-after data.
- For both the treated and the control.

You need to get the control data even before the policy starts.

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差
  - ☞ 貸付がないとき、所得分布は各群で似ているか？

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差
  - ☞ 貸付がないとき、所得分布は各群で似ているか？
  - ☞ ほぼ似ていない（はず）。メンバーは自己選抜 *self-select* して加入するため、メンバーだけの特徴が何かある（はず）。（連帯責任制の場合、メンバーは相互に選び合う）この加入過程は所得稼得能力の分布が各群で異なることを示唆している。そうであれば、介入がないときの所得の分布も各群で異なる（はず）。

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差
  - ☞ 貸付がないとき、所得分布は各群で似ているか？
  - ☞ ほぼ似ていない（はず）。メンバーは自己選抜 *self-select* して加入するため、メンバーだけの特徴が何かある（はず）。（連帯責任制の場合、メンバーは相互に選び合う）この加入過程は所得稼得能力の分布が各群で異なることを示唆している。そうであれば、介入がないときの所得の分布も各群で異なる（はず）。
  - ☞ CF = 借り入れしていない状態のメンバーの所得

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差
  - ▣ 貸付がないとき、所得分布は各群で似ているか？
  - ▣ ほぼ似ていない（はず）。メンバーは自己選抜 *self-select* して加入するため、メンバーだけの特徴が何かある（はず）。（連帯責任制の場合、メンバーは相互に選び合う）この加入過程は所得稼得能力の分布が各群で異なることを示唆している。そうであれば、介入がないときの所得の分布も各群で異なる（はず）。
  - ▣ CF = 借り入れしていない状態のメンバーの所得
- グラミン銀行のメンバーからランダムにサンプル=the treated、非メンバーからランダムにサンプル=the control、貸付の所得への ATE=各群の平均所得の差

## With-without

- グラミン銀行メンバー=the treated、非メンバー=the control、マイクロファイナンス貸付の所得への ATE=各群の所得の平均値の差
  - ▣ 貸付がないとき、所得分布は各群で似ているか？
  - ▣ ほぼ似ていない（はず）。メンバーは自己選抜 *self-select* して加入するため、メンバーだけの特徴が何かある（はず）。（連帯責任制の場合、メンバーは相互に選び合う）この加入過程は所得稼得能力の分布が各群で異なることを示唆している。そうであれば、介入がないときの所得の分布も各群で異なる（はず）。
  - ▣ CF = 借り入れしていない状態のメンバーの所得
- グラミン銀行のメンバーからランダムにサンプル=the treated、非メンバーからランダムにサンプル=the control、貸付の所得への ATE=各群の平均所得の差
  - ▣ ランダム・サンプリング ≠ (貸付の) ランダム割当

## Before-after:

- とある集団で、運動開始後の体重=the treated、運動開始前の体重=the control、運動の体重への ATE=各群の平均体重の差

## Before-after:

- とある集団で、運動開始後の体重=the treated、運動開始前の体重=the control、運動の体重へのATE=各群の平均体重の差
  - 運動をしないとき、運動開始前と運動開始後の体重の分布は似ているか？

## Before-after:

- とある集団で、運動開始後の体重=the treated、運動開始前の体重=the control、運動の体重への ATE=各群の平均体重の差
  - ☞ 運動をしないとき、運動開始前と運動開始後の体重の分布は似ているか？
  - ☞ 体重が時間とともに変化するときは似ていない。冬に運動を開始すれば、たくさん食べることで体重が増えるため、運動の効果は過少推計される underestimated。

## Before-after:

- とある集団で、運動開始後の体重=the treated、運動開始前の体重=the control、運動の体重への ATE=各群の平均体重の差
  - ☞ 運動をしないとき、運動開始前と運動開始後の体重の分布は似ているか？
  - ☞ 体重が時間とともに変化するときは似ていない。冬に運動を開始すれば、たくさん食べることで体重が増えるため、運動の効果は過少推計される underestimated。
  - ☞ CF = 運動する人たちが運動しないときの体重

## Before-after:

- 技術支援前の時期の穀物反収=the control、技術支援後の時期の穀物反収=the treated、技術支援の穀物反収への ATE=各群の平均反収の差

## Before-after:

- 技術支援前の時期の穀物反収=the control、技術支援後の時期の穀物反収=the treated、技術支援の穀物反収への ATE=各群の平均反収の差
  - ☞ 技術支援がないときに、支援前の時期と支援後の時期で反収分布は似ているか？

## Before-after:

- 技術支援前の時期の穀物反収=the control、技術支援後の時期の穀物反収=the treated、技術支援の穀物反収への ATE=各群の平均反収の差
  - ▣ 技術支援がないときに、支援前の時期と支援後の時期で反収分布は似ているか？
  - ▣ おそらく似ていない。穀物生産は気候などの時間を通じて変わることによるショックに曝されるため。反収変化がショックによる部分と技術支援による部分に分離することは難しい。

## Before-after:

- 技術支援前の時期の穀物反収=the control、技術支援後の時期の穀物反収=the treated、技術支援の穀物反収への ATE=各群の平均反収の差
  - ▣ 技術支援がないときに、支援前の時期と支援後の時期で反収分布は似ているか?
  - ▣ おそらく似ていない。穀物生産は気候などの時間を通じて変わることによるショックに曝されるため。反収変化がショックによる部分と技術支援による部分に分離することは難しい。
  - ▣ 支援前のパネル・データ(支援2期前から支援1期前など)を使って反収のトレンドを計算し、このトレンドが続くと仮定すれば、ショック・トレンドによる部分と技術支援による部分に分解できる。

アイサイトの効果...?



## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差.

## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差.
  - ☞ アイサイト非搭載時に、アイサイト搭載車とアイサイト非搭載車で事故の確率分布は似ているか?

## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差。
  - ☞ アイサイト非搭載時に、アイサイト搭載車とアイサイト非搭載車で事故の確率分布は似ているか？
  - ☞ おそらく似ていない。アイサイトはハイエンドの車種に搭載され、そうした車を運転する人たちは系統的に異なる（より高所得のはず）。このため、運転の仕方や事故確率も異なる（はず）。

## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差。
  - ☞ アイサイト非搭載時に、アイサイト搭載車とアイサイト非搭載車で事故の確率分布は似ているか？
  - ☞ おそらく似ていない。アイサイトはハイエンドの車種に搭載され、そうした車を運転する人たちは系統的に異なる（より高所得のはず）。このため、運転の仕方や事故確率も異なる（はず）。
  - ☞ アイサイト非搭載時にハイエンド車種の運転手は安全運転をする傾向があるか？おそらく安全運転をする傾向が強い、失うものが多いから。この場合、アイサイト搭載の事故への効果は過大評価される overestimated。

## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差。
  - ☞ アイサイト非搭載時に、アイサイト搭載車とアイサイト非搭載車で事故の確率分布は似ているか？
  - ☞ おそらく似ていない。アイサイトはハイエンドの車種に搭載され、そうした車を運転する人たちは系統的に異なる（より高所得のはず）。このため、運転の仕方や事故確率も異なる（はず）。
  - ☞ アイサイト非搭載時にハイエンド車種の運転手は安全運転をする傾向があるか？おそらく安全運転をする傾向が強い、失うものが多いから。この場合、アイサイト搭載の事故への効果は過大評価される overestimated。
  - ☞ アイサイト非搭載時にハイエンド車種の運転手が危険な運転をする傾向があるかもしれない、リスクを取って財を成したから。この場合、アイサイト搭載の事故への効果は過小評価される underestimated。

## With-without:

- Eyesight 搭載車=the treated、Eyesight 非搭載車=the control、Eyesight の事故確率への ATE=各群の平均事故率の差。
  - ☞ アイサイト非搭載時に、アイサイト搭載車とアイサイト非搭載車で事故の確率分布は似ているか？
  - ☞ おそらく似ていない。アイサイトはハイエンドの車種に搭載され、そうした車を運転する人たちは系統的に異なる（より高所得のはず）。このため、運転の仕方や事故確率も異なる（はず）。
  - ☞ アイサイト非搭載時にハイエンド車種の運転手は安全運転をする傾向があるか？おそらく安全運転をする傾向が強い、失うものが多いから。この場合、アイサイト搭載の事故への効果は過大評価される overestimated。
  - ☞ アイサイト非搭載時にハイエンド車種の運転手が危険な運転をする傾向があるかもしれない、リスクを取って財を成したから。この場合、アイサイト搭載の事故への効果は過小評価される underestimated。
  - ☞ アイサイトの導入をランダムに割り当てるこどもできる → 実験可能

“Accident rate reduced by 61%.”

New  
SUBARU  
SAFETY

スバル 安全

あなたの大切な人を事故から守りたい

日本国内で販売したスバル車のうち、  
アイサイト<sup>®</sup>搭載車は非搭載車に比べて  
事故の発生率が61%減<sup>①</sup>。さらに、  
車両同士の追突事故では84%減<sup>②</sup>  
という調査結果が出ました。  
詳しくはホームページをご覧ください。

すべてのアイサイト搭載車が安全評価で最高ランク<sup>③</sup>

スバル車 一万台あたりの 事故件数	154件	61件
アイサイト <sup>®</sup> (ver.2) 非搭載車	154件	61件
アイサイト <sup>®</sup> (ver.2) 搭載車	61件	154件

2010~2014年実調失敗台数  
事故件数  
48,085  
741

246,139  
1,493

立派財團法人 交通安全総合分析センター(HTARD)のデータを基に独自算出

WRX  
OUTBACK  
BRZ  
LEGACY  
LEVORG  
SG  
CROSSTOUR  
FORESTER  
IMPREZA  
IMPREZA  
SPORT  
HYBRID  
G4  
XV  
XV  
WRX

“Accident rate reduced by 61%.”

WRONG: Not reduced but “smaller than non-Eyesight cars.”

New  
SUBARU  
SAFETY

スバル 安全

あなたの大切な人を事故から守りたい

日本国内で販売したスバル車のうち、  
アイサイト搭載車は非搭載車に比べて  
事故の発生率が61%減。さらに、  
車両同士の追突事故では84%減<sup>③</sup>  
という調査結果が出ました。  
詳しくはホームページをご覧ください。

すべてのアイサイト搭載車が安全評価で最高ランク<sup>④</sup>

スバル車  
一台あたりの  
事故件数

アイサイト搭載車	非搭載車
154件	61件
2010~2014年度累計台数	48,085
事故件数	741

2010~2014年度累計台数  
48,085  
事故件数  
741

2010~2014年度累計台数  
246,139  
事故件数  
1,493

車種一覧

- OUTBACK
- B4
- LEVORG
- B4
- CROSSTOUR
- FORESTER
- IMPREZA SPORT
- IMPREZA HYBRID
- GR
- XV
- XV

Source: Based on all Subaru cars sold in Japan, computed by the company using ITARDA's vehicle registration data.

## With-without:

- 日常的によく笑う人=the treated、日常的にあまり笑わない人=the control、笑いの健康への ATE=両群の自己申告による健康度合いの平均値の差

## With-without:

- 日常的によく笑う人=the treated、日常的にあまり笑わない人=the control、笑いの健康への ATE=両群の自己申告による健康度合いの平均値の差
  - ☞ 笑う頻度が同じという仮想的状況で、両群の自己申告健康の分布は似ているか？

## With-without:

- 日常的によく笑う人=the treated、日常的にあまり笑わない人=the control、笑いの健康への ATE=両群の自己申告による健康度合いの平均値の差
  - ☞ 笑う頻度が同じという仮想的状況で、両群の自己申告健康の分布は似ているか？
  - ☞ おそらく似ていない。健康を自認する人ほど笑う頻度は高いはずだから。この場合、過大な効果推計になる。

## With-without:

- 日常的によく笑う人=the treated、日常的にあまり笑わない人=the control、笑いの健康への ATE=両群の自己申告による健康度合いの平均値の差
  - ☞ 笑う頻度が同じという仮想的状況で、両群の自己申告健康の分布は似ているか？
  - ☞ おそらく似ていない。健康を自認する人ほど笑う頻度は高いはずだから。この場合、過大な効果推計になる。
  - ☞ 厳密に笑いの効果を示すには、笑う頻度を各人にランダムに割り当てねばならない。そんな実験は可能か？

## With-without:

- 日常的によく笑う人=the treated、日常的にあまり笑わない人=the control、笑いの健康への ATE=両群の自己申告による健康度合いの平均値の差
  - ☞ 笑う頻度が同じという仮想的状況で、両群の自己申告健康の分布は似ているか？
  - ☞ おそらく似ていない。健康を自認する人ほど笑う頻度は高いはずだから。この場合、過大な効果推計になる。
  - ☞ 厳密に笑いの効果を示すには、笑う頻度を各人にランダムに割り当てねばならない。そんな実験は可能か？
  - ☞ ある程度までは可能。コメディや漫才の放送局の無料視聴権を与えて笑いの頻度を変えることはできるかもしれない。しかし、どのくらい実験を続けなくてはいけないか。不明。おそらく、笑いの効果を研究すること自体が野心的すぎるかもしれない。

# 因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当がランダム化

ランダム化統御試験 randomised controlled trial (RCT) 実験の統御ができている  
実験者が被験者を募り、治療群と統御群にランダムに  
割り振る。

## 因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当がランダム化

ランダム化統御試験 randomised controlled trial (RCT) 実験の統御ができている  
実験者が被験者を募り、治療群と統御群にランダムに  
割り振る。

自然実験 natural experiment 実験者なしに発生する実 群の割り当てを観察でき、実施選択がない  
験(自然災害、政策)

# 因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当がランダム化

ランダム化統御試験 randomised controlled trial (RCT) 実験の統御ができている  
実験者が被験者を募り、治療群と統御群にランダムに  
割り振る。

自然実験 natural experiment 実験者なしに発生する実 群の割り当てを観察でき、実施選抜がない  
験(自然災害、政策)

操作変数推計量 instrumental variables estimator ラン  
ダムなショックで参加機会が増えるときに、新たに参加  
する人たちの結果指標変化。2段階最小自乗法 2-stage  
least squares (2SLS) estimator, ウォルド推計量 Wald  
estimatorともいう。

結果に影響(相関)せず参加だけに影響(相  
関)する変数(=操作変数)がある

因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当に自己選抜あり

回帰不連続デザイン regression discontinuity design 境 介入変数  $p$  が不連続に変化する地点が観測界で政策が変わるとき、境界近傍での結果指標変化 できる

因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当に自己選抜あり

回帰不連続デザイン regression discontinuity design 境 介入変数  $p$  が不連続に変化する地点が観測界で政策が変わるとき、境界近傍での結果指標変化 できる

傾向値マッチング推計量 propensity score matching es- 観察可能な変数が似た同士で比較すると、  
timator 治療を受ける確率が等しい統御群標本との比 介入を受ける群と受けない群で参加確率の  
較 範囲が重なる

# 因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当に自己選抜あり

回帰不連続デザイン regression discontinuity design 境 介入変数  $p$  が不連続に変化する地点が観測界で政策が変わるとき、境界近傍での結果指標変化 できる

傾向値マッチング推計量 propensity score matching es- 観察可能な変数が似た同士で比較すると、  
timator 治療を受ける確率が等しい統御群標本との比 介入を受ける群と受けない群で参加確率の  
較 範囲が重なる

2重差分推計量 difference-in-differences estimator 介入 両群で介入前後のデータがあり、介入なし  
前後変化を両群間で差をとる。固定効果推計量 fixed のときに両群の結果変数が共有トレンドを  
effects estimator ともいう。 もつ

因果関係を示す方法: 各方法とも治療群と統御群が似通っていることが前提

方法

条件

割当に自己選抜あり

回帰不連続デザイン regression discontinuity design 境 介入変数  $p$  が不連続に変化する地点が観測界で政策が変わるととき、境界近傍での結果指標変化 できる

傾向値マッチング推計量 propensity score matching es- 観察可能な変数が似た同士で比較すると、  
timator 治療を受ける確率が等しい統御群標本との比 介入を受ける群と受けない群で参加確率の  
較 範囲が重なる

2重差分推計量 difference-in-differences estimator 介入 両群で介入前後のデータがあり、介入なし  
前後変化を両群間で差をとる。固定効果推計量 fixed のときに両群の結果変数が共有トレンドを  
effects estimator ともいう。 もつ

合成統御法(推計量) synthetic control method 統御群 介入前のデータが比較的長期にあり、介入  
(の加重平均値) から治療群各標本の CF を合成、DID を受けない群の標本サイズが大きく、両群  
の拡張 の結果変数が共有トレンドをもつ

# 自然実験

オランダ飢餓の冬(1944-45)への胎内曝露(?)

成人1人当たり配給カロリー(飢餓の冬は1000カロリー/日)

データ:

- 地域: アムステルダム (Wilhelmina Gasthuis 病院) 生まれ、女性のみ
  - 時期: 1944年8月-1946年4月生まれ ( $n = 1116$ 、1987-91年追跡可能 834、参加 700)
- ☞ バーカー仮説 (Barker hypothesis): 飢餓の冬を胎内で過ごした人は成人期に循環器系疾患に罹りやすい

# 自然実験

オランダ飢餓の冬 (1944-45)への胎内曝露 (?)

成人1人当たり配給カロリー (飢餓の冬は 1000 カロリー/日)

データ:

- 地域: アムステルダム (Wilhelmina Gasthuis 病院) 生まれ、女性のみ
- 時期: 1944 年 8 月-1946 年 4 月生まれ ( $n = 1116$ 、1987-91 年追跡可能 834、参加 700)

☞ バーカー仮説 (Barker hypothesis): 飢餓の冬を胎内で過ごした人は成人期に循環器系疾患に罹りやすい

? : 飢餓の冬生まれ vs. その他期間生まれ (Before-after に近い)

# 自然実験

オランダ飢餓の冬 (1944-45)への胎内曝露 (?)

成人1人当たり配給カロリー (飢餓の冬は 1000 カロリー/日)

データ:

- 地域: アムステルダム (Wilhelmina Gasthuis 病院) 生まれ、女性のみ
- 時期: 1944 年 8 月-1946 年 4 月生まれ ( $n = 1116$ 、1987-91 年追跡可能 834、参加 700)

☞ バーカー仮説 (Barker hypothesis): 飢餓の冬を胎内で過ごした人は成人期に循環器系疾患に罹りやすい

? : 飢餓の冬生まれ vs. その他期間生まれ (Before-after に近い)

信頼性のより高い推計方法: アムステルダムでの飢餓の冬前後の変化 vs. その他地域での飢餓の冬前後の変化 (DID)

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

$a_1, a_2, a_3$  で曝露の影響が測れるための識別仮定: 同じ地域生まれであれば、飢餓の冬がなければ、全てのコーホート (1944年8月-1946年4月) の死亡率は一定

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

$a_1, a_2, a_3$  で曝露の影響が測れるための識別仮定: 同じ地域生まれであれば、飢餓の冬がなければ、全てのコーホート (1944年8月-1946年4月) の死亡率は一定

そうかもしれないし、そうじゃないかもしない

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

$a_1, a_2, a_3$  で曝露の影響が測れるための識別仮定: 同じ地域生まれであれば、飢餓の冬がなければ、全てのコーホート (1944年8月-1946年4月) の死亡率は一定

そうかもしれないし、そうじゃないかもしれない

CF=他期間生まれ

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

$a_1, a_2, a_3$  で曝露の影響が測れるための識別仮定: 同じ地域生まれであれば、飢餓の冬がなければ、全てのコーホート (1944年8月-1946年4月) の死亡率は一定

そうかもしれないし、そうじゃないかもしれない

CF=他期間生まれ

- 死亡率が一定と期待する理由は不明なので、推計値の信頼性は不明

第3三半期曝露 1945年2-6月生まれ

第2三半期曝露 1945年5-9月生まれ

第1三半期曝露 1945年8-12月生まれ

統御群 1944年8月-1945年1月、1946年1月-4月生まれ

子どもの周産期 (W22-D7) 死亡率 = 他期間生まれ +  $a_1$  \* 第1三半期曝露

+  $a_2$  \* 第2三半期曝露 +  $a_3$  \* 第3三半期曝露 + 誤差項

$a_1, a_2, a_3$  で曝露の影響が測れるための識別仮定: 同じ地域生まれであれば、飢餓の冬がなければ、全てのコーホート (1944年8月-1946年4月) の死亡率は一定

そうかもしれないし、そうじゃないかもしれない

CF=他期間生まれ

- 死亡率が一定と期待する理由は不明なので、推計値の信頼性は不明
- 戦争の影響を受けなければ、別年同月の死亡率は同じになると期待できるが、これは big if

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	... <sup>c</sup>
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	... <sup>c</sup>
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

7/688 vs. 12/309 という比較:

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	... <sup>c</sup>
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	...
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

7/688 vs. 12/309 という比較:

確かに第3三半期の方は3.82倍 [表のRR=3.71は計算間違い? 下段RRは正しいが上段全て間違い]

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	...
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

7/688 vs. 12/309 という比較:

☞ 確かに第3三半期の方は3.82倍 [表のRR=3.71は計算間違い? 下段RRは正しいが上段全て間違い]

☞ 統御群で1.02%という珍しい症例。しかも、310(=全出産数688-女性総数378)件は2人目以降の出産。おそらく標本サイズが十分ではない。

**Table 4—Perinatal Deaths<sup>a</sup> among 1334 Infants Born to 626 Parous Women, by Cohort of Famine Exposure<sup>b</sup>: Dutch Famine Birth Cohort Study**

	Not Exposed (n = 378)	Exposed in Third Trimester (n = 160)	Exposed in Second Trimester (n = 152)	Exposed in First Trimester (n = 125)
<b>Among singleton offspring</b>				
No.	7/688	12/309	6/294	6/240
%	1.0	3.9	2.0	2.5
RR	1.00	3.71	1.95	2.39
95% CI	(Reference)	1.35, 11.1	0.54, 6.77	0.66, 8.29
<b>Among twins</b>				
No.	3 of 18	3 of 4	2 of 10	0 of 8
%	16.7	75.0	20.0	0.0
RR	1.00	4.50	1.20	...
95% CI	(Reference)	0.60, 33.6	0.10, 10.5	...

Note. RR = relative risk; CI = confidence interval.

<sup>a</sup>Stillbirths plus deaths in the first 7 days of life as a proportion of all births.

<sup>b</sup>Women may have been exposed in more than one trimester, and these categories are therefore not mutually exclusive. There is no overlap between any exposure cohort and the unexposed cohort, or between those exposed in the third trimester and those exposed in the first trimester.

<sup>c</sup>No deaths occurred among twin offspring of women exposed early in pregnancy.

7/688 vs. 12/309 という比較:  
small sample bias?

☞ 確かに第3三半期の方は3.82倍 [表のRR=3.71は計算間違い? 下段RRは正しいが上段全て間違い]

☞ 統御群で1.02%という珍しい症例。しかも、310(=全出産数688-女性総数378)件は2人目以降の出産。おそらく標本サイズが十分ではない。

☞ 例えば3%が4.3%に増えたとき: 誤差、それとも、43%ポイントも増えた、と表現するのか。標本サイズ300人だと9人が13人に増えただけ。誤差。と共に著者に言ったら怒られた。

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- だから、標本サイズはさらに小さくなる
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- だから、標本サイズはさらに小さくなる
  - ※ 最小標本サイズ minimum sample size: 想定した効果と誤差の大きさの下で、一定の確率(「80%」)で効果を検知できる最小標本サイズ。標本サイズが大きいほど誤差(サンプリング・エラー)が減って分散は小さくなり、効果を検知しやすくなる。
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

## この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- だから、標本サイズはさらに小さくなる
  - ☞ 最小標本サイズ minimum sample size: 想定した効果と誤差の大きさの下で、一定の確率(「80%」)で効果を検知できる最小標本サイズ。標本サイズが大きいほど誤差(サンプリング・エラー)が減って分散は小さくなり、効果を検知しやすくなる。
  - ☞ 効果サイズが確率 2.9% ポイント ( $= \frac{7}{688} - \frac{12}{309}$ ) 増える(それだけ treated の分布が動く)という想定で標本を得て効果を推計し、帰無仮説(曝露ありの死亡確率=曝露無しの死亡確率=.0102)が成立する確率  $p$  値が 5%未満であれば棄却判断する。この作業を繰り返したときに、80%の回数で棄却判断するための標本サイズが最小標本サイズ。計算すると 370 人の標本が必要。
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

# この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- だから、標本サイズはさらに小さくなる
  - ⇒ 最小標本サイズ minimum sample size: 想定した効果と誤差の大きさの下で、一定の確率(「80%」)で効果を検知できる最小標本サイズ。標本サイズが大きいほど誤差(サンプリング・エラー)が減って分散は小さくなり、効果を検知しやすくなる。
  - ⇒ 効果サイズが確率 2.9% ポイント ( $= \frac{7}{688} - \frac{12}{309}$ ) 増える(それだけ treated の分布が動く)という想定で標本を得て効果を推計し、帰無仮説(曝露ありの死亡確率=曝露無しの死亡確率=.0102)が成立する確率  $p$  値が 5%未満であれば棄却判断する。この作業を繰り返したときに、80%の回数で棄却判断するための標本サイズが最小標本サイズ。計算すると 370 人の標本が必要。
  - ⇒ 第 1 子統御群  $n_0 = 378$ , 第 1 三半期治療群  $n_1 = 160$  だとサンプルサイズが小さすぎる。
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

この標本サイズ(1病院の出生記録)で真の効果を計測できると考えない方が良い

- 細かなことを言えば、2人目以降の出産は選抜がある
- 1人目で正常出産すると、2人目の妊娠を望む比率は減るはず
- だから、標本サイズはさらに小さくなる
  - ☞ 最小標本サイズ minimum sample size: 想定した効果と誤差の大きさの下で、一定の確率(「80%」)で効果を検知できる最小標本サイズ。標本サイズが大きいほど誤差(サンプリング・エラー)が減って分散は小さくなり、効果を検知しやすくなる。
  - ☞ 効果サイズが確率 2.9% ポイント ( $= \frac{7}{688} - \frac{12}{309}$ ) 増える(それだけ treated の分布が動く)という想定で標本を得て効果を推計し、帰無仮説(曝露ありの死亡確率=曝露無しの死亡確率=.0102)が成立する確率  $p$  値が 5%未満であれば棄却判断する。この作業を繰り返したときに、80%の回数で棄却判断するための標本サイズが最小標本サイズ。計算すると 370 人の標本が必要。
  - ☞ 第1子統御群  $n_0 = 378$ , 第1三半期治療群  $n_1 = 160$  だとサンプルサイズが小さすぎる。
  - ☞ 第1子統御群  $n_0 = 378$ , 全三半期での治療群  $n_1 = 160 + 152 + 125 = 437$ の方が最小標本サイズを満たしやすい。しかし、三半期ごとの治療群を一緒にして異なる効果を混ぜている。
- 2人目以降の妊娠では周産期死亡リスクがより高い人が多い標本
- よって、1人目の妊娠だけで比較すべき

# 信頼性のより高いデザイン・推計方法 (DID):

信頼性のより高いデザイン・推計方法(DID):

全ての地域のコーホート・パネル・データ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
- 時期: 1944 年 8 月-1946 年 4 月生まれ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
  - ➡ 同じ誕生日の人たちを他地域で見つければ良い
- 時期: 1944年8月-1946年4月生まれ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
  - ⇒ 同じ誕生日の人たちを他地域で見つければ良い
  - ⇒ 1987年時点までに死亡している女性はいる
- 時期: 1944年8月-1946年4月生まれ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
  - ⇒ 同じ誕生日の人たちを他地域で見つければ良い
  - ⇒ 1987年時点までに死亡している女性はいる=より健康な女性のみが残っている
- 時期: 1944年8月-1946年4月生まれ

## 信頼性のより高いデザイン・推計方法 (DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
  - ⇒ 同じ誕生日の人たちを他地域で見つけければ良い
  - ⇒ 1987年時点までに死亡している女性はいる=より健康な女性のみが残っている=子どもの周産期死亡率は過小推計されるはず
- 時期: 1944年8月-1946年4月生まれ

## 信頼性のより高いデザイン・推計方法(DID):

全ての地域のコーホート・パネル・データ

- 地域: アムステルダムとその他地域生まれ
  - ⇒ 同じ誕生日の人たちを他地域で見つけければ良い
  - ⇒ 1987年時点までに死亡している女性はいる=より健康な女性のみが残っている=子どもの周産期死亡率は過小推計されるはずだが、アムステルダム標本でも生存選抜について何も対応していないので、他地域標本を加えない理由が見当たらない
- 時期: 1944年8月-1946年4月生まれ

1944年8月-1945年1月、1946年1月-4月を  $ym$ 、このうちの各月(コード)を  $t$  と表記する。 $t \in ym$  は「 $ym$  の要素の  $t$ 」と読む。 $\Delta$  は前月との差分。

1944年8月-1945年1月、1946年1月-4月を  $ym$ 、このうちの各月(コードホート)を  $t$  と表記する。 $t \in ym$  は「 $ym$  の要素の  $t$ 」と読む。 $\Delta$  は前月との差分。

$\Delta$ 子どもの周産期死亡率 =

$$\sum_{t \in ym} b_{0t} * t \text{生まれ} +$$

$$\sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ}$$

$$+ b_{21} * \text{第1三半期}$$

$$+ b_{22} * \text{第2三半期}$$

$$+ b_{23} * \text{第3三半期}$$

$$+ b_{31} * \text{アムステルダム生まれ} * \text{第1三半期}$$

$$+ b_{32} * \text{アムステルダム生まれ} * \text{第2三半期}$$

$$+ b_{33} * \text{アムステルダム生まれ} * \text{第3三半期}$$

$$+\Delta\text{誤差項}$$

1944年8月-1945年1月、1946年1月-4月を $ym$ 、このうちの各月(コードート)を $t$ と表記する。 $t \in ym$ は「 $ym$ の要素の $t$ 」と読む。 $\Delta$ は前月との差分。

$\Delta$ 子どもの周産期死亡率 =

$$\left. \begin{array}{l} \sum_{t \in ym} b_{0t} * t \text{生まれ} + \\ \sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ} \end{array} \right\} \begin{array}{l} \text{その他期間生まれ} \\ \text{の対前月変化平均値} \end{array}$$
$$\begin{aligned} &+ b_{21} * \text{第1三半期} \\ &+ b_{22} * \text{第2三半期} \\ &+ b_{23} * \text{第3三半期} \\ &+ b_{31} * \text{アムステルダム生まれ} * \text{第1三半期} \\ &+ b_{32} * \text{アムステルダム生まれ} * \text{第2三半期} \\ &+ b_{33} * \text{アムステルダム生まれ} * \text{第3三半期} \\ &+ \Delta \text{誤差項} \end{aligned}$$

1944年8月-1945年1月、1946年1月-4月を  $ym$ 、このうちの各月(コードート)を  $t$  と表記する。 $t \in ym$  は「 $ym$  の要素の  $t$ 」と読む。 $\Delta$  は前月との差分。

$\Delta$ 子どもの周産期死亡率 =

$$\left. \begin{aligned} & \sum_{t \in ym} b_{0t} * t \text{生まれ} + \\ & \sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ} \\ & + b_{21} * \text{第1三半期} \\ & + b_{22} * \text{第2三半期} \\ & + b_{23} * \text{第3三半期} \\ & + b_{31} * \text{アムステルダム生まれ} * \text{第1三半期} \\ & + b_{32} * \text{アムステルダム生まれ} * \text{第2三半期} \\ & + b_{33} * \text{アムステルダム生まれ} * \text{第3三半期} \\ & + \Delta \text{誤差項} \end{aligned} \right\}$$

その他期間生まれ  
の対前月変化平均値

飢餓の冬生まれ  
の対前月変化平均値

1944年8月-1945年1月、1946年1月-4月を $ym$ 、このうちの各月(コードート)を $t$ と表記する。 $t \in ym$ は「 $ym$ の要素の $t$ 」と読む。 $\Delta$ は前月との差分。

$\Delta$ 子どもの周産期死亡率 =

$$\left. \begin{array}{l} \sum_{t \in ym} b_{0t} * t \text{生まれ} + \\ \sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ} \\ + b_{21} * \text{第1三半期} \\ + b_{22} * \text{第2三半期} \\ + b_{23} * \text{第3三半期} \\ + b_{31} * \text{アムステルダム生まれ} * \text{第1三半期} \\ + b_{32} * \text{アムステルダム生まれ} * \text{第2三半期} \\ + b_{33} * \text{アムステルダム生まれ} * \text{第3三半期} \\ + \Delta \text{誤差項} \end{array} \right\} \begin{array}{l} \text{その他期間生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれで} \\ \text{アムステルダム生まれ} \\ \text{の対前月変化平均値} \end{array}$$

1944年8月-1945年1月、1946年1月-4月を  $ym$ 、このうちの各月(コード)を  $t$  と表記する。 $t \in ym$  は「 $ym$  の要素の  $t$ 」と読む。 $\Delta$  は前月との差分。

△子どもの周産期死亡率 =

$$\left. \begin{array}{l} \sum_{t \in ym} b_{0t} * t \text{生まれ} + \\ \sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ} \\ + b_{21} * \text{第1三半期} \\ + b_{22} * \text{第2三半期} \\ + b_{23} * \text{第3三半期} \\ + b_{31} * \text{アムステルダム生まれ} * \text{第1三半期} \\ + b_{32} * \text{アムステルダム生まれ} * \text{第2三半期} \\ + b_{33} * \text{アムステルダム生まれ} * \text{第3三半期} \\ + \Delta \text{誤差項} \end{array} \right\} \begin{array}{l} \text{その他期間生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれで} \\ \text{アムステルダム生まれ} \\ \text{の対前月変化平均値} \end{array}$$

$b_{31}, b_{32}, b_{33}$  は飢餓の冬該当地と他地域とのトレンドの差を表す

1944年8月-1945年1月、1946年1月-4月を $ym$ 、このうちの各月(コーホート)を $t$ と表記する。 $t \in ym$ は「 $ym$ の要素の $t$ 」と読む。 $\Delta$ は前月との差分。

$\Delta$ 子どもの周産期死亡率 =

$$\left. \begin{array}{l} \sum_{t \in ym} b_{0t} * t \text{生まれ} + \\ \sum_{t \in ym} b_{1t} * \text{アムステルダム生まれ} * t \text{生まれ} \\ + b_{21} * \text{第1三半期} \\ + b_{22} * \text{第2三半期} \\ + b_{23} * \text{第3三半期} \\ + b_{31} * \text{アムステルダム生まれ} * \text{第1三半期} \\ + b_{32} * \text{アムステルダム生まれ} * \text{第2三半期} \\ + b_{33} * \text{アムステルダム生まれ} * \text{第3三半期} \\ + \Delta \text{誤差項} \end{array} \right\} \begin{array}{l} \text{その他期間生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれ} \\ \text{の対前月変化平均値} \\ \\ \text{飢餓の冬生まれで} \\ \text{アムステルダム生まれ} \\ \text{の対前月変化平均値} \end{array}$$

$b_{31}, b_{32}, b_{33}$ は飢餓の冬該当地と他地域とのトレンドの差を表す

$b_{31}, b_{32}, b_{33}$ で曝露の影響が測れるための識別仮定: 飢餓の冬がなければ、全地域で死亡率トレンドは共通=トレンドに地域差なし

## CF: 他地域の 1944-45 年冬のトレンド

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

$b_{1t} = 0$  を  $t$  すべてについて検定すれば、1944-45 年冬以外でアムステルダムとその他地域で共通トレンドの有無を確認できる

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

$b_{1t} = 0$  を  $t$  すべてについて検定すれば、1944-45 年冬以外でアムステルダムとその他地域で共通トレンドの有無を確認できる

☞ より望ましくは...

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

$b_{1t} = 0$  を  $t$  すべてについて検定すれば、1944-45 年冬以外でアムステルダムとその他地域で共通トレンドの有無を確認できる

- ☞ より望ましくは... 成人 1 人当たり配給カロリー(飢餓の冬は 1000 カロリー/日)を全標本から得られれば、推計式に各期各地域の配給カロリーを加えると、周産期死亡率とコート、周産期死亡率と配給カロリーの関係を分離して計測でき、配給カロリーが影響しているか確認できる

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

$b_{1t} = 0$  を  $t$  すべてについて検定すれば、1944-45 年冬以外でアムステルダムとその他地域で共通トレンドの有無を確認できる

- ☞ より望ましくは... 成人 1 人当たり配給カロリー(飢餓の冬は 1000 カロリー/日)を全標本から得られれば、推計式に各期各地域の配給カロリーを加えると、周産期死亡率とコーホート、周産期死亡率と配給カロリーの関係を分離して計測でき、配給カロリーが影響しているか確認できる

自然実験は個人ごとに実際の「治療」が観察できないので、生まれた時期と地域で曝露を仮定せざるを得ない

CF: 他地域の 1944-45 年冬のトレンド: 地域間の栄養状態格差が大きくなれば信頼性高い

- ☞ 識別仮定そのもの(アムステルダム 1944-45 年冬の CF= 他地域 1944-45 年冬の変化分)は CF が観察できないので検定できない
- ☞ しかし、それ以外の時期で成り立っていれば、1944-45 年冬にも成り立つ蓋然性が高い

$b_{1t} = 0$  を  $t$  すべてについて検定すれば、1944-45 年冬以外でアムステルダムとその他地域で共通トレンドの有無を確認できる

- ☞ より望ましくは... 成人 1 人当たり配給カロリー(飢餓の冬は 1000 カロリー/日)を全標本から得られれば、推計式に各期各地域の配給カロリーを加えると、周産期死亡率とコーホート、周産期死亡率と配給カロリーの関係を分離して計測でき、配給カロリーが影響しているか確認できる

自然実験は個人ごとに実際の「治療」が観察できないので、生まれた時期と地域で曝露を仮定せざるを得ない

仮説の原因(栄養不足)とより近接した配給カロリーを用いる方が検定結果解釈の確度は増す

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

Good news:

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

Consider a poverty reduction policy that gives a subsidy to the people below the poverty line.

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

Consider a poverty reduction policy that gives a subsidy to the people below the poverty line.

- “BPL” card in India.

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが (=DID ができる) ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

Consider a poverty reduction policy that gives a subsidy to the people below the poverty line.

☞ “BPL” card in India.

Suppose poverty line is USD 1.25 per day and this criteria is strictly enforced. So if your income is USD 1.24 per day, you get the money. If your income is USD 1.25, you don't.

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが(=DIDができる)ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

Consider a poverty reduction policy that gives a subsidy to the people below the poverty line.

↳ “BPL” card in India.

Suppose poverty line is USD 1.25 per day and this criteria is strictly enforced. So if your income is USD 1.24 per day, you get the money. If your income is USD 1.25, you don't.

People with daily income of USD 1.24 and USD 1.25 are similar.

# Regression discontinuity design

全ての特徴を観察できれば結果の差をすべて説明できる。が、観察できない。  
実験でなく、パネル・データが(=DIDができる)ないとき、どうすれば良いのか?

Good news: インパクト評価の範囲を狭くすれば、推計可能。

Consider a poverty reduction policy that gives a subsidy to the people below the poverty line.

↳ “BPL” card in India.

Suppose poverty line is USD 1.25 per day and this criteria is strictly enforced. So if your income is USD 1.24 per day, you get the money. If your income is USD 1.25, you don't.

People with daily income of USD 1.24 and USD 1.25 are similar.

Estimate impacts by comparing BPL and APL near the poverty line.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

Applications: Cutoffs, geographical boundaries.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

Applications: Cutoffs, geographical boundaries.

Policies are full of cutoffs. So almost every policy has a chance of estimating its impacts near the cutoff.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

Applications: Cutoffs, geographical boundaries.

Policies are full of cutoffs. So almost every policy has a chance of estimating its impacts near the cutoff.

Identifying assumption:

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

Applications: Cutoffs, geographical boundaries.

Policies are full of cutoffs. So almost every policy has a chance of estimating its impacts near the cutoff.

Identifying assumption: There is nothing other than policy cutoff that has a discrete jump around the cutoff point. So a jump in the outcome can be attributed only to the policy.

The narrower focus around cutoff gives us a “matched pair” or forms a counterfactual.

Interpretation of estimates: Policy impacts on the subpopulation near the cutoff. It is a local impact we are estimating, not a global impact such as ATE (and ATT, ATC).

Applications: Cutoffs, geographical boundaries.

Policies are full of cutoffs. So almost every policy has a chance of estimating its impacts near the cutoff.

Identifying assumption: There is nothing other than policy cutoff that has a discrete jump around the cutoff point. So a jump in the outcome can be attributed only to the policy.

But there is a catch: (Because we fit the line locally around the cutoff neighbourhood)  
It takes a large sample to use RDD estimator with the order of 10,000.

# 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

## 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

右辺の変数が急に変化する状況を見つけ、その前後で左辺の変化を観察する

## 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

右辺の変数が急に変化する状況を見つけ、その前後で左辺の変化を観察する

teacher-pupil ratio ⇒ exam score

## 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

右辺の変数が急に変化する状況を見つけ、その前後で左辺の変化を観察する

teacher-pupil ratio ⇒ exam score

メモニデス (中世のトーラー学者) による戒律

## 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

右辺の変数が急に変化する状況を見つけ、その前後で左辺の変化を観察する

teacher-pupil ratio ⇒ exam score

メモニデス (中世のトーラー学者) による戒律

“Only up to 40 students in one class....”

# 因果関係を示す方法: 回帰不連続 regression discontinuity design (RDD)

右辺の変数が急に変化する状況を見つけ、その前後で左辺の変化を観察する

teacher-pupil ratio  $\Rightarrow$  exam score

メモニデス (中世のトーラー学者) による戒律

“Only up to 40 students in one class....”

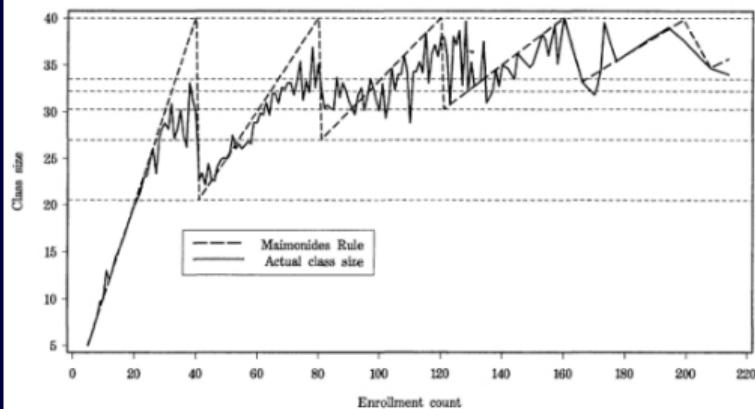




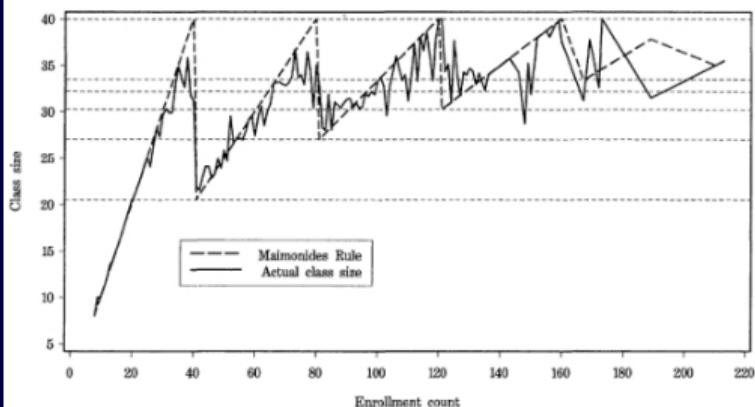
The Government of Israel still holds it.

# The Government of Israel still holds it.

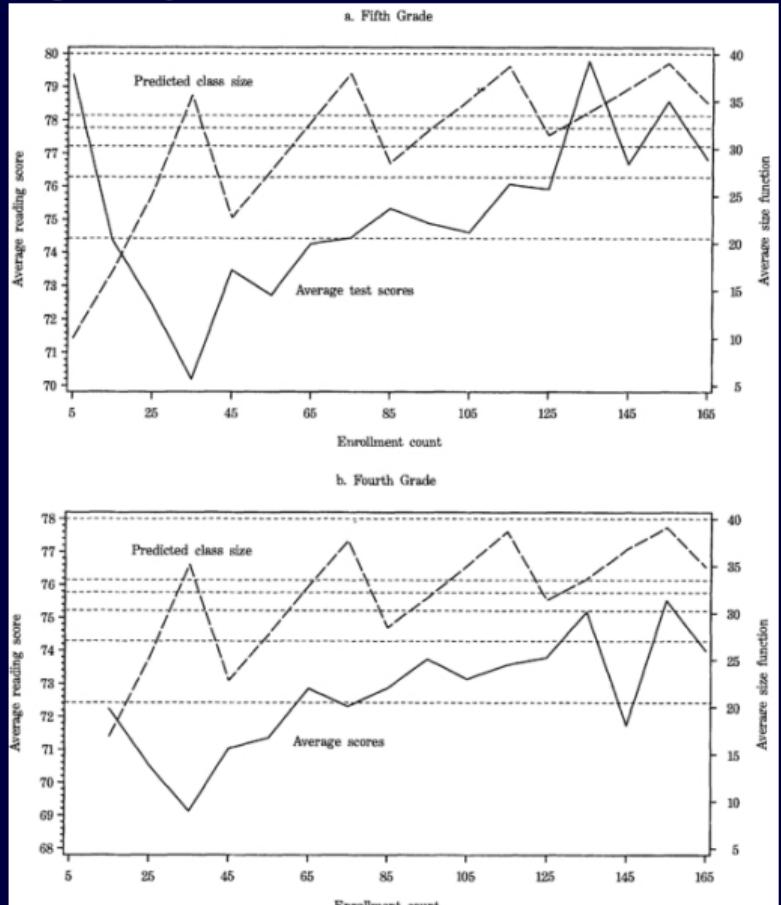
a. Fifth Grade



b. Fourth Grade



# Ingenuity of ?: Predicted class size vs. exam scores.



Impact is more evident in smaller enrollment counts. Average score is increasing after 60 regardless of predicted class size. Possible reasons: Greater deviation of actual class size from predicted class size, different pedagogical methods in large schools or more competition/learning among peers.

# 4th graders

TABLE V  
2SLS ESTIMATES FOR 1991 (FOURTH GRADERS)

	Reading comprehension						Math					
	Full sample				+/- 5 Discontinuity sample		Full sample				+/- 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score (s.d.)	72.5 (8.0)				72.5 (7.8)		67.3 (9.6)				68.7 (9.1)	
Regressors												
Class size	-.110 (.040)	-.133 (.059)	-.074 (.067)	-.147 (.084)	-.098 (.090)	-.150 (.128)	.049 (.048)	-.050 (.070)	-.033 (.081)	-.098 (.092)	.095 (.114)	.023 (.160)
Percent disadvantaged	-.346 (.014)	-.345 (.014)	-.346 (.014)		-.354 (.034)	-.347 (.034)	-.290 (.017)	-.284 (.017)	-.284 (.017)	-.284 (.017)	-.299 (.042)	-.290 (.043)
Enrollment		.005 (.008)	-.040 (.024)			.017 (.022)		-.020 (.010)	.007 (.029)		.023 (.028)	
Enrollment squared/100			.021 (.011)						.006 (.014)			
Piecewise linear trend				.100 (.026)						.130 (.028)		
Root MSE	6.65	6.66	6.63	8.02	6.64	6.69	7.82	7.82	7.82	8.65	8.23	8.24
N	2049		2001		415		2049		2001		415	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes.  
All 2SLS estimates use  $f_{sc}$  as an instrument for class size.

Using all sample, there is a negative relationship between class size and learning only in reading but not in maths. For 4th graders, no impacts of class size at discontinuity neighbourhood sample.

# 4th and 5th graders

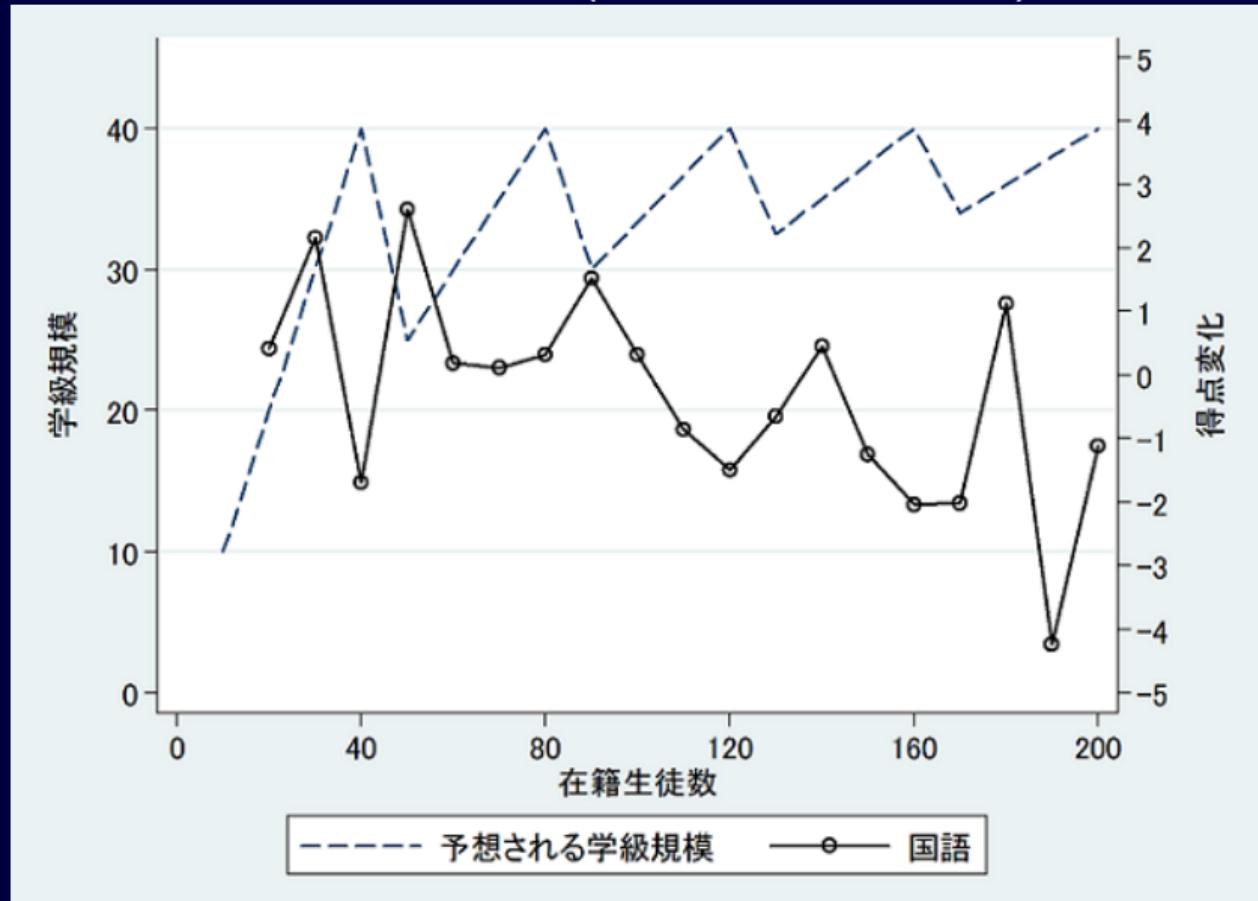
TABLE VI  
DUMMY-INSTRUMENT RESULTS FOR DISCONTINUITY SAMPLES

	5th grade						4th grade					
	Reading comprehension			Math			Reading comprehension			Math		
	+/- 5 Sample	+/- 3 Sample	+/- 5 Sample	+/- 3 Sample	+/- 5 Sample	+/- 3 Sample	+/- 5 Sample	+/- 3 Sample	+/- 5 Sample	+/- 3 Sample	+/- 5 Sample	+/- 3 Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Regressors</i>												
Class size	-.687 (.197)	-.588 (.198)	-.451 (.236)	-.596 (.254)	-.395 (.254)	-.270 (.281)	-.175 (.130)	-.234 (.157)	-.380 (.205)	.018 (.162)	-.118 (.202)	-.247 (.234)
Percent disadvantaged	-.464 (.039)	-.452 (.045)		-.433 (.050)	-.416 (.058)		-.350 (.034)	-.372 (.043)		-.291 (.043)	-.323 (.055)	
Segment 1 (enrollment 36–45)	-5.09 (2.40)	-4.54 (2.59)	-10.7 (3.19)	-7.54 (3.07)	-6.94 (3.34)	-12.6 (3.80)	-1.62 (1.77)	-2.67 (2.23)	-6.94 (2.90)	-1.89 (2.21)	-3.57 (2.87)	-7.31 (3.31)
Segment 2 (enrollment 76–85)	-1.64 (1.41)	-2.18 (1.64)	-2.96 (2.00)	-1.57 (1.83)	-2.17 (2.14)	-2.89 (2.41)	-1.52 (1.24)	-2.16 (1.59)	-3.83 (2.10)	-1.15 (1.56)	-2.50 (2.07)	-3.96 (2.39)
Root MSE	7.46	7.24	8.67	9.41	9.14	10.2	6.72	6.70	8.30	8.25	8.53	9.52
N	471	302		471	302		415	265		415	265	

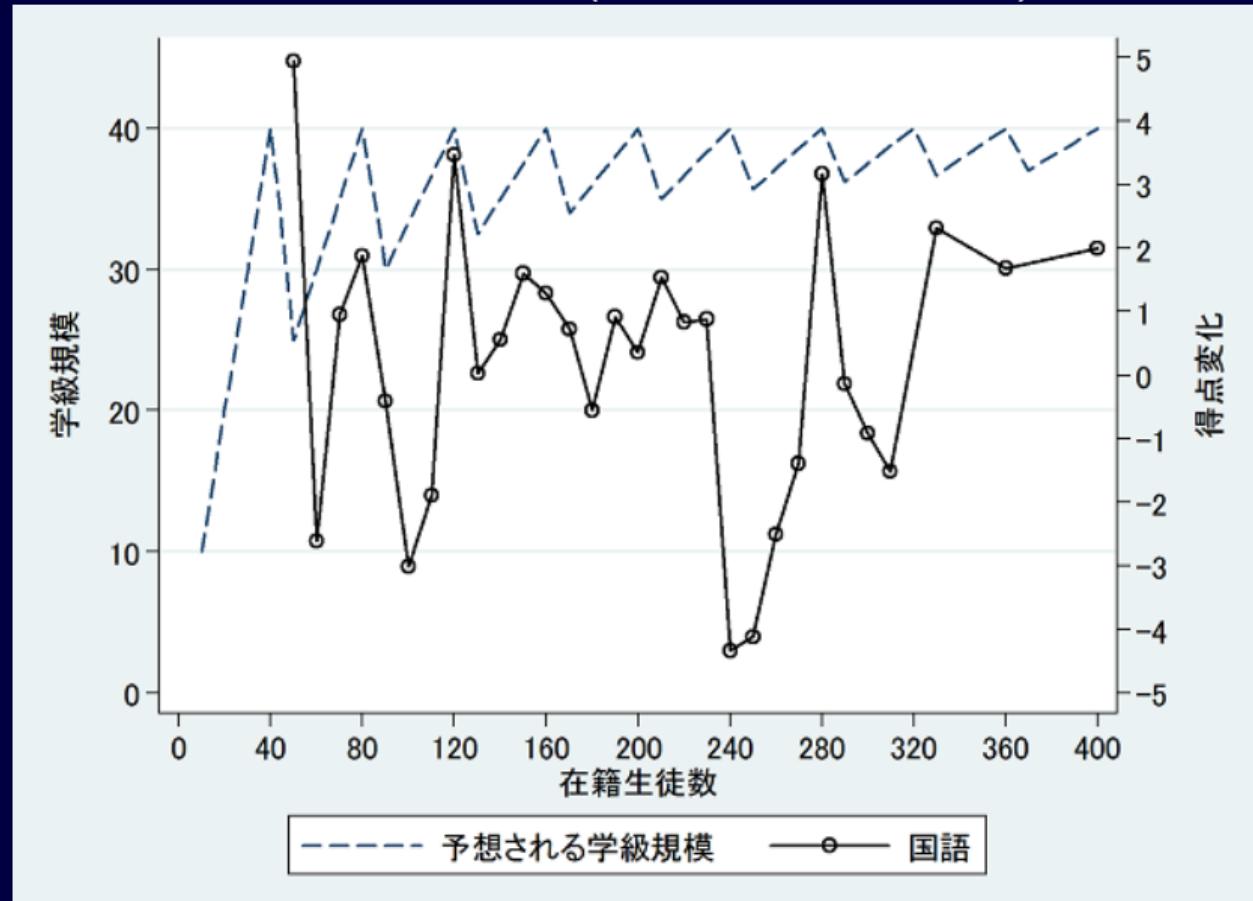
The table reports results from a sample of classes in schools with enrollment close to points of discontinuity. The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use  $I[f_{sc} < 32]$  and interactions with dummies for enrollment segments as instruments for class size. Since there are three segments, there are three instruments. The models include dummies for the first two segments to control for segment main effects.

For 5th graders, using discontinuity neighbourhood sample, there is a negative causal relationship between class size and learning in both reading and maths.

# 横浜市 2009 年 40 人学級: 小 6 国語 (共通試験点数偏差値)



# 横浜市 2009 年 40 人学級: 中 3 国語 (共通試験点数偏差値)



赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない  
教員の指導能力、意欲、効率性重視方針も必要

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

教員の指導能力、意欲、効率性重視方針も必要

教員の数が増えれば能力が低い、意欲が弱い人も雇用される

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

教員の指導能力、意欲、効率性重視方針も必要

教員の数が増えれば能力が低い、意欲が弱い人も雇用される

伸び代が高い生徒に注力(効率性重視)するか、成績の低い生徒に注力(公平性重視)するか

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

教員の指導能力、意欲、効率性重視方針も必要

教員の数が増えれば能力が低い、意欲が弱い人も雇用される

伸び代が高い生徒に注力(効率性重視)するか、成績の低い生徒に注力(公平性重視)するか

どうも地価の高い地域では成績が上がったようです(論文アクセスないので未読)

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

教員の指導能力、意欲、効率性重視方針も必要

教員の数が増えれば能力が低い、意欲が弱い人も雇用される

伸び代が高い生徒に注力(効率性重視)するか、成績の低い生徒に注力(公平性重視)するか

どうも地価の高い地域では成績が上がったようです(論文アクセスないので未読)

情報開示請求で得た学校平均点数と学年生徒数のデータなので、実際のクラス数(よって、実際のクラス・サイズ)は分からぬ模様

赤林さんは中3国語で期待した成果が出なかったことに驚くが、以下のように解釈  
<https://synodos.jp/education/12530>

少人数学級=きめ細やかな指導が可能、なので、自動的に成績が上がるわけではない

教員の指導能力、意欲、効率性重視方針も必要

教員の数が増えれば能力が低い、意欲が弱い人も雇用される

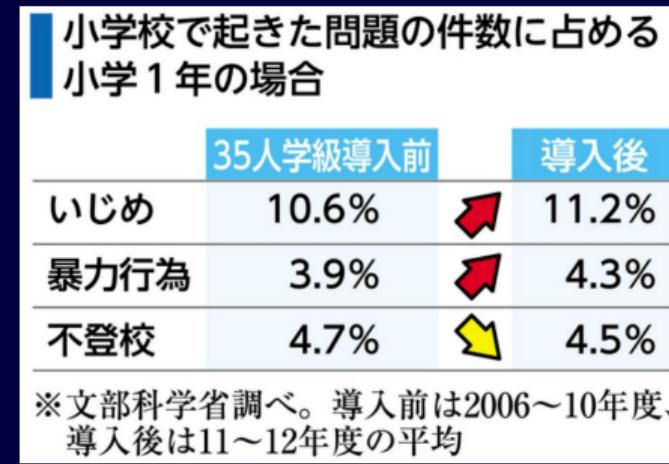
伸び代が高い生徒に注力(効率性重視)するか、成績の低い生徒に注力(公平性重視)するか

どうも地価の高い地域では成績が上がったようです(論文アクセスないので未読)

情報開示請求で得た学校平均点数と学年生徒数のデータなので、実際のクラス数(よって、実際のクラス・サイズ)は分からぬ模様

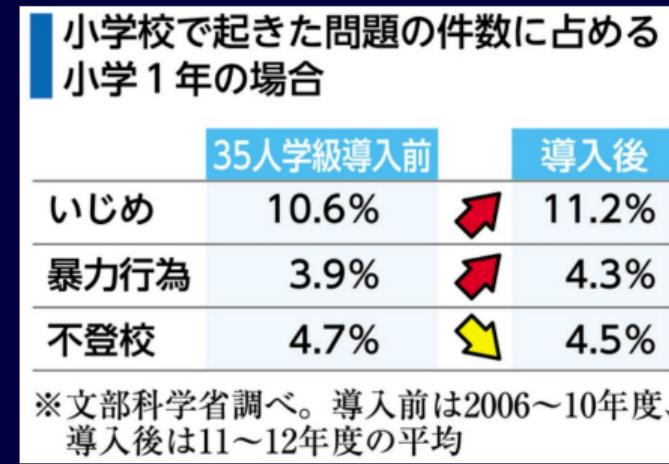
点数がより変化しやすそうな英語や数学の結果が知りたい

中室牧子さん：35人学級は、2011年に公立小学校の1年生に対してのみ導入されました。財務省は、2011年以前と以後で、いじめ、暴力行為、不登校の平均値を比べると、いじめや暴力、不登校には大きな変化が見られないので、少人数学級には効果がない。したがって、「40人学級に戻すべき」と主張したのです。



<https://diamond.jp/articles/-/66992>

中室牧子さん: 35人学級は、2011年に公立小学校の1年生に対してのみ導入されました。財務省は、2011年以前と以後で、いじめ、暴力行為、不登校の平均値を比べると、いじめや暴力、不登校には大きな変化が見られないので、少人数学級には効果がない。したがって、「40人学級に戻すべき」と主張したのです。



<https://diamond.jp/articles/-/66992>

この効果推計で不適切な点: 帰無仮説(効果ゼロ)の  $p$  値がない、推計方法(識別仮定)の信頼性が低い、プラシーボ検定(私立1年生や公立2年生)もできる

公立小、1 學級 35 人以下

政府調整 25年度までに全学年で

政府は公立小学校の1学級あたりの上限人数を引き下げ2025年度までに全学年で35人にする調整に入った。小2から段階的に引き下げる計画で21年度予算案に関する財務相が17日に折衝して21年度予算案に関する合意をめざす。政府関係者への取材で分かった。現在の上限は小1が35人、小2から中3までは40人と義務教育標準法で決まっている。文科省は「児童生徒に個別に学習できる環境をつくる」として、各学年で3密を回避でき、対策で3密を回避でき、教員が児童生徒に目配りをしやすくなり学力向上にもつながると主張。小中全学年で30人にするよう求めている。

、級35人以下までに全学年で、人減らしても感染リスクには差がないなどとして30人学級に慎重姿勢を示していた。少子化により、全国で小学校の9割、中学校の7割で1学級の児童生徒数が35人以下の児童生徒数が35人以下となる現状のまま消極的な理由のひとつに挙げた。計画の実現には、21年送り、検討課題とすべき上限人数は198度に45人から40人に減る。11年度からは小1年生に引き下げられることとなり、民主党政権下の人数学級導入は今後も実現しない可能性がある。

# 歐州向けEV、日本から日産、英工場で生産せず

日産自動車は欧州で2021年に発売する新型本から輸出する。英国の両地域間の関税増が

## 欧洲向けEV、日本から

日産、英工場で生産せず

日産自動車は欧州で2021年に発売する新型電気自動車（EV）を日本から輸出する。英国の両地域間の関税増が

### 3密回避が動機

今回も(35人が良いとする)エビデンスに依拠せず

全国一律だと効果推計は難しい

- ?: In Quebec, unemployment benefits are increased once reaching the age of 30 for adults with no child. This should have disincentives to work for age 30 and older. If this is true, at around 30, work outcomes will be reduced.

- ?: In Quebec, unemployment benefits are increased once reaching the age of 30 for adults with no child. This should have disincentives to work for age 30 and older. If this is true, at around 30, work outcomes will be reduced.
  - ☞ Will there be a jump in employment rates at 30 to the below?

- ?: In Quebec, unemployment benefits are increased once reaching the age of 30 for adults with no child. This should have disincentives to work for age 30 and older. If this is true, at around 30, work outcomes will be reduced.
  - ☞ Will there be a jump in employment rates at 30 to the below?
- ?: Being an incumbent can give an additional benefit in the next election. If this is true, at the vote share margin close to zero, an incumbent vs. non-incumbent contrast gives effects of this benefit. Most suitable data comes from the US state gubernatorial elections where there are effectively only two candidates/parties.

- ?: In Quebec, unemployment benefits are increased once reaching the age of 30 for adults with no child. This should have disincentives to work for age 30 and older. If this is true, at around 30, work outcomes will be reduced.
  - ☞ Will there be a jump in employment rates at 30 to the below?
- ?: Being an incumbent can give an additional benefit in the next election. If this is true, at the vote share margin close to zero, an incumbent vs. non-incumbent contrast gives effects of this benefit. Most suitable data comes from the US state gubernatorial elections where there are effectively only two candidates/parties.
  - ☞ Will there be a jump in winning probability at zero vote margin to the above?

## ? Age and unemployment benefits in Quebec.

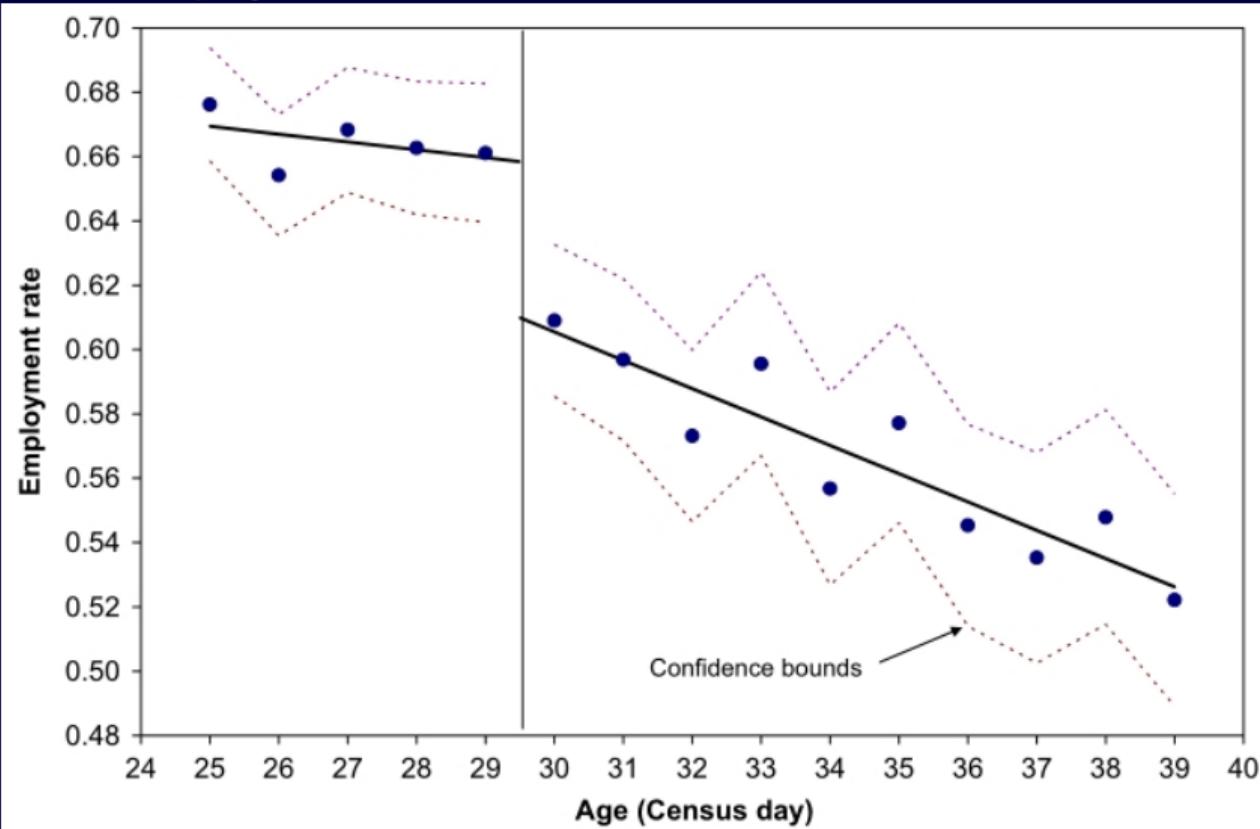
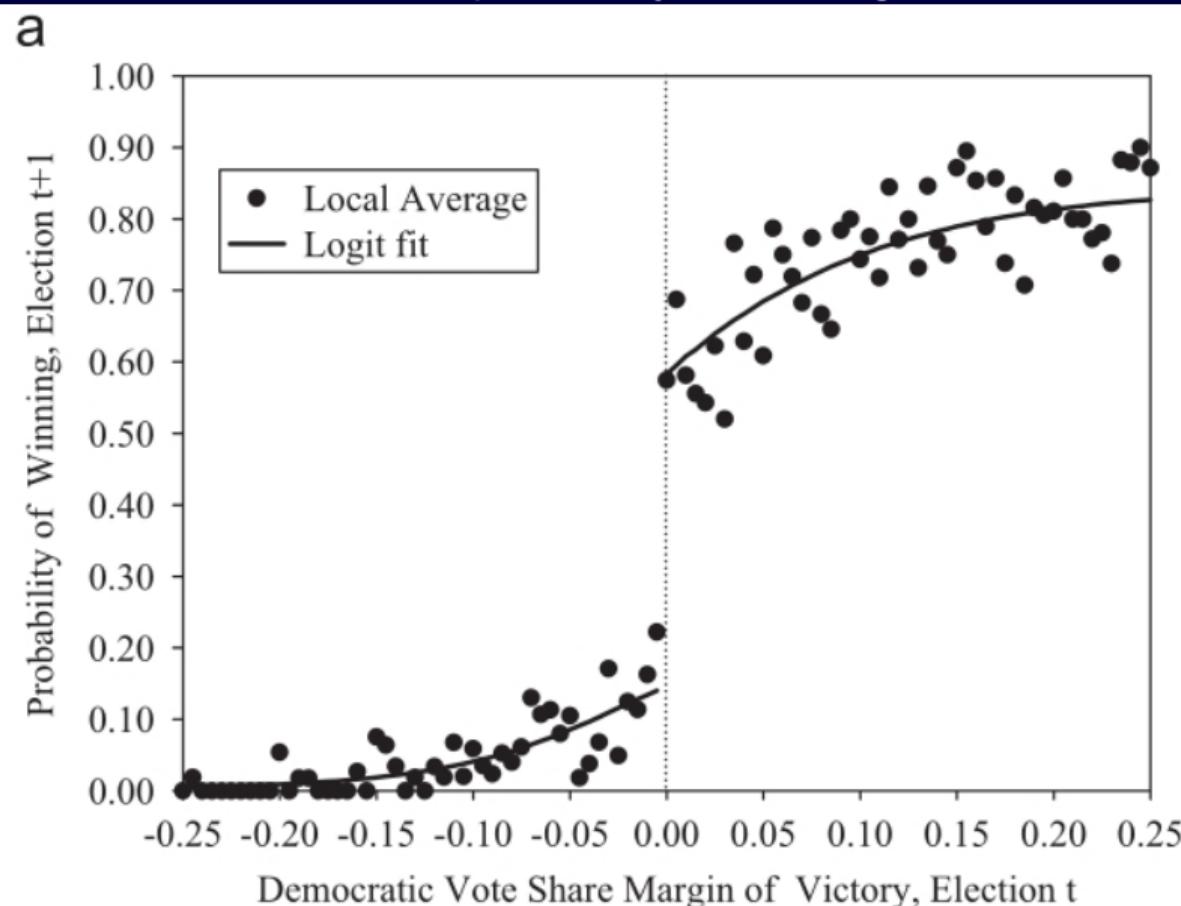


Fig. 3. Employment rate in Census week, Quebec 1986.

? Margin of vote share in  $t$  and probability of winning in  $t + 1$ .



There is an indirect yet nice use of data to assess the logical coherence of the results.

There is an indirect yet nice use of data to assess the logical coherence of the results.

Take a sample that should not be affected by the policy and test if there is an impact.

There is an indirect yet nice use of data to assess the logical coherence of the results.

Take a sample that should not be affected by the policy and test if there is an impact.

If there is an impact, then your main result is likely to be also picking up something different from the policy.

There is an indirect yet nice use of data to assess the logical coherence of the results.

Take a sample that should not be affected by the policy and test if there is an impact.

If there is an impact, then your main result is likely to be also picking up something different from the policy.

This is called a *placebo test*.

There is an indirect yet nice use of data to assess the logical coherence of the results.

Take a sample that should not be affected by the policy and test if there is an impact.

If there is an impact, then your main result is likely to be also picking up something different from the policy.

This is called a *placebo test*.

A placebo test is run only when the main estimation indicates a non-zero effect.

There is an indirect yet nice use of data to assess the logical coherence of the results.

Take a sample that should not be affected by the policy and test if there is an impact.

If there is an impact, then your main result is likely to be also picking up something different from the policy.

This is called a *placebo test*.

A placebo test is run only when the main estimation indicates a non-zero effect.

If the main estimate has a low  $p$  value (probability of null hypothesis ‘zero effect’ being true), one tries to run a placebo test, in a hope of finding a large  $p$  value in it, to get a further confidence that the result is not an accident.

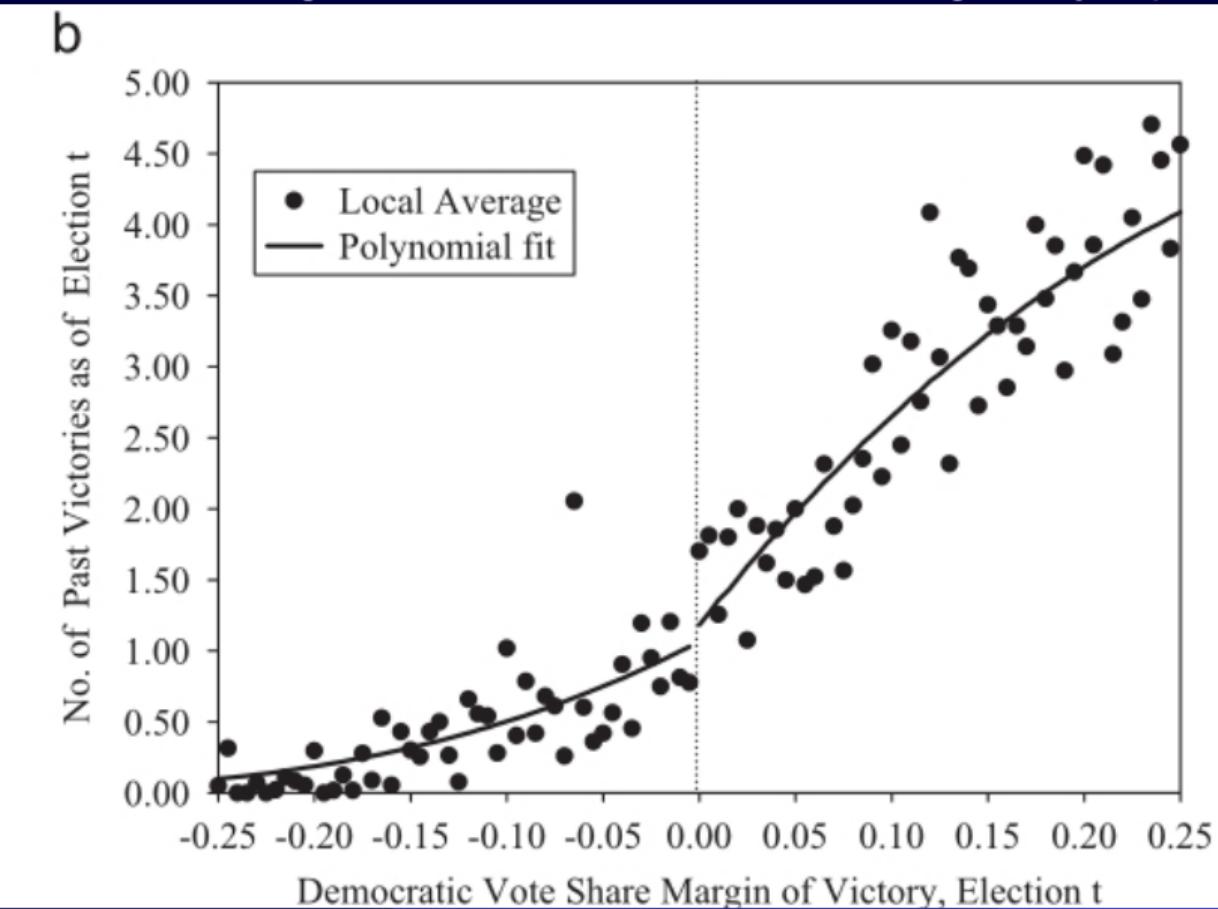
# Placebo tests in ?. Rest of Canada or post 1991 should not detect effects.

Table 3

Falsification test: comparing labour supply response in Quebec and rest of Canada in 1986 and 1991

Specification for age	Quebec, 1986	Rest of Canada, 1986	Quebec, 1991	Rest of Canada, 1991
<i>Regression discontinuity estimates: employment rate on Census week</i>				
Linear	-0.041*** (0.012)	-0.013** (0.006)	0.041* (0.022)	0.005 (0.011)
Quadratic	-0.051*** (0.012)	-0.013* (0.007)	0.012 (0.023)	-0.017*** (0.006)
Cubic	-0.048*** (0.014)	-0.009 (0.007)	0.037** (0.015)	-0.016** (0.007)
Linear spline	-0.049*** (0.011)	-0.014* (0.006)	0.010 (0.017)	-0.010 (0.007)
Quadratic spline	-0.056** (0.018)	-0.007 (0.010)	0.042* (0.022)	-0.007 (0.007)
<i>Regression discontinuity estimates: difference in employment rate</i>				
Linear	-0.029** (0.011)	-0.009 (0.007)	0.022* (0.011)	-0.007 (0.006)
Quadratic	-0.031** (0.012)	-0.006 (0.007)	0.022 (0.013)	-0.005 (0.006)
Cubic	-0.030** (0.013)	-0.004 (0.006)	0.020 (0.014)	-0.002 (0.006)
Linear spline	-0.032** (0.013)	-0.004 (0.008)	0.021 (0.014)	-0.003 (0.006)
Quadratic spline	-0.035* (0.016)	0.001 (0.009)	0.012 (0.016)	-0.005 (0.008)

Placebo tests in ?. Margin of vote share in  $t$  should not give a jump of events before  $t$ .



現在ではカットオフから遠い標本を使った多項式(曲線)をフィットさせることは誤りとされている

- Ⓐ カットオフから遠い標本は推計に重用されるべきではない
- Ⓑ 多項式はカットオフから遠い標本の僅かな値の変動によって推計値が大きく変化する

現在は直線、もしくは、local linear regression という遠い標本のウェイトを小さくして直線を推計する手法を使う

When applied to geographical boundaries, such as school zones, there may be parents who move across border for better education.

When applied to geographical boundaries, such as school zones, there may be parents who move across border for better education.

Then the cutoff becomes nondeterministic, or “fuzzy.” If the jump does not happen exactly at the predetermined cutoff, RDD is said to have a *fuzzy design*.

When applied to geographical boundaries, such as school zones, there may be parents who move across border for better education.

Then the cutoff becomes nondeterministic, or “fuzzy.” If the jump does not happen exactly at the predetermined cutoff, RDD is said to have a *fuzzy design*.

- ☞ There can be many instances of fuzzy RDD if one can fabricate the eligibility. In India's BPL example, the BPL card can be issued to a resident who pays a bribe. This introduces a noise in estimates, hence estimates become less precise.

When applied to geographical boundaries, such as school zones, there may be parents who move across border for better education.

Then the cutoff becomes nondeterministic, or “fuzzy.” If the jump does not happen exactly at the predetermined cutoff, RDD is said to have a *fuzzy design*.

- ☞ There can be many instances of fuzzy RDD if one can fabricate the eligibility. In India's BPL example, the BPL card can be issued to a resident who pays a bribe. This introduces a noise in estimates, hence estimates become less precise.
- ☞ Nonetheless, a fuzzy RDD can also give a consistent estimate of local impacts, as long as we have a large enough sample that complies with the cutoff.

When applied to geographical boundaries, such as school zones, there may be parents who move across border for better education.

Then the cutoff becomes nondeterministic, or “fuzzy.” If the jump does not happen exactly at the predetermined cutoff, RDD is said to have a *fuzzy design*.

- ☞ There can be many instances of fuzzy RDD if one can fabricate the eligibility. In India's BPL example, the BPL card can be issued to a resident who pays a bribe. This introduces a noise in estimates, hence estimates become less precise.
- ☞ Nonetheless, a fuzzy RDD can also give a consistent estimate of local impacts, as long as we have a large enough sample that complies with the cutoff.

If there is not cross cutoff movement, RDD is said to have a *sharp design*.

RDD は局地的な実験と捉えられる

境界内外で越境があったとしても、treatment assignment について人々の意向が不正確にしか反映されなければ、ランダムな割当の要素があるから

実験と同様、density test はランダム化を達成できたか検定

# RDD での識別仮定の信頼性チェック

RDD での識別仮定の信頼性チェック

placebo test 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定

RDD での識別仮定の信頼性チェック

placebo test 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる:  
*forcing variable* の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCary test)

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる:  
*forcing variable* の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCary test)...棄却 ⇒ 越境あり

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる:  
*forcing variable* の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCary test)...棄却 ⇒ 越境あり
- 治療確率 (*propensity score* という) の差が境界内外でゼロ (=帰無仮説: 越境が完全で治療確率に対し境界が無意味になる) か検定

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる:  
*forcing variable* の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCary test)...棄却 ⇒ 越境あり
- 治療確率 (*propensity score* という) の差が境界内外でゼロ (=帰無仮説: 越境が完全で治療確率に対し境界が無意味になる) か検定...棄却 ⇒ 越境なし

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる: forcing variable の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCrary test)...棄却 ⇒ 越境あり
- 治療確率 (propensity score という) の差が境界内外でゼロ (=帰無仮説: 越境が完全で治療確率に対し境界が無意味になる) か検定...棄却 ⇒ 越境なし
- 特定の特徴を持つ人たちが越境していないか: その他変数の平均値が境界内外で等しい (=帰無仮説: 越境がなく平均値の差が境界内外でゼロ) か検定

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる: forcing variable の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCrary test)...棄却 ⇒ 越境あり
- 治療確率 (propensity score という) の差が境界内外でゼロ (=帰無仮説: 越境が完全で治療確率に対し境界が無意味になる) か検定...棄却 ⇒ 越境なし
- 特定の特徴を持つ人たちが越境していないか: その他変数の平均値が境界内外で等しい (=帰無仮説: 越境がなく平均値の差が境界内外でゼロ) か検定...棄却 ⇒ 越境あり

## RDD での識別仮定の信頼性チェック

**placebo test** 欠落変数の影響を確認: 政策変化のない点の推計値がゼロを検定...棄却  
⇒ 推計した効果に疑問

**density test** 非越境 (local randomisation) の確認:

- 越境者が多ければ境界で標本が多くなり、分布に凹凸ができる: forcing variable の分布密度が境界内外で等しい (=帰無仮説: 越境がなく分布の差が境界内外でゼロ) か検定 (density test or McCary test)...棄却 ⇒ 越境あり
- 治療確率 (propensity score という) の差が境界内外でゼロ (=帰無仮説: 越境が完全で治療確率に対し境界が無意味になる) か検定...棄却 ⇒ 越境なし
- 特定の特徴を持つ人たちが越境していないか: その他変数の平均値が境界内外で等しい (=帰無仮説: 越境がなく平均値の差が境界内外でゼロ) か検定...棄却 ⇒ 越境あり

PT は論理的なチェックで傍証。DT はランダム化の検定。

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$  →  $\alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1}$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

- ☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$  →  $\alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1}$
- ☞ leakage のみがあるとき ( $p_0 > 0, p_1 = 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $\alpha - p_0\alpha$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$  →  $\alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1}$

☞ leakage のみがあるとき ( $p_0 > 0, p_1 = 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $\alpha - p_0\alpha \rightarrow \alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{1-p_0}$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

- ☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$  →  $\alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1}$
- ☞ leakage のみがあるとき ( $p_0 > 0, p_1 = 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $\alpha - p_0\alpha \rightarrow \alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{1-p_0}$
- ☞ exclusion と leakage があるとき ( $p_0 > 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $p_1\alpha - p_0\alpha$

fuzzy design の (越境者がいる) 場合: propensity score が 0 と 1 ではなく  
 $p_0, p_1 \in (0, 1), p_0 < p_1$  の場合

確率差  $p_1 - p_0$  でカットオフでの平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  を補正

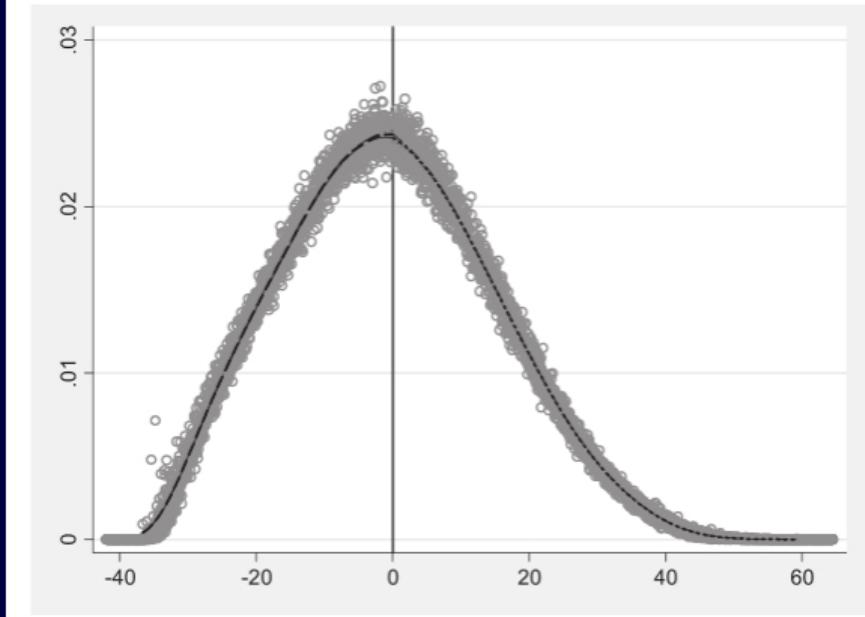
$$\text{カットオフ近辺での効果} = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}.$$

sharp design だったら  $p_1 = 1, p_0 = 0$

- ☞ exclusion のみがあるとき ( $p_0 = 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は真の効果  $\alpha$  の  $100 * p_1\%$  →  $\alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1}$
- ☞ leakage のみがあるとき ( $p_0 > 0, p_1 = 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $\alpha - p_0\alpha \rightarrow \alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{1 - p_0}$
- ☞ exclusion と leakage があるとき ( $p_0 > 0, p_1 < 1$ ): 計測される平均値差  $\bar{y}_{1+} - \bar{y}_{0-}$  は  $p_1\alpha - p_0\alpha \rightarrow \alpha = \frac{\bar{y}_{1+} - \bar{y}_{0-}}{p_1 - p_0}$

?, Figure 2: density test の例

McCary Test: -0.007(006)



確率密度は境界でジャンプしていないことが視覚的に分かる (検定結果の統計値

Sacred Heart, IDE

# ? 習熟度別学級実験(ケニア)

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が单一学級の小学校121校

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が単一学級の小学校 121 校

統御群 61 校: 単一学級 → 2つの学級 (契約教員 1 名を雇用)、生徒の割り振りは  
ランダム

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が単一学級の小学校 121 校

**統御群** 61 校: 単一学級 → 2つの学級 (契約教員 1 名を雇用)、生徒の割り振りはランダム

**治療群** 60 校: 単一学級 → 2つの習熟度学級 (契約教員 1 名を雇用)、1 学期試験点数で割り振り (成績百分位 50% がカットオフ)、契約教員と正規教員を習熟度学級についてランダムに配置

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が単一学級の小学校 121 校

**統御群** 61校: 単一学級 → 2つの学級(契約教員1名を雇用)、生徒の割り振りはランダム

**治療群** 60校: 単一学級 → 2つの習熟度学級(契約教員1名を雇用)、1学期試験点数で割り振り(成績百分位50%がカットオフ)、契約教員と正規教員を習熟度学級についてランダムに配置

ランダム化

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が単一学級の小学校 121 校

**統御群** 61校: 単一学級 → 2つの学級(契約教員1名を雇用)、生徒の割り振りはランダム

**治療群** 60校: 単一学級 → 2つの習熟度学級(契約教員1名を雇用)、1学期試験点数で割り振り(成績百分位50%がカットオフ)、契約教員と正規教員を習熟度学級についてランダムに配置

ランダム化

tracking schools: 習熟度学級有無 \* 教員雇用形態

## ?: 習熟度別学級実験(ケニア)

2005年2学期-2006年学年終了、ケニア西部、1年生が单一学級の小学校 121 校

**統御群** 61 校: 単一学級 → 2つの学級 (契約教員 1 名を雇用)、生徒の割り振りはランダム

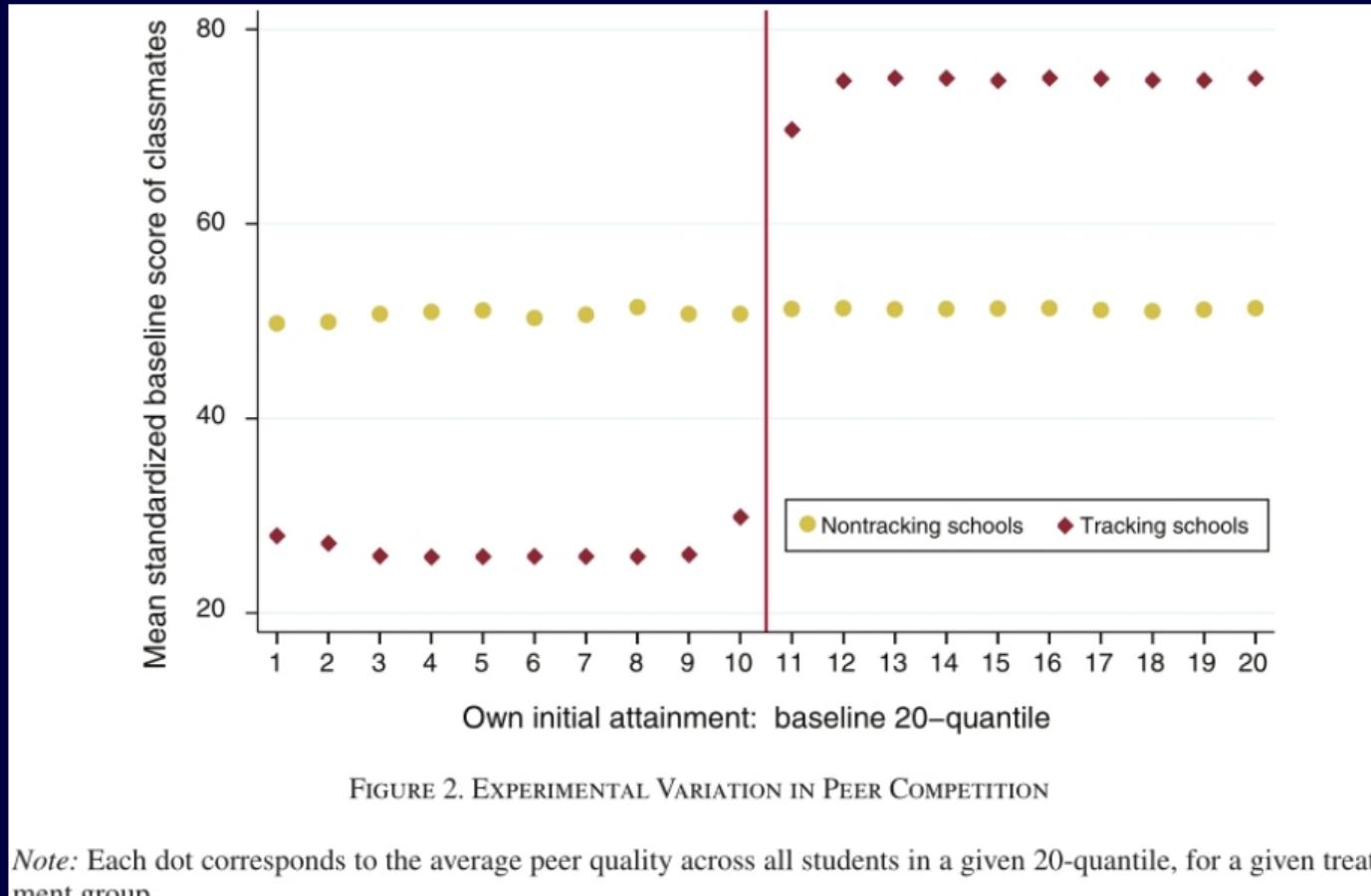
**治療群** 60 校: 単一学級 → 2つの習熟度学級 (契約教員 1 名を雇用)、1 学期試験点数で割り振り (成績百分位 50% がカットオフ)、契約教員と正規教員を習熟度学級についてランダムに配置

ランダム化

tracking schools: 習熟度学級有無 \* 教員雇用形態

nontracking schools: 級友 (peer) の成績 \* 教員雇用形態

# 習熟度学級の学校は平均点数が分かれる=実験は概ね正しく実施



Note: Each dot corresponds to the average peer quality across all students in a given 20-quantile, for a given treatment group.

# Local linear regression: 境界 (50) 点で 2006 年試験点数に変化なし

Panel B. Fan locally weighted regression

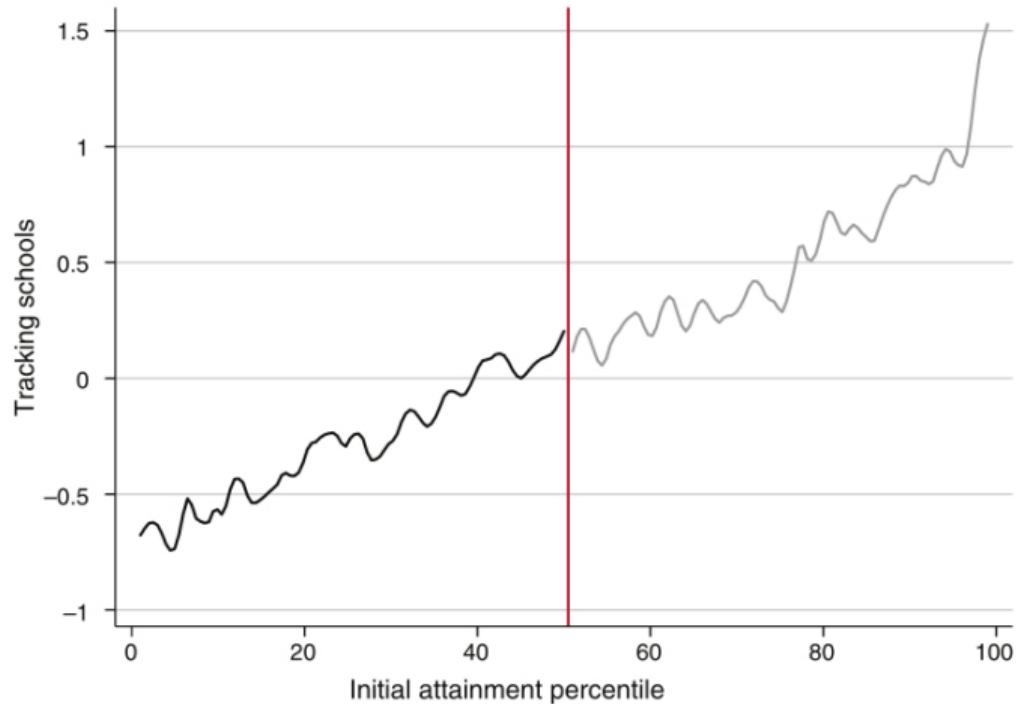


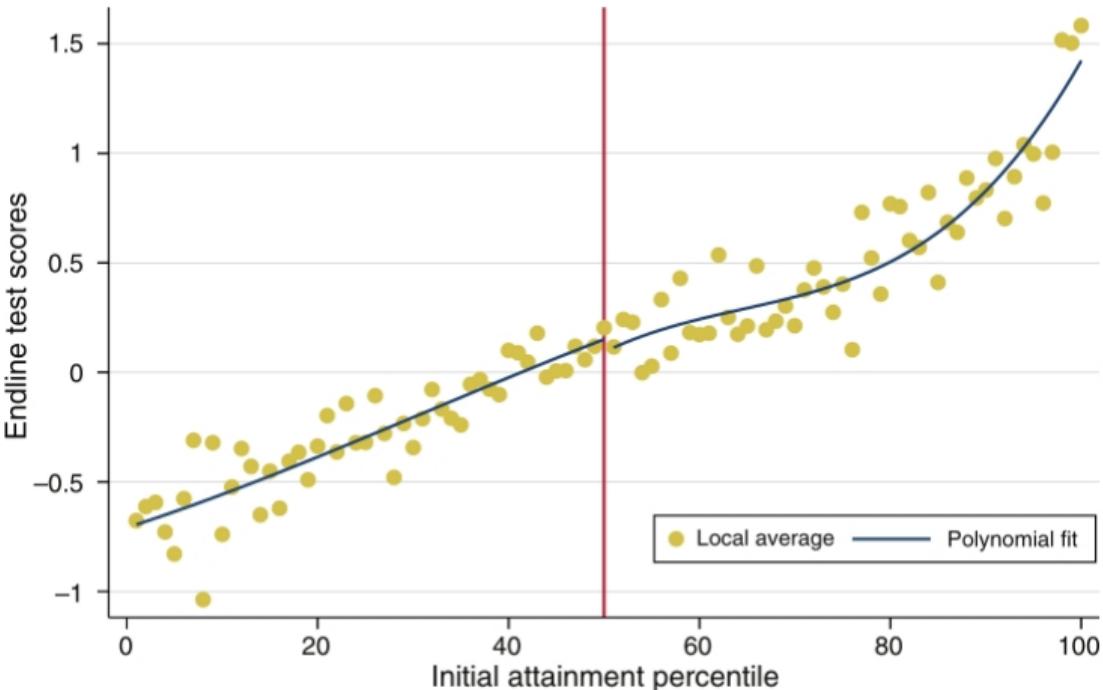
FIGURE A1. PEER QUALITY AND ENDLINE SCORES IN TRACKING SCHOOLS

Note: Fitted values from Fan's locally weighted regressions with quartic (biweight) kernels and a bandwidth of 2.0.

# Local average と global quadratic (2次関数、なぜやったのか不明)

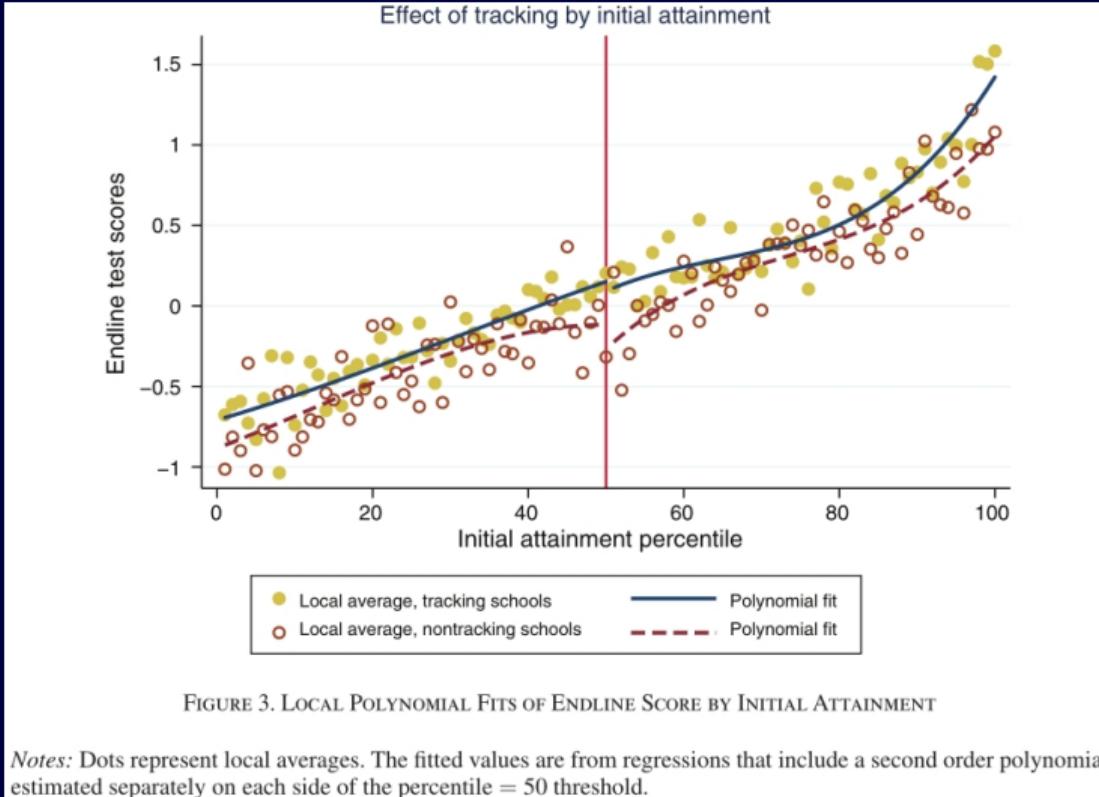
## APPENDIX

Panel A. Quadratic fit



Note: The dots are the average score. The fitted values are from regressions that include a second order polynomial estimated separately on each side of the percentile = 50 threshold.

# 習熟度学級の学校は分布全体で 2006 年末試験点数が上がる



“local polynomial fits” は、注から類推するとおそらく 2 次関数なので信頼性はあり、位置は上に来そうだが各点での検定 (nontracking との差がゼロ) の  $p$  値は示されてい

TABLE 2—OVERALL EFFECT OF TRACKING

	Total score				Math score		Literacy score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. Short-run effects (after 18 months in program)</i>								
(1) Tracking school	0.139 (0.078)*	0.176 (0.077)**	0.192 (0.093)**	0.182 (0.093)*	0.139 (0.073)*	0.156 (0.083)*	0.198 (0.108)*	0.166 (0.098)*
(2) In bottom half of initial distribution × tracking school			-0.036 (0.07)		0.04 (0.07)		-0.091 (0.08)	
(3) In bottom quarter × tracking school				-0.045 (0.08)		0.012 (0.09)		-0.083 (0.08)
(4) In second-to-bottom quarter × tracking school				-0.013 (0.07)		0.026 (0.08)		-0.042 (0.07)
(5) In top quarter × tracking school				0.027 (0.08)		-0.026 (0.07)		0.065 (0.08)
(6) Assigned to contract teacher	0.181 (0.038)***	0.18 (0.038)***	0.18 (0.038)***	0.16 (0.038)***	0.161 (0.037)***	0.16 (0.038)***	0.16 (0.038)***	0.16 (0.038)***
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,795	5,279	5,279	5,279	5,280	5,280	5,280	5,280
<i>Total effects on bottom half and bottom quarter</i>								
Coeff (Row 1) + Coeff (Row 2)			0.156		0.179		0.107	
Coeff (Row 1) + Coeff (Row 3)				0.137		0.168		0.083
F-test: total effect = 0			4.40	2.843	5.97	3.949	2.37	1.411
p-value (total effect for bottom = 0)			0.038	0.095	0.016	0.049	0.127	0.237
p-value (effect for top quarter = effect for bottom quarter)				0.507		0.701		0.209
<i>Panel B. Longer-run effects (a year after program ended)</i>								
(1) Tracking school	0.163 (0.069)**	0.178 (0.073)**	0.216 (0.079)***	0.235 (0.088)***	0.143 (0.064)**	0.168 (0.075)**	0.231 (0.089)**	0.241 (0.096)**

TABLE 2—OVERALL EFFECT OF TRACKING

	Total score				Math score		Literacy score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. Short-run effects (after 18 months in program)</i>								
(1) Tracking school	0.139 (0.078)*	0.176 (0.077)**	0.192 (0.093)**	0.182 (0.093)*	0.139 (0.073)*	0.156 (0.083)*	0.198 (0.108)*	0.166 (0.098)*
(2) In bottom half of initial distribution × tracking school			-0.036 (0.07)		0.04 (0.07)		-0.091 (0.08)	
(3) In bottom quarter × tracking school				-0.045 (0.08)		0.012 (0.09)		-0.083 (0.08)
(4) In second-to-bottom quarter × tracking school				-0.013 (0.07)		0.026 (0.08)		-0.042 (0.07)
(5) In top quarter × tracking school				0.027 (0.08)		-0.026 (0.07)		0.065 (0.08)
(6) Assigned to contract teacher	0.181 (0.038)***	0.18 (0.038)***	0.18 (0.038)***	0.16 (0.038)***	0.161 (0.037)***	0.16 (0.038)***	0.16 (0.038)***	0.16 (0.038)***
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,795	5,279	5,279	5,279	5,280	5,280	5,280	5,280
<i>Total effects on bottom half and bottom quarter</i>								
Coeff (Row 1) + Coeff (Row 2)			0.156		0.179		0.107	
Coeff (Row 1) + Coeff (Row 3)				0.137		0.168		0.083
F-test: total effect = 0			4.40	2.843	5.97	3.949	2.37	1.411
p-value (total effect for bottom = 0)			0.038	0.095	0.016	0.049	0.127	0.237
p-value (effect for top quarter = effect for bottom quarter)				0.507		0.701		0.209
<i>Panel B. Longer-run effects (a year after program ended)</i>								
(1) Tracking school	0.163 (0.069)**	0.178 (0.073)**	0.216 (0.079)***	0.235 (0.088)***	0.143 (0.064)**	0.168 (0.075)**	0.231 (0.089)**	0.241 (0.096)**

全体の平均値の差がゼロという帰無仮説の *p* 値

TABLE 2—OVERALL EFFECT OF TRACKING

	Total score				Math score		Literacy score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. Short-run effects (after 18 months in program)</i>								
(1) Tracking school	0.139 (0.078)*	0.176 (0.077)**	0.192 (0.093)**	0.182 (0.093)*	0.139 (0.073)*	0.156 (0.083)*	0.198 (0.108)*	0.166 (0.098)*
(2) In bottom half of initial distribution × tracking school			-0.036 (0.07)		0.04 (0.07)		-0.091 (0.08)	
(3) In bottom quarter × tracking school				-0.045 (0.08)		0.012 (0.09)		-0.083 (0.08)
(4) In second-to-bottom quarter × tracking school				-0.013 (0.07)		0.026 (0.08)		-0.042 (0.07)
(5) In top quarter × tracking school				0.027 (0.08)		-0.026 (0.07)		0.065 (0.08)
(6) Assigned to contract teacher	0.181 (0.038)***	0.18 (0.038)***	0.18 (0.038)***	0.16 (0.038)***	0.161 (0.037)***	0.16 (0.038)***	0.16 (0.038)***	0.16 (0.038)***
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,795	5,279	5,279	5,279	5,280	5,280	5,280	5,280
<i>Total effects on bottom half and bottom quarter</i>								
Coeff (Row 1) + Coeff (Row 2)			0.156		0.179		0.107	
Coeff (Row 1) + Coeff (Row 3)				0.137		0.168		0.083
F-test: total effect = 0			4.40	2.843	5.97	3.949	2.37	1.411
p-value (total effect for bottom = 0)			0.038	0.095	0.016	0.049	0.127	0.237
p-value (effect for top quarter = effect for bottom quarter)				0.507		0.701		0.209
<i>Panel B. Longer-run effects (a year after program ended)</i>								
(1) Tracking school	0.163 (0.069)**	0.178 (0.073)**	0.216 (0.079)***	0.235 (0.088)***	0.143 (0.064)**	0.168 (0.075)**	0.231 (0.089)**	0.241 (0.096)**

全体の平均値の差がゼロという帰無仮説の *p* 値

2006 年末: 10%以下

TABLE 2—OVERALL EFFECT OF TRACKING

	Total score				Math score		Literacy score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. Short-run effects (after 18 months in program)</i>								
(1) Tracking school	0.139 (0.078)*	0.176 (0.077)**	0.192 (0.093)**	0.182 (0.093)*	0.139 (0.073)*	0.156 (0.083)*	0.198 (0.108)*	0.166 (0.098)*
(2) In bottom half of initial distribution × tracking school		-0.036 (0.07)			0.04 (0.07)		-0.091 (0.08)	
(3) In bottom quarter × tracking school			-0.045 (0.08)		0.012 (0.09)		-0.083 (0.08)	
(4) In second-to-bottom quarter × tracking school			-0.013 (0.07)		0.026 (0.08)		-0.042 (0.07)	
(5) In top quarter × tracking school			0.027 (0.08)		-0.026 (0.07)		0.065 (0.08)	
(6) Assigned to contract teacher	0.181 (0.038)***	0.18 (0.038)***	0.18 (0.038)***	0.16 (0.038)***	0.161 (0.037)***	0.16 (0.038)***	0.16 (0.038)***	0.16 (0.038)***
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,795	5,279	5,279	5,279	5,280	5,280	5,280	5,280
<i>Total effects on bottom half and bottom quarter</i>								
Coeff (Row 1) + Coeff (Row 2)		0.156			0.179		0.107	
Coeff (Row 1) + Coeff (Row 3)			0.137		0.168		0.083	
F-test: total effect = 0		4.40	2.843	5.97	3.949	2.37	1.411	
p-value (total effect for bottom = 0)		0.038	0.095	0.016	0.049	0.127	0.237	
p-value (effect for top quarter = effect for bottom quarter)			0.507		0.701		0.209	
<i>Panel B. Longer-run effects (a year after program ended)</i>								
(1) Tracking school	0.163 (0.069)**	0.178 (0.073)**	0.216 (0.079)***	0.235 (0.088)***	0.143 (0.064)**	0.168 (0.075)**	0.231 (0.089)**	0.241 (0.096)**

全体の平均値の差がゼロという帰無仮説の *p* 値

2006 年末: 10%以下  
2007 年末: 5%以下

TABLE 4—PEER QUALITY: EXOGENOUS VARIATION IN PEER QUALITY (NONTRACKING SCHOOLS ONLY)

	All			25th–75th percentiles only	Bottom 25th percentiles	Top 25th percentiles only
	Total score (1)	Math score (2)	Lit score (3)	Total score (4)	Total score (5)	Total score (6)
<i>Panel A. Reduced form</i>						
Average baseline score of classmates <sup>a</sup>	0.346 (0.150)**	0.323 (0.160)**	0.293 (0.131)**	-0.052 (0.227)	0.505 (0.199)**	0.893 (0.330)***
Observations	2,188	2,188	2,188	2,188	2,188	2,188
School fixed effects	x	x	x	x	x	x
<i>Panel B. IV</i>						
Average endline score of classmates (predicted)	0.445 (0.117)***	0.470 (0.124)***	0.423 (0.120)***	-0.063 (0.306)	0.855 (0.278)***	1.052 (0.368)***
Observations	2,188	2,188	2,189	1,091	524	573
School fixed effects	x	x	x	x	x	x
<i>Panel C. First-Stage for IV: average endline score of classmates</i>						
	Average total score	Average math score	Average lit score	Average total score	Average total Score	Average total score
Average (standardized) baseline score of classmates <sup>‡</sup>	0.768 (0.033)***	0.680 (0.033)***	0.691 (0.030)***	0.795 (0.056)***	0.757 (0.066)***	0.794 (0.070)***

Notes: Sample restricted to the 61 nontracking schools (where students were randomly assigned to a section). Individual controls included but not shown: gender, age, being assigned to the contract teacher, and own baseline score. Robust standard errors clustered at the school level in parentheses.

<sup>a</sup>This variable has a mean of 0.0009 and a standard deviation of 0.1056. We define classmates as follows: two students in the same section are classmates; two students in the same grade but different sections are not classmates.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

TABLE 4—PEER QUALITY: EXOGENOUS VARIATION IN PEER QUALITY (NONTRACKING SCHOOLS ONLY)

	All			25th–75th percentiles only	Bottom 25th percentiles	Top 25th percentiles only
	Total score (1)	Math score (2)	Lit score (3)	Total score (4)	Total score (5)	Total score (6)
<i>Panel A. Reduced form</i>						
Average baseline score of classmates <sup>a</sup>	0.346 (0.150)**	0.323 (0.160)**	0.293 (0.131)**	-0.052 (0.227)	0.505 (0.199)**	0.893 (0.330)***
Observations	2,188	2,188	2,188	2,188	2,188	2,188
School fixed effects	x	x	x	x	x	x
<i>Panel B. IV</i>						
Average endline score of classmates (predicted)	0.445 (0.117)***	0.470 (0.124)***	0.423 (0.120)***	-0.063 (0.306)	0.855 (0.278)***	1.052 (0.368)***
Observations	2,188	2,188	2,189	1,091	524	573
School fixed effects	x	x	x	x	x	x
<i>Panel C. First-Stage for IV: average endline score of classmates</i>						
	Average total score	Average math score	Average lit score	Average total score	Average total Score	Average total score
Average (standardized) baseline score of classmates <sup>‡</sup>	0.768 (0.033)***	0.680 (0.033)***	0.691 (0.030)***	0.795 (0.056)***	0.757 (0.066)***	0.794 (0.070)***

Notes: Sample restricted to the 61 nontracking schools (where students were randomly assigned to a section). Individual controls included but not shown: gender, age, being assigned to the contract teacher, and own baseline score. Robust standard errors clustered at the school level in parentheses.

<sup>a</sup>This variable has a mean of 0.0009 and a standard deviation of 0.1056. We define classmates as follows: two students in the same section are classmates; two students in the same grade but different sections are not classmates.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

Nontracking 学校は peer effects が平均的に正

TABLE 4—PEER QUALITY: EXOGENOUS VARIATION IN PEER QUALITY (NONTRACKING SCHOOLS ONLY)

	All			25th–75th percentiles only	Bottom 25th percentiles	Top 25th percentiles only
	Total score (1)	Math score (2)	Lit score (3)	Total score (4)	Total score (5)	Total score (6)
<i>Panel A. Reduced form</i>						
Average baseline score of classmates <sup>a</sup>	0.346 (0.150)**	0.323 (0.160)**	0.293 (0.131)**	-0.052 (0.227)	0.505 (0.199)**	0.893 (0.330)***
Observations	2,188	2,188	2,188	2,188	2,188	2,188
School fixed effects	x	x	x	x	x	x
<i>Panel B. IV</i>						
Average endline score of classmates (predicted)	0.445 (0.117)***	0.470 (0.124)***	0.423 (0.120)***	-0.063 (0.306)	0.855 (0.278)***	1.052 (0.368)***
Observations	2,188	2,188	2,189	1,091	524	573
School fixed effects	x	x	x	x	x	x
<i>Panel C. First-Stage for IV: average endline score of classmates</i>						
	Average total score	Average math score	Average lit score	Average total score	Average total Score	Average total score
Average (standardized) baseline score of classmates <sup>‡</sup>	0.768 (0.033)***	0.680 (0.033)***	0.691 (0.030)***	0.795 (0.056)***	0.757 (0.066)***	0.794 (0.070)***

Notes: Sample restricted to the 61 nontracking schools (where students were randomly assigned to a section). Individual controls included but not shown: gender, age, being assigned to the contract teacher, and own baseline score. Robust standard errors clustered at the school level in parentheses.

<sup>a</sup>This variable has a mean of 0.0009 and a standard deviation of 0.1056. We define classmates as follows: two students in the same section are classmates; two students in the same grade but different sections are not classmates.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

Nontracking 学校は peer effects が平均的に正

とくに下位と上位の生徒において正

TABLE 4—PEER QUALITY: EXOGENOUS VARIATION IN PEER QUALITY (NONTRACKING SCHOOLS ONLY)

	All			25th–75th percentiles only	Bottom 25th percentiles	Top 25th percentiles only
	Total score (1)	Math score (2)	Lit score (3)	Total score (4)	Total score (5)	Total score (6)
<i>Panel A. Reduced form</i>						
Average baseline score of classmates <sup>a</sup>	0.346 (0.150)**	0.323 (0.160)**	0.293 (0.131)**	-0.052 (0.227)	0.505 (0.199)**	0.893 (0.330)***
Observations	2,188	2,188	2,188	2,188	2,188	2,188
School fixed effects	x	x	x	x	x	x
<i>Panel B. IV</i>						
Average endline score of classmates (predicted)	0.445 (0.117)***	0.470 (0.124)***	0.423 (0.120)***	-0.063 (0.306)	0.855 (0.278)***	1.052 (0.368)***
Observations	2,188	2,188	2,189	1,091	524	573
School fixed effects	x	x	x	x	x	x
<i>Panel C. First-Stage for IV: average endline score of classmates</i>						
	Average total score	Average math score	Average lit score	Average total score	Average total Score	Average total score
Average (standardized) baseline score of classmates <sup>‡</sup>	0.768 (0.033)***	0.680 (0.033)***	0.691 (0.030)***	0.795 (0.056)***	0.757 (0.066)***	0.794 (0.070)***

Notes: Sample restricted to the 61 nontracking schools (where students were randomly assigned to a section). Individual controls included but not shown: gender, age, being assigned to the contract teacher, and own baseline score. Robust standard errors clustered at the school level in parentheses.

<sup>a</sup>This variable has a mean of 0.0009 and a standard deviation of 0.1056. We define classmates as follows: two students in the same section are classmates; two students in the same grade but different sections are not classmates.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

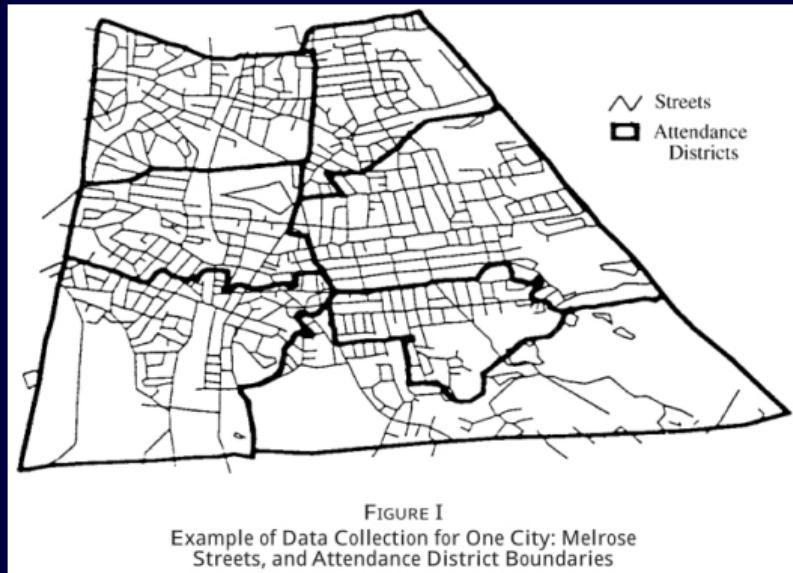
\* Significant at the 10 percent level.

Nontracking 学校は peer effects が平均的に正

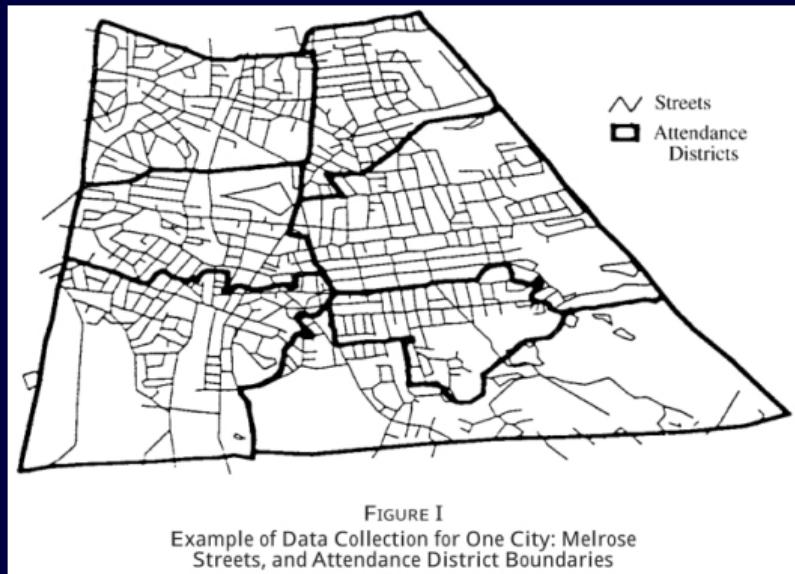
とくに下位と上位の生徒において正

下位が正なのは想像に難くないが、上位が正なのは解釈が複雑

# 学区と成績と地価(??)

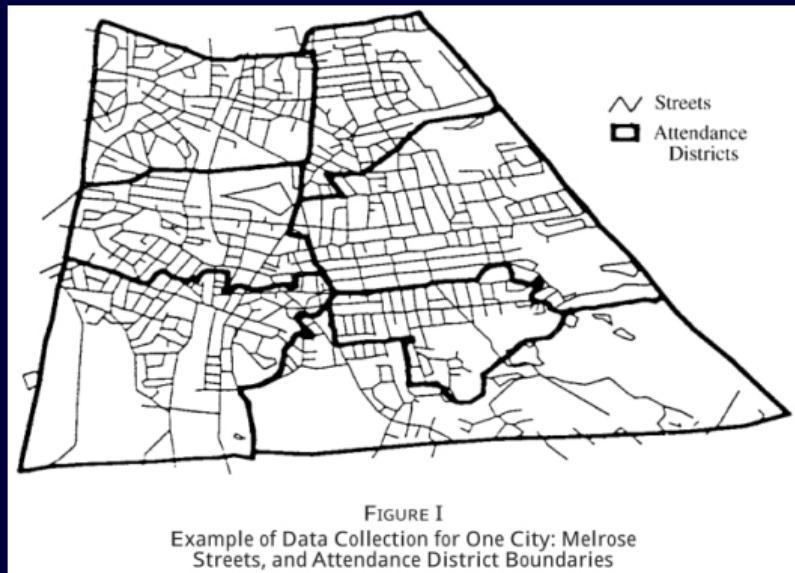


# 学区と成績と地価(??)



学区境界を挟んだ家計グループを比較

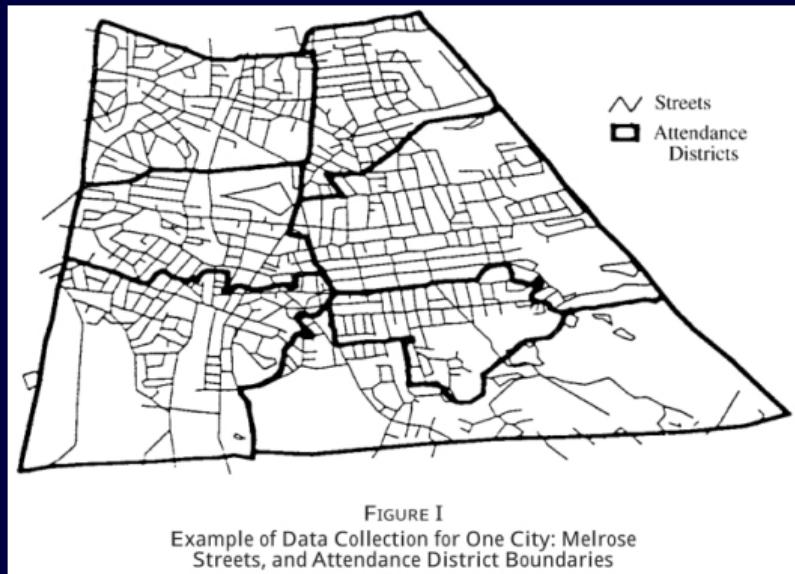
# 学区と成績と地価(??)



学区境界を挟んだ家計グループを比較

forcing variable=(垂直距離で)最も近い  
境界までの距離

# 学区と成績と地価(??)

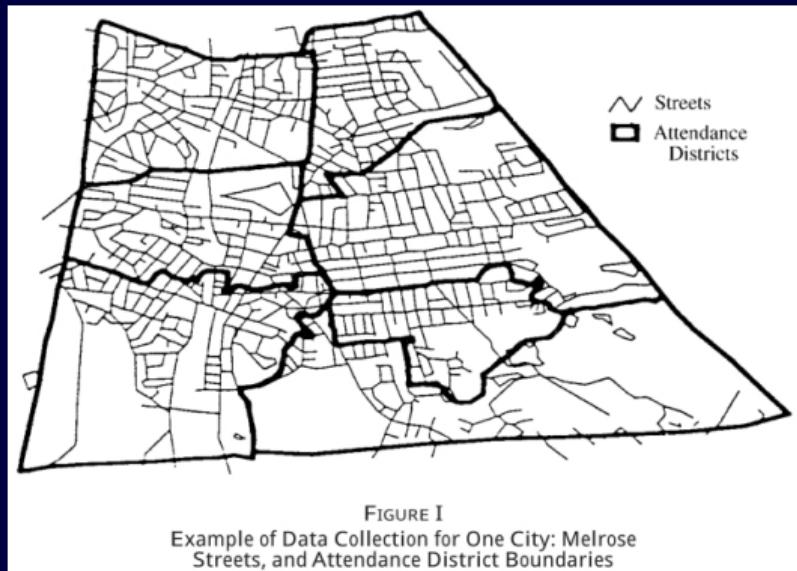


学区境界を挟んだ家計グループを比較

forcing variable=(垂直距離で)最も近い  
境界までの距離

outcome variables=試験点数、不動産  
価格

# 学区と成績と地価(??)



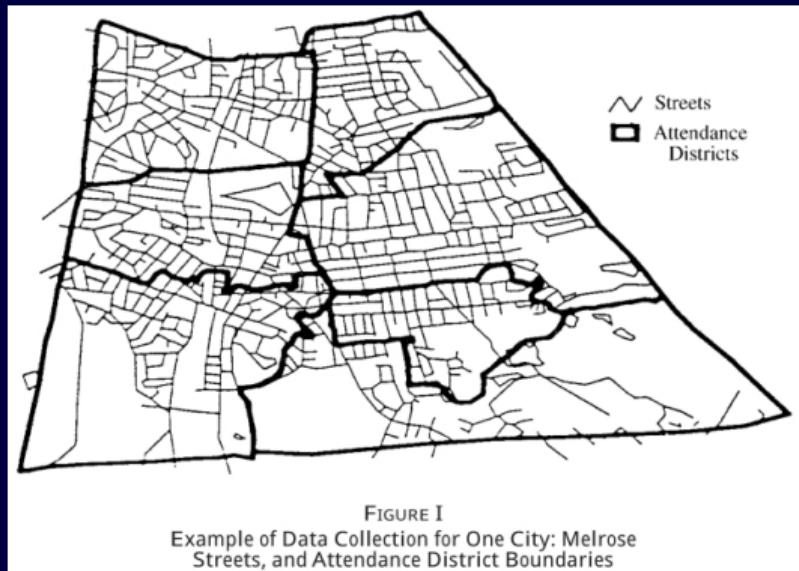
?: 不動産価格(不動産取引データ)、学区と各校試験点数(MEAP)、その他変数(センサス)

学区境界を挟んだ家計グループを比較

forcing variable=(垂直距離で)最も近い境界までの距離

outcome variables=試験点数、不動産価格

# 学区と成績と地価(??)



?: 不動産価格(不動産取引データ)、学区と各校試験点数(MEAP)、その他変数(センサス)

?: 不動産価格と試験点数(センサス)、不動産価格(不動産取引データ)、その他変数(センサス)

学区境界を挟んだ家計グループを比較

forcing variable=(垂直距離で)最も近い境界までの距離

outcome variables=試験点数、不動産価格

# 学区や学校単位での比較

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていた可能性あり

## 学区境界を挟んだ狭い地域同士の比較

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

△ 学区境界のどちら側かがランダムと考えることはできない

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

- 学区境界のどちら側かがランダムと考えることはできない... 親は学区を気にして家を買う・借りる

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

- Ⓐ 学区境界のどちら側かがランダムと考えることはできない... 親は学区を気にして家を買う・借りる
- Ⓑ 親の能力や資産などが同じとは思えない

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

- Ⓐ 学区境界のどちら側かがランダムと考えることはできない... 親は学区を気にして家を買う・借りる
- Ⓑ 親の能力や資産などが同じとは思えない... しかし、観察可能な変数(学歴や所得など)はPT、DTで検定可能

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えていたり可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

- Ⓐ 学区境界のどちら側かがランダムと考えることはできない... 親は学区を気にして家を買う・借りる
- Ⓑ 親の能力や資産などが同じとは思えない... しかし、観察可能な変数(学歴や所得など)はPT、DTで検定可能
- Ⓒ 川、公園、ゴルフ場境界などで区切られた境界だと、境界内外で観察できない変数は同じと思えないで標本から除外

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の状況、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えている可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い(のはず)

- Ⓐ 学区境界のどちら側かがランダムと考えることはできない... 親は学区を気にして家を買う・借りる
- Ⓑ 親の能力や資産などが同じとは思えない... しかし、観察可能な変数(学歴や所得など)はPT、DTで検定可能
- Ⓒ 川、公園、ゴルフ場境界などで区切られた境界だと、境界内外で観察できない変数は同じと思ないので標本から除外
- Ⓓ 少なくとも家庭の外部環境要因(治安、騒音、公害など)は揃えて推計する

## 学区や学校単位での比較

観察できない学区間の違い(欠落変数 omitted variable、親の知的能力、家庭の構成、教員の質や熱意、教授法の違い、学級崩壊度、治安、騒音、環境汚染など)が結果に影響を与えている可能性あり

## 学区境界を挟んだ狭い地域同士の比較

学区境界に共通する観察できない変数の一部は同じと想定できる  
⇒ 結果の差の一部は学区の違い（のはず）

- 学区境界のどちら側かがランダムと考えることはできない…親は学区を気にして家を買う・借りる
  - 親の能力や資産などが同じとは思えない…しかし、観察可能な変数(学歴や所得など)はPT、DTで検定可能
  - 川、公園、ゴルフ場境界などで区切られた境界だと、境界内外で観察できない変数は同じと思えないので標本から除外
  - 少なくとも家庭の外部環境要因(治安、騒音、公害など)は撇えて推計する

TABLE IV  
MAGNITUDE OF RESULTS<sup>a</sup>

	(1) Basic hedonic regression <sup>d</sup>	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score <sup>b</sup>	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores) <sup>c</sup>	4.9%	2.3%	1.8%	2.1%
\$ Value (at mean tax-adjusted house price of \$188,000 in \$1993)	\$9212	\$4324	\$3384	\$3948
\$ Value (at median tax-adjusted house price of \$158,000 in \$1993)	\$7742	\$3634	\$2844	\$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). *Source:* Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix I for more complete results.

TABLE IV  
MAGNITUDE OF RESULTS<sup>a</sup>

	(1) Basic hedonic regression <sup>d</sup>	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score <sup>b</sup>	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores) <sup>c</sup>	4.9%	2.3%	1.8%	2.1%
\$ Value (at mean tax-adjusted house price of \$188,000 in \$1993)	\$9212	\$4324	\$3384	\$3948
\$ Value (at median tax-adjusted house price of \$158,000 in \$1993)	\$7742	\$3634	\$2844	\$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). *Source:* Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix I for more complete results.

区間比較 4.9%よりも  
地価に与える効果は小  
さい

TABLE IV  
MAGNITUDE OF RESULTS<sup>a</sup>

	(1) Basic hedonic regression <sup>d</sup>	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score <sup>b</sup>	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores) <sup>c</sup>	4.9%	2.3%	1.8%	2.1%
\$ Value (at mean tax-adjusted house price of \$188,000 in \$1993)	\$9212	\$4324	\$3384	\$3948
\$ Value (at median tax-adjusted house price of \$158,000 in \$1993)	\$7742	\$3634	\$2844	\$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). *Source:* Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix I for more complete results.

区間比較 4.9%よりも  
地価に与える効果は小  
さい

観察できない変数を揃  
えると地価に与える影  
響は半分以下(2.3, 1.8,  
2.1)になる

TABLE IV  
MAGNITUDE OF RESULTS<sup>a</sup>

	(1) Basic hedonic regression <sup>d</sup>	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score <sup>b</sup>	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores) <sup>c</sup>	4.9%	2.3%	1.8%	2.1%
\$ Value (at mean tax-adjusted house price of \$188,000 in \$1993)	\$9212	\$4324	\$3384	\$3948
\$ Value (at median tax-adjusted house price of \$158,000 in \$1993)	\$7742	\$3634	\$2844	\$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). *Source:* Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix I for more complete results.

区間比較 4.9%よりも  
地価に与える効果は小  
さい

観察できない変数を揃  
えると地価に与える影  
響は半分以下(2.3, 1.8,  
2.1)になる



TABLE IV  
MAGNITUDE OF RESULTS<sup>a</sup>

	(1) Basic hedonic regression <sup>d</sup>	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score <sup>b</sup>	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores) <sup>c</sup>	4.9%	2.3%	1.8%	2.1%
\$ Value (at mean tax-adjusted house price of \$188,000 in \$1993)	\$9212	\$4324	\$3384	\$3948
\$ Value (at median tax-adjusted house price of \$158,000 in \$1993)	\$7742	\$3634	\$2844	\$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). Source: Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix I for more complete results.

区間比較 4.9%よりも  
地価に与える効果は小  
さい

観察できない変数を揃  
えると地価に与える影  
響は半分以下(2.3, 1.8,  
2.1)になる

⇒ 学区間比較での効  
果: 観察できない変数  
の効果が半分以上

? 左列はセンサス(回答者情報)データ、右列は不動産取引データ

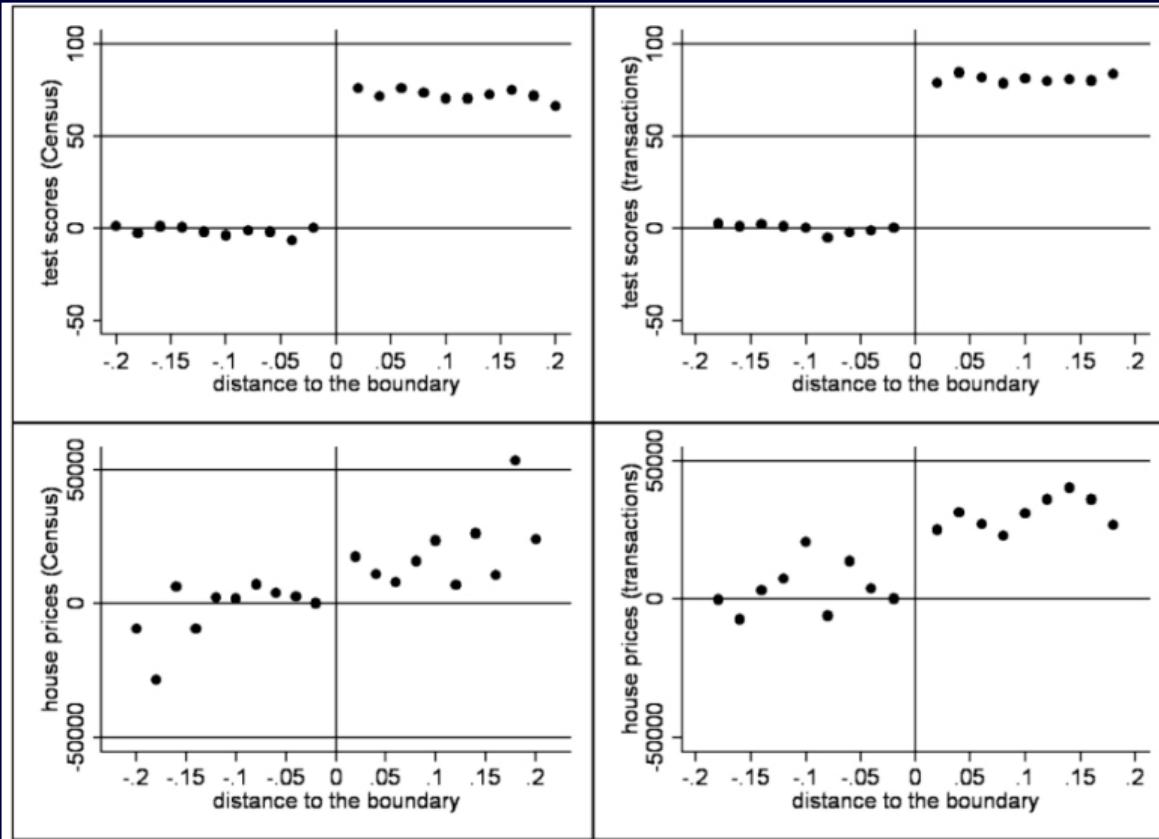


FIG. 1.—Test scores and house prices around the boundary. Each panel is constructed  
Ito (IDE, Sacred Heart)

? 物件の特徴は境界内外で連続

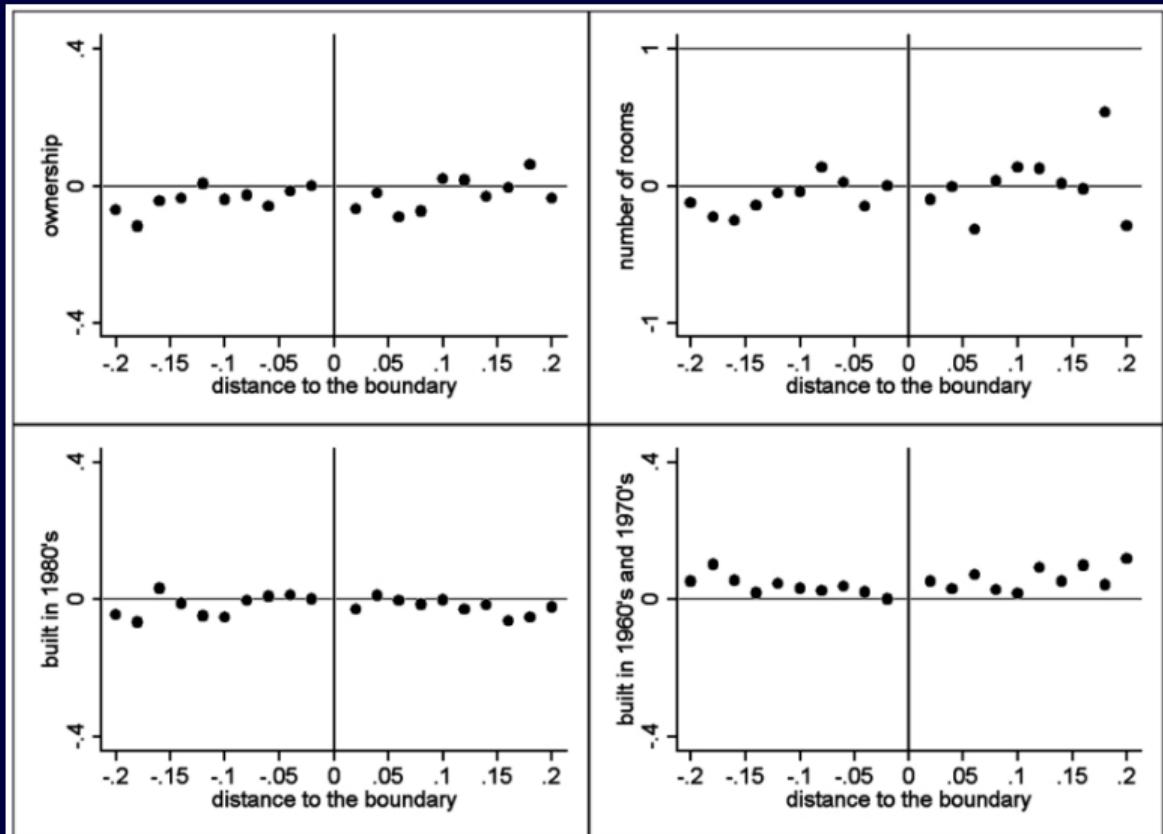


FIG. 2.—Census housing characteristics around the boundary. Each panel is constructed  
Ito (IDE, Sacred Heart)

? 物件の特徴は境界内外で連続

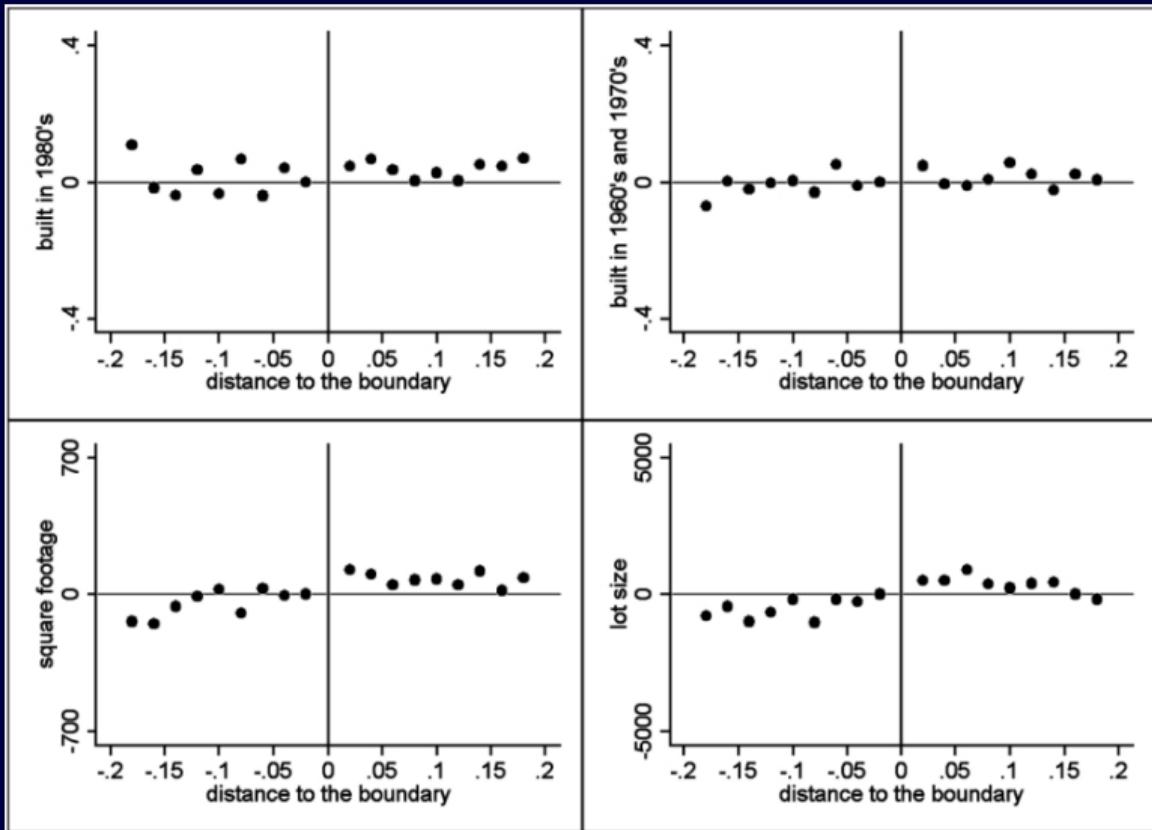


FIG. 3.—Transactions data housing characteristics around the boundary. Each panel is

? 居住者の特徴は境界内外で相違

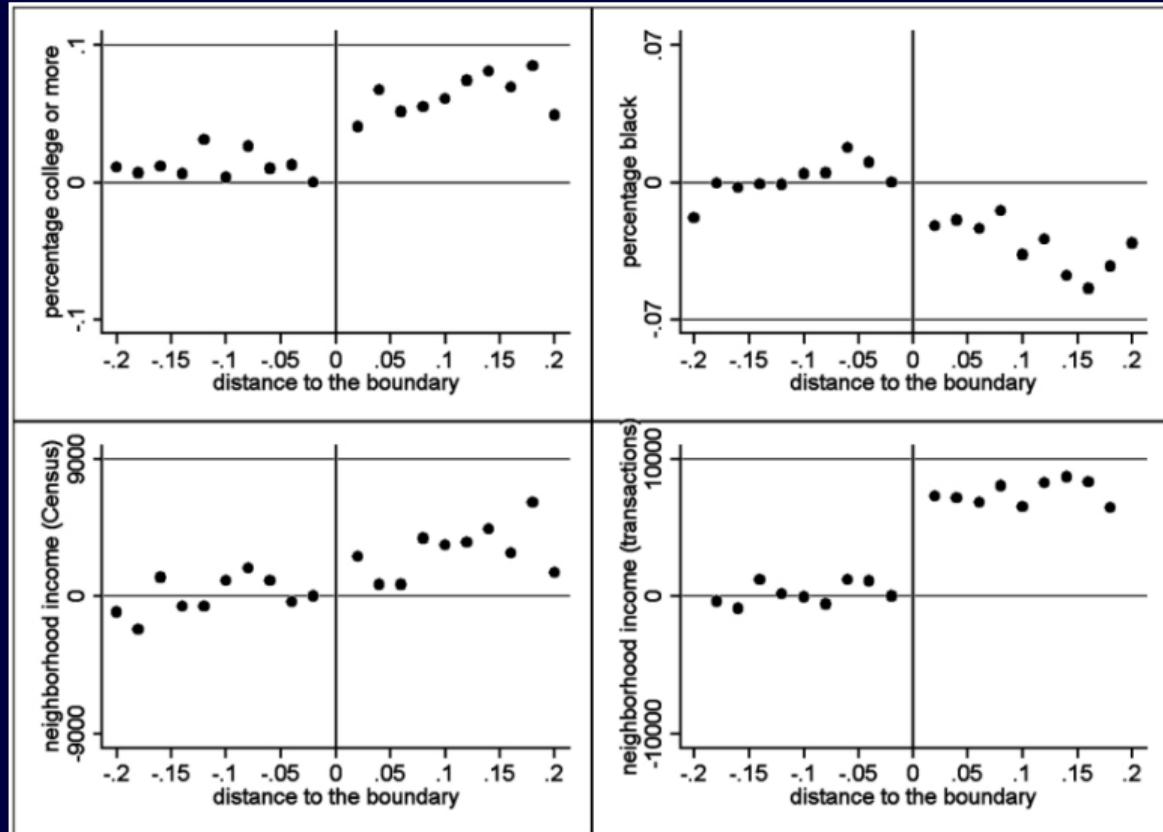


FIG. 4.—Neighborhood sociodemographics around the boundary. Each panel is constructed by Ito (IDE, Sacred Heart)

境界を挟んだ家計同士の比較で親の学歴、所得、人種を共変数に加えて統御、家の使用者費用/月に与える差の効果を推計

# 境界を挟んだ家計同士の比較で親の学歴、所得、人種を共変数に加えて統御、家の使用者費用/月に与える差の効果を推計

TABLE 3

KEY COEFFICIENTS FROM BASELINE HEDONIC PRICE REGRESSIONS

Boundary fixed effects included	SAMPLE			
	Within 0.20 Mile of Boundary (N = 27,548)		Within 0.10 Mile of Boundary (N = 15,122)	
	No	Yes	No	Yes
A. Excluding Neighborhood Sociodemographic Characteristics				
	(1)	(2)	(5)	(6)
Average test score (in standard deviations)	123.7 (13.2)	33.1 (7.6)	126.5 (12.4)	26.1 (6.6)
R <sup>2</sup>	.54	.62	.54	.62
B. Including Neighborhood Sociodemographic Characteristics				
	(3)	(4)	(7)	(8)
Average test score (in standard deviations)	34.8 (8.1)	17.3 (5.9)	44.1 (8.5)	14.6 (6.3)
% census block group black	-99.8 (33.4)	1.5 (38.9)	-123.1 (32.5)	4.3 (39.1)
% block group with college degree or more	220.1 (39.9)	89.9 (32.3)	204.4 (40.8)	80.8 (39.7)
Average block group income (/10,000)	60.0 (4.0)	45.0 (4.6)	55.6 (4.3)	42.9 (6.1)
R <sup>2</sup>	.59	.64	.59	.63

NOTE.—All regressions shown in the table also include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, %

境界を挟んだ家計同士の比較で親の学歴、所得、人種を共変数に加えて統御、家の使用者費用/月に与える差の効果を推計

各境界線の平均値  
(boundary fixed effects) を差し引くと試験点数の効果が 1/4 に減少 [A(2)]

TABLE 3 KEY COEFFICIENTS FROM BASELINE HEDONIC PRICE REGRESSIONS				
Boundary fixed effects included	SAMPLE			
	Within 0.20 Mile of Boundary (N = 27,548)		Within 0.10 Mile of Boundary (N = 15,122)	
	No	Yes	No	Yes
A. Excluding Neighborhood Sociodemographic Characteristics				
	(1)	(2)	(5)	(6)
Average test score (in standard deviations)	123.7 (13.2)	33.1 (7.6)	126.5 (12.4)	26.1 (6.6)
R <sup>2</sup>	.54	.62	.54	.62
B. Including Neighborhood Sociodemographic Characteristics				
	(3)	(4)	(7)	(8)
Average test score (in standard deviations)	34.8 (8.1)	17.3 (5.9)	44.1 (8.5)	14.6 (6.3)
% census block group black	-99.8 (33.4)	1.5 (38.9)	-123.1 (32.5)	4.3 (39.1)
% block group with college degree or more	220.1 (39.9)	89.9 (32.3)	204.4 (40.8)	80.8 (39.7)
Average block group income (/10,000)	60.0 (4.0)	45.0 (4.6)	55.6 (4.3)	42.9 (6.1)
R <sup>2</sup>	.59	.64	.59	.63
NOTE.—All regressions shown in the table also include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, %				

境界を挟んだ家計同士の比較で親の学歴、所得、人種を共変数に加えて統御、家の使用者費用/月に与える差の効果を推計

各境界線の平均値  
(boundary fixed effects) を差し引くと試験点数の効果が  $1/4$  に減少 [A(2)]

観察可能な境界線地域の特徴を共変数に加えるとさらに  $1/2$  に減少 [B(4)]

TABLE 3 KEY COEFFICIENTS FROM BASELINE HEDONIC PRICE REGRESSIONS				
Boundary fixed effects included	SAMPLE			
	Within 0.20 Mile of Boundary (N = 27,548)		Within 0.10 Mile of Boundary (N = 15,122)	
	No	Yes	No	Yes
A. Excluding Neighborhood Sociodemographic Characteristics				
	(1)	(2)	(5)	(6)
Average test score (in standard deviations)	123.7 (13.2)	33.1 (7.6)	126.5 (12.4)	26.1 (6.6)
R <sup>2</sup>	.54	.62	.54	.62
B. Including Neighborhood Sociodemographic Characteristics				
	(3)	(4)	(7)	(8)
Average test score (in standard deviations)	34.8 (8.1)	17.3 (5.9)	44.1 (8.5)	14.6 (6.3)
% census block group black	-99.8 (33.4)	1.5 (38.9)	-123.1 (32.5)	4.3 (39.1)
% block group with college degree or more	220.1 (39.9)	89.9 (32.3)	204.4 (40.8)	80.8 (39.7)
Average block group income (/10,000)	60.0 (4.0)	45.0 (4.6)	55.6 (4.3)	42.9 (6.1)
R <sup>2</sup>	.59	.64	.59	.63

NOTE.—All regressions shown in the table also include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, %

境界を挟んだ家計同士の比較で親の学歴、所得、人種を共変数に加えて統御、家の使用者費用/月に与える差の効果を推計

各境界線の平均値  
(boundary fixed effects) を差し引くと試験点数の効果が  $1/4$  に減少 [A(2)]

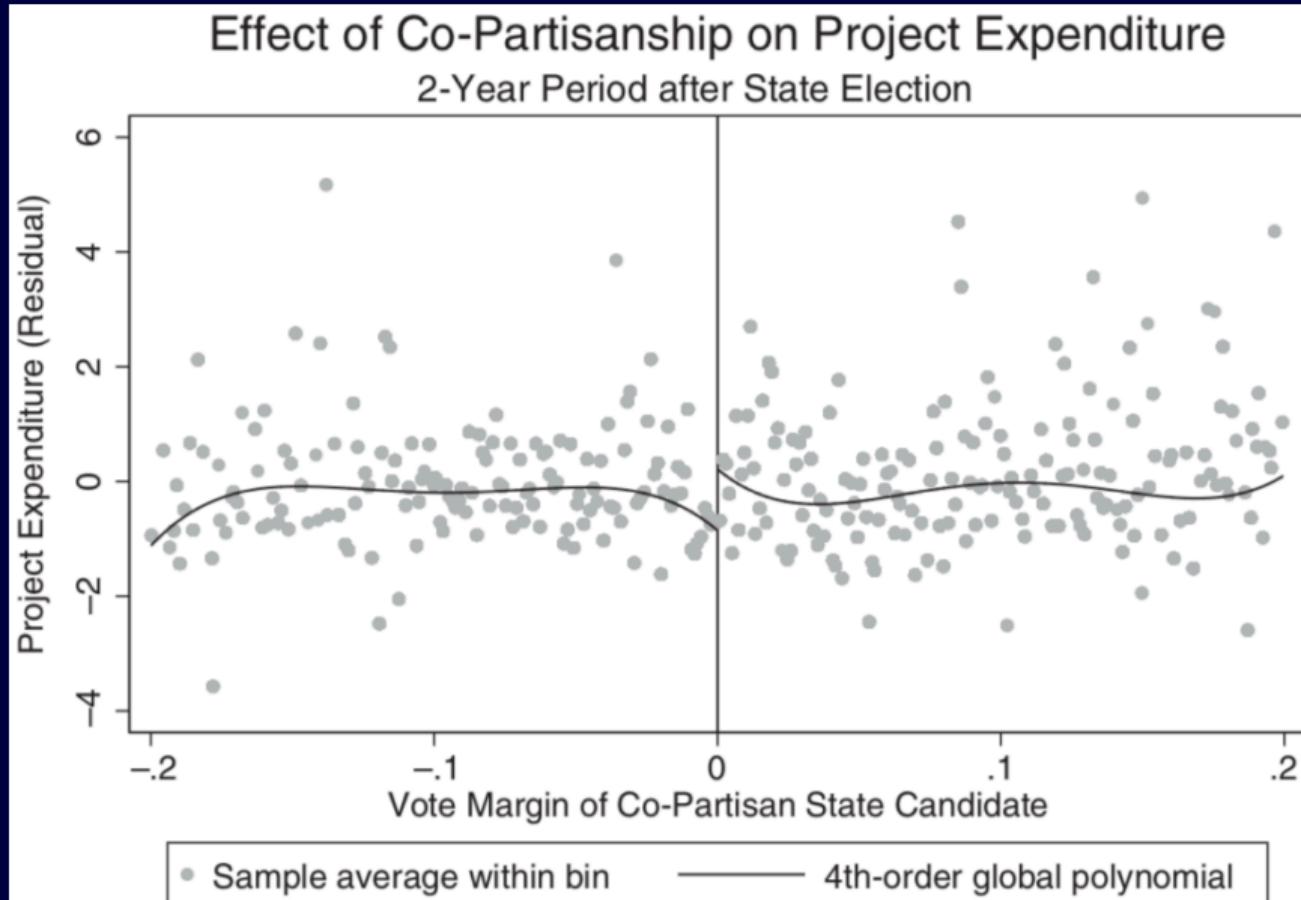
観察可能な境界線地域の特徴を共変数に加えるとさらに  $1/2$  に減少 [B(4)]

学歴や所得などを統御すると、黒人居住者比率の影響はほぼゼロ [D(4)]

TABLE 3 KEY COEFFICIENTS FROM BASELINE HEDONIC PRICE REGRESSIONS				
Boundary fixed effects included	SAMPLE			
	Within 0.20 Mile of Boundary (N = 27,548)		Within 0.10 Mile of Boundary (N = 15,122)	
	No	Yes	No	Yes
A. Excluding Neighborhood Sociodemographic Characteristics				
	(1)	(2)	(5)	(6)
Average test score (in standard deviations)	123.7 (13.2)	33.1 (7.6)	126.5 (12.4)	26.1 (6.6)
R <sup>2</sup>	.54	.62	.54	.62
B. Including Neighborhood Sociodemographic Characteristics				
	(3)	(4)	(7)	(8)
Average test score (in standard deviations)	34.8 (8.1)	17.3 (5.9)	44.1 (8.5)	14.6 (6.3)
% census block group black	-99.8 (33.4)	1.5 (38.9)	-123.1 (32.5)	4.3 (39.1)
% block group with college degree or more	220.1 (39.9)	89.9 (32.3)	204.4 (40.8)	80.8 (39.7)
Average block group income (/10,000)	60.0 (4.0)	45.0 (4.6)	55.6 (4.3)	42.9 (6.1)
R <sup>2</sup>	.59	.64	.59	.63

NOTE.—All regressions shown in the table also include controls for whether the house is owner-occupied, the number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and land use (% industrial, %

? インドで与党議員得票率と予算配分



統計学者 Andrew Gelman (コロンビア大学、ブログ魔):

*A global fourth-degree polynomial, huh? This is almost a parody of how to make spurious discoveries.* (4次関数? 見せかけの発見のパロディ [つまり、意図的な冗談] みたいだ。)

統計学者 Andrew Gelman (コロンビア大学、ブログ魔):

*A global fourth-degree polynomial, huh? This is almost a parody of how to make spurious discoveries. (4次関数? 見せかけの発見のパロディ[つまり、意図的な冗談]みたいだ。)*

*Those of us who develop advanced statistical methods have to be aware that, once a method is out there, it can be used “off-label” by anybody, and lots of those uses will be mistaken in some way or another. No, my problem is the false sense of certainty that appears to be engendered by the use of high-tech statistics. (推計方法が広まると「目的外利用」をする人が必ず出る。にもかかわらず、ハイテク統計学を利用したから推計結果は正しい、と誤った確信を深めていそうなのが問題だ。)*

統計学者 Andrew Gelman (コロンビア大学、ブログ魔):

*A global fourth-degree polynomial, huh? This is almost a parody of how to make spurious discoveries. (4次関数? 見せかけの発見のパロディ[つまり、意図的な冗談]みたいだ。)*

*Those of us who develop advanced statistical methods have to be aware that, once a method is out there, it can be used “off-label” by anybody, and lots of those uses will be mistaken in some way or another. No, my problem is the false sense of certainty that appears to be engendered by the use of high-tech statistics. (推計方法が広まると「目的外利用」をする人が必ず出る。にもかかわらず、ハイテク統計学を利用したから推計結果は正しい、と誤った確信を深めていそうなのが問題だ。)*

現在は曲線を伴う関数で回帰することは厳禁とされ、分断近傍で直線を使うことが多い

▶ RDD 推計の図解 (別ファイル)

Sacred Heart, IDE

?: メキシコのオアハカ州で民選か世襲の伝統的指導者による地方自治が混在したときの比較 (RDD)

?: メキシコのオアハカ州で民選か世襲の伝統的指導者による地方自治が混在したときの比較 (RDD)

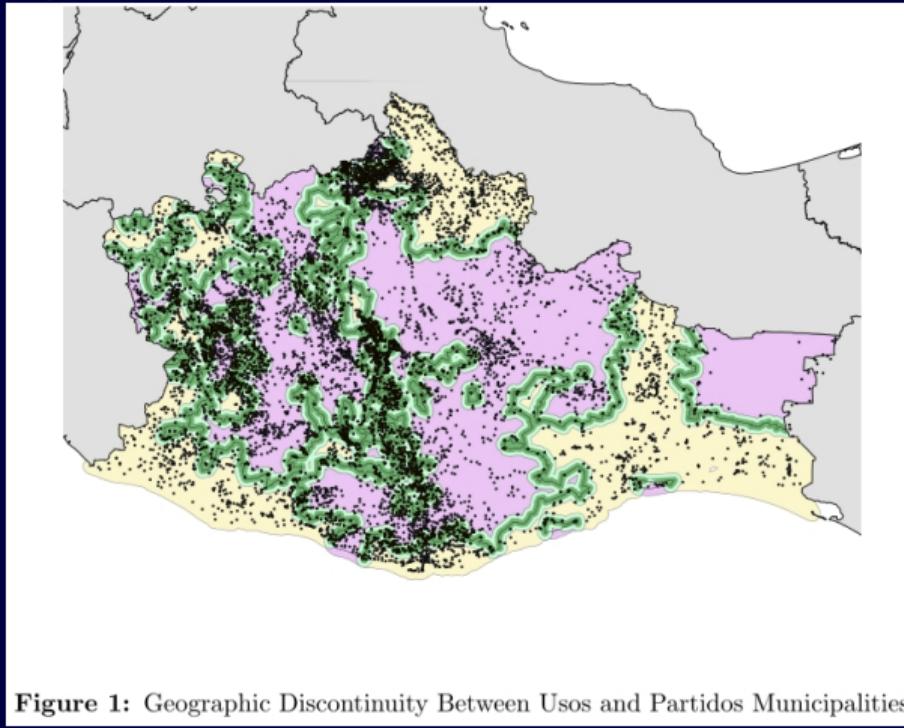
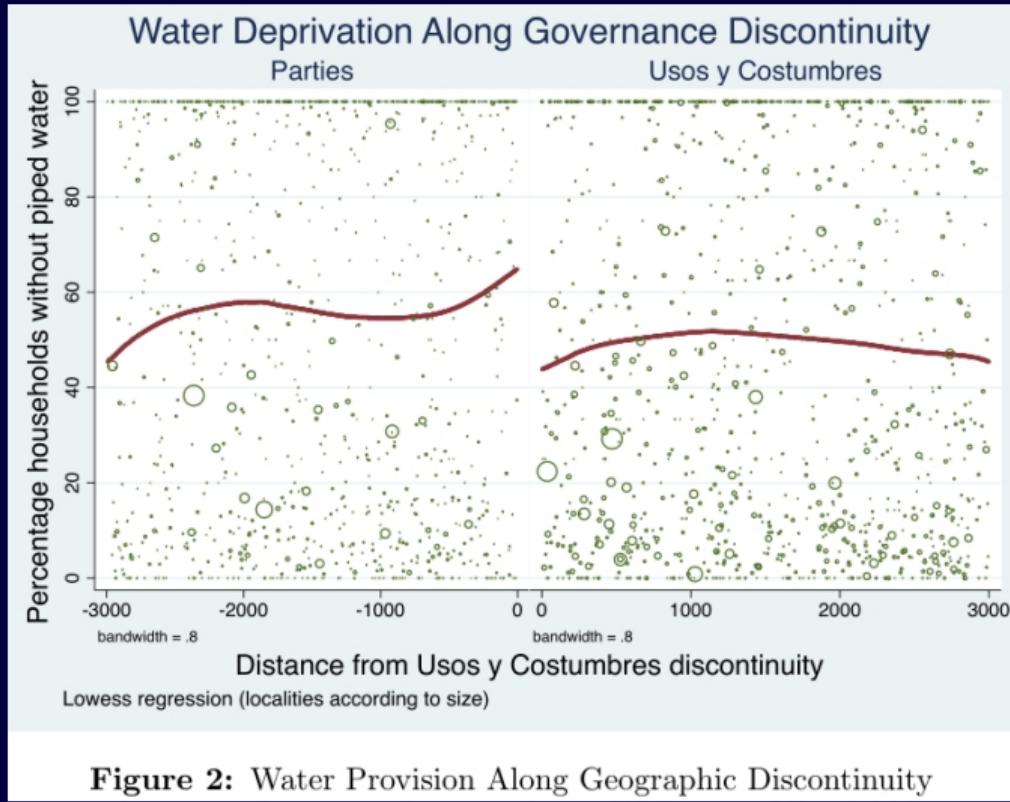
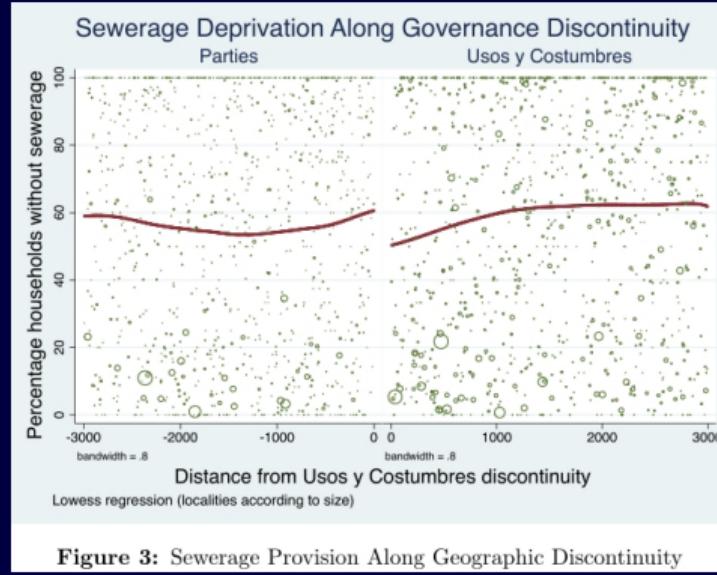


Figure 1: Geographic Discontinuity Between Usos and Partidos Municipalities

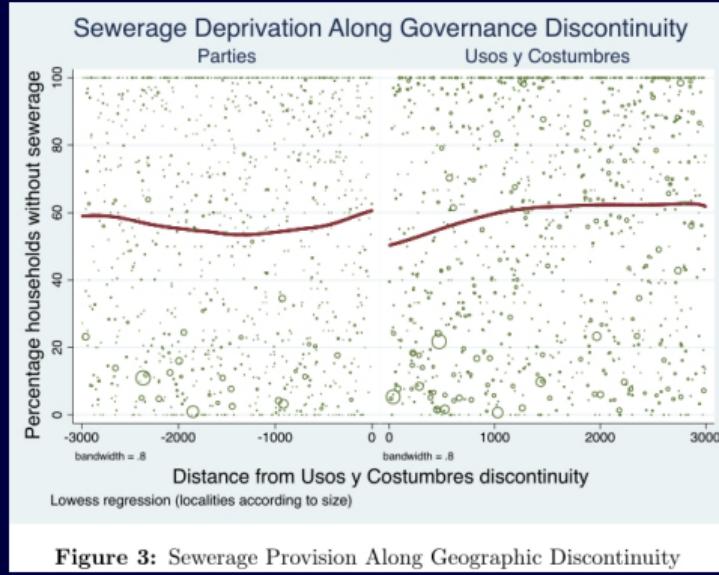
# 上水道未整備率、民選首長 vs. 伝統的指導者



# 下水道未整備率、民選首長 vs. 伝統的指導者



# 下水道未整備率、民選首長 vs. 伝統的指導者



何が何だか分からぬ図

# 下水道未整備率、民選首長 vs. 伝統的指導者

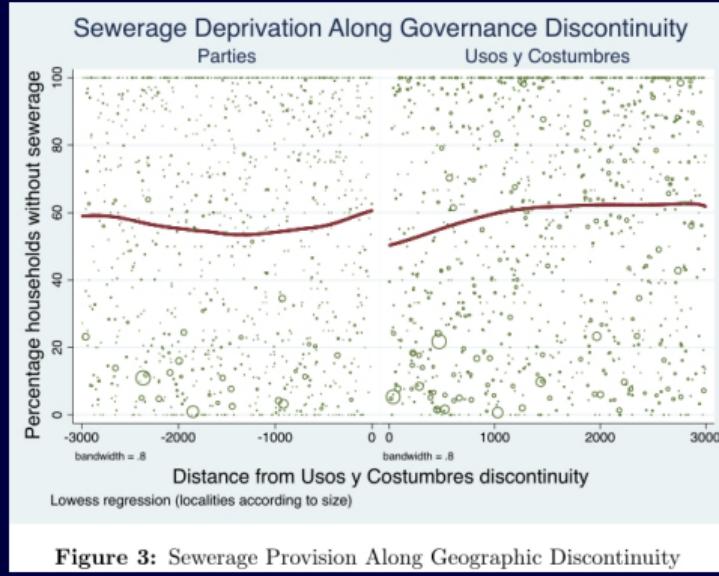


Figure 3: Sewerage Provision Along Geographic Discontinuity

何が何だか分からぬ図

「境界からの距離」だけでは多様な地域同士を比較、政策効果か異質なグループ比較によるノイズか分からぬ

# 下水道未整備率、民選首長 vs. 伝統的指導者

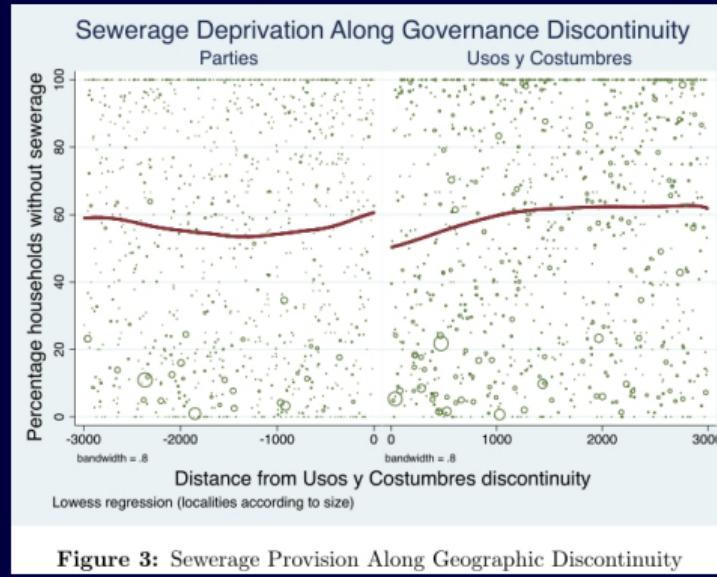


Figure 3: Sewerage Provision Along Geographic Discontinuity

何が何だか分からぬ図

「境界からの距離」だけでは多様な地域同士を比較、政策効果か異質なグループ比較によるノイズか分からぬ

境界のどの部分か限定する必要: 地形、水源、人口規模、上水道行政単位などが違うと工事の難易度(ノイズ)が入り込む

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid

境界(緑の帯):

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid CC BY SA

境界(緑の帯):

- 沿岸部平地や河川湖沼の有無を含んで多様

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid

境界(緑の帯):

- ☞ 沿岸部平地や河川湖沼の有無を含んで多様
- ☞ 地形による境界(河川湖沼、山脈)と一致している部分も多くある

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid

境界(緑の帯):

- ☛ 沿岸部平地や河川湖沼の有無を含んで多様
- ☛ 地形による境界(河川湖沼、山脈)と一致している部分も多くある
- ☛ 自治形態ではなく地形が原因で上水道普及率がジャンプしている可能性あり

# オアハカ(メキシコ)の地図との比較



出所: 左?, Figure 1, 右 Mario Fuente Cid

境界(緑の帯):

- ☛ 沿岸部平地や河川湖沼の有無を含んで多様
- ☛ 地形による境界(河川湖沼、山脈)と一致している部分も多くある
- ☛ 自治形態ではなく地形が原因で上水道普及率がジャンプしている可能性あり

use your common sense と Gelman に怒られそう...

## Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

# G o T o 見直し 募る懸念



## 飲食店店長「危機感覚える」

漏らした。

11

3連休中日に東京・大仏公園で開かれたいベ

A large outdoor event in Japan featuring Spanish culture, with many people in traditional dress and flags.

キャンペーンの実験結果を疑問視する意見があつた。緊急事態宣言の「自粛生活」が頭に「立喰いそぼろ」と書かれた札幌市の繁華街で、すきのは売り上げが減少の一途を辿る。古町や大通商店街などでも盛り場が死んでしまった。一方で、居酒屋や飲食店では、客の半分以上が「[GO TO]利用者(飲食店関係者)」だ。大衆的感覚が、度が引き上げられ、営業の要請も出てきた。

新規本部長の西田哲也によると、政府が「GO TO キャンペーン」の見直し（48）は、「観光客が動く中、新型コロナウイルスの感染が急拡大し、不安の声がある」として、確認。感染対策を悪化を懸念する声たい」と話した。

文部省

Ito (IDE, Sacred Heart)

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊



## Go To 見直し募る懸念

政府が「Go To キャンペーン」の見直しに動く中、新型コロナウイルスの感染が急拡大している地域では22日、経済の悪化を懸念する声や、キャンペーン自体の効果を疑問視する意見が上がった。緊急事態宣言下の「自粛生活」が頭によぎる人もいた。

札幌市の繁華街・すきの。3連休中のこの日は雨も重なり、人通りはまばらだった。これほどに盛ったいくら丼が「Go To」利用者。

「客の半分が『Go To』利用者

統括本部長の西田等さん

統括本部長の西田等さん

屋「酒蔵館」を営む吉田貴志さん(47)は「感染者が急増した今月は、客

が普段の1割程度、閉店も考えない」といれない」と肩を落とす。地域共通クーポンや「イート」の対象店舗だが、利用はまばらで制度を見直して、苦しい店を直接支援してほしい」と訴える。

感染状況に関する警戒度を最高レベルにしてい

る東京都。代々木公園では、スペイン料理店が出店する大規模イベントが開かれていた。出店数を

昨年の半分以下に減らし、会場の入場は最大500人に制限。

実行委員会は「取り扱う対策を

全て講じた」と強調する。

今回の最後に「Go To」は、急速な感染拡大に参加

者の中には複雑だ。大田区の無職、奥田和彦さん

(48)は「屋外だったが、

参加は少しめたが、

くなつて感染が拡大する

と、どこにも行けなくな

るものはない。今後は

どうすればいいのか」と漏らした。

## 飲食店店長「危機感覚える」

3連休中に東京・代々木公園で開かれたイベ

ントで、スペイン料理を楽しむ人たち(22日)

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

## Go To 見直し募る懸念



政府が「Go To キャンペーン」の見直しに動く中、新型コロナウイルスの感染が急拡大している地域では22日、経済の悪化を懸念する声や、キャンペーン自体の効果を疑問視する意見が上がった。緊急事態宣言下の「自粛生活」が頭によぎる人もいた。

札幌市の繁華街・すきの。3連休中のこの日は雨も重なり、人通りはまばらだった。これほどに盛ったいくら丼が「Go To」利用者が「毎来はちきょう」は、「客の半分」と店舗関係者が「飲食店店長」「危機感覚える」

「客の半分が『Go To』利用者」

「『Go To』利用者」=Go To Eat の効果でしょうか?

統括本部長の西田等さんは、「酒蔵館」を営む吉田貴志さん(47)は、「感染者が急増した今月は、客が普段の1割程度、閉店も考えない」といれない」と肩を落とす。地域共通クーポンや「イート」の対象店舗だが、利用はまばらで制度を見直して、苦しい店を直接支援してほしい」と訴える。

感染状況に関する吉田貴志さんは、「危機感を覚える」。名古屋市の繁華街・栄周辺は家族連れなどでぎわつたものの、「例年に比べれば人が少ない(飲食店関係者)」。大衆酒場は「取り扱う対策を全て講じた」と強調する。今回を最後にイベントは急遽な感染拡大に参加者の思いは複雑だ。大田区の無職、奥田和彦さんは、「やはりいいのか」と漏らした。

（左）3連休中日に東京・代々木公園で開かれた出版社大塚太さん（48）も「寒くなつて感染が拡大する」と、どこにも行けなくななるかもしれない。今は

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

## G o T o 見直し 募る懸念



# 飲食店店長「危機感覚える」

「いいやつはいいのか」と漏らした。

「客の半分が『Go To』利用者」

「『Go To』利用者」=Go To Eat の効果でしょうか?

GoToなしでも、もともと行くつもりだったかも。

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

## G o T o 見直し 募る懸念



## 飲食店店長「危機感覚える」

「いいやつはいいのか」と漏らした。

Ito (IDE, Sacred Heart)

2022 年秋

123 / 1

「客の半分が『Go To』利用者」

「『Go To』利用者」=Go To Eat の効果でしょうか?

GoToなしでも、もともと行くつもりだったかも。

## Go To Eat の効果を測定できる研究 デザインとは?

A large crowd of people gathered at a food festival booth for 'Fiesta de Espana 2020'. The booth features a red banner with the text 'FIESTA DE ESPAÑA 2020'. Many people are seen eating and interacting with the staff at the booth.

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

## Go To 見直し募る懸念



### 飲食店店長「危機感覚える」

3連休中白天に東京・代々木公園で開かれたイベ  
ントで、スペイン料理を楽しむ人たち(22日)

漏らした。

政府が「Go To キャンペーン」の見直し(48)は、櫻花館を営む吉田に動く中、新型コロナウイルスの感染が急拡大し、不安の声がある。半径22kmの地域では22日、経済の悪化を懸念する声や、キャンペーン自体の効果を疑問視する意見が上がった。緊急事態宣言下の「自粛生活」が頭によぎる人もいた。札幌市の繁華街・すすきの。3連休中のこの日は雨も重なり、人通りはまばらだった。これほどに盛ったくら人が多いと酒場は「毎来ばかり」といふ。客の半分が「Go To」利用者。

「客の半分が『Go To』利用者」

「『Go To』利用者」=Go To Eat の効果でしょうか?

Go Toなしでも、もともと行くつもりだったかも。

Go To Eat の効果を測定できる研究デザインとは?  
ランダム化比較試験、  
SCM(synthetic control method)など

# Go To Eat の効果?

日経新聞 2020年11月23日(月)朝刊

## Go To 見直し募る懸念



飲食店店長「危機感覚える」

3連休中白天に東京・代々木公園で開かれたイベ  
ントで、スペイン料理を楽しむ人たち(22日)

「Go To」見直し募る懸念

飲食店店長「危機感覚える」

貴志さん(47)は「感染者が急増した今月は、客  
が普段の1割程度、閉店も考えない」とい  
ふ。3連休中のこの日は雨も重なり、人通り  
はまばらだった。これほどに盛ったくら丼  
の「Go To」利用者。

札幌市の繁華街・すすきの。3連休中のこの  
日は雨も重なり、人通りは「危機感覚える」。  
名古屋市の繁華街・栄周辺は家族連れなどで  
ぎわったものの「例年に比べれば人が少ない(飲  
食店関係者)」。大衆居酒屋の司酒選「毎来は  
ちきょう」は、「客の半分が『Go To』利用者」  
と見直しを最後に、4歳の皇子塚雷太さん(48)  
も「寒くなつて感染が拡大する」と言つたが、  
参加は止められた。今後は「冬本音の寒さで、  
とにかく参加は止められたが、冬本音の寒さで、  
どうでも行けなくなつたかもしれない。今後は  
どうすればいいのか」と漏らした。

統括本部長の西田等さんは、「酒蔵館」を営む吉田屋「酒蔵館」を営む吉田屋  
の営業を要請も出している。立喰いそば・うどんす  
すきの」は売り上げが例年よりも落としている。店舗は、制度を見直して、苦しい店を直接支援して  
ほしい」と訴える。感染状況に関する警戒度を最高レベルにしている。出店数を500人に制限。実行委員会は「取り扱う対策を全て講じた」と強調する。昨年の半分以下に減らし、会場の入場は最大500人。

「客の半分が『Go To』利用者」

「『Go To』利用者」 = Go To Eat の効果でしょうか?

Go To なしでも、もともと行くつもりだったかも。

Go To Eat の効果を測定できる研究  
デザインとは?  
ランダム化比較試験、  
SCM(synthetic control method)など

どんな方法かは後で説明します

Sacred Heart, IDE

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

- 科学的な方法: 反証可能な命題が成立するか検定する方法

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

- 科学的な方法: 反証可能な命題が成立するか検定する方法

エビデンスがないのは事実

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

☞ 科学的な方法: 反証可能な命題が成立するか検定する方法

エビデンスがないのは事実

☞ 政府は効果検証用にデータを収集していないため

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

- 科学的な方法: 反証可能な命題が成立するか検定する方法

エビデンスがないのは事実

- 政府は効果検証用にデータを収集していないため
- 検証するつもりなし、と言っているに等しい

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

FNN プライムニュース 2020年11月25日(水)



エビデンス=科学的な方法で明らかになった事象

- 科学的な方法: 反証可能な命題が成立するか検定する方法

エビデンスがないのは事実

- 政府は効果検証用にデータを収集していないため
- 検証するつもりなし、と言っているに等しい

利用可能なデータを前提にすると、どのような効果検証が可能か?

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

[http://www.kantei.go.jp/jp/99\\_suga/statement/2022/1214kaiken02.html](http://www.kantei.go.jp/jp/99_suga/statement/2022/1214kaiken02.html)

記者「GoTo トラベルに感染拡大のエビデンスがないとの認識はあったか?」  
(NHK サイトでは「GoTo トラベルに感染拡大のエビデンスはないという認識は変わったのか」)

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

[http://www.kantei.go.jp/jp/99\\_suga/statement/2022/1214kaiken02.html](http://www.kantei.go.jp/jp/99_suga/statement/2022/1214kaiken02.html)

記者「GoTo トラベルに感染拡大のエビデンスがないとの認識はあったか?」  
(NHK サイトでは「GoTo トラベルに感染拡大のエビデンスはないという認識は変わったのか」)

菅首相

「そこは、医師会の会長が申し上げているのではないでしょうか。それと、当時は移動によっては、感染を拡大しないということ、ここも提言もあります。そこについては変わりません。ただ、今回そうしたことの専門家の委員の先生方からそういう指摘をいただきましたので、この3000人、現実的に患者の方が出ていますから、年末年始、集中的に対応できる、そういうチャンスだと、そういう思いの中で私は判断しました。」

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

[http://www.kantei.go.jp/jp/99\\_suga/statement/2022/1214kaiken02.html](http://www.kantei.go.jp/jp/99_suga/statement/2022/1214kaiken02.html)

記者「GoTo トラベルに感染拡大のエビデンスがないとの認識はあったか?」  
(NHK サイトでは「GoTo トラベルに感染拡大のエビデンスはないという認識は変わったのか」)

菅首相

「そこは、医師会の会長が申し上げているのではないでしょうか。それと、当時は移動によっては、感染を拡大しないということ、ここも提言もあります。そこについては変わりません。ただ、今回そうしたことの専門家の委員の先生方からそういう指摘をいただきましたので、この3000人、現実的に患者の方が出ていますから、年末年始、集中的に対応できる、そういうチャンスだと、そういう思いの中で私は判断しました。」

エビデンスはなく(と会長が言っている)、移動は感染を広げないと(誰から?)聞いていたが、感染者数が増えたから停止、という主張

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

2020年11月18日(水)日本医師会会長中川氏

「GoTo トラベル自体から感染者が急増したというエビデンス（根拠）はなかなかはっきりしないが、きっかけになったことは間違いないと私は思っている。感染者が増えたタイミングを考えると関与は十分しているだろう」

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

2020年11月18日(水)日本医師会会長中川氏

「GoTo トラベル自体から感染者が急増したというエビデンス（根拠）はなかなかはっきりしないが、きっかけになったことは間違いないと私は思っている。感染者が増えたタイミングを考えると関与は十分しているだろう」

☞ 原因ではないがきっかけ、というのは意味不明

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

2020年11月18日(水)日本医師会会長中川氏

「GoTo トラベル自体から感染者が急増したというエビデンス（根拠）はなかなかはっきりしないが、きっかけになったことは間違いないと私は思っている。感染者が増えたタイミングを考えると関与は十分しているだろう」

⇒ 原因ではないがきっかけ、というのは意味不明

2020年12月16日(水)衆院内閣閉院中審査: 新型コロナウィルス感染症対策分科会尾身会長

「50歳以下の人人が移動して二次感染を起こしていることがはっきりしきたので、人の動きを止めることが重要で、その一環のなかで GoTo トラベルもある。」「本質は意図せず重症化が出るので、そのような文脈のなかで GoTo トラベルも考えるべきと再三申し上げている。」

⇒ 「移動は感染を拡大させない」という提言はない。

⇒ 50歳以下が感染伝播の高リスク・グループと判明、この移動を制限すべき、という提言。GoTo トラベルによる感染拡大のエビデンスではないが、論理的帰結によって制限を結論。

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

因果関係を検定できるデータを集めることが大事

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

因果関係を検定できるデータを集めることが大事  
ただし、政策担当者の協力なしには効果推計に限界がある

# GoTo トラベルの効果?

2020年12月14日(月)GoTo トラベル全国一斉一時停止

因果関係を検定できるデータを集めることが大事  
ただし、政策担当者の協力なしには効果推計に限界がある

- ▣ GoTo トラベル利用者には COCOA 導入と登録を条件付けてもよかったですかもしれない
- ▣ GoTo トラベル利用中に感染や濃厚接触をするリスクを計算できる
- ▣ GoTo トラベル期間ごとに、たとえば、11月第1週に予算を幾ら支出したら、利用者がどれだけいて、こうしたリスクがこれだけ高まった、と計算できる
  - △ それでも、補助金に関係なく行く人たちもいるので、GoToTravel の効果だけではない、過大推計
  - △ 補助金に関係なく行く人たちの分を差し引く必要あり

# GoTo トラベルの効果?

## 効果推計の例

- 旅客人数: 前年度同月値との比較
- 旅行支出: 推計した需要関数が正しいと前提に補助金の効果を試算
- (有症率比較: GoTo トラベル利用者 vs. 非利用者)
- ビッグ・データによる人の移動:  
GoTo トラベル開始前後の変化-昨年同時期の変化
- グレンジャー因果性: 航空旅行客と感染者数の関係
- 合成統御法: GTT 東京追加と東京圏感染者数の関係

# GoTo トラベルの効果?

よくある比較

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

昨年度と比較=GoTo トラベルなしだと昨年と同じという(暗黙の?)仮定をしている

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

昨年度と比較=GoTo トラベルなしだと昨年と同じという(暗黙の?)仮定をしている

被説明変数 旅客人数

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

昨年度と比較=GoTo トラベルなしだと昨年と同じという(暗黙の?)仮定をしている

被説明変数 旅客人数

データ 国土交通省データ

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

昨年度と比較=GoTo トラベルなしだと昨年と同じという(暗黙の?)仮定をしている

被説明変数 旅客人数

データ 国土交通省データ

識別仮定 **identification assumption** 「GoTo トラベルがなければ、今年も昨年と同じ人の移動だった」

# GoTo トラベルの効果?

よくある比較

GoTo トラベル実施月の旅客人数を前年度同月値と比較して「x%多かった」と示す

昨年度と比較=GoTo トラベルなしだと昨年と同じという(暗黙の?)仮定をしている

被説明変数 旅客人数

データ 国土交通省データ

識別仮定 **identification assumption** 「GoTo トラベルがなければ、今年も昨年と同じ人の移動だった」

☞ このデザインでの効果推計値の信頼性=この識別仮定の現実妥当性

# GoTo トラベルの効果?

# GoTo トラベルの効果?

- 課題 ① GoTo トラベルなしのとき、移動人数が当該年と前年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと GoTo トラベル実施年は景気後退でその前年より少なかったはず。(分母=なしのときの水準が過大なので) 水準との比による評価は過小評価になる。

# GoTo トラベルの効果?

## 課題

- ① GoTo トラベルなしのとき、移動人数が当該年と前年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと GoTo トラベル実施年は景気後退でその前年より少なかったはず。(分母=なしのときの水準が過大なので) 水準との比による評価は過小評価になる。
- ② 2020 年 (GoTo トラベル実施時) に入国制限されたインバウンド客を除外して 2019 年のデータを作成できるか?

# GoTo トラベルの効果?

## 課題

- ① GoTo トラベルなしのとき、移動人数が当該年と前年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと GoTo トラベル実施年は景気後退でその前年より少なかったはず。(分母=なしのときの水準が過大なので) 水準との比による評価は過小評価になる。
- ② 2020 年 (GoTo トラベル実施時) に入国制限されたインバウンド客を除外して 2019 年のデータを作成できるか?

誰にでもできる分析なので、何と比較すべきかを考えずにやっている人が多いはず

# GoTo トラベルの効果?

## 課題

- ① GoTo トラベルなしのとき、移動人数が当該年と前年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと GoTo トラベル実施年は景気後退でその前年より少なかったはず。(分母=なしのときの水準が過大なので) 水準との比による評価は過小評価になる。
- ② 2020 年 (GoTo トラベル実施時) に入国制限されたインバウンド客を除外して 2019 年のデータを作成できるか?

誰にでもできる分析なので、何と比較すべきかを考えずにやっている人が多いはず

「今年はコロナウィルス流行によって大きく減っているという事情はあります  
が」などという数字以外の補正をして説明するはず

# GoTo トラベルの効果?

## 課題

- ① GoTo トラベルなしのとき、移動人数が当該年と前年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと GoTo トラベル実施年は景気後退でその前年より少なかったはず。(分母=なしのときの水準が過大なので) 水準との比による評価は過小評価になる。
- ② 2020 年 (GoTo トラベル実施時) に入国制限されたインバウンド客を除外して 2019 年のデータを作成できるか?

誰にでもできる分析なので、何と比較すべきかを考えずにやっている人が多いはず

「今年はコロナウィルス流行によって大きく減っているという事情はあります  
が」などという数字以外の補正をして説明するはず

効果があったのかはっきりしないし、聞き手の主観が入り込む

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」  
<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が 1 %低下すると実質サービス消費は 0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり 50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を 40%増加させる計算だ（支援部分も含む）。」

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が1%低下すると実質サービス消費は0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を40%増加させる計算だ（支援部分も含む）。」

$$\text{☞ } 50\% \times .8 = 40\%$$

▶ 需要の価格彈力性

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が 1 %低下すると実質サービス消費は 0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり 50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を 40%増加させる計算だ（支援部分も含む）。」

$$\text{☞ } 50\% \times .8 = 40\%$$

▶ 需要の価格彈力性

「観光庁によれば、日本人の国内旅行の関連消費額は、2017 年に 21.5 兆円と、個人消費全体の 7.1 %を占めている。これが 40%増加すれば、1 年間で消費を 8.7 兆円増加させる。」

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が1%低下すると実質サービス消費は0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を40%増加させる計算だ（支援部分も含む）。」

「観光庁によれば、日本人の国内旅行の関連消費額は、2017年に21.5兆円と、個人消費全体の7.1%を占めている。これが40%増加すれば、1年間で消費を8.7兆円増加させる。」

$$\Rightarrow 50\% \times .8 = 40\%$$

▶ 需要の価格彈力性

$$\Rightarrow 21.5 \text{兆円} \times .4 \\ = 8.60 \text{兆円}$$

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が 1 %低下すると実質サービス消費は 0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり 50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を 40%増加させる計算だ（支援部分も含む）。」

「観光庁によれば、日本人の国内旅行の関連消費額は、2017 年に 21.5 兆円と、個人消費全体の 7.1 %を占めている。これが 40%増加すれば、1 年間で消費を 8.7 兆円増加させる。」

「内閣府の県民経済計算（平成 28 年度）によると、東京都の都民所得は全国の 17.8%である。この比率分…(中略)...（感染への警戒から、東京着の旅行は現在かなり少ないとみられるため、ここでは東京発の旅行のみ考慮する）、今回の東京の除外によって、「Go To トラベル」の消費押し上げ効果は 1 年間で 1.54 兆円減少し、1 年間の GDP 成長率の押し上げ効果を 0.28%減らす計算となる。」

$$\Rightarrow 50\% \times .8 = 40\%$$

▶ 需要の価格彈力性

$$\Rightarrow 21.5 \text{ 兆円} \times .4 \\ = 8.60 \text{ 兆円}$$

# Go To トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

「内閣府の分析によると、サービス消費の価格弹性値は-0.8である。これは、価格が 1 %低下すると実質サービス消費は 0.8%増加する傾向にある、ということを意味している。」「ところで、『Go To トラベル』では旅行費用が半分になる、つまり 50%の値下げが実施されるに等しくなるため（上限を超える支出部分は考慮しない）、それは支援の対象となる旅行関連消費を 40%増加させる計算だ（支援部分も含む）。」

「観光庁によれば、日本人の国内旅行の関連消費額は、2017 年に 21.5 兆円と、個人消費全体の 7.1 %を占めている。これが 40%増加すれば、1 年間で消費を 8.7 兆円増加させる。」

「内閣府の県民経済計算（平成 28 年度）によると、東京都の都民所得は全国の 17.8%である。この比率分…(中略)...（感染への警戒から、東京着の旅行は現在かなり少ないとみられるため、ここでは東京発の旅行のみ考慮する）、今回の東京の除外によって、「Go To トラベル」の消費押し上げ効果は 1 年間で 1.54 兆円減少し、1 年間の GDP 成長率の押し上げ効果を 0.28%減らす計算となる。」

☞  $50\% \times .8 = 40\%$

▶ 需要の価格彈力性

☞  $21.5 \text{ 兆円} \times .4 = 8.60 \text{ 兆円}$

☞  $8.6 \text{ 兆円} \times .178 = 1.53 \text{ 兆円}$

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」  
<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

需要関数を使って影響を考えるのはよいことだが、2020 年以前のデータを使って推計された COVID 以前の需要関数を COVID 下の 2020 年の需要関数として利用

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

需要関数を使って影響を考えるのはよいことだが、2020 年以前のデータを使って推計された COVID 以前の需要関数を COVID 下の 2020 年の需要関数として利用

2020 年は旅行需要が価格と無関係に萎縮しているはず

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

需要関数を使って影響を考えるのはよいことだが、2020 年以前のデータを使って推計された COVID 以前の需要関数を COVID 下の 2020 年の需要関数として利用

2020 年は旅行需要が価格と無関係に萎縮しているはず

萎縮した 2020 年需要関数の価格弾力性は COVID 以前の価格弾力性と同じ保証はない

# Go To トラベルの効果?

需要関数を使った試算

野村総研木内氏「東京除外で減少する Go To トラベルの消費押し上げ効果は 1.5 兆円程度か」<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/0717>

需要関数を使って影響を考えるのはよいことだが、2020 年以前のデータを使って推計された COVID 以前の需要関数を COVID 下の 2020 年の需要関数として利用

2020 年は旅行需要が価格と無関係に萎縮しているはず

萎縮した 2020 年需要関数の価格弾力性は COVID 以前の価格弾力性と同じ保証はない

情報が無いながらもやってみる、というスタンスならば、その留意を明記すべき

# GoTo トラベルの効果?

需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

## ■ 感染への警戒がGOTO トラベルの利用を4分の1にとどめる

GOTO トラベル事業の経済効果を確認してみよう。国土交通省の発表によれば、7月22日から10月31日までに、GOTO トラベル事業のもとで宿泊・旅行代金が割引された総額は、少なくとも約1,886億円、10月1日から11月9日までの地域共通クーポンの付与額は少なくとも約201億円である。

☞ GoTo トラベル利用総額 1886 億円

# GoTo トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

### ■ 感染への警戒がGOTO トラベルの利用を4分の1にとどめる

GOTO トラベル事業の経済効果を確認してみよう。国土交通省の発表によれば、7月22日から10月31日までに、GOTO トラベル事業のもとで宿泊・旅行代金が割引された総額は、少なくとも約1,886億円、10月1日から11月9日までの地域共通クーポンの付与額は少なくとも約201億円である。

☞ GoTo トラベル利用総額 1886 億円

「宿泊・旅行代金の 35% に相当する割引額から、その間の GOTO トラベルを利用した全体の旅行支出額を計算すると、5,389 億円となる。必ずしも正確ではないが、(中略) さらに東京が除外されていなかった場合(中略)、推定で 6,150 億円、月間平均で 1,809 億円、年換算で 2 兆 1,708 億円となる。」

# GoTo トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

### ■ 感染への警戒がGOTO トラベルの利用を4分の1にとどめる

GOTO トラベル事業の経済効果を確認してみよう。国土交通省の発表によれば、7月22日から10月31日までに、GOTO トラベル事業のもとで宿泊・旅行代金が割引された総額は、少なくとも約1,886億円、10月1日から11月9日までの地域共通クーポンの付与額は少なくとも約201億円である。

☞ GoTo トラベル利用総額 1886 億円

「宿泊・旅行代金の 35% に相当する割引額から、その間の GOTO トラベルを利用した全体の旅行支出額を計算すると、5,389 億円となる。必ずしも正確ではないが、(中略) さらに東京が除外されていなかった場合(中略)、推定で 6,150 億円、月間平均で 1,809 億円、年換算で 2 兆 1,708 億円となる。」

☞ 1886 億円/.35  
=5389 億円

# GoTo トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

### ■ 感染への警戒がGOTO トラベルの利用を4分の1にとどめる

GOTO トラベル事業の経済効果を確認してみよう。国土交通省の発表によれば、7月22日から10月31日までに、GOTO トラベル事業のもとで宿泊・旅行代金が割引された総額は、少なくとも約1,886億円、10月1日から11月9日までの地域共通クーポンの付与額は少なくとも約201億円である。

☞ GoTo トラベル利用総額 1886 億円

「宿泊・旅行代金の 35% に相当する割引額から、その間の GOTO トラベルを利用した全体の旅行支出額を計算すると、5,389 億円となる。必ずしも正確ではないが、(中略) さらに東京が除外されていなかった場合(中略)、推定で 6,150 億円、月間平均で 1,809 億円、年換算で 2 兆 1,708 億円となる。」

☞  $1886 \text{ 億円} / .35 = 5389 \text{ 億円}$

☞ 「東京が除外されていなかった場合、推定で 6150 億円」をどうやって計算したのか不明。効果推計で最も大事な部分を解説していないのはまともではない。

# GoTo トラベルの効果?

需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

「ところで、GOTO トラベル事業は、国内旅行需要を1年間で8.7兆円増加させると試算される（コラム「東京除外で減少する Go To トラベルの消費押し上げ効果は1.5兆円程度か」、2020年7月17日）。この試算値に対して、実績値は4分1にとどまっている計算だ。その分、感染への警戒が、GOTO トラベルを利用した旅行需要を減少させている、と考えることができる。」

# GoTo トラベルの効果?

需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

「ところで、GOTO トラベル事業は、国内旅行需要を1年間で8.7兆円増加させると試算される（コラム「東京除外で減少する Go To トラベルの消費押し上げ効果は1.5兆円程度か」、2020年7月17日）。この試算値に対して、実績値は4分1にとどまっている計算だ。その分、感染への警戒が、GOTO トラベルを利用した旅行需要を減少させている、と考えることができる。」

「東京除外でない場合の GoTo トラベルによる增加試算額8.7兆円」は絶対値で1よりも小さい点彈力性を使っている

☞ 需要曲線の左側=直線により近い

# GoTo トラベルの効果?

需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

「ところで、GOTO トラベル事業は、国内旅行需要を1年間で8.7兆円増加させると試算される（コラム「東京除外で減少する Go To トラベルの消費押し上げ効果は1.5兆円程度か」、2020年7月17日）。この試算値に対して、実績値は4分1にとどまっている計算だ。その分、感染への警戒が、GOTO トラベルを利用した旅行需要を減少させている、と考えることができる。」

「東京除外でない場合の GoTo トラベルによる增加試算額8.7兆円」は絶対値で1よりも小さい点彈力性を使っている

- ☞ 需要曲線の左側=直線により近い  
⇒ 過大推計になりやすい

# GoTo トラベルの効果?

## 需要関数を使った試算

野村総研木内氏「GOTO トラベル見直しとその経済効果の試算」

<https://www.nri.com/jp/knowledge/blog/lst/2020/fis/kiuchi/1124>

「ところで、GOTO トラベル事業は、国内旅行需要を1年間で8.7兆円増加させると試算される（コラム「東京除外で減少する Go To トラベルの消費押し上げ効果は1.5兆円程度か」、2020年7月17日）。この試算値に対して、実績値は4分1にとどまっている計算だ。その分、感染への警戒が、GOTO トラベルを利用した旅行需要を減少させている、と考えることができる。」

「東京除外でない場合の GoTo トラベルによる增加試算額8.7兆円」は絶対値で1よりも小さい点彈力性を使っている

「除外されていなかった場合の試算額」 - 実績額=感染への警戒による減少額、は乱暴な議論。試算額が過大だっただけかも。

- ☞ 需要曲線の左側=直線により近い  
⇒ 過大推計になりやすい

# GoTo トラベルの効果?

需要関数を使った試算

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

しかし、COVID以前の推計結果を使って良いわけがない

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

しかし、COVID以前の推計結果を使って良いわけがない

さらに、価格弹性値の扱いが不適切

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

しかし、COVID以前の推計結果を使って良いわけがない

さらに、価格弾性値の扱いが不適切

- ☞ 内閣府推計値は点弾力性 point elasticity で非弾力的、過大推計になりやすい

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

しかし、COVID以前の推計結果を使って良いわけがない

さらに、価格弹性値の扱いが不適切

- ☞ 内閣府推計値は点弾力性 point elasticity で非弾力的、過大推計になりやすい
- ☞ 乗じる 50% は大きな価格変化  $\Delta p$  なので、推計された需要変化は一層過大

# GoTo トラベルの効果?

需要関数を使った試算

用いたデータと推計方法が適切ならば、理想的な効果試算方法

- ☞ 需要関数とは消費者の効用最大化から求められるので、経済理論の裏付けがある
- ☞ 原典は内閣府「日本経済 2018-2019」白書第3章付注2-3、推計方法 (QUAIDS 消費関数を FGNLS)、データは総務省「家計調査」
- ☞ データは理想的、推計方法も標準的

しかし、COVID以前の推計結果を使って良いわけがない

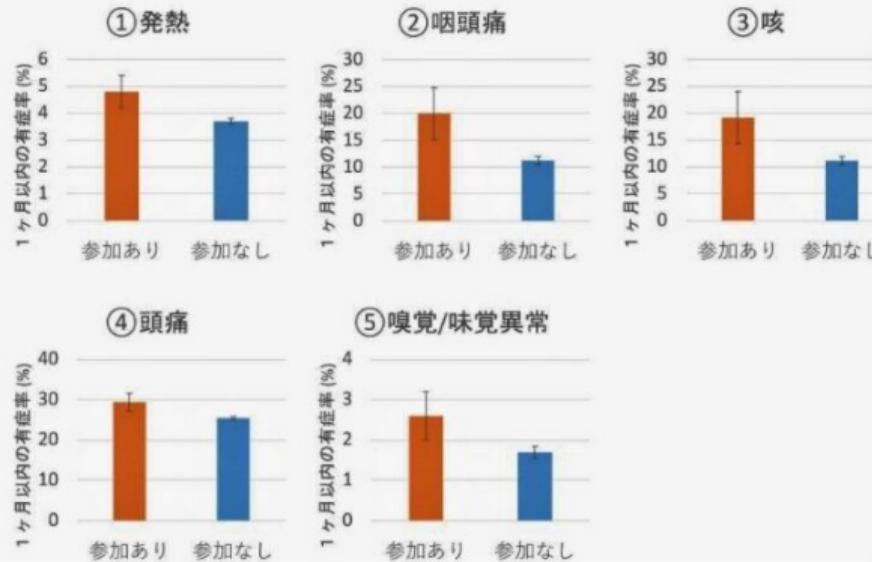
さらに、価格弹性値の扱いが不適切

- ☞ 内閣府推計値は点弾力性 point elasticity で非弾力的、過大推計になりやすい
- ☞ 乗じる 50% は大きな価格変化  $\Delta p$  なので、推計された需要変化は一層過大

経済理論を使って厳密そうな印象だが、印象で判断してはいけない

# Go To トラベルの効果?

図1 Go To トラベル利用の有無別の新型コロナを示唆する5つの症状の有症率



「Go To トラベル利用者感染リスク高い」  
東京大学など研究チーム  
2020年12月8日(火)  
論文: ?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

## 研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

## 研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

GoTo トラベル利用者は非利用者よりも有症率が2倍

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

GoTo トラベル利用者は非利用者よりも有症率が2倍

解釈: 以下の可能性がある

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

GoTo トラベル利用者は非利用者よりも有症率が2倍

解釈: 以下の可能性がある

- ① GoTo トラベルで罹患

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

GoTo トラベル利用者は非利用者よりも有症率が2倍

解釈: 以下の可能性がある

- ① GoTo トラベルで罹患
- ② 有症率(=罹患確率)の高い人が GoTo トラベル利用

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム  
研究デザイン

- インターネット調査(楽天インサイト社実施)、22万4389人中2万8000人回答、  
2万5482人有効回答
- 症状有無、GoTo トラベル利用有無、社会経済変数、既往症

GoTo トラベル利用者は非利用者よりも有症率が2倍

解釈: 以下の可能性がある

- ① GoTo トラベルで罹患
- ② 有症率(=罹患確率)の高い人が GoTo トラベル利用

結論: 「リスクの低い人に経済活動の誘因を与え、高い人は自宅待機を促すべき」

Sacred Heart, IDE

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

閣僚の反応 [https://news.tbs.co.jp/newseye/tbs\\_newseye4146048.htm?1607587275057](https://news.tbs.co.jp/newseye/tbs_newseye4146048.htm?1607587275057)

田村憲久厚労相 「ちょっとエビデンスといえるものなのかどうかどうなのか、ちょっと査読も終わっていないという話ですし。評価のしようがないというのが、正直なところでありますので」

赤羽一嘉国交相 「この論文についても、ちょっと正式に査読前という話もありましたし。現時点では全くコメントする段階でないと思っている」

加藤勝信官房長官 「著者自らもですね、研究方法の限界として、GoTo トラベルの利用が直接的に新型コロナ症状の増加につながったという因果関係は断定できないこと」

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし...

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

何でしょうか?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

何でしょうか?

空いたスペースで政治家の反応への疑問

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

何でしょうか?

空いたスペースで政治家の反応への疑問

☞ 査読通ったら意見、対応するのか?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

何でしょうか?

空いたスペースで政治家の反応への疑問

☞ 査読通ったら意見、対応するのか?

☞ 実験は存在しないので因果関係は示せない。では、どんな事実があれば対応するのか?エビデンスなんて出てこないことを知っているながらの、何もしない言い訳では?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示した解釈が面白いし、明確な政策提言になっている

結果の提示も慎重「因果関係を示している訳ではない」

しかし... この研究には方法論として弱点があると思います

何でしょうか?

空いたスペースで政治家の反応への疑問

- ☞ 査読通ったら意見、対応するのか?
- ☞ 実験は存在しないので因果関係は示せない。では、どんな事実があれば対応するのか?エビデンスなんて出てこないことを知っているながらの、何もしない言い訳では?
- ☞ 単にケチ付けているだけでは?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム  
示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

C の例: 外出量

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

C の例: 外出量

外出好きな人は罹患確率 [ $\propto$  有症率 (A)] が高い

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

C の例: 外出量

外出好きな人は罹患確率 [ $\propto$  有症率 (A)] が高い

外出好きな人は旅行をよくする (行くのだったら割引を使う)

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

C の例: 外出量

外出好きな人は罹患確率 [ $\propto$  有症率 (A)] が高い

外出好きな人は旅行をよくする (行くのだったら割引を使う)

外出好きな人は罹患しやすく、旅行も頻繁にする、というだけでは?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

示したこと: 有症率 (A) と GoTo トラベル利用 (B) の正の相関

検討していない可能性: みせかけの相関

その他の現象 (C) が (A) と (B) を同時に動かしているのでは?

C の例: 外出量

外出好きな人は罹患確率 [ $\propto$  有症率 (A)] が高い

外出好きな人は旅行をよくする (行くのだったら割引を使う)

外出好きな人は罹患しやすく、旅行も頻繁にする、というだけでは?

もしもそうだったら、結論は常識の範囲内

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

- ☞ 年齢: 年齢差別になりかねず、違憲の可能性あり

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

☞ 年齢: 年齢差別になりかねず、違憲の可能性あり

もしも可能だったら... ランダムに GoTo トラベル資格を配布する実験をする

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

☞ 年齢: 年齢差別になりかねず、違憲の可能性あり

もしも可能だったら... ランダムに GoTo トラベル資格を配布する実験をする

有資格者と無資格者の有症率の違いを検定する: 明確な因果関係

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

☞ 年齢: 年齢差別になりかねず、違憲の可能性あり

もしも可能だったら... ランダムに GoTo トラベル資格を配布する実験をする

有資格者と無資格者の有症率の違いを検定する: 明確な因果関係

でも、有資格者と周辺者が危険に曝される可能性があるので、研究倫理審査委員会が却下するかも

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

政策提言が実施困難: どうやって個人の感染させるリスクを判断するのか?

☞ 年齢: 年齢差別になりかねず、違憲の可能性あり

もしも可能だったら... ランダムに GoTo トラベル資格を配布する実験をする

有資格者と無資格者の有症率の違いを検定する: 明確な因果関係

でも、有資格者と周辺者が危険に曝される可能性があるので、研究倫理審査委員会が却下するかも

GoTo トラベル=研究倫理審査委員会が却下しかねない政策

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

普段の外出頻度を尋ねていれば、外出頻度を制御して比較

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

普段の外出頻度を尋ねていれば、外出頻度を制御して比較

$$\Delta \Pr(\text{有症} \mid \text{外出多}) = \Pr[\text{有症} \mid \text{外出多, GoTo 利用}] - \Pr[\text{有症} \mid \text{外出多, GoTo 非利用}].$$

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

普段の外出頻度を尋ねていれば、外出頻度を制御して比較

$$\Delta \Pr(\text{有症} | \text{外出多}) = \Pr[\text{有症} | \text{外出多, GoTo 利用}] - \Pr[\text{有症} | \text{外出多, GoTo 非利用}].$$

外出頻度ごとに帰無仮説を検定

$$H_{01} : \Delta \Pr(\text{有症} | \text{外出少}) = 0$$

$$H_{02} : \Delta \Pr(\text{有症} | \text{外出多}) = 0$$

仮に、それぞれの帰無仮説において  $p$  value が小さい

⇒ 外出頻度同程度の人で GoTo トラベルと有症率に正の相関関係

⇒ GoTo トラベルが有症率を高める因果関係と矛盾しない

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

普段の外出頻度を尋ねていれば、外出頻度を制御して比較

$$\Delta \Pr(\text{有症} | \text{外出多}) = \Pr[\text{有症} | \text{外出多, GoTo 利用}] - \Pr[\text{有症} | \text{外出多, GoTo 非利用}].$$

外出頻度ごとに帰無仮説を検定

$$H_{01} : \Delta \Pr(\text{有症} | \text{外出少}) = 0$$

$$H_{02} : \Delta \Pr(\text{有症} | \text{外出多}) = 0$$

仮に、それぞれの帰無仮説において  $p$  value が小さい

⇒ 外出頻度同程度の人で GoTo トラベルと有症率に正の相関関係

⇒ GoTo トラベルが有症率を高める因果関係と矛盾しない

残念ながら、外出頻度は尋ねていない模様。

# GoTo トラベルの効果?

「GoTo トラベル利用者感染リスク高い」 東京大学など研究チーム

普段の外出頻度を尋ねていれば、外出頻度を制御して比較

$$\Delta \Pr(\text{有症} | \text{外出多}) = \Pr[\text{有症} | \text{外出多, GoTo 利用}] - \Pr[\text{有症} | \text{外出多, GoTo 非利用}].$$

外出頻度ごとに帰無仮説を検定

$$H_{01} : \Delta \Pr(\text{有症} | \text{外出少}) = 0$$

$$H_{02} : \Delta \Pr(\text{有症} | \text{外出多}) = 0$$

仮に、それぞれの帰無仮説において  $p$  value が小さい

⇒ 外出頻度同程度の人で GoTo トラベルと有症率に正の相関関係

⇒ GoTo トラベルが有症率を高める因果関係と矛盾しない

残念ながら、外出頻度は尋ねていない模様。デザイン段階でもう少し考えるべきだったかも？

# GoTo トラベルの効果?

ビッグ・データによる人の移動

被説明変数 移動人数

データ スマートフォン・データ

効果  $\underbrace{\text{GoTo トラベル開始前後の変化}}_{\text{2020年の8月前後変化}} - \underbrace{\text{昨年同時期の変化}}_{\text{2019年の8月前後変化}}$

推計量 二重差分推計量 double difference (difference-in-differences, DID)

estimator: 今年の変化と比較対象の変化の差(差分と差分の差=二重差分)を政策の効果と見なす。

# GoTo トラベルの効果?

ビッグ・データによる人の移動

**長所** 被説明変数(出典ビッグ・データ)が正確、高頻度、即時。昨年同時期比較で季節性を制御。

**短所** 信頼性の低さ。感染者数ではなく移動人数という間接指標を使っていること。

識別仮定 **identification assumption** 「GoTo トラベルがなければ、今年も昨年と同じ人の移動人数変化だった」

☞ このデザインでの効果推計値の信頼性=この識別仮定の現実妥当性

**課題** ① GoTo トラベルなしのとき、移動人数の変化(傾き)が今年と昨年で同じと期待する理由はない。おそらく、GoTo トラベルなしだと今年は増加幅が去年より小さかったはず。過小評価になる。

② インバウンド客を除外した昨年のデータを作成できるか? インバウンド客を除外することで移動人数の変化が増えるのか減るのか不明。

# GoTo トラベルの効果?

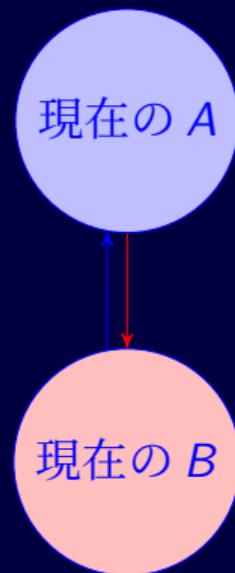
グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



矢印は現在の  $B$  に関するもののみ表示(他は無視)

# GoTo トラベルの効果?

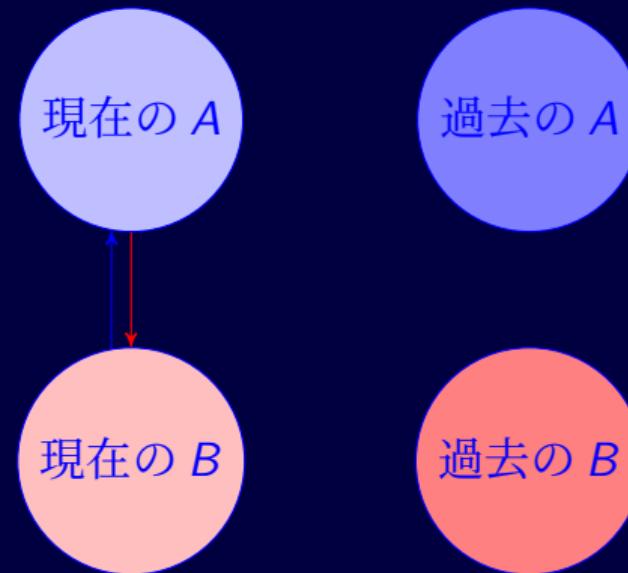
グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



矢印は現在の  $B$   
に関わるもののみ表示(他は  
無視)

# GoTo トラベルの効果?

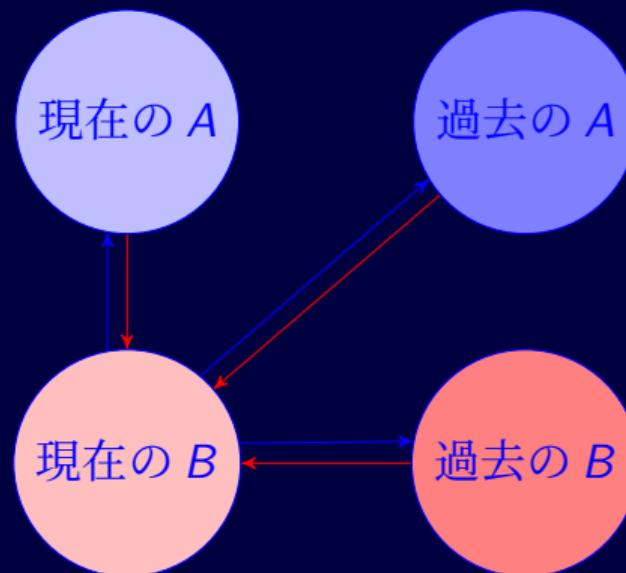
グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



矢印は現在の  $B$   
に関わるもののみ表示(他は  
無視)

# GoTo トラベルの効果?

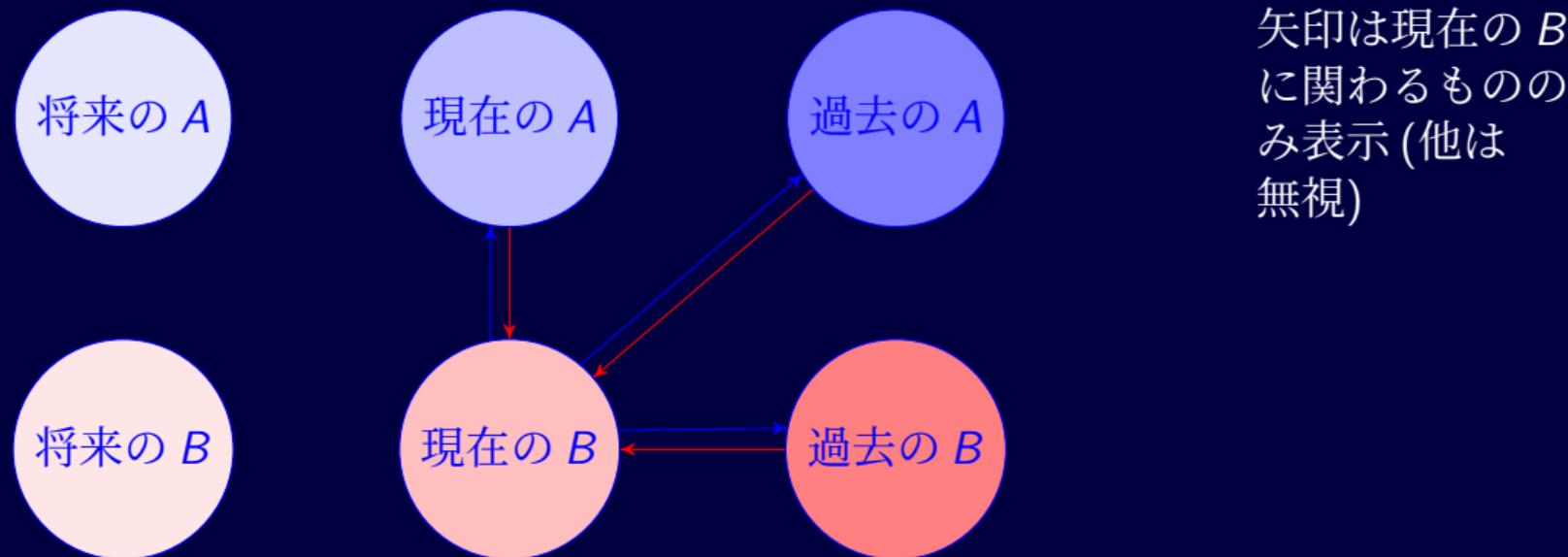
グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



矢印は現在の  $B$   
に関わるもののみ表示(他は  
無視)

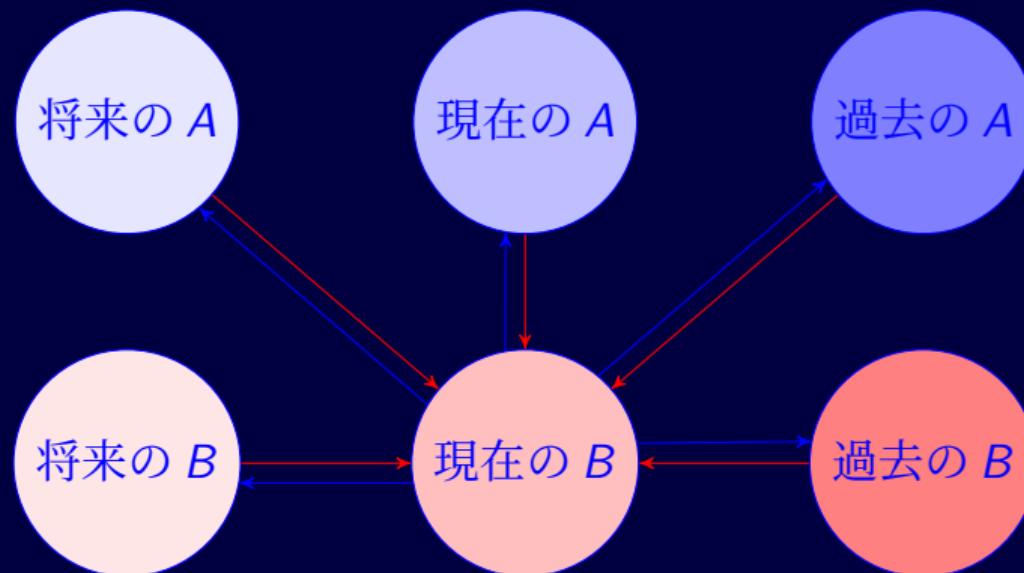
# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



# GoTo トラベルの効果?

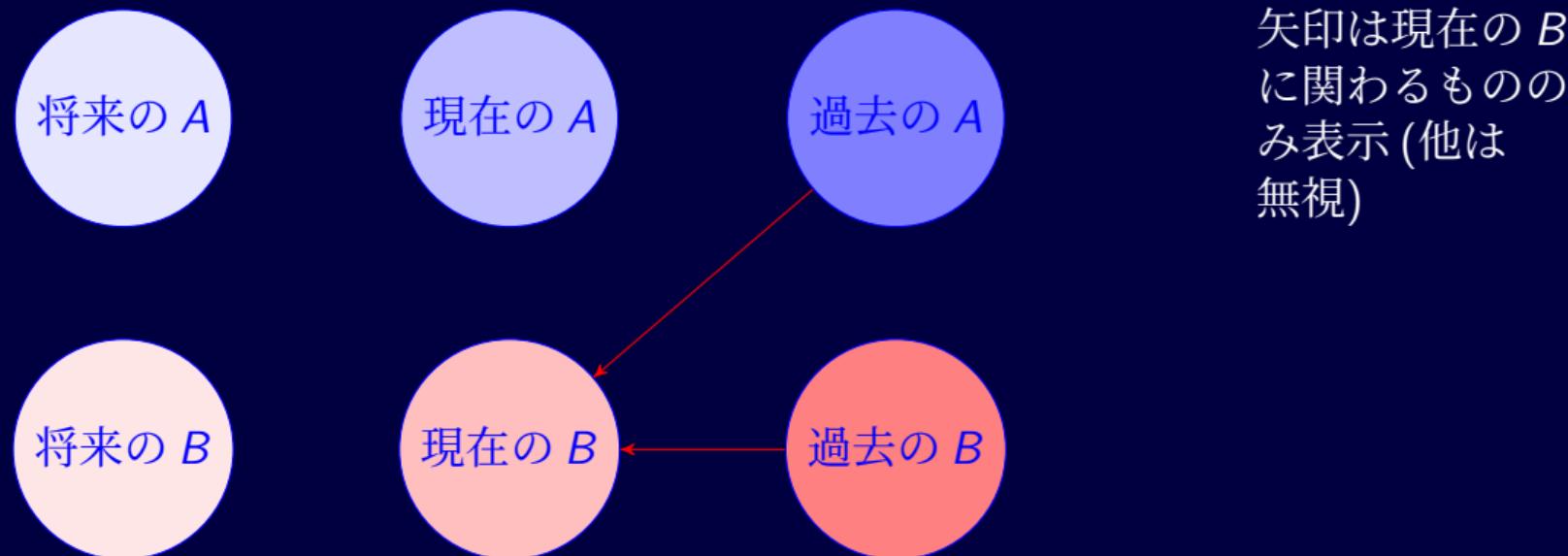
グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



矢印は現在の  $B$   
に関わるもののみ表示(他は無視)

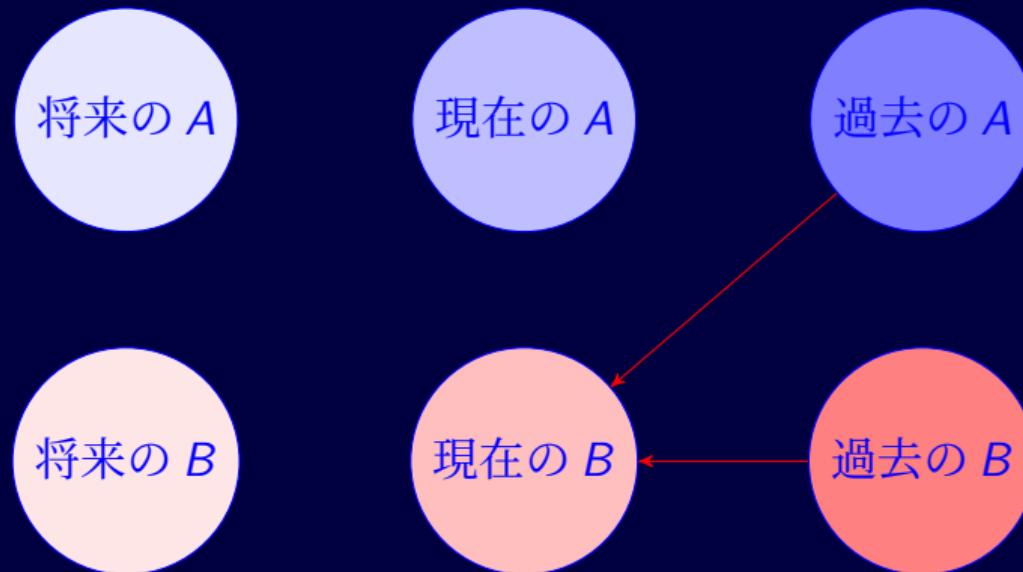
# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$



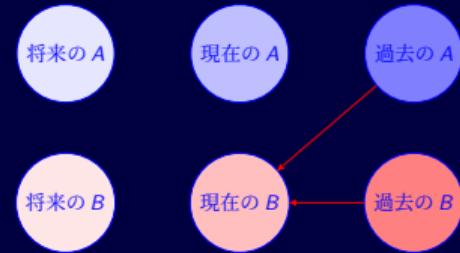
矢印は現在の  $B$  に関するもののみ表示(他は無視)

過去の  $B$  を考慮にした上で過去の  $A$  が現在の  $B$  と相関があること

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

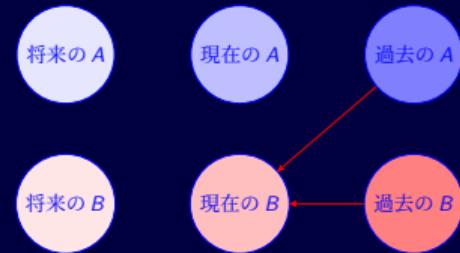
☞ 現在の  $A$  は?



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

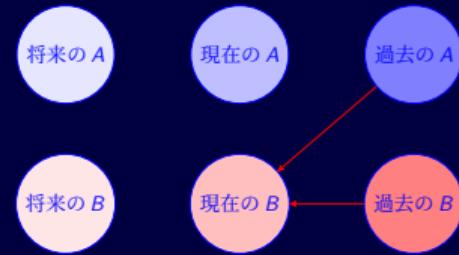
- ☛ 現在の  $A$  は?
- ☛ 将来の  $B$  を見越して現在の  $B$  が変化している場合は? 現在の  $B$  を見越して過去の  $B$  が変化している場合は?



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

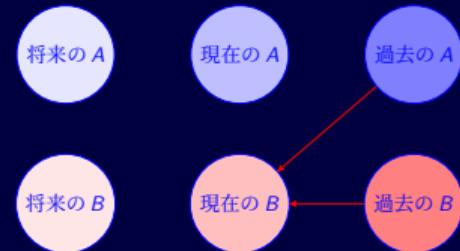
- ☛ 現在の  $A$  は?
- ☛ 将来の  $B$  を見越して現在の  $B$  が変化している場合は? 現在の  $B$  を見越して過去の  $B$  が変化している場合は?
- ☛ 将来の  $B$  を見越して現在の  $A$  が変化している場合は? 現在の  $B$  を見越して過去の  $A$  が変化している場合は?



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

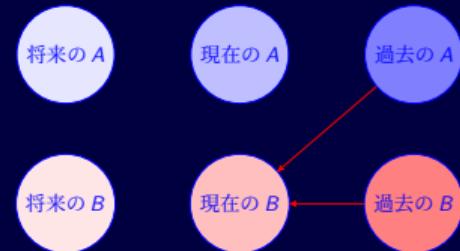
- ☛ 現在の  $A$  は?
- ☛ 将来の  $B$  を見越して現在の  $B$  が変化している場合は? 現在の  $B$  を見越して過去の  $B$  が変化している場合は?
- ☛ 将来の  $B$  を見越して現在の  $A$  が変化している場合は? 現在の  $B$  を見越して過去の  $A$  が変化している場合は?
- ☛ 本来の G 因果性ならば遡れるだけの過去の  $A$  との相関を推計するが、どこまで遡るべきかは考慮している現象次第



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

- ☛ 現在の  $A$  は?
  - ☛ 将来の  $B$  を見越して現在の  $B$  が変化している場合は? 現在の  $B$  を見越して過去の  $B$  が変化している場合は?
  - ☛ 将来の  $B$  を見越して現在の  $A$  が変化している場合は? 現在の  $B$  を見越して過去の  $A$  が変化している場合は?
  - ☛ 本来の G 因果性ならば遡れるだけの過去の  $A$  との相関を推計するが、どこまで遡るべきかは考慮して
- } 無いと仮定



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

厚生労働省感染症アドバイザリーボード 2020年11月19日

(木) 資料3(参考資料): 東京と沖縄、福岡、北海道

第13回新型コロナウイルス感染症対策分科会(令和2年10月29日)事務局提出資料を基に内閣官房・内閣府作成

参考資料

航空旅客数と感染者数の増加には統計的な因果関係は確認できない

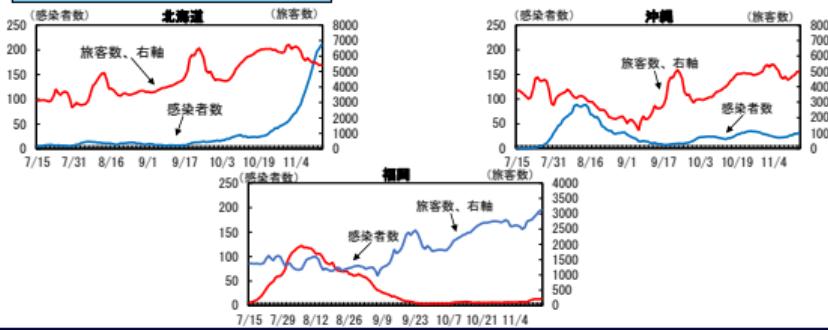
東京発の航空旅客数が北海道・沖縄県・福岡県の感染者数に与える影響について、統計的観点から検証したところ、いずれについてもグレンジャー因果性は確認できなかった。  
(期間は7月15日～11月15日。航空便は、羽田発・新千歳着、羽田発・那覇着、羽田発・福岡着)

グレンジャー検定の結果

航空便 (旅客数)	感染者数	因果性 (7/15～11/15)
羽田発 新千歳着	⇒ 北海道	×
羽田発 那覇着	⇒ 沖縄県	×
羽田発 福岡着	⇒ 福岡県	×

(備考)福岡着はJAL、他はJALとANAの合計。

航空旅客数と感染者数の動向



# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

厚生労働省感染症アドバイザリーボード 2020年11月19日 「統計的な因果関係は確認できない」  
(木) 資料3(参考資料): 東京と沖縄、福岡、北海道

第13回新型コロナウイルス感染症対策分科会(令和2年10月29日)事務局提出資料を基に内閣官房・内閣府作成

参考資料

航空旅客数と感染者数の増加には統計的な因果関係は確認できない

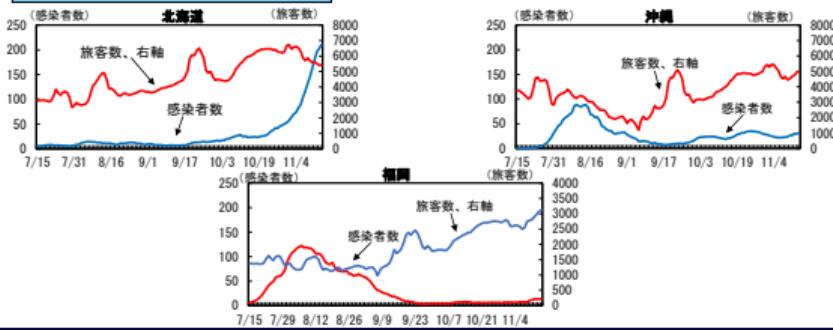
東京発の航空旅客数が北海道・沖縄県・福岡県の感染者数に与える影響について、統計的観点から検証したところ、いずれについてもグレンジャー因果性は確認できなかった。  
(期間は7月15日～11月15日。航空便は、羽田発・新千歳着、羽田発・那覇着、羽田発・福岡着)

## グレンジャー検定の結果

航空便(旅客数)	感染者数	因果性(7/15～11/15)
羽田発 新千歳着	⇒ 北海道	×
羽田発 那覇着	⇒ 沖縄県	×
羽田発 福岡着	⇒ 福岡県	×

(備考)福岡着はJAL、他のJALとANAの合計。

## 航空旅客数と感染者数の動向



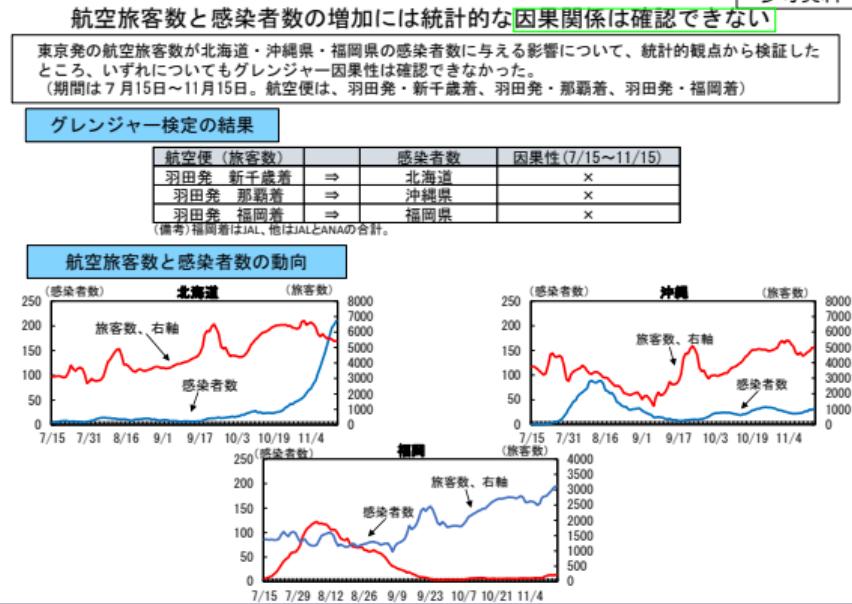
# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

厚生労働省感染症アドバイザリーボード 2020年11月19日  
(木) 資料3(参考資料): 東京と沖縄、福岡、北海道

第13回新型コロナウイルス感染症対策分科会(令和2年10月29日)事務局提出資料を基に内閣官房・内閣府作成

参考資料



「統計的な因果関係は確認できない」 → 「グレンジャー因果関係は」

上昇局面が関係を検出しやすい:  
旅客数が先行、感染者数が追従

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

厚生労働省感染症アドバイザリーボード 2020年11月19日  
(木) 資料3(参考資料): 東京と沖縄、福岡、北海道

第13回新型コロナウイルス感染症対策分科会(令和2年10月29日)事務局提出資料を基に内閣官房・内閣府作成

参考資料

航空旅客数と感染者数の増加には統計的な因果関係は確認できない

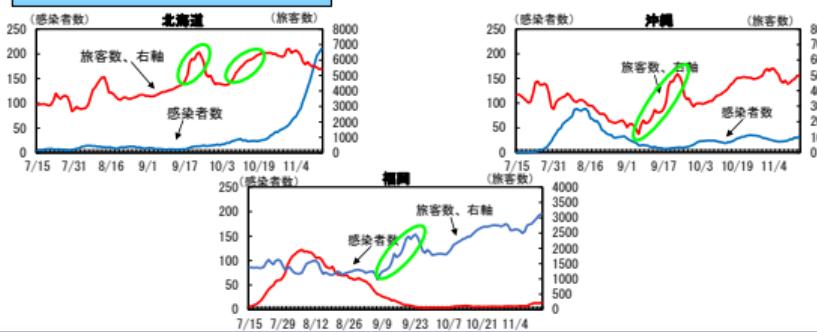
東京発の航空旅客数が北海道・沖縄県・福岡県の感染者数に与える影響について、統計的観点から検証したところ、いずれについてもグレンジャー因果性は確認できなかった。  
(期間は7月15日～11月15日。航空便は、羽田発・新千歳着、羽田発・那覇着、羽田発・福岡着)

## グレンジャー検定の結果

航空便 (旅客数)	感染者数	因果性 (7/15～11/15)
羽田発・新千歳着	⇒ 北海道	×
羽田発・那覇着	⇒ 沖縄県	×
羽田発・福岡着	⇒ 福岡県	×

(備考)福岡着はJAL、他のはJALとANAの合計。

## 航空旅客数と感染者数の動向



「統計的な因果関係は確認できない」 → 「グレンジャー因果関係は」

上昇局面が関係を検出しやすい:  
旅客数が先行、感染者数が追従

⇒ 沖縄と福岡: 推計するには感染  
ピーク前のデータが短かすぎ  
る、もっと遡ってデータを使  
うべき

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

厚生労働省感染症アドバイザリーボード 2020年11月19日  
(木) 資料3(参考資料): 東京と沖縄、福岡、北海道

第13回新型コロナウイルス感染症対策分科会(令和2年10月29日)事務局提出資料を基に内閣官房・内閣府作成

参考資料

航空旅客数と感染者数の増加には統計的な因果関係は確認できない

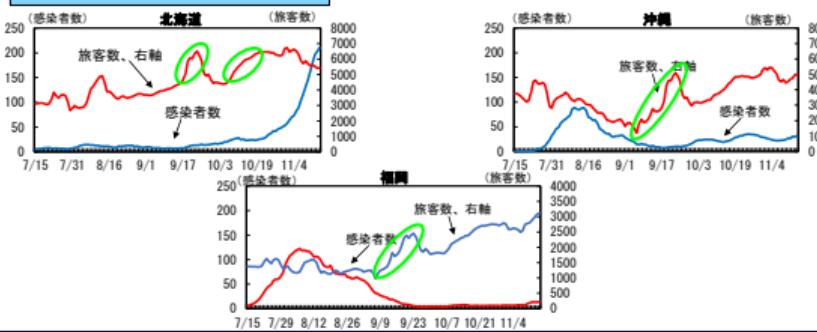
東京発の航空旅客数が北海道・沖縄県・福岡県の感染者数に与える影響について、統計的観点から検証したところ、いずれについてもグレンジャー因果性は確認できなかった。  
(期間は7月15日～11月15日。航空便は、羽田発・新千歳着、羽田発・那覇着、羽田発・福岡着)

## グレンジャー検定の結果

航空便 (旅客数)	感染者数	因果性 (7/15～11/15)
羽田発 新千歳着	⇒ 北海道	✗
羽田発 那覇着	⇒ 沖縄県	✗
羽田発 福岡着	⇒ 福岡県	✗

(備考)福岡着はJAL、他のはJALとANAの合計。

## 航空旅客数と感染者数の動向



「統計的な因果関係は確認できない」 → 「グレンジャー因果関係は」

上昇局面が関係を検出しやすい:  
旅客数が先行、感染者数が追従

⇒ 沖縄と福岡: 推計するには感染  
ピーク前のデータが短かすぎ  
る、もっと遡ってデータを使  
うべき

下降局面は検出しにくい: 旅客数が  
減っても、モメンタム(市中感染に  
による自然増)が減るとは限らない

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

ラグ数: どこまで過去に遡るか? 最低2週間、3-4週間くらい?

- ☞ 北海道: 10月1日以降の旅客数の増加によって1ヶ月後に感染者数が増えているようにも見える。推計でどこまで前を考慮しているか不明。

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

ラグ数: どこまで過去に遡るか? 最低2週間、3-4週間くらい?

☞ 北海道: 10月1日以降の旅客数の増加によって1ヶ月後に感染者数が増えている  
ように見える。推計でどこまで前を考慮しているか不明。

なぜこの3道県だけ?

# GoTo トラベルの効果?

グレンジャー因果性 Granger causality:  $A$  Granger-causes  $B$

ラグ数: どこまで過去に遡るか? 最低2週間、3-4週間くらい?

☞ 北海道: 10月1日以降の旅客数の増加によって1ヶ月後に感染者数が増えている  
ように見える。推計でどこまで前を考慮しているか不明。

なぜこの3道県だけ?

東京の GoTo トラベルは秋以降なので、北海道や沖縄よりも、近隣の紅葉のきれいな  
ところや温泉地に行く人が多そう... 静岡、山梨、栃木、福島とかを検討すればいい  
のでは?

# Go To トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

# Go To トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

被説明変数 各都道府県の新規感染者数

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

被説明変数 各都道府県の新規感染者数

データ 厚労省の日次データ

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

被説明変数 各都道府県の新規感染者数

データ 厚労省の日次データ

効果 東京追加前後での: 静岡の変化-合成静岡(=非東京圏加重平均値)の変化

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

被説明変数 各都道府県の新規感染者数

データ 厚労省の日次データ

効果 東京追加前後での: 静岡の変化-合成静岡(=非東京圏加重平均値)の変化

推計量 合成統御法(synthetic control method): 非東京圏のデータで「政策なしの静岡」の値を合成し、実際の静岡の値と合成値との差を政策の効果と見なす

- 被説明変数と相関のある変数で、静岡の値と他の非東京圏道府県の加重平均値の差(の2乗和)を最小化するように加重平均ウェイトを選ぶ。このウェイトと非東京圏の被説明変数データを使って合成静岡の被説明変数値を計算する。

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

被説明変数 各都道府県の新規感染者数

データ 厚労省の日次データ

効果 東京追加前後での: 静岡の変化-合成静岡(=非東京圏加重平均値)の変化

推計量 合成統御法(synthetic control method): 非東京圏のデータで「政策なしの静岡」の値を合成し、実際の静岡の値と合成値との差を政策の効果と見なす

- 被説明変数と相関のある変数で、静岡の値と他の非東京圏道府県の加重平均値の差(の2乗和)を最小化するように加重平均ウェイトを選ぶ。このウェイトと非東京圏の被説明変数データを使って合成静岡の被説明変数値を計算する。

▶ SCMに適したデータ

▶ SCM

# Go To トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

長所 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

識別仮定 **identification assumption** 「GoTo トラベルによって東京との旅客移動の影  
響が無視し得る非東京圏(道府県)が複数あり、GoTo トラベル東京追加  
なしの場合に、感染者数トレンドは静岡とこれら非東京圏で等しい」

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

識別仮定 **identification assumption** 「GoTo トラベルによって東京との旅客移動の影  
響が無視し得る非東京圏(道府県)が複数あり、GoTo トラベル東京追加  
なしの場合に、感染者数トレンドは静岡とこれら非東京圏で等しい」

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

識別仮定 **identification assumption** 「GoTo トラベルによって東京との旅客移動の影  
響が無視し得る非東京圏(道府県)が複数あり、GoTo トラベル東京追加  
なしの場合に、感染者数トレンドは静岡とこれら非東京圏で等しい」

課題 ① 非東京圏が存在するか。

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

識別仮定 **identification assumption** 「GoTo トラベルによって東京との旅客移動の影響が無視し得る非東京圏(道府県)が複数あり、GoTo トラベル東京追加なしの場合に、感染者数トレンドは静岡とこれら非東京圏で等しい」

- 課題**
- ① 非東京圏が存在するか。
  - ② 非東京圏の道府県を決めるのに一定の恣意性があり、変えると結果が変わること可能性あり。

# GoTo トラベルの効果?

東京追加による東京圏(例: 静岡県)の新規感染者数の変化

**長所** 感染者数をそのものを取り上げている。データは長期のリードタイムが  
あってこの推計方法に向いている。

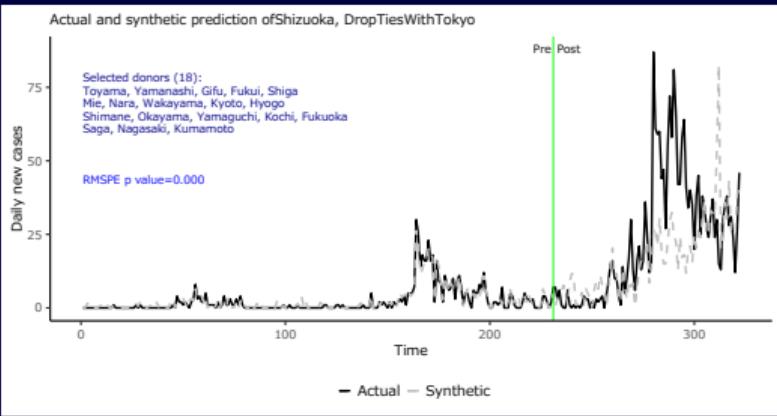
**短所** トレンドが同じになる疫学的根拠なしには、信頼性が高いとはいえない。  
→ 独自トレンドの県をドナー・プールから除外すればいい。

識別仮定 **identification assumption** 「GoTo トラベルによって東京との旅客移動の影響が無視し得る非東京圏(道府県)が複数あり、GoTo トラベル東京追加なしの場合に、感染者数トレンドは静岡とこれら非東京圏で等しい」

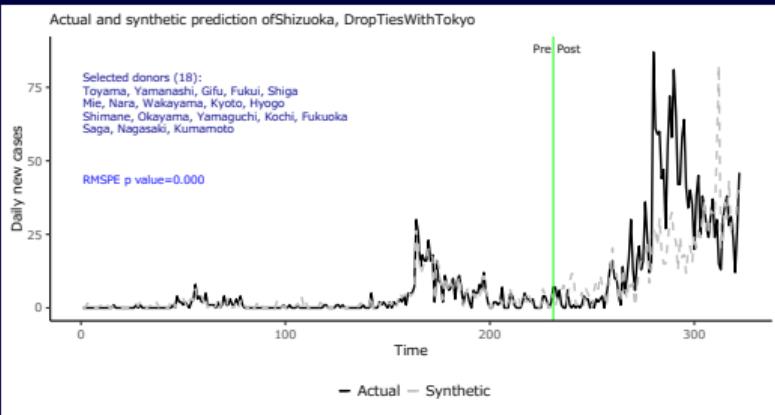
- 課題**
- ① 非東京圏が存在するか。
  - ② 非東京圏の道府県を決めるのに一定の恣意性があり、変えると結果が変わること可能性あり。
  - ③ 加重平均ウェイトを選ぶときに観察可能な違い(観察可能な変数)のみを考慮している。

# GoTo トラベルの効果?

GTT+東京による静岡県の新規感染者数変化



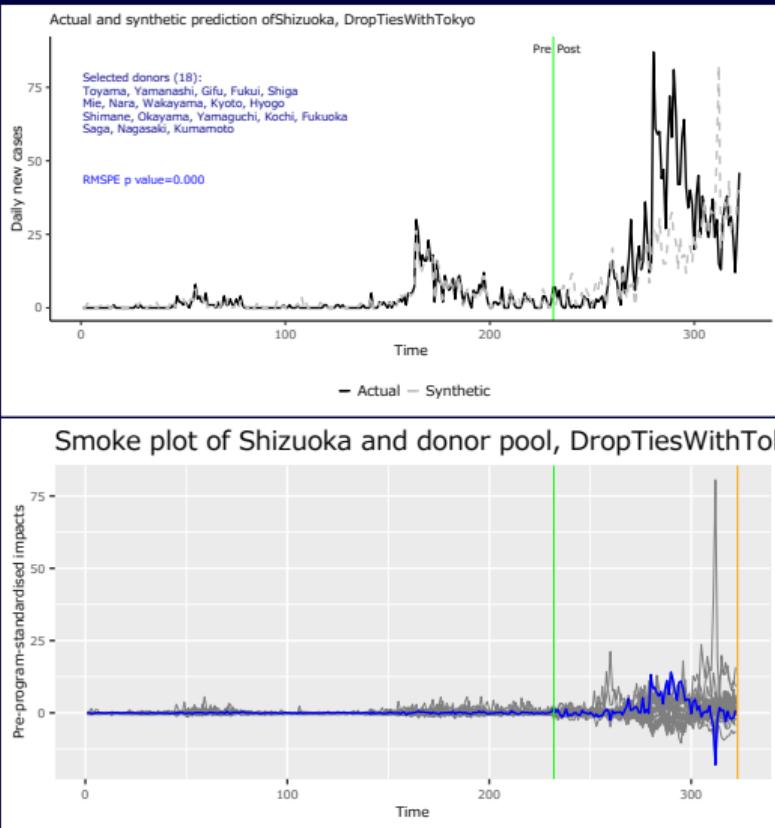
# GoTo トラベルの効果?



GTT+東京による静岡県の新規感染者数変化

非東京圏: 東京からの宿泊者数が全国平均よりも少ない道府県。

# GoTo トラベルの効果?

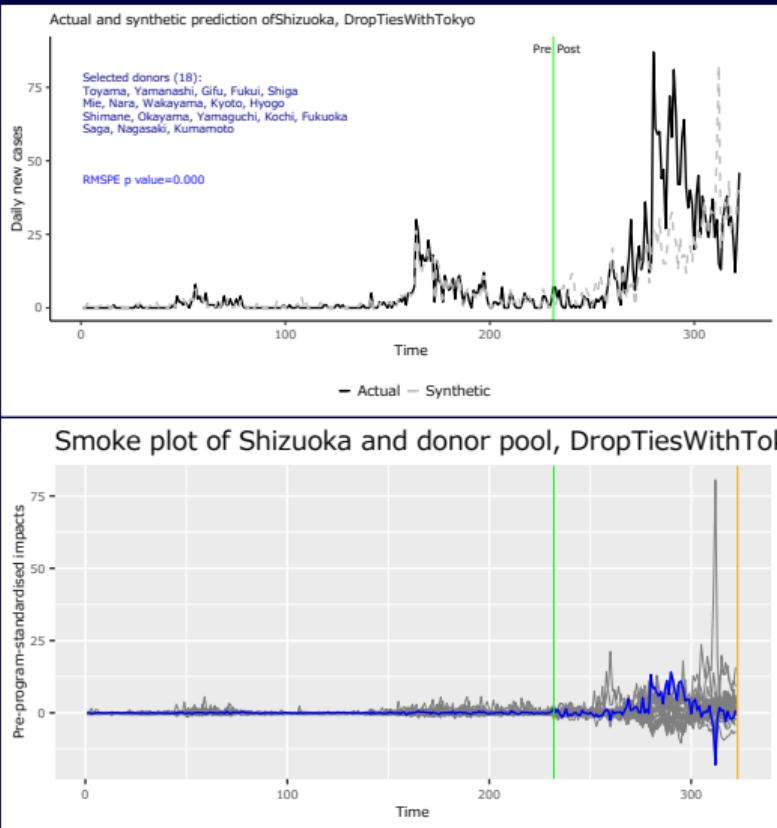


GTT+東京による静岡県の新規感染者数変化

非東京圏: 東京からの宿泊者数が全国平均よりも少ない道府県。

スモーク図 smoke plot

# GoTo トラベルの効果?



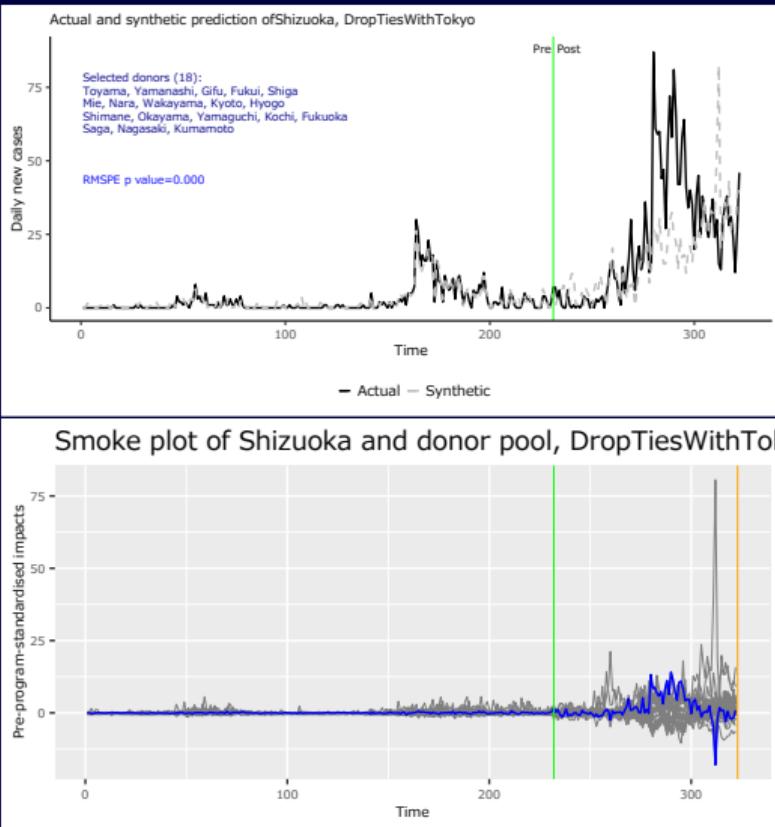
GTT+東京による静岡県の新規感染者数変化

非東京圏: 東京からの宿泊者数が全国平均よりも少ない道府県。

スモーク図 smoke plot

静岡と東京以外すべての道府県でインパクト推計し、灰色の線で示した。

# GoTo トラベルの効果?



GTT+東京による静岡県の新規感染者数変化

非東京圏: 東京からの宿泊者数が全国平均よりも少ない道府県。

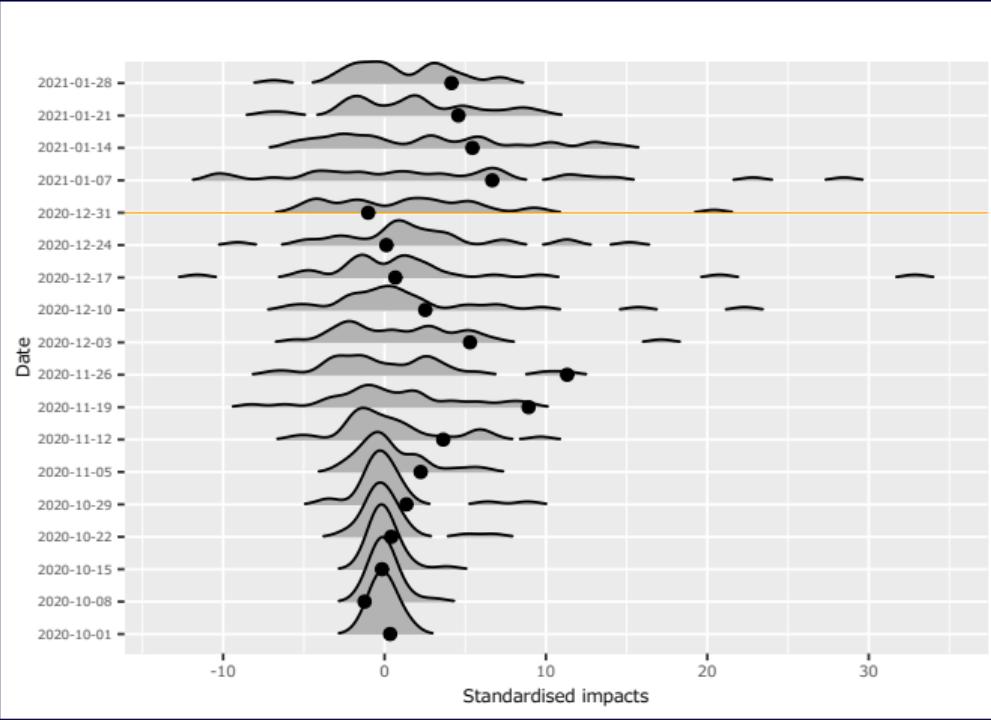
スモーク図 smoke plot

静岡と東京以外すべての道府県でインパクト推計し、灰色の線で示した。

GTT+東京に曝露されたと仮定: 他道府県のインパクト推計値と静岡県のインパクト推計値の比較。静岡の外れ度を示す。

# GoTo トラベルの効果?

リッジ密度図

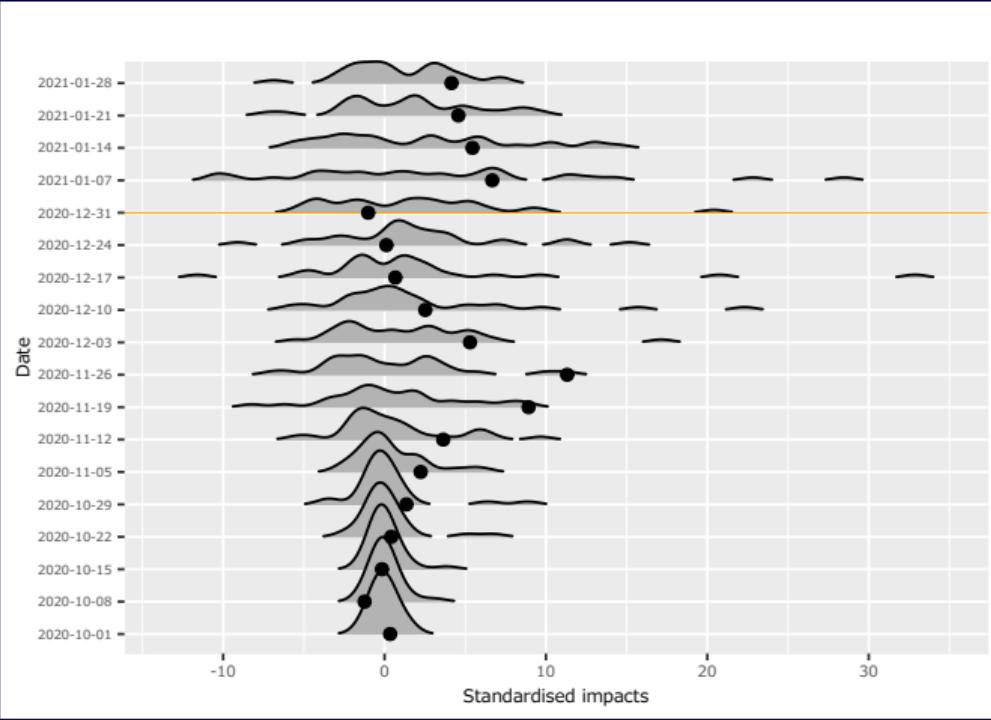


スモーク図と同じ情報を別の視覚化をした。

点が静岡、分布がその他道府県のインパクト推計値。

# GoTo トラベルの効果?

リッジ密度図



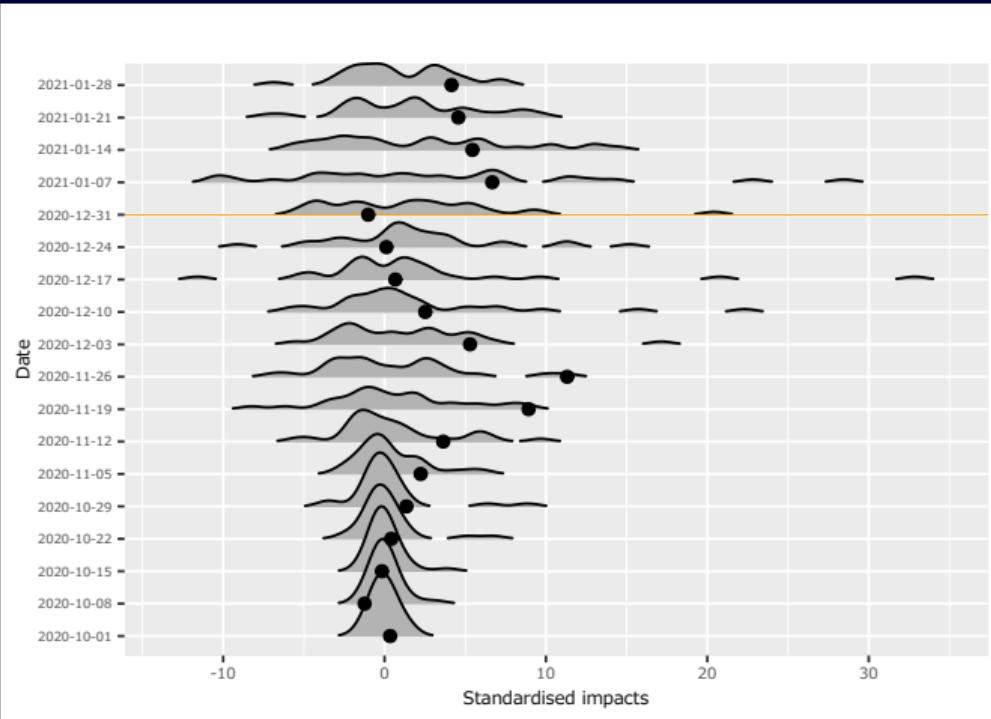
スモーク図と同じ情報を別の視覚化をした。

点が静岡、分布がその他道府県のインパクト推計値。

静岡インパクト推計値の外れ度を示す。

# GoTo トラベルの効果?

リッジ密度図



スモーク図と同じ情報を別の視覚化をした。

点が静岡、分布がその他道府県のインパクト推計値。

静岡インパクト推計値の外れ度を示す。

11月後半-12月初旬の3週間で他道府県に比べてインパクトが大きい。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

- 社会科学分野の政策は大きな単位で実施することが多い。貿易自由化、経済自由化、アベノミクス、量的金融緩和政策、マイナス金利など。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

- 社会科学分野の政策は大きな単位で実施することが多い。貿易自由化、経済自由化、アベノミクス、量的金融緩和政策、マイナス金利など。

問題: 統御群 (CF となる国や地方)がない。インパクト評価の手法を使えない。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

- 社会科学分野の政策は大きな単位で実施することが多い。貿易自由化、経済自由化、アベノミクス、量的金融緩和政策、マイナス金利など。

問題: 統御群(CFとなる国や地方)がない。インパクト評価の手法を使えない。

- マクロ経済学: マクロ経済モデル(マクロ経済を表す数理統計モデル、 $Y = a + bX + e$ が複数ある連立方程式)を作り、データを使ってモデルのパラメタ  $a, b$  を推計し、推計値  $\hat{a}, \hat{b}$  をモデルに代入して影響を計算  $Y_1 = \hat{a} + \hat{b}X_1$  する。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

- 社会科学分野の政策は大きな単位で実施することが多い。貿易自由化、経済自由化、アベノミクス、量的金融緩和政策、マイナス金利など。

問題: 統御群(CFとなる国や地方)がない。インパクト評価の手法を使えない。

- マクロ経済学: マクロ経済モデル(マクロ経済を表す数理統計モデル、 $Y = a + bX + e$ が複数ある連立方程式)を作り、データを使ってモデルのパラメタ  $a, b$  を推計し、推計値  $\hat{a}, \hat{b}$  をモデルに代入して影響を計算  $Y_1 = \hat{a} + \hat{b}X_1$  する。
- CFはモデルで作っている。金利を下げると銀行貸出が(下げなかったときよりも)増えて企業の投資が(下げなかったときよりも)増え...などという設定は、CF(金利を下げなかったとき)との比較で示されている。

# Synthetic control method

国や(関東など大きな)地方単位でインパクトを推計したいときもある。

- 社会科学分野の政策は大きな単位で実施することが多い。貿易自由化、経済自由化、アベノミクス、量的金融緩和政策、マイナス金利など。

問題: 統御群(CFとなる国や地方)がない。インパクト評価の手法を使えない。

- マクロ経済学: マクロ経済モデル(マクロ経済を表す数理統計モデル、 $Y = a + bX + e$ が複数ある連立方程式)を作り、データを使ってモデルのパラメタ  $a, b$  を推計し、推計値  $\hat{a}, \hat{b}$  をモデルに代入して影響を計算  $Y_1 = \hat{a} + \hat{b}X_1$  する。
- CFはモデルで作っている。金利を下げると銀行貸出が(下げなかったときよりも)増えて企業の投資が(下げなかったときよりも)増え...などという設定は、CF(金利を下げなかったとき)との比較で示されている。 $\hat{a}, \hat{b}$  やモデルは正しい?

実は、CFがなくてもインパクト評価を可能にする方法がある。

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能=観察可能な特徴を考慮すれば、統御群と治療群への割り振りがランダム(選抜がない)と考えられる

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能=観察可能な特徴を考慮すれば、統御群と治療群への割り振りがランダム(選抜がない)と考えられる

ここで「ランダム」とは?

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能=観察可能な特徴を考慮すれば、統御群と治療群への割り振りがランダム(選抜がない)と考えられる

ここで「ランダム」とは?

観察可能な特徴を考慮すれば、効果の大きさと治療状態が相関していないこと

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能=観察可能な特徴を考慮すれば、統御群と治療群への割り振りがランダム(選抜がない)と考えられる

ここで「ランダム」とは?

観察可能な特徴を考慮すれば、効果の大きさと治療状態が相関していないこと  
=効果の大きい/小さい人ほど治療群になりやすい、がないこと

実は、CFがなくてもインパクト評価を可能にする方法がある。

用いる仮定: 観察可能な変数による選抜 selection on the observable

観察可能な変数で統御群と治療群への選抜が説明可能=観察可能な特徴を考慮すれば、統御群と治療群への割り振りがランダム(選抜がない)と考えられる

ここで「ランダム」とは?

観察可能な特徴を考慮すれば、効果の大きさと治療状態が相関していないこと

=効果の大きい/小さい人ほど治療群になりやすい、がないこと

観察可能な特徴を考慮すれば、治療状態と結果が無相関/統計的に独立を条件付無相関 conditional orthogonality/条件付独立 conditional independence という

selection on the observable は滅多に満たされない

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

手法開発者の意図

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

手法開発者の意図

- 叙述研究は CF を意識しないので信頼性が低い。数量研究は CF について一定の仮定を満たさないと作業しない。

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

手法開発者の意図

- ☞ 叙述研究は CF を意識しないので信頼性が低い。数量研究は CF について一定の仮定を満たさないと作業しない。
- ☞ 中間がない。叙述研究と数量研究の橋渡しをしたい。

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

手法開発者の意図

- ☞ 叙述研究は CF を意識しないので信頼性が低い。数量研究は CF について一定の仮定を満たさないと作業しない。
- ☞ 中間がない。叙述研究と数量研究の橋渡しをしたい。
- ☞ 仮定を明示して、何が言えるかを示すことにも意義がある。

selection on the observable は滅多に満たされない

人間は自分にとって得なことを選択する (=選抜がある) し、選抜に関わる観察不可能な情報が必ずある

では、なぜこんな仮定をおいてインパクト評価をするのか？

何もやらないよりも、仮定と限界を明示して作業する方が意義があるため

手法開発者の意図

- ☞ 叙述研究は CF を意識しないので信頼性が低い。数量研究は CF について一定の仮定を満たさないと作業しない。
- ☞ 中間がない。叙述研究と数量研究の橋渡しをしたい。
- ☞ 仮定を明示して、何が言えるかを示すことにも意義がある。

オリジナルの開発者の崇高な意志に反して仮定もよく考えずにインパクト評価をやってしまう例は多いが…

?: テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

?: テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

? : テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

$$ATT_{basque,t} = \underbrace{y_{basque,t}}_{\text{バスク}} - \underbrace{\sum_{j=1}^J w_j y_{j,t}}_{\substack{\text{バスク以外の} \\ \text{加重平均}}} , \quad 0 \leq w_j \leq 1, \quad \sum_{j=1}^J w_j = 1.$$

? : テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

$$ATT_{basque,t} = \underbrace{y_{basque,t}}_{\text{バスク}} - \underbrace{\sum_{j=1}^J w_j y_{j,t}}_{\substack{\text{バスク以外の} \\ \text{加重平均}}} , \quad 0 \leq w_j \leq 1, \quad \sum_{j=1}^J w_j = 1.$$

テロ開始前データを使って、テロ無しバスクと加重平均の誤差の 2 乗和が最小化するようにウェイト  $w_j$  を選ぶ

? : テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

$$ATT_{basque,t} = \underbrace{y_{basque,t}}_{\text{バスク}} - \underbrace{\sum_{j=1}^J w_j y_{j,t}}_{\substack{\text{バスク以外の} \\ \text{加重平均}}} , \quad 0 \leq w_j \leq 1, \quad \sum_{j=1}^J w_j = 1.$$

テロ開始前データを使って、テロ無しバスクと加重平均の誤差の 2 乗和が最小化するようにウェイト  $w_j$  を選ぶ

△ なぜ和ではなく 2 乗和?

? : テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

$$ATT_{basque,t} = \underbrace{y_{basque,t}}_{\text{バスク}} - \underbrace{\sum_{j=1}^J w_j y_{j,t}}_{\substack{\text{バスク以外の} \\ \text{加重平均}}} , \quad 0 \leq w_j \leq 1, \quad \sum_{j=1}^J w_j = 1.$$

テロ開始前データを使って、テロ無しバスクと加重平均の誤差の 2 乗和が最小化するようにウェイト  $w_j$  を選ぶ

なぜ和ではなく 2 乗和? 2 乗和が最小化されたら、(その正の平方根の) 和も最小化されるから問題ない

? : テロがバスク郡の経済成長に与える影響を推計。スペインでテロはバスク郡以外にはない。

アイディア: それ以外の郡すべて ("donor pool") から CF を合成 synthesise

$$ATT_{basque,t} = \underbrace{y_{basque,t}}_{\text{バスク}} - \underbrace{\sum_{j=1}^J w_j y_{j,t}}_{\substack{\text{バスク以外の} \\ \text{加重平均}}} , \quad 0 \leq w_j \leq 1, \quad \sum_{j=1}^J w_j = 1.$$

テロ開始前データを使って、テロ無しバスクと加重平均の誤差の 2 乗和が最小化するようにウェイト  $w_j$  を選ぶ

- Ⓐ なぜ和ではなく 2 乗和? 2 乗和が最小化されたら、(その正の平方根の) 和も最小化されるから問題ない
- Ⓑ 2 乗和にするのは数学上の都合: 2 乗するとウェイトの関数 (=誤差) を最小化 (微分) して解を求められるから

SPAIN



## $w_j$ の計算方法 (テロ前 $T$ 期、テロ発生以降 $T + 1, T + 2, \dots$ )

- ① 計  $J$  郡 ( $j = 1, \dots, J$ ) について結果変数 (成長) に影響する  $I$  個 の共変数 predictors  $x_{1jt}, \dots, x_{Ijt}$  をテロ前の時期  $t = [1, \dots, T]$  について集める。結果変数を全期間  $t = [1, \dots, T], T + 1, T + 2, \dots$  について集める。

## $w_j$ の計算方法 (テロ前 $T$ 期、テロ発生以降 $T + 1, T + 2, \dots$ )

- ① 計  $J$  郡 ( $j = 1, \dots, J$ ) について結果変数 (成長) に影響する  $I$  個 の共変数 predictors  $x_{1jt}, \dots, x_{Ijt}$  をテロ前の時期  $t = 1, \dots, T$  について集める。結果変数を全期間  $t = [1, \dots, T], T + 1, T + 2, \dots$  について集める。
- ② テロ前  $T$  期間データ (結果変数と共変数) を使い、バスクとバスク以外の加重平均の差の 2 乗和を最小化する各変数 (結果変数と共変数)  $z_{ibt}$  共通の郡ウェイト  $w_1^*, \dots, w_J^*$  を選ぶ。

☞ 2 乗和関数  $\sum_{i=1}^{I+1} \sum_{t=1}^T \left( z_{ibt} - \sum_{j=1}^J w_j z_{jxt} \right)^2$  に最小化の一階条件を使う

## $w_j$ の計算方法 (テロ前 $T$ 期、テロ発生以降 $T+1, T+2, \dots$ )

- ① 計  $J$  郡 ( $j = 1, \dots, J$ ) について結果変数(成長)に影響する  $I$  個の共変数 predictors  $x_{1jt}, \dots, x_{Ijt}$  をテロ前の時期  $t = 1, \dots, T$  について集める。結果変数を全期間  $t = [1, \dots, T], [T+1, T+2, \dots]$  について集める。
- ② テロ前  $T$  期間データ(結果変数と共変数)を使い、バスクとバスク以外の加重平均の差の2乗和を最小化する各変数(結果変数と共変数)  $z_{ibt}$  共通の郡ウェイト  $w_1^*, \dots, w_J^*$  を選ぶ。

☞ 2乗和関数  $\sum_{i=1}^{I+1} \sum_{t=1}^T \left( z_{ibt} - \sum_{j=1}^J w_j z_{jxt} \right)^2$  に最小化の一階条件を使う

- ③  $\sum_{j=1}^J w_j^* y_j[T+1] = \hat{y}_{T+1}, \sum_{j=1}^J w_j^* y_j[T+2] = \hat{y}_{T+2}, \dots$  がテロ以降のテロ無し合成バスク。

## 2乗和関数とは

$$\begin{aligned} e_{it}(w_1, w_2, \dots, w_J)^2 &= \{t \text{ 期の変数 } i \text{ の誤差 } (w_1, w_2, \dots, w_J)\}^2, \\ &= (z_{i, \text{バスク } t} - w_1 z_{i1t} - w_2 z_{i2t} - \cdots - w_J z_{iJt})^2 \\ e_{i1}(w_1, w_2, \dots, w_J)^2 + e_{i2}(w_1, w_2, \dots, w_J)^2 + \cdots + e_{iT}(w_1, w_2, \dots, w_J)^2 \\ &= \{ \text{各期の変数 } i \text{ の誤差 } (w_1, w_2, \dots, w_J) \}^2 \text{ の和} \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^{I+1} \sum_{t=1}^T \left( z_{ibt} - \sum_{j=1}^J w_j z_{ijt} \right)^2 &= e_{11}(w_1, w_2, \dots, w_J)^2 + \cdots + e_{1T}(w_1, w_2, \dots, w_J)^2 \\ &\quad + e_{21}(w_1, w_2, \dots, w_J)^2 + \cdots + e_{2T}(w_1, w_2, \dots, w_J)^2 \\ &\quad \cdots + e_{I+11}(w_1, w_2, \dots, w_J)^2 + \cdots + e_{I+1T}(w_1, w_2, \dots, w_J)^2 \\ &= \{ \text{各期の変数 } 1 \text{ の誤差 } (w_1, w_2, \dots, w_J) \}^2 \text{ の和} \\ &\quad + \{ \text{各期の変数 } 2 \text{ の誤差 } (w_1, w_2, \dots, w_J) \}^2 \text{ の和} \\ &\quad \cdots + \{ \text{各期の変数 } I+1 \text{ の誤差 } (w_1, w_2, \dots, w_J) \}^2 \text{ の和} \end{aligned}$$

## SCM: Terrorism impacts on growth: T=1969

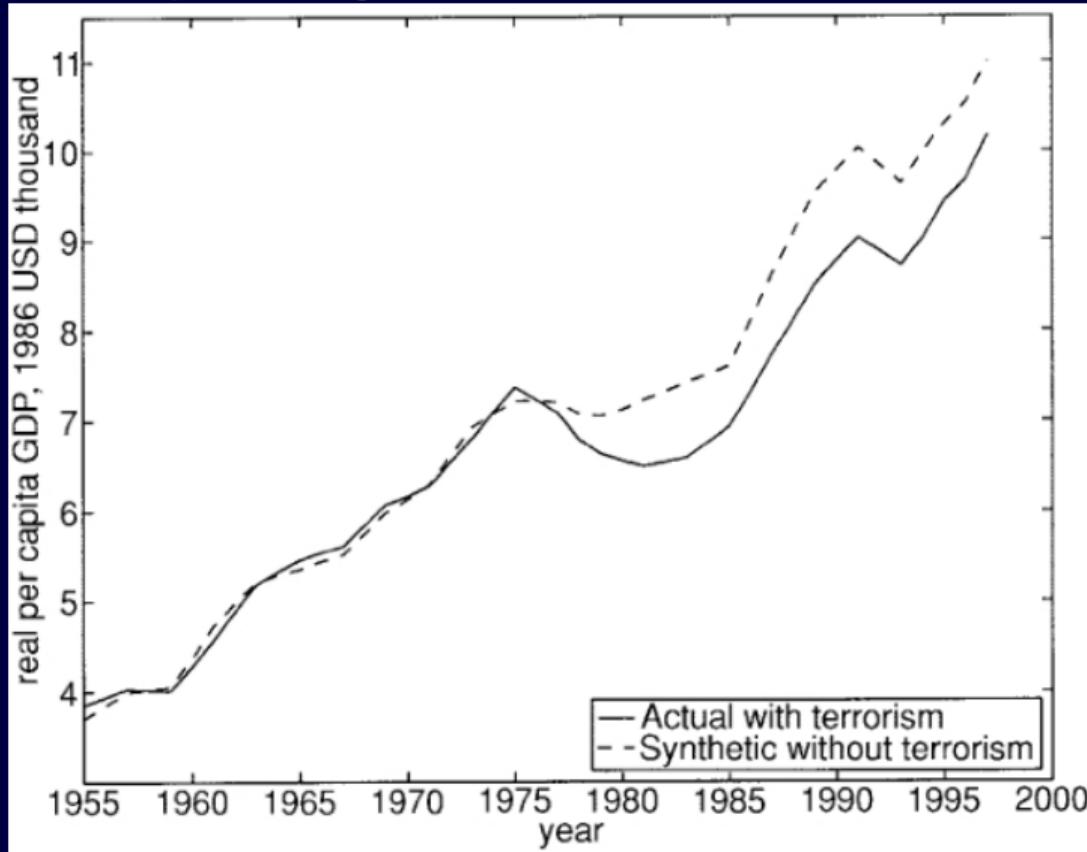


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

# A placebo study: Catalonia: バスクに似ているがテロがない

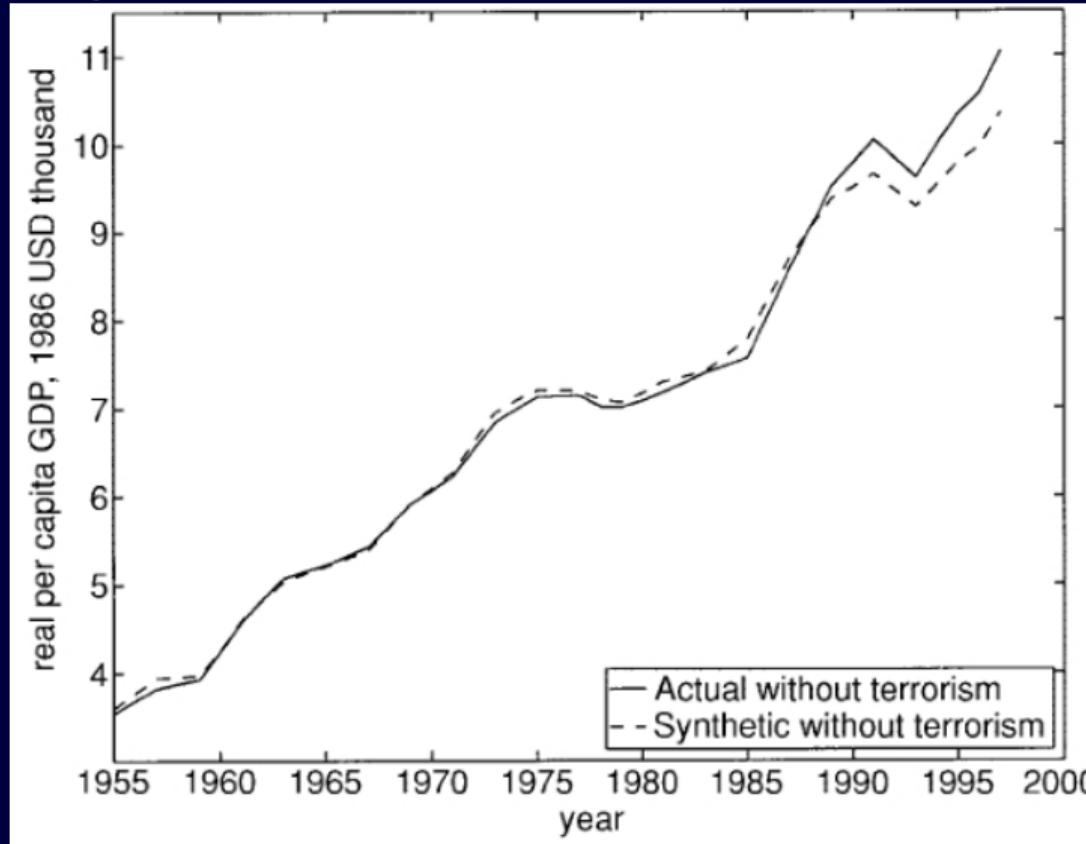


FIGURE 4. A "PLACEBO STUDY," PER CAPITA GDP FOR CATALUNYA

## SCM が適しているデータ

- 治療群が少なく統御群(ドナー・プール)が多い
- 「推計標本 estimation sample」(=イベント発生前)の期間  $T$  が比較的長い  
(10-30 期間など)

推計期間での予測誤差が大きいとイベント後の予測も精度が低いので、使えない

短所 どの程度の誤差なら良いのか現在は基準がない、共変数選定の基準がない、イベント前データが長期に必要

長所 比較対象の選択基準を客観化、プラセボ分析が可能

## SCM が適しているデータ

- 治療群が少なく統御群(ドナー・プール)が多い
- 「推計標本 estimation sample」(=イベント発生前)の期間  $T$  が比較的長い  
(10-30 期間など)

推計期間での予測誤差が大きいとイベント後の予測も精度が低いので、使えない

短所 どの程度の誤差なら良いのか現在は基準がない、共変数選定の基準がない、イベント前データが長期に必要

長所 比較対象の選択基準を客観化、プラセボ分析が可能

## SCM が適しているデータ

- 治療群が少なく統御群(ドナー・プール)が多い
- 「推計標本 estimation sample」(=イベント発生前)の期間  $T$  が比較的長い  
(10-30 期間など)

推計期間での予測誤差が大きいとイベント後の予測も精度が低いので、使えない

**短所** どの程度の誤差なら良いのか現在は基準がない、共変数選定の基準がない、イベント前データが長期に必要

**長所** 比較対象の選択基準を客観化、プラセボ分析が可能

## SCM が適しているデータ

- 治療群が少なく統御群(ドナー・プール)が多い
- 「推計標本 estimation sample」(=イベント発生前)の期間  $T$  が比較的長い  
(10-30 期間など)

推計期間での予測誤差が大きいとイベント後の予測も精度が低いので、使えない

**短所** どの程度の誤差なら良いのか現在は基準がない、共変数選定の基準がない、イベント前データが長期に必要

**長所** 比較対象の選択基準を客観化、プラセボ分析が可能

▶ Go To Travel

# 条件付き直交の意味

条件付き直交の意味

治療群と統御群は観察可能な変数のみを基準に選抜された

## 条件付き直交の意味

治療群と統御群は観察可能な変数のみを基準に選抜された

「テロリストが(カタルーニャやマドリよりも)バスクを選んだのは成長率が下がって不満を持った市民や支持を得られそうと思った(=観察不可能な思い込み)、からではない」

## 条件付き直交の意味

治療群と統御群は観察可能な変数のみを基準に選抜された

「テロリストが(カタルーニャやマドリよりも)バスクを選んだのは成長率が下がって不満を持った市民や支持を得られそうと思った(=観察不可能な思い込み)、からではない」

「成長を低めテロを育てるような要因(例: 分離独立運動)はない」

## 条件付き直交の意味

治療群と統御群は観察可能な変数のみを基準に選抜された

「テロリストが(カタルーニャやマドリよりも)バスクを選んだのは成長率が下がって不満を持った市民や支持を得られそうと思った(=観察不可能な思い込み)、からではない」

「成長を低めテロを育てるような要因(例: 分離独立運動)はない」

「共変数  $x$  を考慮すれば、テロがなければ成長はほぼ同じだった」

## 条件付き直交の意味

治療群と統御群は観察可能な変数のみを基準に選抜された

「テロリストが(カタルーニャやマドリよりも)バスクを選んだのは成長率が下がって不満を持った市民や支持を得られそうと思った(=観察不可能な思い込み)、からではない」

「成長を低めテロを育てるような要因(例: 分離独立運動)はない」

「共変数  $x$  を考慮すれば、テロがなければ成長はほぼ同じだった」

「テロが起こるとすればバスクしかなかったが、テロの原因と成長とは関係がない」

# SCMの考え方: 共変数が似ていれば結果も似ているはず

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

叙述的研究と比べた長所

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

叙述的研究と比べた長所

- 比較対象を選ぶ基準がある程度客観的

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

叙述的研究と比べた長所

- 比較対象を選ぶ基準がある程度客観的
  - ☞ ただし、共変数を変えるとウェイトも変わるのである程度操作可能

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

叙述的研究と比べた長所

- 比較対象を選ぶ基準がある程度客観的
  - ☞ ただし、共変数を変えるとウェイトも変わるのである程度操作可能
  - ☞ よって、理論に照らし合わせて共変数を選ばねば (=共変数選択を理論によって制約しなければ) ならない=理論に即した選択で推計結果の信頼性を高められる

SCMの考え方: 共変数が似ていれば結果も似ているはず

暗に仮定している共変数と結果の関係

- 横断面に均質
- 安定的 (=時系列に均質)

強い仮定に思えるが、これらは叙述的研究も同じ

叙述的研究と比べた長所

- 比較対象を選ぶ基準がある程度客観的
  - ☞ ただし、共変数を変えるとウェイトも変わるのである程度操作可能
  - ☞ よって、理論に照らし合わせて共変数を選ばねば (=共変数選択を理論によって制約しなければ) ならない=理論に即した選択で推計結果の信頼性を高められる
- 効果 (=比較対象との差) を数値化して、その値がゼロに等しいか仮説検定が可能

## ?: Tobacco tax in California and per capita cigarette sales

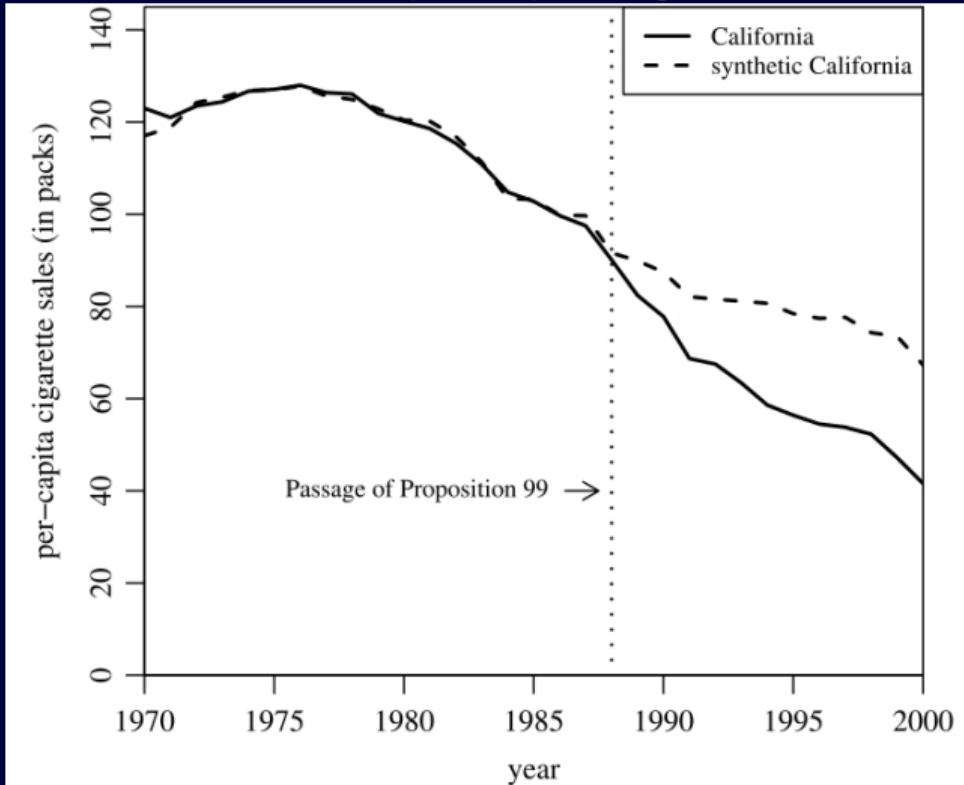


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

?: Tobacco tax in California, placebo studies on all control states, lines show gap = actual - synthetic control

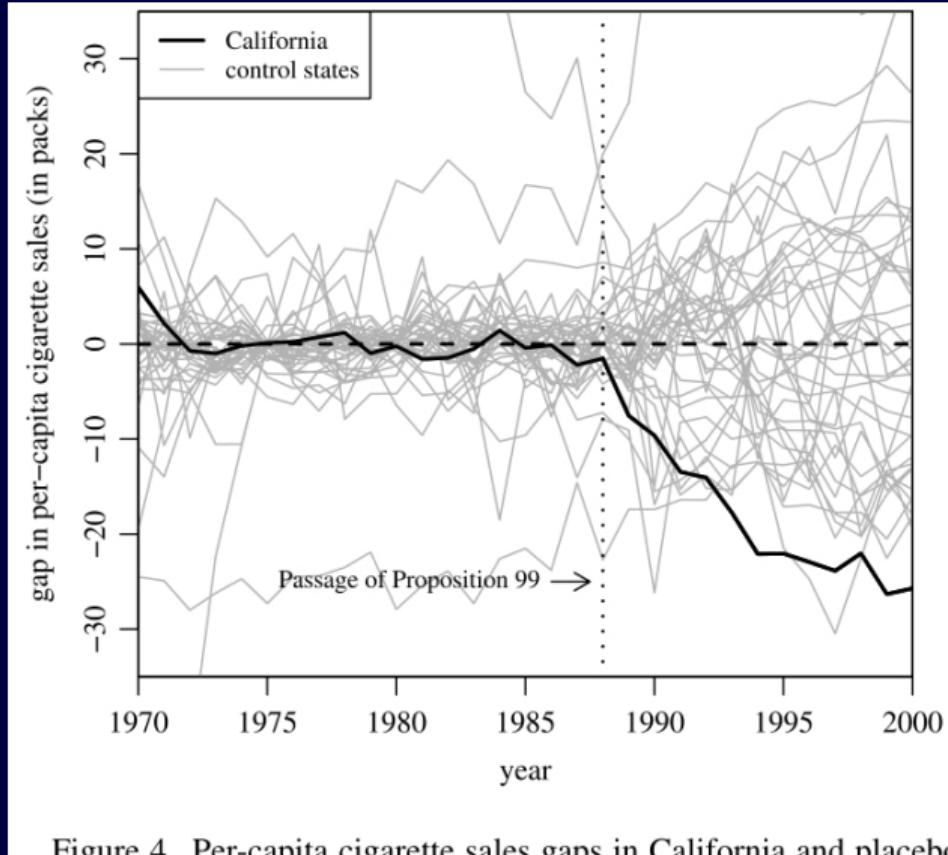


Figure 4. Per-capita cigarette sales gaps in California and placebo

? : Madrid bombings on incumbent votes.

2004 年 3 月 11 日：爆破事件

2004 年 3 月 14 日：投票日

? : Madrid bombings on incumbent votes.

2004年3月11日：爆破事件

2004年3月14日：投票日

治療群 通常の投票者

? : Madrid bombings on incumbent votes.

2004年3月11日：爆破事件

2004年3月14日：投票日

治療群 通常の投票者

統御群 在外不在者、期日（マドリドの爆破事件）前投票

? : Madrid bombings on incumbent votes.

2004年3月11日：爆破事件

2004年3月14日：投票日

治療群 通常の投票者

統御群 在外不在者、期日（マドリドの爆破事件）前投票

2つの方法で効果を推計

- ① DID
- ② SCM

? : Madrid bombings on incumbent votes.

2004年3月11日：爆破事件

2004年3月14日：投票日

治療群 通常の投票者

統御群 在外不在者、期日（マドリドの爆破事件）前投票

2つの方法で効果を推計

① DID

② SCM

☞ 平均的通常投票者の CF = 在外居住者の加重平均、は適切か？

? : Madrid bombings on incumbent votes.

2004年3月11日：爆破事件

2004年3月14日：投票日

治療群 通常の投票者

統御群 在外不在者、期日(マドリドの爆破事件)前投票

2つの方法で効果を推計

① DID

② SCM

☞ 平均的通常投票者の CF=在外居住者の加重平均、は適切か?

52 provinces  $\times$  5 national elections (1989, 1993, 1996, 2000, 2004)  $\times$  2 residency statuses (domestic/abroad)=520

Year 2004	0.80 [0.04] *** (0.03) *** {0.06} ***	
Year 2000	0.68 [0.04] *** (0.03) *** {0.05} ***	0.69 [0.05] *** (0.03) *** {0.07} ***
Year 1996	0.28 [0.02] *** (0.04) *** {0.05} ***	0.28 [0.02] *** (0.04) *** {0.06} **
Year 1993	0.17 [0.01] *** (0.03) *** {0.05} **	0.17 [0.01] *** (0.03) *** {0.06} *
Resident	0.57 [0.04] *** (0.02) *** {0.03} ***	0.58 [0.04] *** (0.03) *** {0.05} ***
Resident × Year 2004	-0.61 [0.04] *** (0.02) *** {0.08} ***	
Resident × Year 2000		-0.01 [0.05] (0.03) {0.09}
Constant	0.24 [0.03] *** (0.03) *** {0.04} ***	0.24 [0.03] *** (0.04) *** {0.04} ***
$R^2$	0.46	0.51
$H_0$ : ratio < 1   $\delta = 0$	$p = 0.00$	
N	520	416

In brackets, pooled OLS clustered (at the province level) robust standard error, and in parentheses, pooled OLS clustered (at the group/time level) robust standard error. In braces, Donald and Lang's (2007) estimator of the standard error. The coefficient estimated is the same for all three procedures.

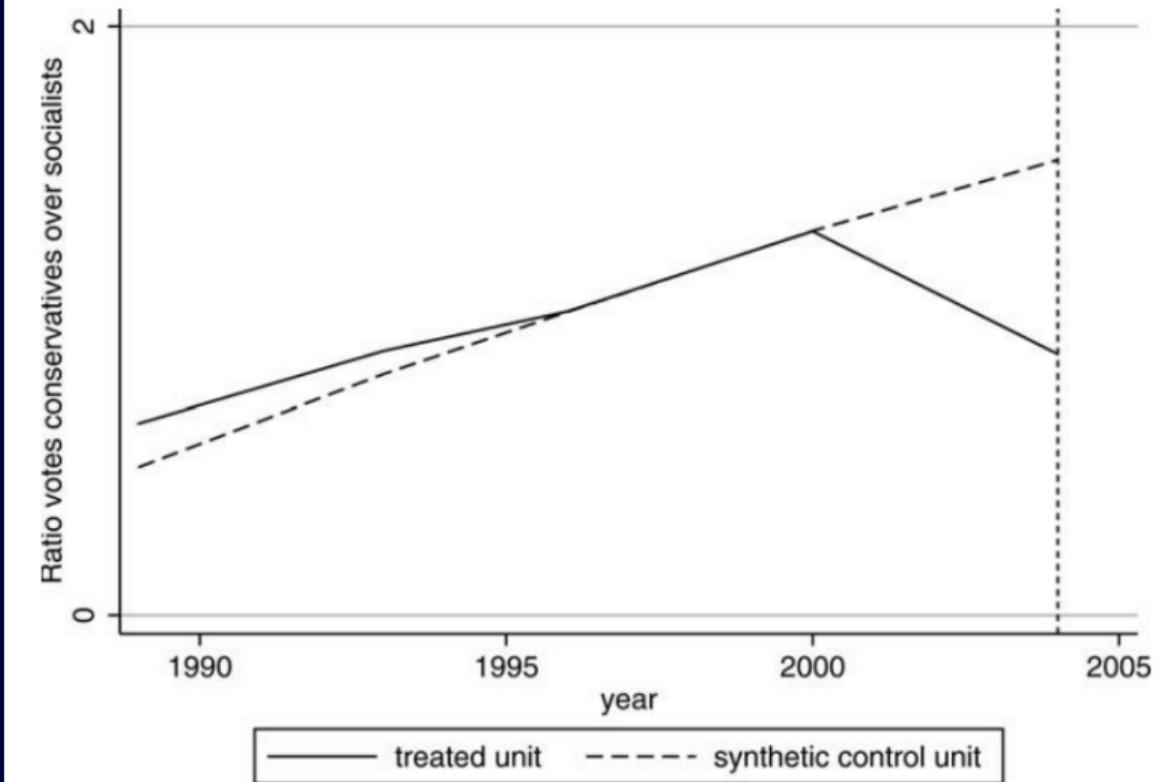
\*\*\*Significant at 1%. \*\*Significant at 5%. \*Significant at 10%.

Resident × 2004 = 治療群  
2004 = 統御群

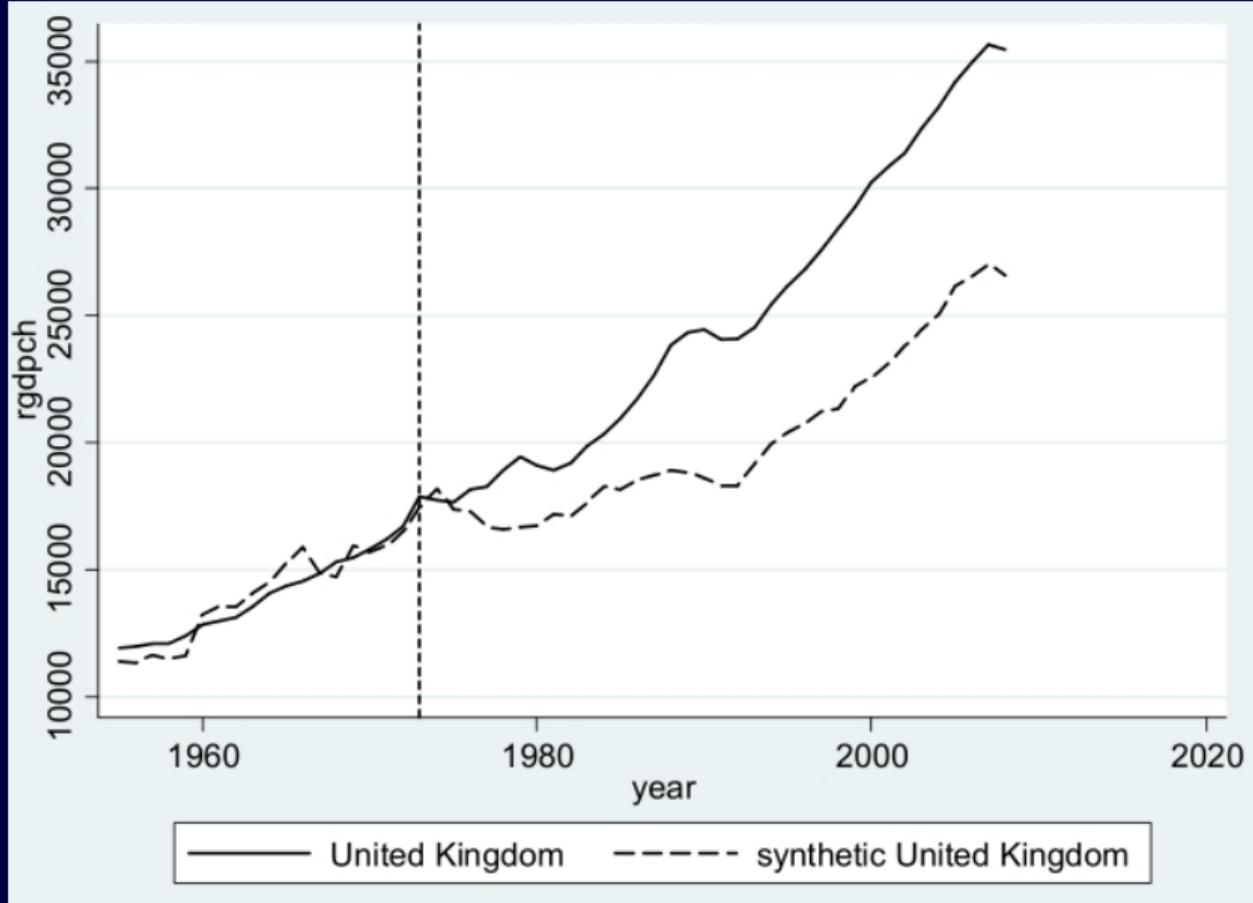
1993, 1996, 2000 も Resident との交差項=イベント前居住者の投票傾向、を作り、すべてがゼロ (=非居住者と変わらない) という検定をすべき  
テロ前からの傾向という解釈を否定できない

## ?: Madrid bombings on incumbent votes (SCM).

FIGURE 3.—TRENDS IN THE RATIO OF CONSERVATIVE TO SOCIALIST VOTES: TREATED AGGREGATE GROUP VERSUS SYNTHETIC CONTROL GROUP



# ?: UK の EU 加盟効果



サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後

EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない

SCM の危うさ: (DID も同じ欠点あり)

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

- イベント後の何かが結果に影響した可能性を排除できない

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

- イベント後の何かが結果に影響した可能性を排除できない
  - ☞ イベント 1 の SCM、イベント 2 の SCM: 2 個の placebo UK を使い 2 イベントの効果推計可能、ただし、イベント間期間が短いとイベント 2 ウェイト推計値が不正確になる

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

- イベント後の何かが結果に影響した可能性を排除できない
  - ☞ イベント 1 の SCM、イベント 2 の SCM: 2 個の placebo UK を使い 2 イベントの効果推計可能、ただし、イベント間期間が短いとイベント 2 ウェイト推計値が不正確になる
- イベントが donor pool に影響を与えていると、比較対象として不適切

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

- イベント後の何かが結果に影響した可能性を排除できない
  - ☞ イベント 1 の SCM、イベント 2 の SCM: 2 個の placebo UK を使い 2 イベントの効果推計可能、ただし、イベント間期間が短いとイベント 2 ウェイト推計値が不正確になる
- イベントが donor pool に影響を与えていると、比較対象として不適切 (イギリスの EU 加盟がその他の国の成長を変えた)

サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

- イベント後の何かが結果に影響した可能性を排除できない
  - ☞ イベント 1 の SCM、イベント 2 の SCM: 2 個の placebo UK を使い 2 イベントの効果推計可能、ただし、イベント間期間が短いとイベント 2 ウェイト推計値が不正確になる
- イベントが donor pool に影響を与えていると、比較対象として不適切 (イギリスの EU 加盟がその他の国の成長を変えた)

SUTVA (stable unit treatment value assumption) とある標本への治療が他の標本 (とくに統御群) の結果に影響しない=外部性なし、効果均質

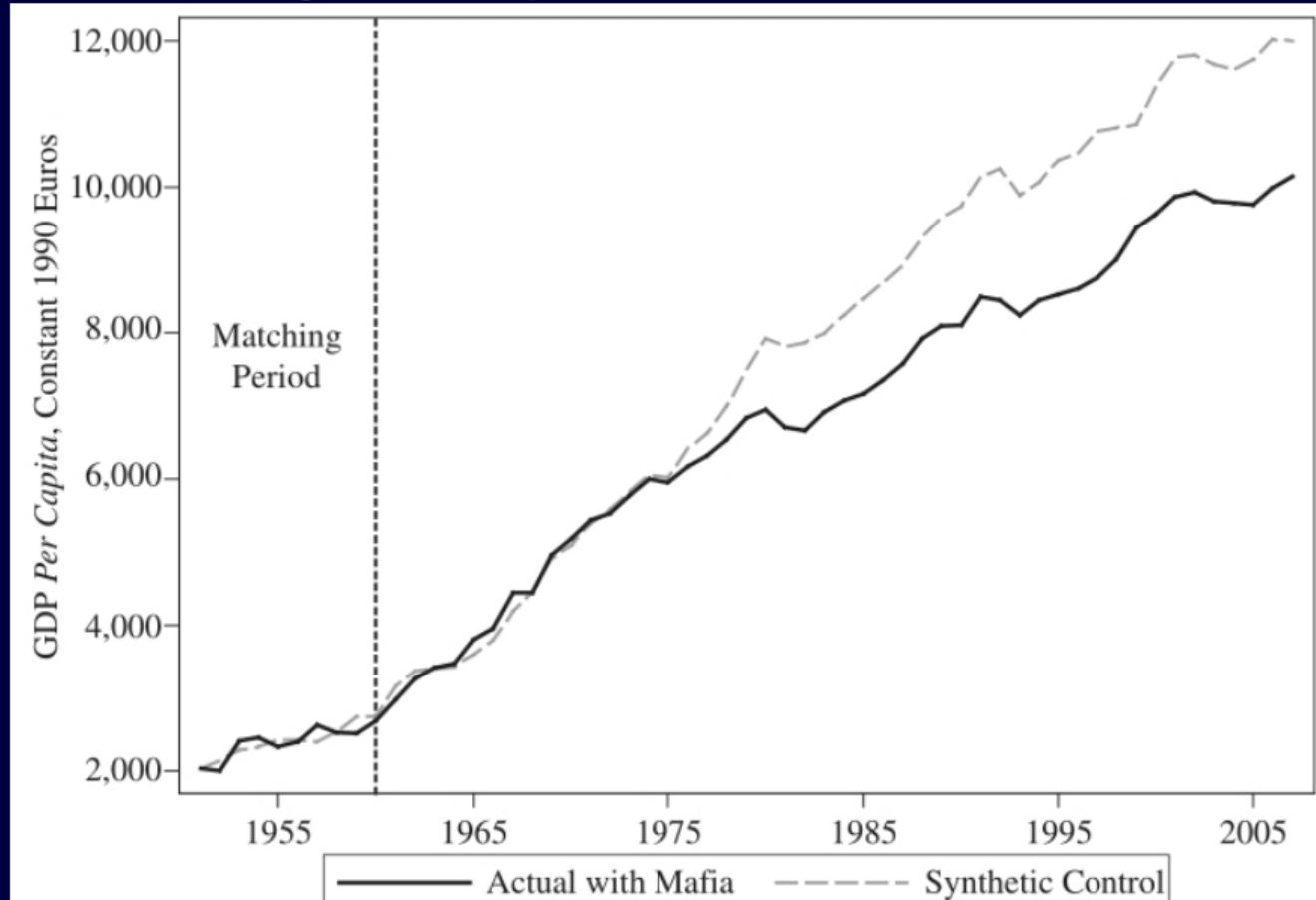
サッチャー首相 (Mrs. Thatcher) の改革: 1980 年代初め。痛みの伴う政策だったが、多くの人が成長の土台となったと議論。

サッチャー改革=EU 加盟後  
EU 加盟なのか、サッチャー改革なのか、タイミングを見ても判定できない  
SCM の危うさ: (DID も同じ欠点あり)

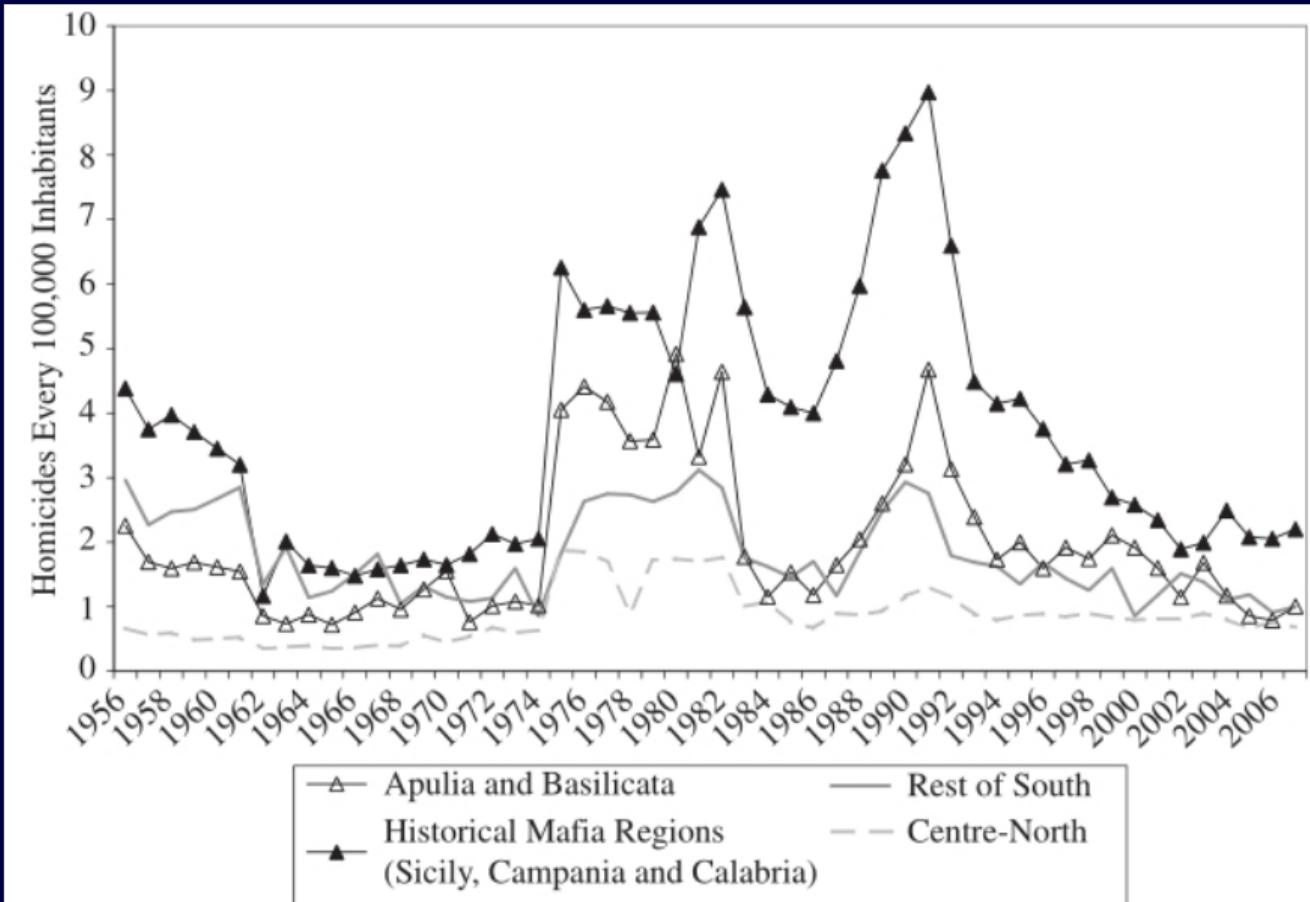
- イベント後の何かが結果に影響した可能性を排除できない
  - ☞ イベント 1 の SCM、イベント 2 の SCM: 2 個の placebo UK を使い 2 イベントの効果推計可能、ただし、イベント間期間が短いとイベント 2 ウェイト推計値が不正確になる
- イベントが donor pool に影響を与えていると、比較対象として不適切 (イギリスの EU 加盟がその他の国の成長を変えた)

SUTVA (stable unit treatment value assumption) とある標本への治療が他の標本 (とくに統御群) の結果に影響しない=外部性なし、効果均質  
数少ない例外を除き、RCT も含めた全てのインパクト評価で前提にする仮定

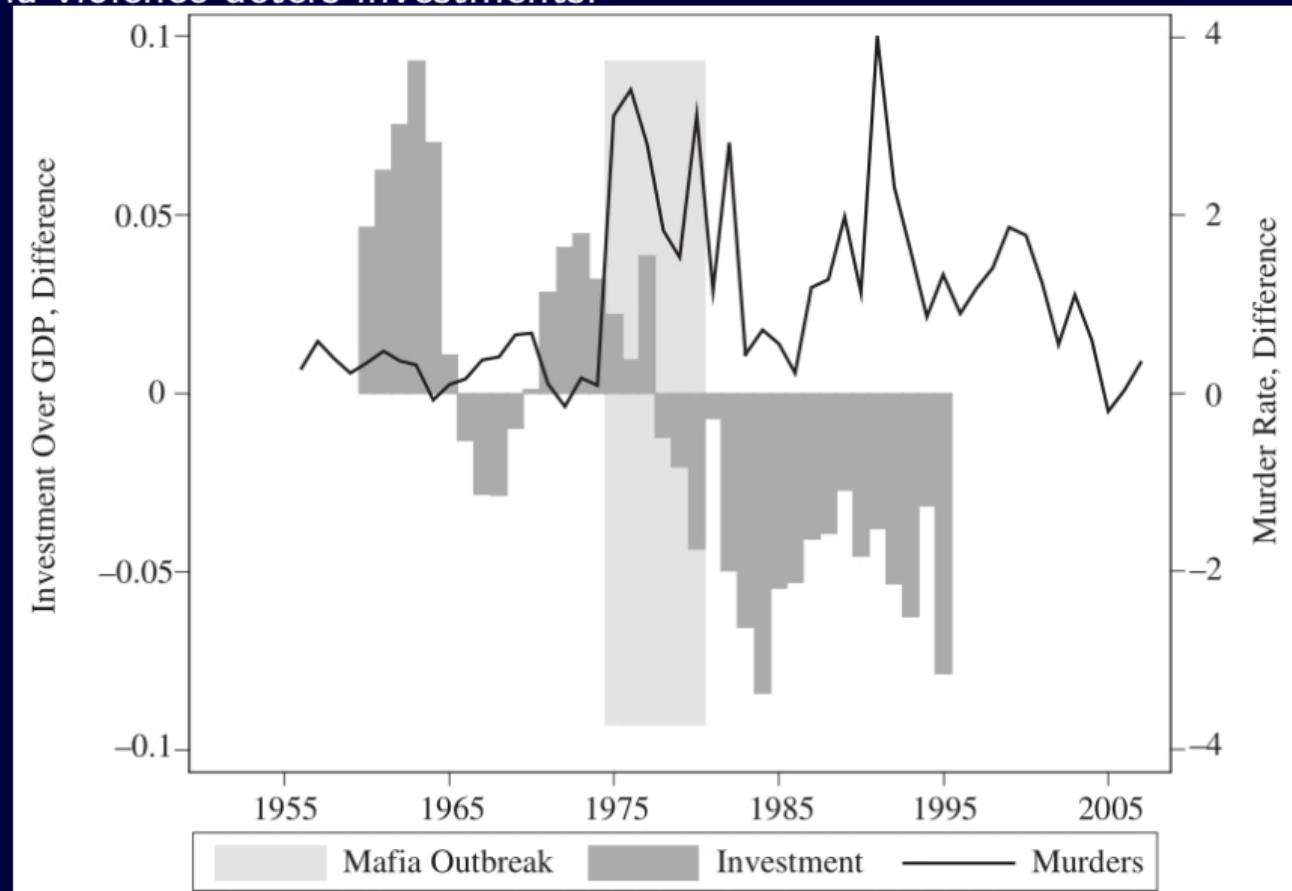
## ?: Mafia violence on growth in Apulia and Basilicata.



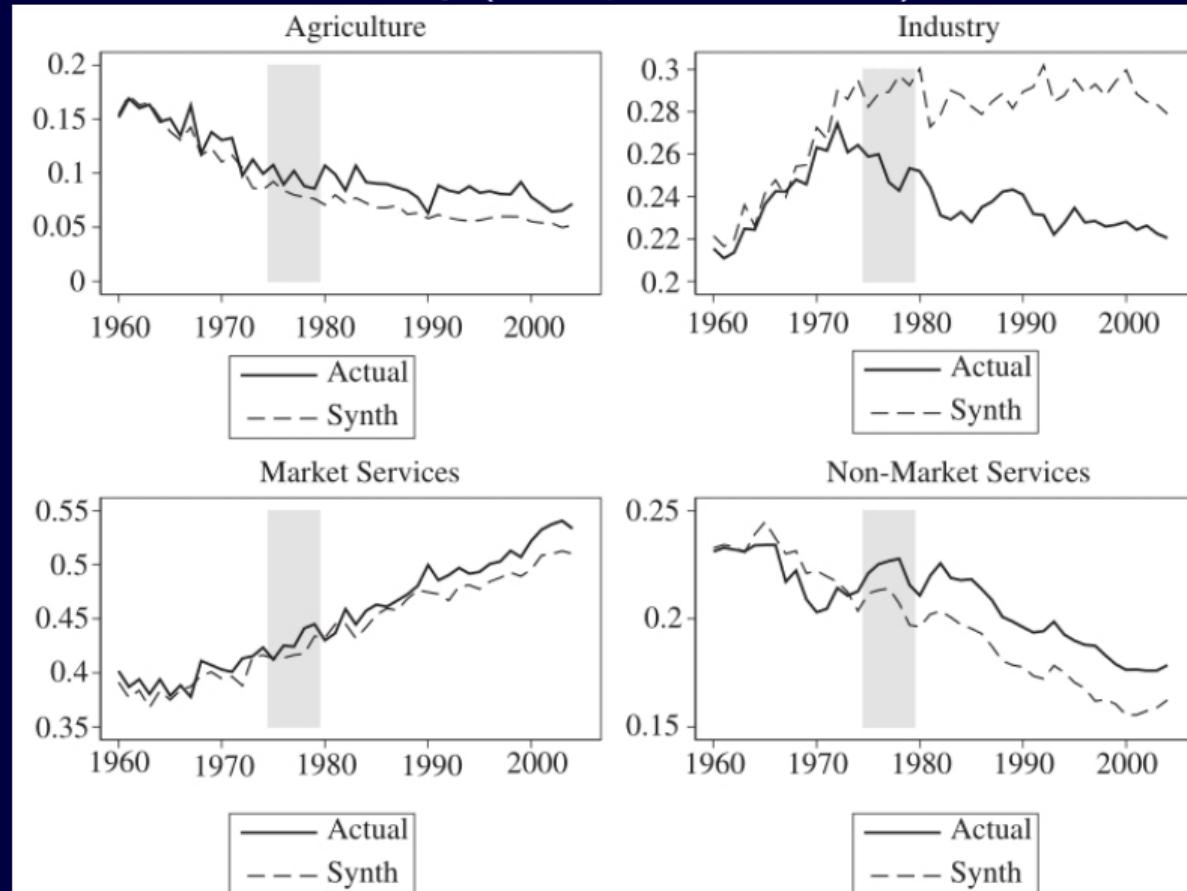
? Apulia and Basilicata were new to violence.



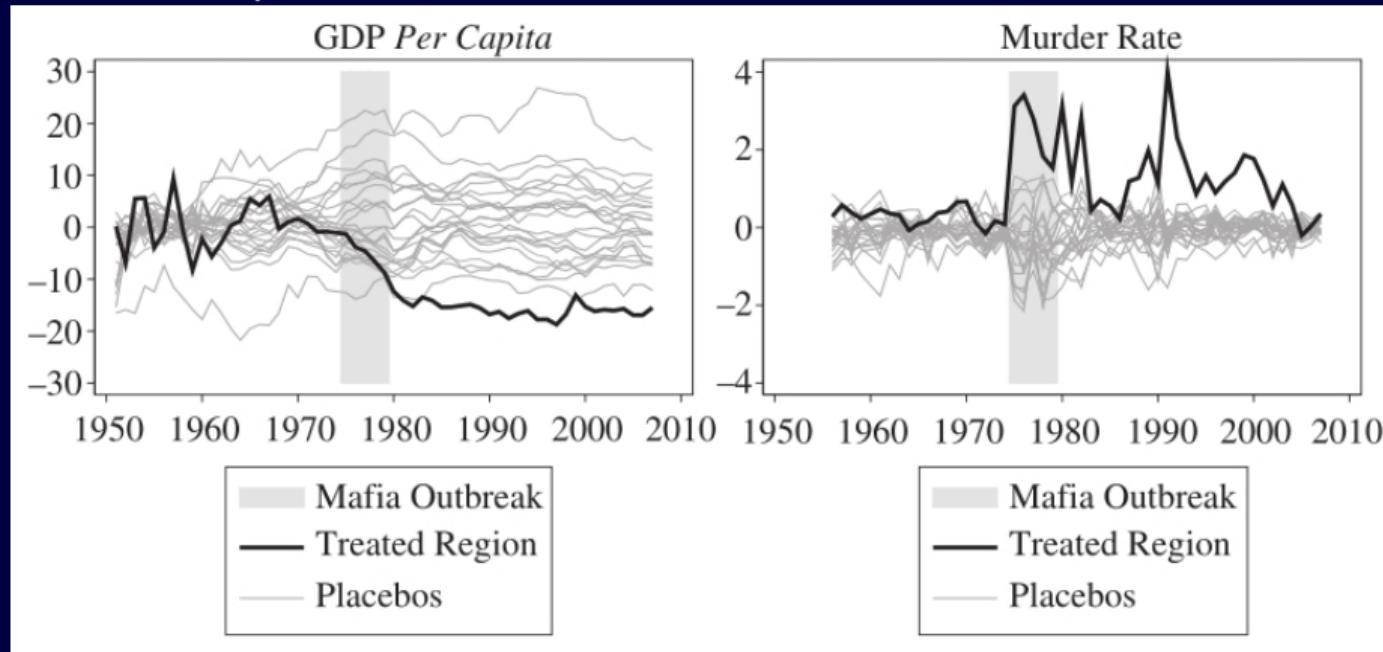
# ?: Mafia violence deters investments.



# ?: Mafia violence slows industry (% of provincial GDP).



# ?: Mafia violence placebo studies.



It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.

It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.
- Public investment increased.

It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.
- Public investment increased.
- Private employment reduced.

It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.
- Public investment increased.
- Private employment reduced.
- Public employment increased but did not offset the reduction in private employment.

It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.
- Public investment increased.
- Private employment reduced.
- Public employment increased but did not offset the reduction in private employment.

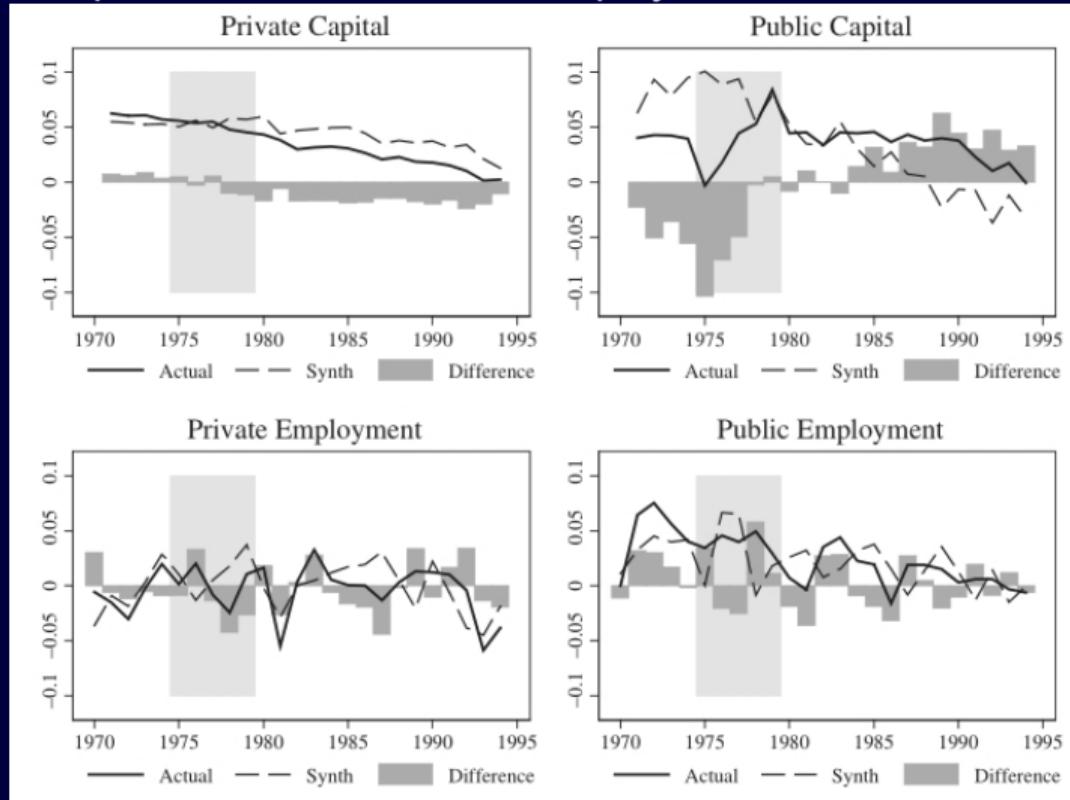
It is still possible that something other than mafia violence caused the slowdown of industrial growth of Apulia and Basilicata, starting from early 1970's. But what is it? If none is found, SCM (may not be very credible but) is convincing.

? also shows that;

- Private investment dried up.
- Public investment increased.
- Private employment reduced.
- Public employment increased but did not offset the reduction in private employment.

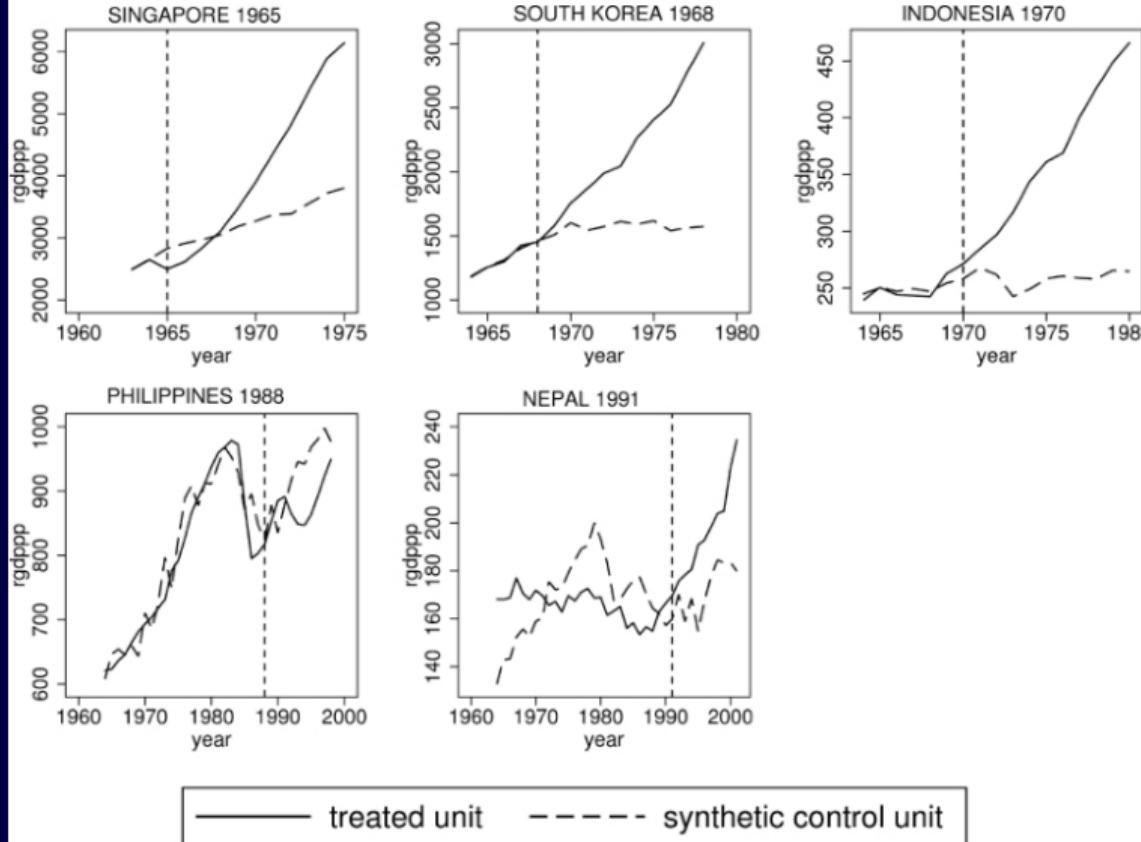
He argues that increased public investments were channeled to mafia activities through corrupt politicians, and quality of elected politicians reduced (shown in the companion paper).

# ?: Private and public investment and employment.



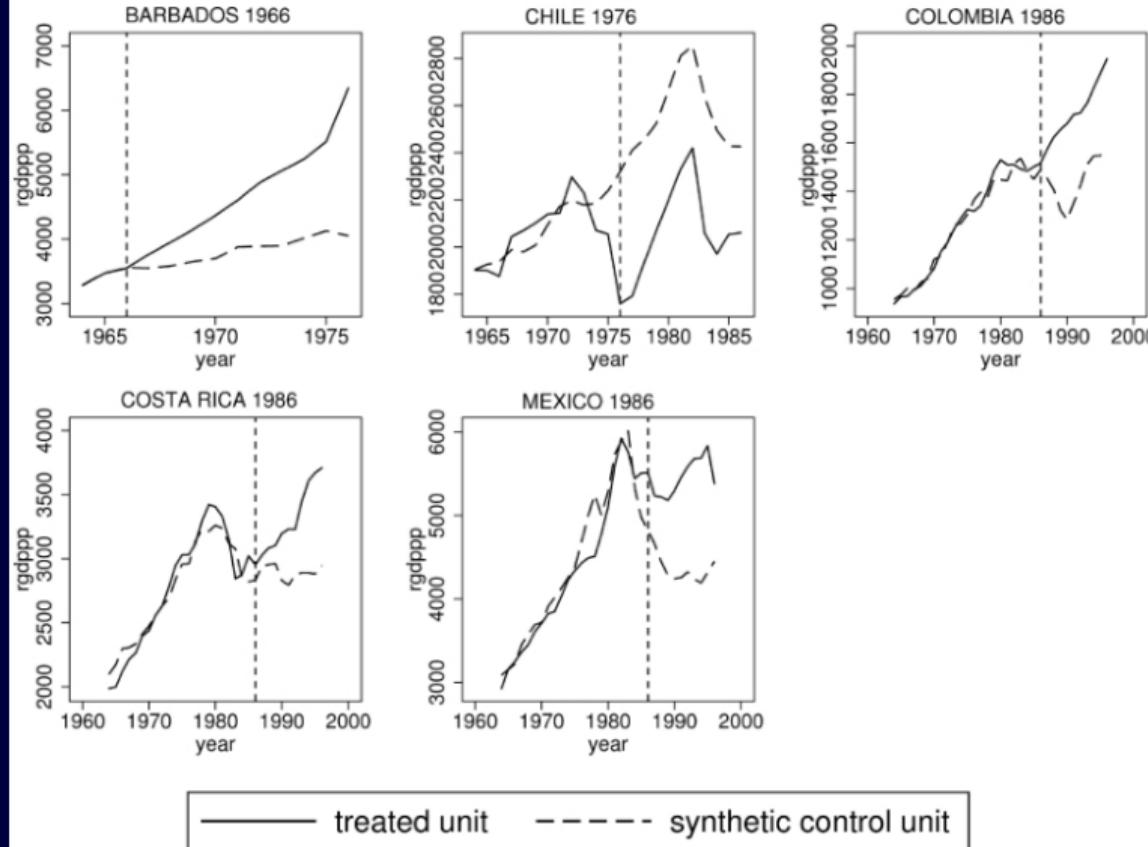
# ?: 貿易自由化の効果、Asia

FIGURE 1.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—ASIA



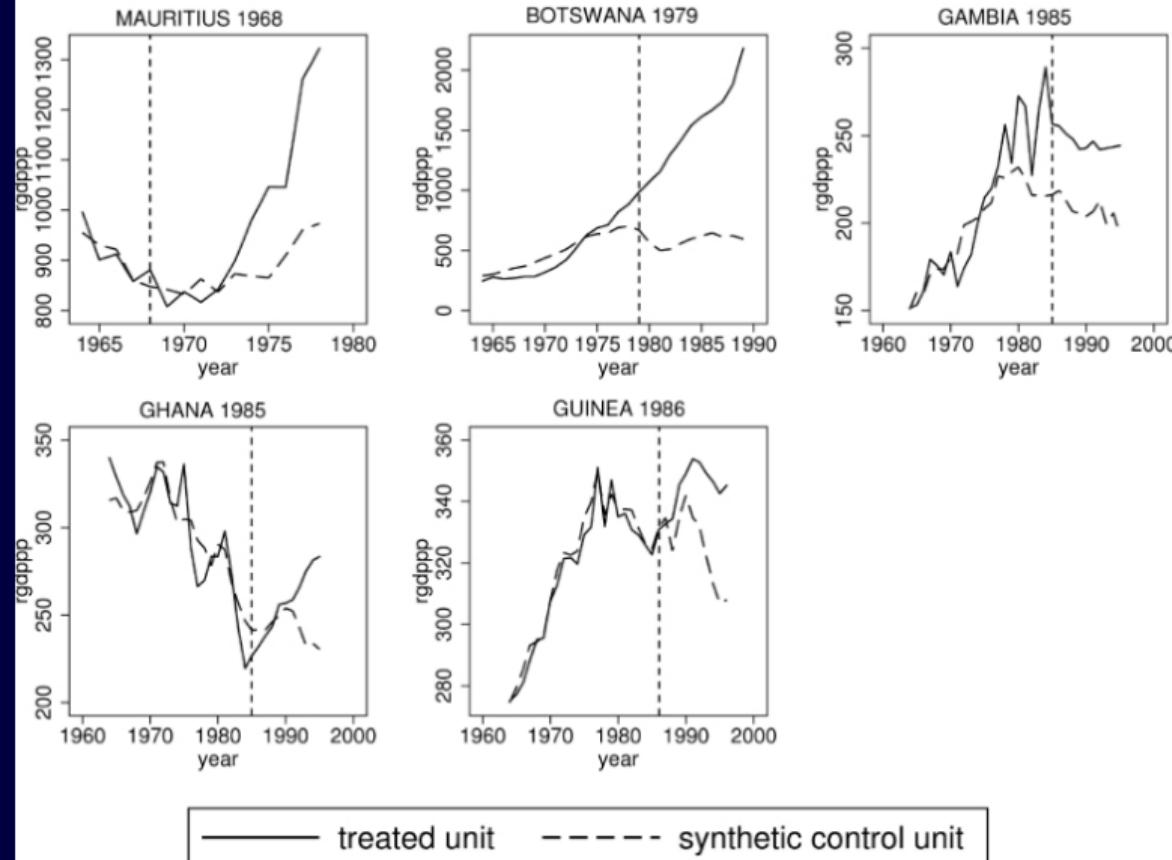
# ? Latin America

FIGURE 2.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—LATIN AMERICA



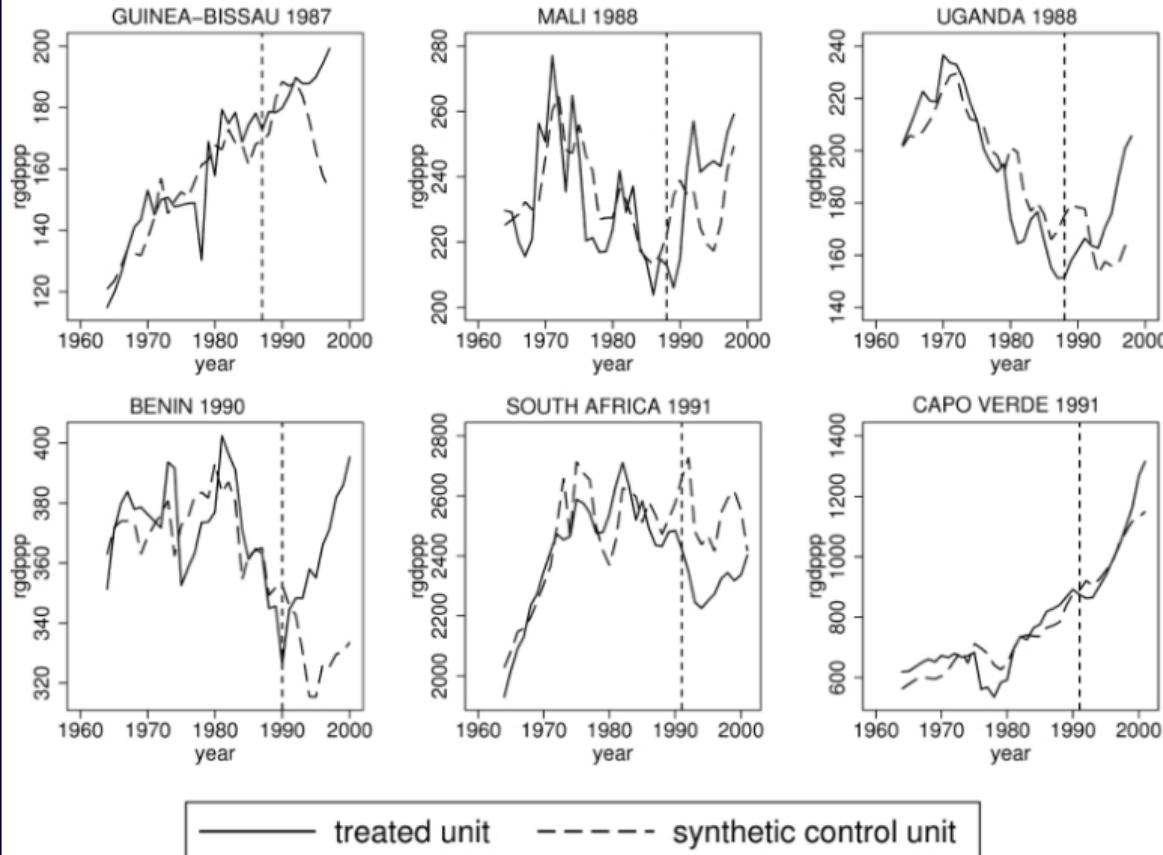
# ?: Africa before 1987

FIGURE 3.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—AFRICA BEFORE 1987



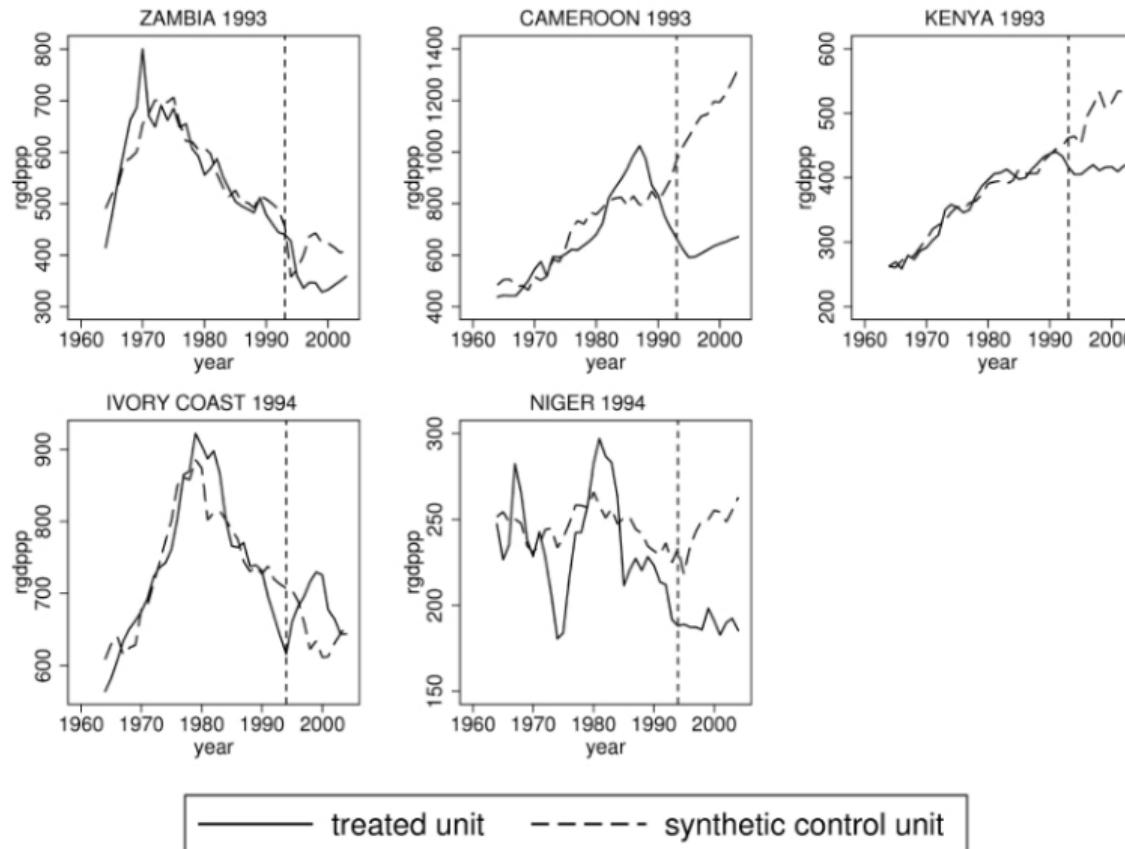
# ?: Africa 1987-91

FIGURE 4.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—AFRICA BETWEEN 1987 AND 1991



# ?: Africa after 1991

FIGURE 5.—GDP TRENDS, TREATED COUNTRY VERSUS SYNTHETIC CONTROL—AFRICA AFTER 1991





? : 怪しい結果に思える。貿易自由化していない donor pool を使って加重平均を合成

共変数: 中等就学率、人口増加率、投資/GDP、インフレ率、民主主義インデックス。

? : 怪しい結果に思える。貿易自由化していない donor pool を使って加重平均を合成

共変数: 中等就学率、人口増加率、投資/GDP、インフレ率、民主主義インデックス。

共変数が似ていて貿易自由化していなくても、成長率は人口規模、天然資源、気候変動、紛争などによっても影響される。

? : 怪しい結果に思える。貿易自由化していない donor pool を使って加重平均を合成

共変数: 中等就学率、人口増加率、投資/GDP、インフレ率、民主主義インデックス。

共変数が似ていて貿易自由化していなくても、成長率は人口規模、天然資源、気候変動、紛争などによっても影響される。

どこまで共変数として含めれば良いのか、誰も説得的な議論はできない

? : 怪しい結果に思える。貿易自由化していない donor pool を使って加重平均を合成

共変数: 中等就学率、人口増加率、投資/GDP、インフレ率、民主主義インデックス。

共変数が似ていて貿易自由化していなくても、成長率は人口規模、天然資源、気候変動、紛争などによっても影響される。

どこまで共変数として含めれば良いのか、誰も説得的な議論はできない

しかし、叙述的研究はそういうことをやっている

選定基準を明確化して数値化していることは貢献

Yet another motivating example.

Yet another motivating example.

About 15 years ago, I attended an evaluation workshop at DAC of OECD.

Yet another motivating example.

About 15 years ago, I attended an evaluation workshop at DAC of OECD.

- Impact evaluation.

Yet another motivating example.

About 15 years ago, I attended an evaluation workshop at DAC of OECD.

- Impact evaluation. Program evaluation. Treatment effect estimation.

Yet another motivating example.

About 15 years ago, I attended an evaluation workshop at DAC of OECD.

- Impact evaluation. Program evaluation. Treatment effect estimation.
- Project evaluation.

▶ DAC 5

- Almost all attendants looked perplexed, because they were asked to do “impact evaluation.”
- Majority of attendants were evaluation officers from donor agencies of OECD.
- Evaluation officers perform “project evaluation.”
- “Project evaluation” is qualitative in nature and usually employs before-after or with-without comparison.

Then a gentleman took the floor and explained what the impact evaluation is.

I was a bit bewildered, because he was hardly an expert on the topic and he had been doing “project evaluation” for a long time.

“Project evaluation” that these people did has its place in policy implementation. But it was not designed to evaluate the policy’s contribution on outcome measures. It describes goals and measures of projects, and also suggests the impacts using with-without or before-after comparisons in the absence of randomisation. Such estimation usually has a fundamental flaw as we see later.

It is fair to say that *ex post* evaluation uses a poor method in estimating impacts which should not be taken too seriously.

*We believe the policy had an impact, however, it is difficult to know. (honest!)*

*Based on the responses from beneficiaries, it is confirmed ... was improved.*  
(subjective)

*...one can judge that the effect was an increase in income. (unsubstantiated)*

*...it is perceivable that the policy provided a support for growth. (unsubstantiated)*

It is fair to say that *ex post* evaluation uses a poor method in estimating impacts which should not be taken too seriously.

*We believe the policy had an impact, however, it is difficult to know. (honest!)*

*Based on the responses from beneficiaries, it is confirmed ... was improved.*  
(subjective)

*...one can judge that the effect was an increase in income. (unsubstantiated)*

*...it is perceivable that the policy provided a support for growth. (unsubstantiated)*

We cannot learn which policy was effective.



CGD document: *When will we ever learn? (?)*

CGD document: *When will we ever learn? (?)*

Asking, do we know how to improve outcomes of the poor? After more than 50 years of aid business?

CGD document: *When will we ever learn? (?)*

Asking, do we know how to improve outcomes of the poor? After more than 50 years of aid business?

Donor agencies' evaluation professionals did not have a good answer.

CGD document: *When will we ever learn? (?)*

Asking, do we know how to improve outcomes of the poor? After more than 50 years of aid business?

Donor agencies' evaluation professionals did not have a good answer.

We need to produce knowledge from past policies for future policies.

## Millenium Villages (2004-2015)

ビッグプッシュ実験。サブサハラ・アフリカの14箇所(80ヵ村)で農業、保健、教育、インフラ、生産への投資。一人当たり1年120ドル×50万人×12=総額72億ドル(8640億円)。コロンビア大学地球研究所ジェフリー・サックス教授主導、国連、世銀、各国政府が支援、民間企業も協賛。PRビデオ(MTV)

[https://www.youtube.com/watch?v=uUHf\\_k0UM74](https://www.youtube.com/watch?v=uUHf_k0UM74)

## Millenium Villages (2004-2015)

ビッグプッシュ実験。サブサハラ・アフリカの14箇所(80ヵ村)で農業、保健、教育、インフラ、生産への投資。一人当たり1年120ドル×50万人×12=総額72億ドル(8640億円)。コロンビア大学地球研究所ジェフリー・サックス教授主導、国連、世銀、各国政府が支援、民間企業も協賛。PRビデオ(MTV)

[https://www.youtube.com/watch?v=uUHf\\_kOUM74](https://www.youtube.com/watch?v=uUHf_kOUM74)

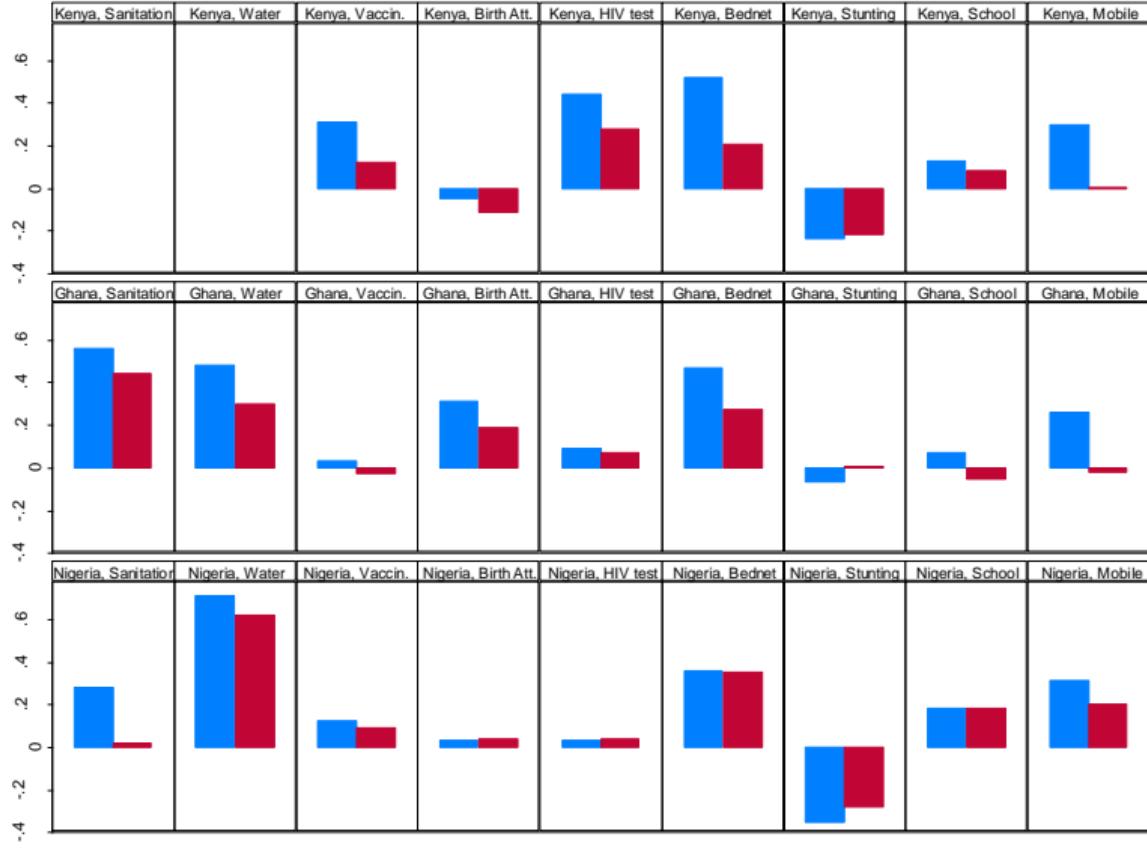
RCTではないこと、統御群データがないことなどから、実施したインパクト評価(before-after)の信頼性credibilityは低く、成果がよく分からぬまま終わった。壮大な実験であり得たのに、評価の準備をしていなかった。

# ?: Before-after 比較

**Box 2: The Millennium Villages Project at a glance**

	Ethiopia	Ghana	Malawi	Uganda
Agriculture Production increases from pre-MVP levels	Cereal crops by 122% (0.9 t/ha to 2.0 t/ha)	Maize by 85% (2.2 t/ha to 4.11 t/ha)	Maize by 350% (0.8 t/ha to 3.6 t/ha)	Maize by 108% (1.9 t/ha to 3.9 t/ha).
Distribution of long-lasting insecticide-treated bed-nets cover 100% of MV sites.				
Malaria	Almost 50% reduction in malaria cases since inception.	N/A	N/A	79% reduction in malaria since inception.
Maternal, infant and child health	Proportion of deliveries attended by professionals increased from 35% to 51% between 2006 and 2007. 89% of under-5 children are fully immunised.	Between 2006 and 2007, number of women giving birth in health facilities increased by 146%; number of additional women seeking antenatal care increased by 129%.	45 community health workers provide child health services and antenatal care.	Proportion of deliveries attended by professionals increased from 8% to 70%. 80% of pregnant mothers access antenatal care.
Health services	De-worming campaign reached 46,435 residents. Utilisation of Koraro health care facilities increased by 528%.	Number of residents using modern family planning techniques increased by 670% (from 296 to 2,278) between 2007 and 2008.	Since introduction of HIV and TB screening, more than 550 patients started AIDS treatment.	Utilisation rates of cluster health care facilities increased by 230% between 2006 and 2008.

Source: MVP, 2008



?: DID と before-after を比較  
 B-A 比較が意外に DID と類似...B-A 比較が良いのか? DID もダメか? B-A は DID よりも効果推計値が上振れ

なぜ信頼性の高いインパクト評価が可能な準備をしなかったのか?

A short answer: インパクト評価に関する勘違い(もしくはその振りをした)

A short answer: インパクト評価に関する勘違い(もしくはその振りをした)

Sachs らによる説明と David MacKenzie (世銀) による反論

(<https://blogs.worldbank.org/impactevaluations/jeff-sachs-the-millennium-villages-project-and-misconceptions-about-impact-evaluation>)

- ① 村は常に変化しているから、村レベルでの実験に統御群は作れない。

- ① 村は常に変化しているから、村レベルでの実験に統御群は作れない。
- ② 介入は村が自ら学んで改善するという過程を経る。この過程は RCT で評価できない。

- ① 村は常に変化しているから、村レベルでの実験に統御群は作れない。
  - ⇒ 村が常に変化しているからこそ、実際の村が統御群に必要になる。
- ② 介入は村が自ら学んで改善するという過程を経る。この過程は RCT で評価できない。

- ① 村は常に変化しているから、村レベルでの実験に統御群は作れない。
  - ⇒ 村が常に変化しているからこそ、実際の村が統御群に必要になる。
- ② 介入は村が自ら学んで改善するという過程を経る。この過程は RCT で評価できない。
  - ⇒ (率直に何を言っているのか理解できないほど) 勘違いしている。

- ① 村は常に変化しているから、村レベルでの実験に統御群は作れない。
  - ☞ 村が常に変化しているからこそ、実際の村が統御群に必要になる。
- ② 介入は村が自ら学んで改善するという過程を経る。この過程は RCT で評価できない。
  - ☞ (率直に何を言っているのか理解できないほど) 勘違いしている。
  - ☞ 学んで改善を記録するとはプロセス評価 process evaluation のことで、必要性を誰も否定していない。学んで改善してもインパクトはまだ分からぬから、インパクト評価は必要。プロセス評価とインパクト評価を混同。

- ③ 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。

- ③ 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。
- ➡ 最初に候補を作って治療群と統御群に割り振れば良いだけ(忙しいのは事実だけど)

③

- 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。
- 最初に候補を作って治療群と統御群に割り振れば良いだけ(忙しいのは事実だけど)
  - 既存の全国データには含まれないような情報を治療群では集めているので既存データでは不十分

③

- 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。
- 最初に候補を作って治療群と統御群に割り振れば良いだけ(忙しいのは事実だけど)
  - 既存の全国データには含まれないような情報を治療群では集めているので既存データでは不十分
  - 変わるからこそ統御群が必要

- ③ 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。
- 最初に候補を作って治療群と統御群に割り振れば良いだけ(忙しいのは事実だけど)
  - 既存の全国データには含まれないような情報を治療群では集めているので既存データでは不十分
  - 変わるからこそ統御群が必要

真の理由?

- 倫理的ではないから統御群を作らない、と 2007 年時点で?が書いていた

- ③ 初年度は忙しくて統御地域を設定できなかった。既存の全国データがあるのに、少数の統御地域だけと比べる意味はあるのか。統御地域とペアにして比較しても、統御地域が経済的に変わったら比較する意味はあるのか。
- ☛ 最初に候補を作って治療群と統御群に割り振れば良いだけ(忙しいのは事実だけど)
  - ☛ 既存の全国データには含まれないような情報を治療群では集めているので既存データでは不十分
  - ☛ 変わるからこそ統御群が必要

真の理由?

- ☛ 倫理的ではないから統御群を作らない、と 2007 年時点で?が書いていた
- ☛ しかし、2008 年には比較村が誕生

# $p$ hacking

全ての推計方法への批判 1: multiple testing ( $p$  hacking)

# p hacking

## 全ての推計方法への批判 1: multiple testing (*p* hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

# $p$ hacking

## 全ての推計方法への批判 1: multiple testing ( $p$ hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。  
 $p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

# p hacking

## 全ての推計方法への批判 1: multiple testing (p hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

$p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

$p = 5\% \Rightarrow$  効果の平均値がゼロのとき、 $n = 100$  回に  $np = 5$  回は発生する結果、ということ

# p hacking

## 全ての推計方法への批判 1: multiple testing (p hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

$p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

$p = 5\% \Rightarrow$  効果の平均値がゼロのとき、 $n = 100$  回に  $np = 5$  回は発生する結果、ということ

少しずつ異なる推計式定式化を 100 回試すと、真の効果がゼロでも 5 回は推計値が外れ値になる

# p hacking

## 全ての推計方法への批判 1: multiple testing (p hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

$p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

$p = 5\% \Rightarrow$  効果の平均値がゼロのとき、 $n = 100$  回に  $np = 5$  回は発生する結果、ということ

少しずつ異なる推計式定式化を 100 回試すと、真の効果がゼロでも 5 回は推計値が外れ値になる

外れ値を得るまで何度も推計式定式化を変えると、望む結果が(理論的には)必ず得られる

# p hacking

## 全ての推計方法への批判 1: multiple testing (p hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

$p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

$p = 5\% \Rightarrow$  効果の平均値がゼロのとき、 $n = 100$  回に  $np = 5$  回は発生する結果、ということ

少しずつ異なる推計式定式化を 100 回試すと、真の効果がゼロでも 5 回は推計値が外れ値になる

外れ値を得るまで何度も推計式定式化を変えると、望む結果が(理論的には)必ず得られる

似た結果指標(疲労と労働時間)への効果推計=multiple testing を無意識にやっている

# p hacking

## 全ての推計方法への批判 1: multiple testing (p hacking)

$p$  値=帰無仮説(効果ゼロ)が正しい確率。効果の平均値がゼロの母集団から  $n$  回標本を集めて検定すると、推計結果は  $p$  の確率 ( $=np$  回) しか発生しない、ということ。

$p$  が小さいほど「外れ値」なので効果ゼロを疑問視して良い

$p = 5\% \Rightarrow$  効果の平均値がゼロのとき、 $n = 100$  回に  $np = 5$  回は発生する結果、ということ

少しずつ異なる推計式定式化を 100 回試すと、真の効果がゼロでも 5 回は推計値が外れ値になる

外れ値を得るまで何度も推計式定式化を変えると、望む結果が(理論的には)必ず得られる

似た結果指標(疲労と労働時間)への効果推計=multiple testing を無意識にやっている

対策: 同じ結果指標グループの  $p$  値平均値 family wise error rate を使う

# small sample bias

全ての推計方法への批判 2: small sample bias

# small sample bias

## 全ての推計方法への批判 2: small sample bias

- ? Jamaica: 早期児童発達介入で成人時所得が 25%増加 (報告当初は 42%だったが訂正), 治療群+統御群=109 人
- ? Perry preschool: 内部収益率 7-10% (研究当初は 15-17%だったが訂正), 治療群+統御群=123
- ? Abecedarian 成人時血圧など改善、治療群+統御群=111→33(35 歳)

# small sample bias

## 全ての推計方法への批判 2: small sample bias

? Jamaica: 早期児童発達介入で成人時所得が 25%増加 (報告当初は 42%だったが訂正), 治療群+統御群=109 人

? Perry preschool: 内部収益率 7-10% (研究当初は 15-17%だったが訂正), 治療群+統御群=123

? Abecedarian 成人時血圧など改善、治療群+統御群=111→33(35 歳)

先入観・常識: そんなに大きいはずがない

<https://statmodeling.stat.columbia.edu/2013/11/05/how-much-do-we-trust-this-claim-that-early-childhood-stimulation-raised-earnings-by-42/>

# small sample bias

## 全ての推計方法への批判 2: small sample bias

? Jamaica: 早期児童発達介入で成人時所得が 25%増加 (報告当初は 42%だったが訂正), 治療群+統御群=109 人

? Perry preschool: 内部収益率 7-10% (研究当初は 15-17%だったが訂正), 治療群+統御群=123

? Abecedarian 成人時血圧など改善、治療群+統御群=111→33(35 歳)

先入観・常識: そんなに大きいはずがない

<https://statmodeling.stat.columbia.edu/2013/11/05/how-much-do-we-trust-this-claim-that-early-childhood-stimulation-raised-earnings-by-42/>

小標本=誤差が大きい=分散が大きい=分布の裾野が広い

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない  
 $p$  hacking をすればいつかは大きな効果を推計できる  
するではなく偶然大きな値になった可能性もある  
placebo test でも  $p$  hacking 可能 (プラシーボの効果がゼロを探索)

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない

$p$  hacking をすればいつかは大きな効果を推計できる

するではなく偶然大きな値になった可能性もある

placebo test でも  $p$  hacking 可能 (プラシーボの効果がゼロを探索)

小標本で大きな効果: 効果ゼロの可能性が高い

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない

$p$  hacking をすればいつかは大きな効果を推計できる

するではなく偶然大きな値になった可能性もある

placebo test でも  $p$  hacking 可能 (プラシーボの効果がゼロを探索)

小標本で大きな効果: 効果ゼロの可能性が高い

小標本の研究は信頼性が低い

よほどのメリットがない限りやらない方が良い

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない  
 $p$  hacking をすればいつかは大きな効果を推計できる  
するではなく偶然大きな値になった可能性もある  
placebo test でも  $p$  hacking 可能 (プラシーボの効果がゼロを探索)  
小標本で大きな効果: 効果ゼロの可能性が高い  
小標本の研究は信頼性が低い  
よほどのメリットがない限りやらない方が良い  
Heckman curve は本当にあの形か? 水平線かも?

▶ Figure

平均ゼロ、大きな分散の分布で  $p$  値 5%: 推計値が大きくないと実現しない

$p$  hacking をすればいつかは大きな効果を推計できる

するではなく偶然大きな値になった可能性もある

placebo test でも  $p$  hacking 可能 (プラシーボの効果がゼロを探索)

小標本で大きな効果: 効果ゼロの可能性が高い

小標本の研究は信頼性が低い

よほどのメリットがない限りやらない方が良い

Heckman curve は本当にあの形か? 水平線かも?

▶ Figure

- ◆ ヘックマンらの研究では認知能力 (IQ) と非認知能力 (中身=executive function、忍耐力、やる気 grit、対人スキル) の両方が所得と関わっており、前者は 3-5 歳くらいまでに決まるが後者はそれ以降も変化する、としている

# まとめ

まとめ

効果推計の議論にはレベルがある:

まとめ

効果推計の議論にはレベルがある：初心者

まとめ

効果推計の議論にはレベルがある：初心者、プロ

まとめ

効果推計の議論にはレベルがある：初心者、プロ、応用

まとめ

効果推計の議論にはレヴェルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

まとめ

効果推計の議論にはレヴェルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

まとめ

効果推計の議論にはレヴェルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

社会科学データで因果関係を直接観察できることは非常に稀

まとめ

効果推計の議論にはレヴェルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

社会科学データで因果関係を直接観察できることは非常に稀

(⇐ 社会経済は相互依存：原因も何かの結果)

## まとめ

効果推計の議論にはレベルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

社会科学データで因果関係を直接観察できることは非常に稀

(←社会経済は相互依存：原因も何かの結果)

多くの「効果」の議論：相関関係を因果関係に解釈している可能性

## まとめ

効果推計の議論にはレベルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

社会科学データで因果関係を直接観察できることは非常に稀

(←社会経済は相互依存：原因も何かの結果)

多くの「効果」の議論：相関関係を因果関係に解釈している可能性

(←因果関係と解釈した方が進化で生存可能性が高かった（説）)

## まとめ

効果推計の議論にはレベルがある：初心者、プロ、応用

相関関係：3通りの因果関係、見せかけの相関

因果関係：方向性のある関係

社会科学データで因果関係を直接観察できることは非常に稀

(←社会経済は相互依存：原因も何かの結果)

多くの「効果」の議論：相関関係を因果関係に解釈している可能性

(←因果関係と解釈した方が進化で生存可能性が高かった(説))

(回帰)式は左右が等しいことを示すだけ、因果の方向は示さない

効果 = 影響を受けたときの結果 – 影響を受けなかったときの結果

効果 = 影響を受けたときの結果 – 影響を受けなかったときの結果

CF(多くの場合は統御群)を作り出す仮定が必要

効果 = 影響を受けたときの結果 – 影響を受けなかったときの結果

CF(多くの場合は統御群)を作り出す仮定が必要

この仮定が現実的=信頼性の高い効果推計値

効果 = 影響を受けたときの結果 – 影響を受けなかったときの結果

CF(多くの場合は統御群)を作り出す仮定が必要

この仮定が現実的=信頼性の高い効果推計値

多くの「効果」の議論: CF=before や without、信頼性が低い

効果 = 影響を受けたときの結果 – 影響を受けなかったときの結果

CF(多くの場合は統御群)を作り出す仮定が必要

この仮定が現実的=信頼性の高い効果推計値

多くの「効果」の議論: CF=before や without、信頼性が低い

⇐ 選抜があるため

# 因果関係を測定する推計方法

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

SCM 大きな観察単位、長い事前データ必要、仮定=観察可能な選抜

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

SCM 大きな観察単位、長い事前データ必要、仮定=観察可能な選抜  
(他にもある...Wald (instrumental variables) estimator、matching estimator)

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

SCM 大きな観察単位、長い事前データ必要、仮定=観察可能な選抜  
(他にもある...Wald (instrumental variables) estimator、matching estimator)

with-without は RCT のみ、before-after は統御群も必要

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

SCM 大きな観察単位、長い事前データ必要、仮定=観察可能な選抜  
(他にもある...Wald (instrumental variables) estimator、matching estimator)

with-without は RCT のみ、before-after は統御群も必要

placebo test: 効果があり得ない部分で効果計測、効果検知すると推計値を疑問視、  
しかし、「出ないでほしい」という思いを叶えるような  $p$  hacking 可能

## 因果関係を測定する推計方法

RCT 内的整合性り、対象が限定され高価、仮定=実験が緻密に制御

DID 実験不要、両群のパネル・データが必要、仮定=共通トレンド

RDD 実験不要、ローカルな効果、仮定=政策のみ離散的变化

SCM 大きな観察単位、長い事前データ必要、仮定=観察可能な選抜  
(他にもある...Wald (instrumental variables) estimator、matching estimator)

with-without は RCT のみ、before-after は統御群も必要

placebo test: 効果があり得ない部分で効果計測、効果検知すると推計値を疑問視、  
しかし、「出ないでほしい」という思いを叶えるような  $p$  hacking 可能

信頼性の高い推計方法でも、 $p$  hacking, small sample bias は存在することに注意

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からぬ

しかし、メカニズムを示唆する検討が可能なこともある

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例: マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例: マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下  
イタリアで起こったことは日本でも起こるか?

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム（理論）を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例：マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下  
イタリアで起こったことは日本でも起こるか？

現代日本で暴力団抗争が激化しても、同じメカニズムが働くとは限らない

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例: マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下  
イタリアで起こったことは日本でも起こるか?

現代日本で暴力団抗争が激化しても、同じメカニズムが働くとは限らない  
理由: 警察が強いから。抗争に一般人を巻き込むと警察が弾圧。

「賃金中央値までなら最低賃金引き上げは雇用を減らさない」「余地の大きい日本でも引き上げよう」

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例: マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下  
イタリアで起こったことは日本でも起こるか?

現代日本で暴力団抗争が激化しても、同じメカニズムが働くとは限らない  
理由: 警察が強いから。抗争に一般人を巻き込むと警察が弾圧。

「賃金中央値までなら最低賃金引き上げは雇用を減らさない」「余地の大きい日本でも引き上げよう」

アメリカやイギリスでそう(DID)でも、日本でそうかは分からない

ハンガリー: 引き上げの75%が消費者に転嫁、10%雇用減(DID)

仮に信頼性の高い効果推計値を得ても、外的整合性があるかは別問題

全ての推計方法は効果発現メカニズム(理論)を明らかにしない

メカニズムが分からないと適用可能性が分からない

しかし、メカニズムを示唆する検討が可能なこともある

例: マフィア ⇒ 民間の投資・雇用低下、政府の投資・雇用増加 ⇒ 成長低下  
イタリアで起こったことは日本でも起こるか?

現代日本で暴力団抗争が激化しても、同じメカニズムが働くとは限らない  
理由: 警察が強いから。抗争に一般人を巻き込むと警察が弾圧。

「賃金中央値までなら最低賃金引き上げは雇用を減らさない」「余地の大きい日本でも引き上げよう」

アメリカやイギリスでそう(DID)でも、日本でそうかは分からない

ハンガリー: 引き上げの75%が消費者に転嫁、10%雇用減(DID)

理由: 産業構成(とくに貿易財比率)、雇用慣行、労働規制が違う

# key words

Poverty trap A vicious cycle that keeps the economy in poverty.

# key words

**Poverty trap** A vicious cycle that keeps the economy in poverty.

**Big Push** A set of policies that moves the economy to a convergent path to “high” equilibria.

# key words

**Poverty trap** A vicious cycle that keeps the economy in poverty.

**Big Push** A set of policies that moves the economy to a convergent path to “high” equilibria.

**Millennium Development Villages** An attempt to mimic a Big Push that intervenes the African villages in every aspect of life.

# key words

**Poverty trap** A vicious cycle that keeps the economy in poverty.

**Big Push** A set of policies that moves the economy to a convergent path to “high” equilibria.

**Millennium Development Villages** An attempt to mimic a Big Push that intervenes the African villages in every aspect of life.

**J-PAL** A research centre focusing on randomised controlled trials (RCTs) to produce practical policy lessons.

# key words

**Poverty trap** A vicious cycle that keeps the economy in poverty.

**Big Push** A set of policies that moves the economy to a convergent path to “high” equilibria.

**Millennium Development Villages** An attempt to mimic a Big Push that intervenes the African villages in every aspect of life.

**J-PAL** A research centre focusing on randomised controlled trials (RCTs) to produce practical policy lessons.

**internal validity** An unbiased causal inference.

# key words

Poverty trap A vicious cycle that keeps the economy in poverty.

Big Push A set of policies that moves the economy to a convergent path to “high” equilibria.

Millennium Development Villages An attempt to mimic a Big Push that intervenes the African villages in every aspect of life.

J-PAL A research centre focusing on randomised controlled trials (RCTs) to produce practical policy lessons.

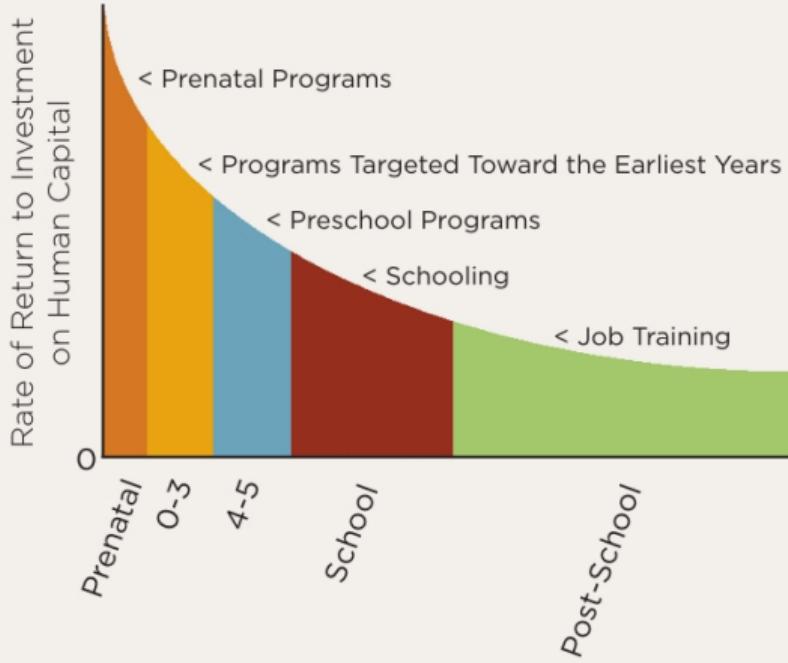
internal validity An unbiased causal inference.

external validity An unbiased causal inference with applicability beyond studied subjects.

# References I

### 3: RETURNS TO A UNIT DOLLAR INVESTED

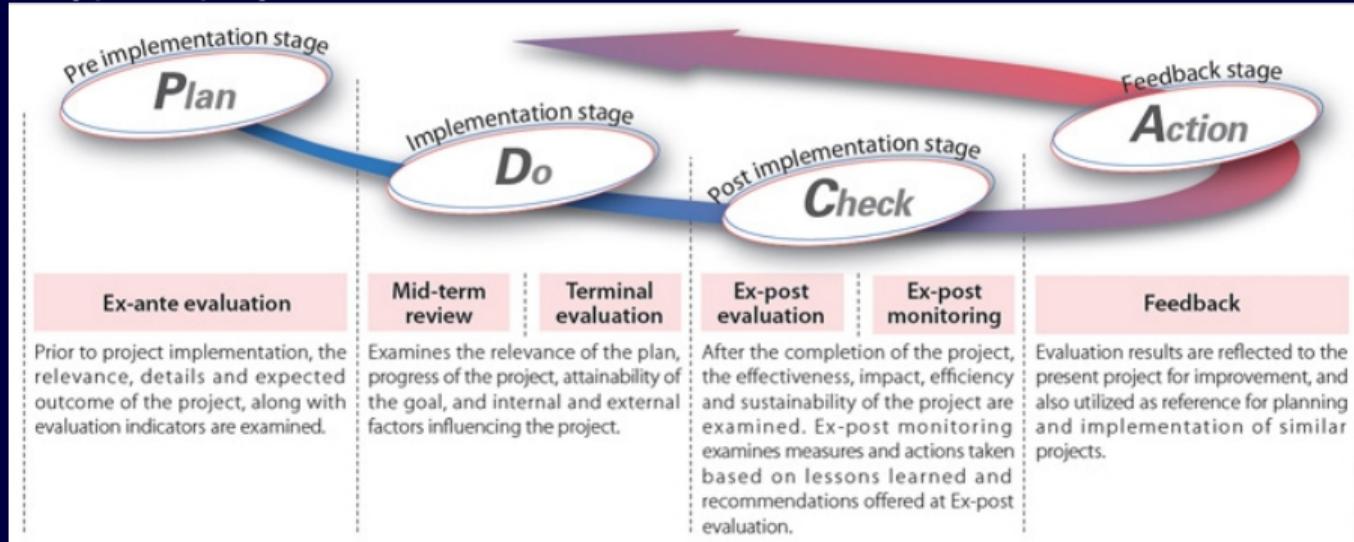
こんなイメージ(このような形か誰も確かめていない)



Source: James J. Heckman, 2008. "Schools, Skills, and Synapses," *Economic Inquiry*, Western Economic Association International, vol. 46(3), pp. 289-324, 07.

▶ go back

# A typical project evaluation framework: JICA.



[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html)

**Ex-ante** Policy goals and policy consistency, F/S.

**Interim** Implementation progress.

**Ex-post** Achievements, lessons learned. Internal evaluation (2 - 5 million), external evaluation (over 10 million).

Looks decent. Unfortunately, this ignores the fact that *ex post* needs to be planned at

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Examines the extent to which the aid activity is suited to the priorities donor: Does the goal of the aid activity meet the needs of beneficiaries? Are the activities and outputs of the program consistent with the overall goal and the attainment of its objectives?

**Effectiveness** Measures the extent to which a program or a project attains its objectives.

**Impact** Examines positive and negative changes as a result of the project. This includes direct and indirect effects and expected and unexpected effects.

**Efficiency** Measures the outputs in relation to the inputs to determine whether the aid uses the least costly resources possible to achieve the desired results.

**Sustainability** Sustainability relates to whether the benefits of the project are likely to continue after the closure of the project.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Measures the extent to which a program or a project attains its objectives.

**Impact** Examines positive and negative changes as a result of the project. This includes direct and indirect effects and expected and unexpected effects.

**Efficiency** Measures the outputs in relation to the inputs to determine whether the aid uses the least costly resources possible to achieve the desired results.

**Sustainability** Sustainability relates to whether the benefits of the project are likely to continue after the closure of the project.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Capacity was built?

**Impact** Examines positive and negative changes as a result of the project. This includes direct and indirect effects and expected and unexpected effects.

**Efficiency** Measures the outputs in relation to the inputs to determine whether the aid uses the least costly resources possible to achieve the desired results.

**Sustainability** Sustainability relates to whether the benefits of the project are likely to continue after the closure of the project.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Capacity was built?

**Impact** Impacts.

**Efficiency** Measures the outputs in relation to the inputs to determine whether the aid uses the least costly resources possible to achieve the desired results.

**Sustainability** Sustainability relates to whether the benefits of the project are likely to continue after the closure of the project.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Capacity was built?

**Impact** Impacts.

**Efficiency** Cost-benefit analysis.

**Sustainability** Sustainability relates to whether the benefits of the project are likely to continue after the closure of the project.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Capacity was built?

**Impact** Impacts.

**Efficiency** Cost-benefit analysis.

**Sustainability** Self sustainability.

▶ go back

# Elements of project evaluation

JICA's *ex post* evaluation reference: DAC evaluation criteria

[http://www.jica.go.jp/english/our\\_work/evaluation/about.html](http://www.jica.go.jp/english/our_work/evaluation/about.html),

[http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/project/ex\\_post/about.html](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/project/ex_post/about.html)

**Relevance** Consistency with other policies.

**Effectiveness** Capacity was built?

**Impact** Impacts.

**Efficiency** Cost-benefit analysis.

**Sustainability** Self sustainability.

They look nice and show “what” but not “how.”

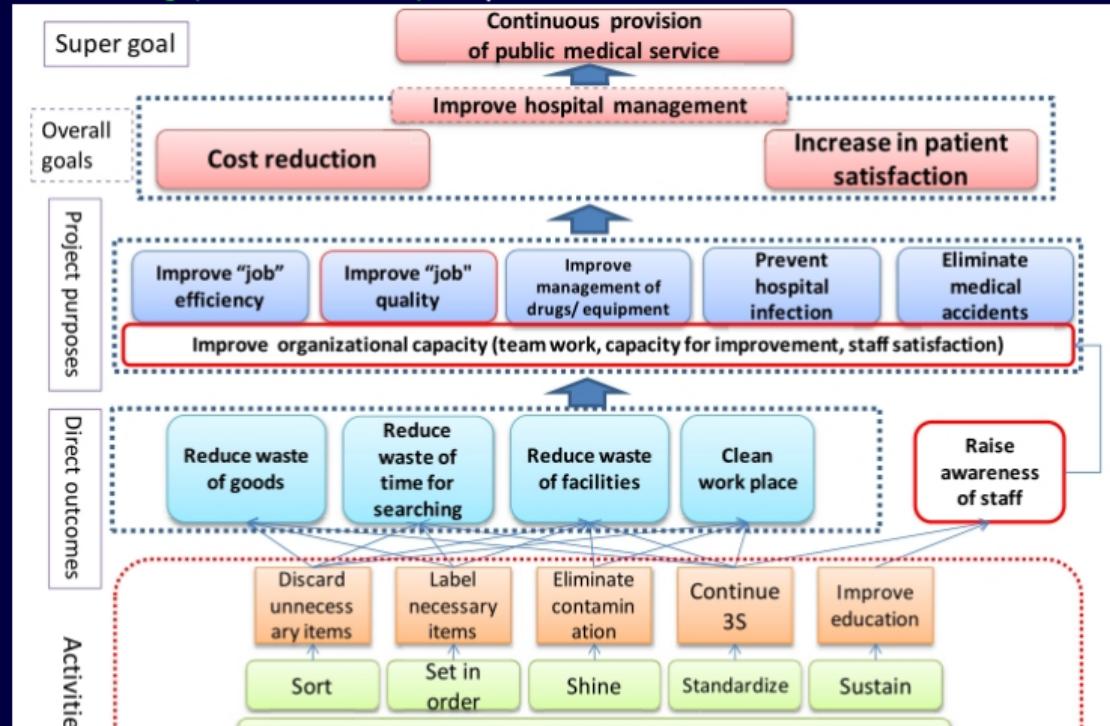
▶ go back

## JICA evaluation has elements of

- Adherence to the agreed principles: DAC evaluation criteria.
  - ☞ Fine as points of reference. But just referencing them “we will do these” leads to nowhere. Need “how” to implement these.
  - ☞ This is in a stark contrast to DFID. In an interview of DFID’s director of evaluations (<http://blogs.worldbank.org/impactevaluations/dfids-approach-to-impact-evaluation-part-i>), he talks about the substance (i.e., how DFID runs evaluation studies, how DFID learns from it, and how lessons are disseminated) but not the principles.
- Categorical framing of evaluation: *ex ante*, *midterm*, *terminal*, *ex post* may not suit well with consistent estimation of impacts.
  - ☞ Bad news: Lack of methodology often leads evaluators to choose with-without or before-after comparisons in *ex post* evaluations in the absence of randomisation.
  - ☞ *Ex post* evaluations has its role: Checking nothing bad happened, producing logistical knowledge. Limited usefulness.

# Logic model as an appraisal tool

An example of logic model. [http://www.jica.go.jp/english/our\\_work/evaluation/tech\\_and\\_grant/program/thematic/c8h0vm000001rgwp-att/2013\\_01.pdf](http://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/program/thematic/c8h0vm000001rgwp-att/2013_01.pdf), p.15.



# Logic model as an appraisal tool

Advantages:

- Can glance the implementation steps and accompanying tasks, allowing easy sharing of logistical procedure and goals with others.
- “If one expresses the basic plans with logic models, one can easily make changes and point the areas of downside and upside” (MOAF Information Centre, 2003, *A guide to writing a logic model*, p.6, in Japanese.)

# Logic model as an appraisal tool

Disadvantages:

- It does not show how one can implement each task.
- There is no way to examine if the logic model is true.
  - Improved hospital management ⇒ Continuous provision of public medical service?  
Reduction in the waste of goods, search time, facilities and a cleaner workplace ⇒ Improved organizational capacity? What are the conditions and how?

Given these relative merits, they use logic models primarily for logistical description and classification in the project appraisal, not in evaluation (Ref: Ministry of Internal Affairs and Communications [www.soumu.go.jp/main\\_content/000017619.pdf](http://www.soumu.go.jp/main_content/000017619.pdf)).

# 弾力性を使う理由

$$\text{需要の価格弾力性} = \frac{\text{需要の\%変化}}{\text{価格 1\%の変化}}$$

$$\text{需要の所得弾力性} = \frac{\text{需要の\%変化}}{\text{所得 1\%の変化}}$$

$$\text{需要の交差価格弾力性} = \frac{\text{需要の\%変化}}{\text{違う財価格 1\%の変化}}$$

price elasticity of demand, income elasticity of demand, cross price elasticity of demand

# 弾力性を使う理由

需要曲線は価格(円)と数量(グラム、cc、その他)の関係

両方とも単位があるので、単位が違うと比較できない

- グラム売りの豚肉の需要曲線の傾きと面積売りの土地の需要曲線の傾きの大小関係は判断できない
- 同じ豚肉でも、通貨単位が異なると全く同じ取引をしても傾きが変わるので、比較できない

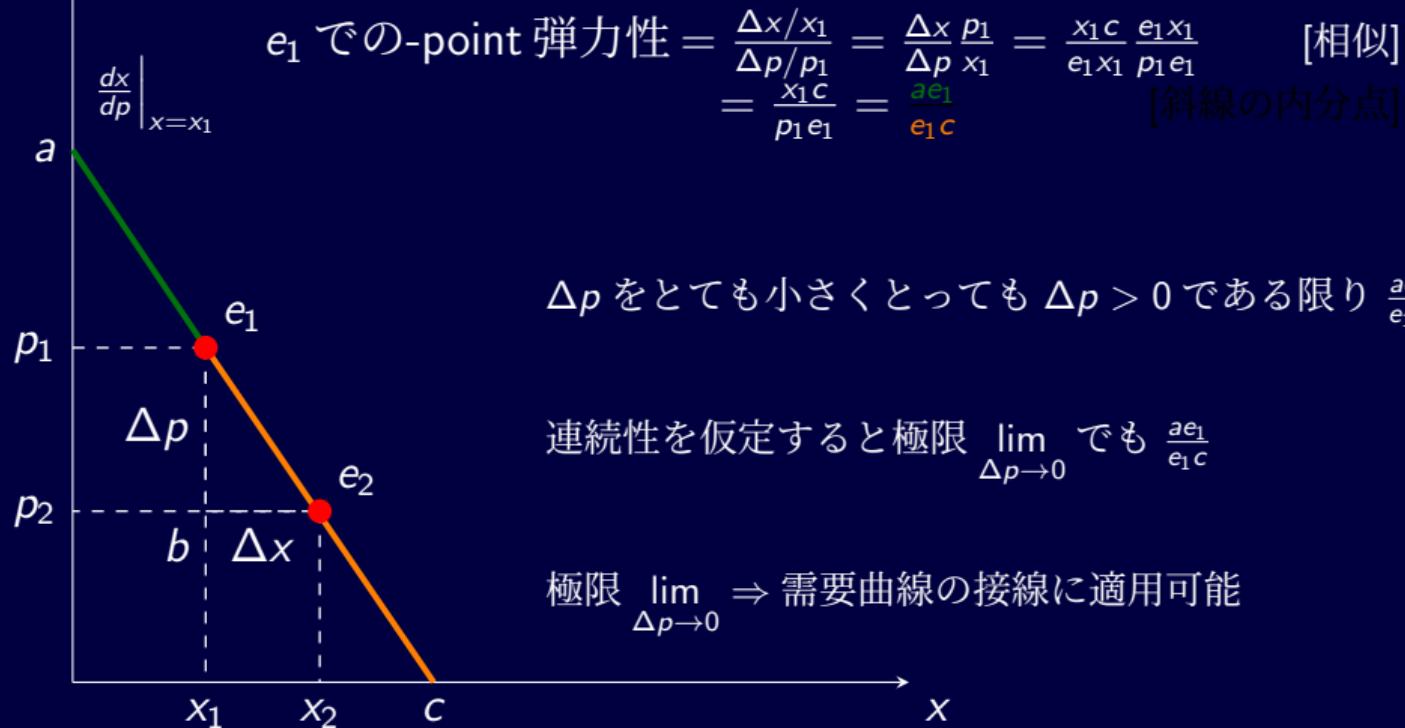
%は比率なので無単位

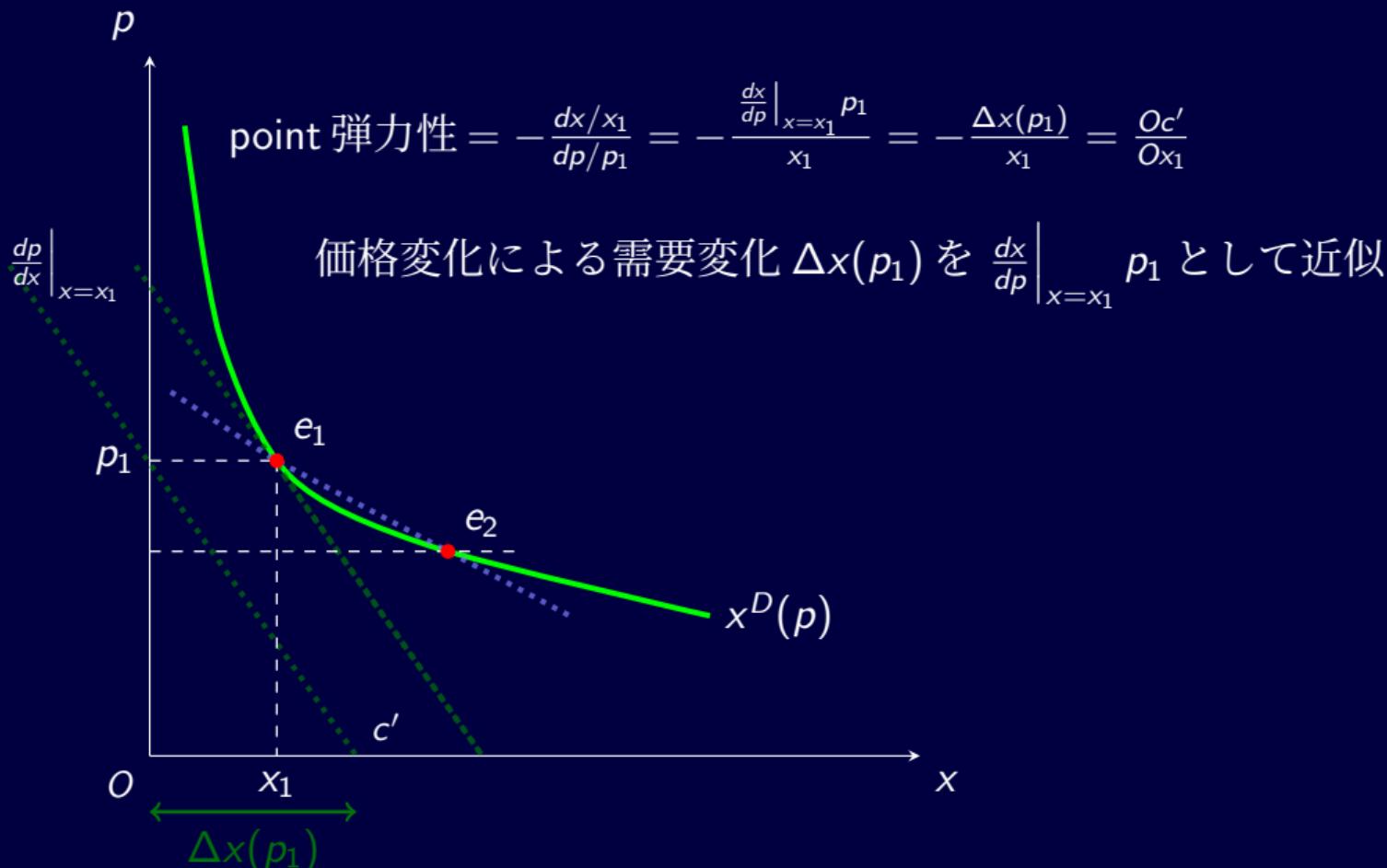
弾力性は%と%の比なので財や通貨が違っても比較可能

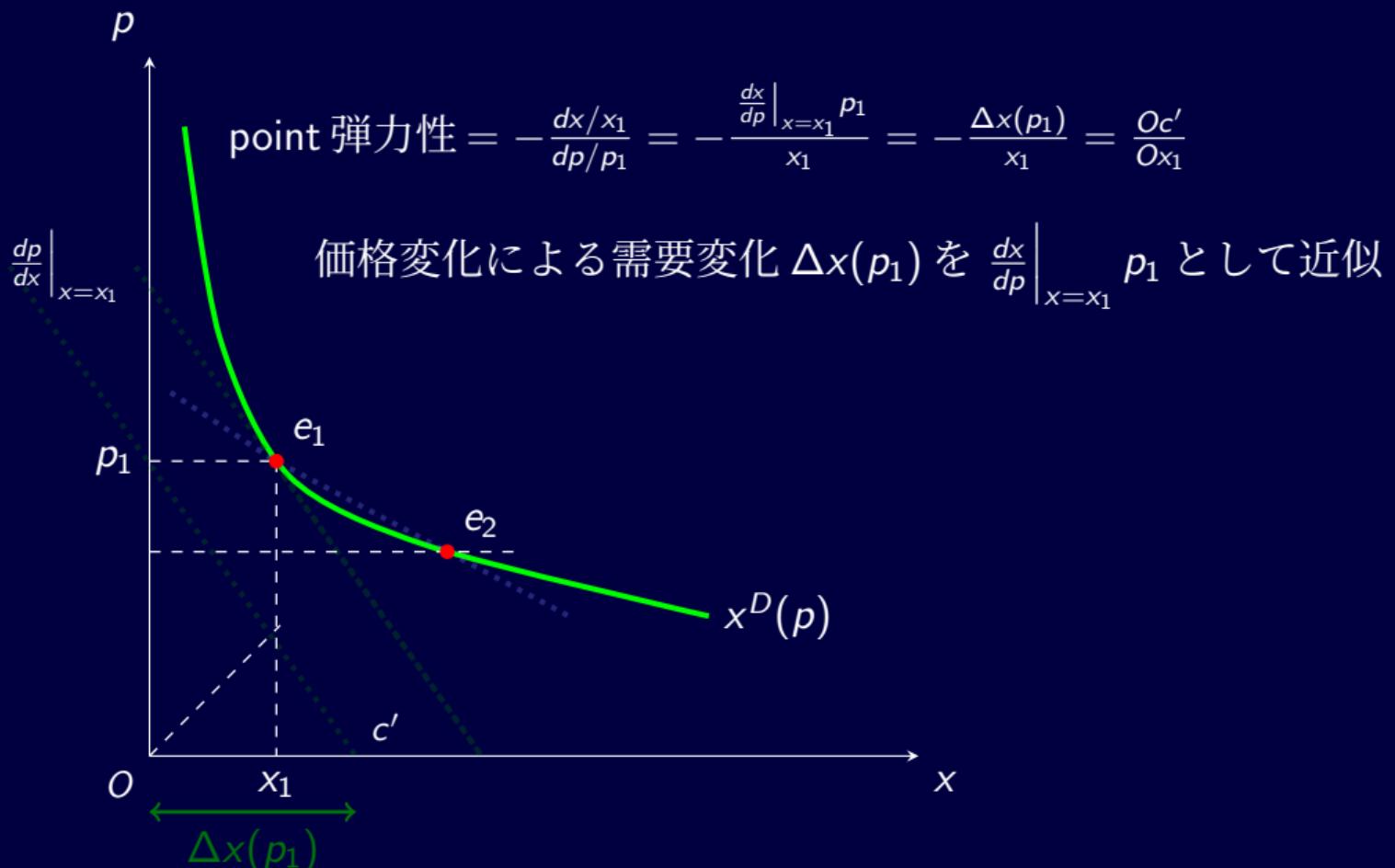
$$\text{elasticity} \leq 1 \Rightarrow \text{demand is } \begin{cases} \text{inelastic} \\ \text{elastic} \end{cases}$$

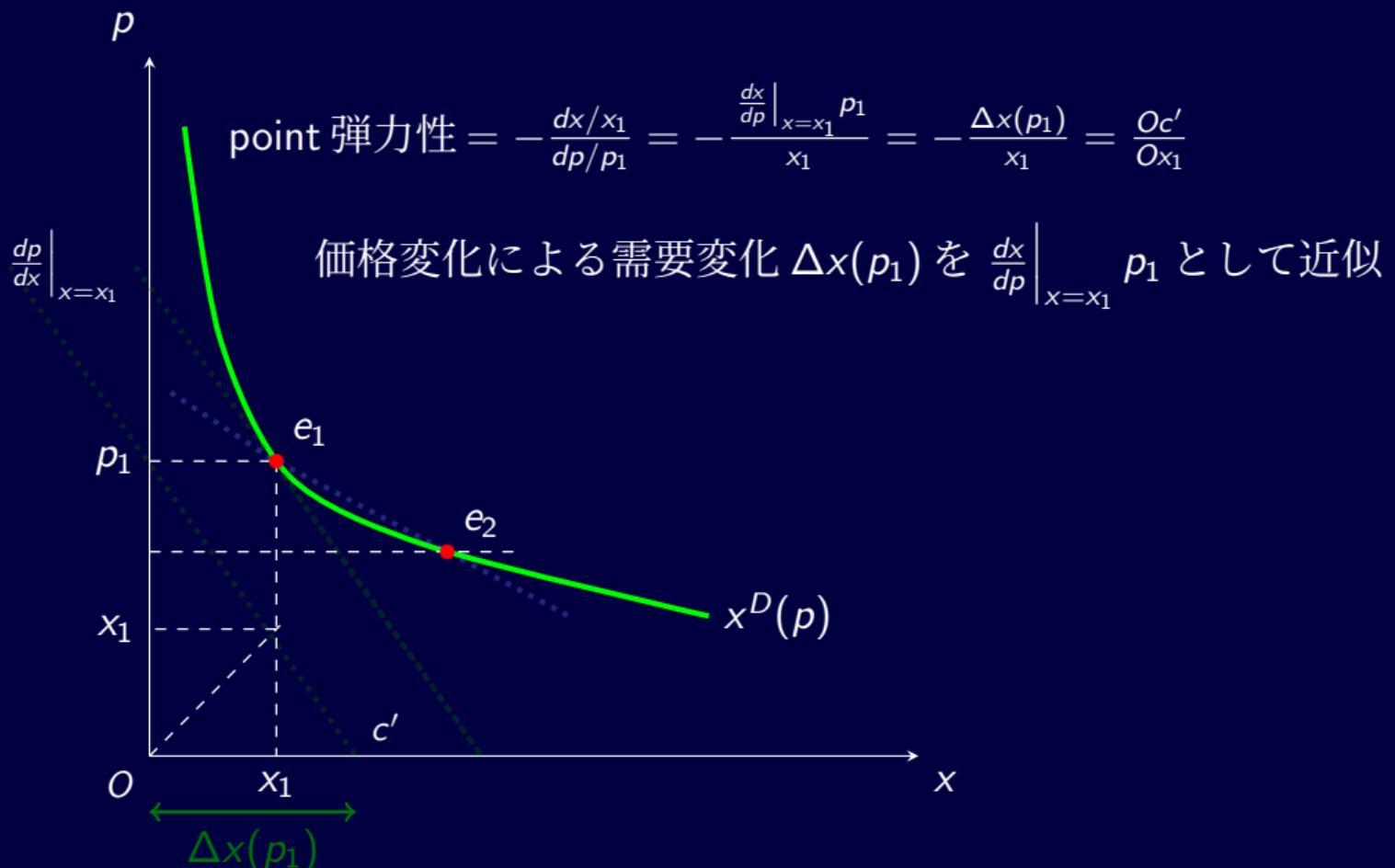
$p$

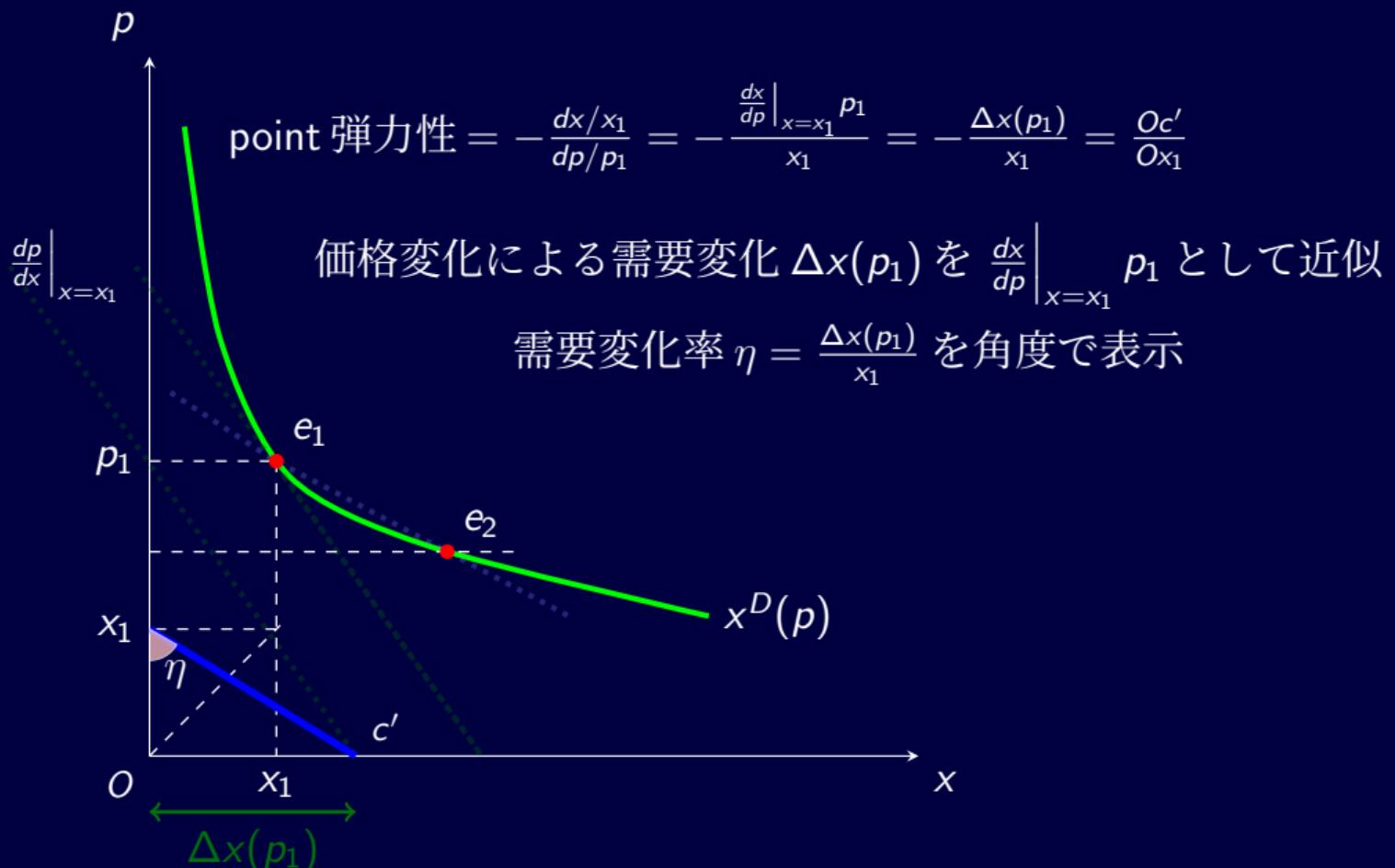
$e_1$  での需要の価格弾力性 =  $-\frac{\text{需要の } e_1 \text{ からの \% 变化}}{e_1 \text{ での 価格 } 1\% \text{ の 变化}}$

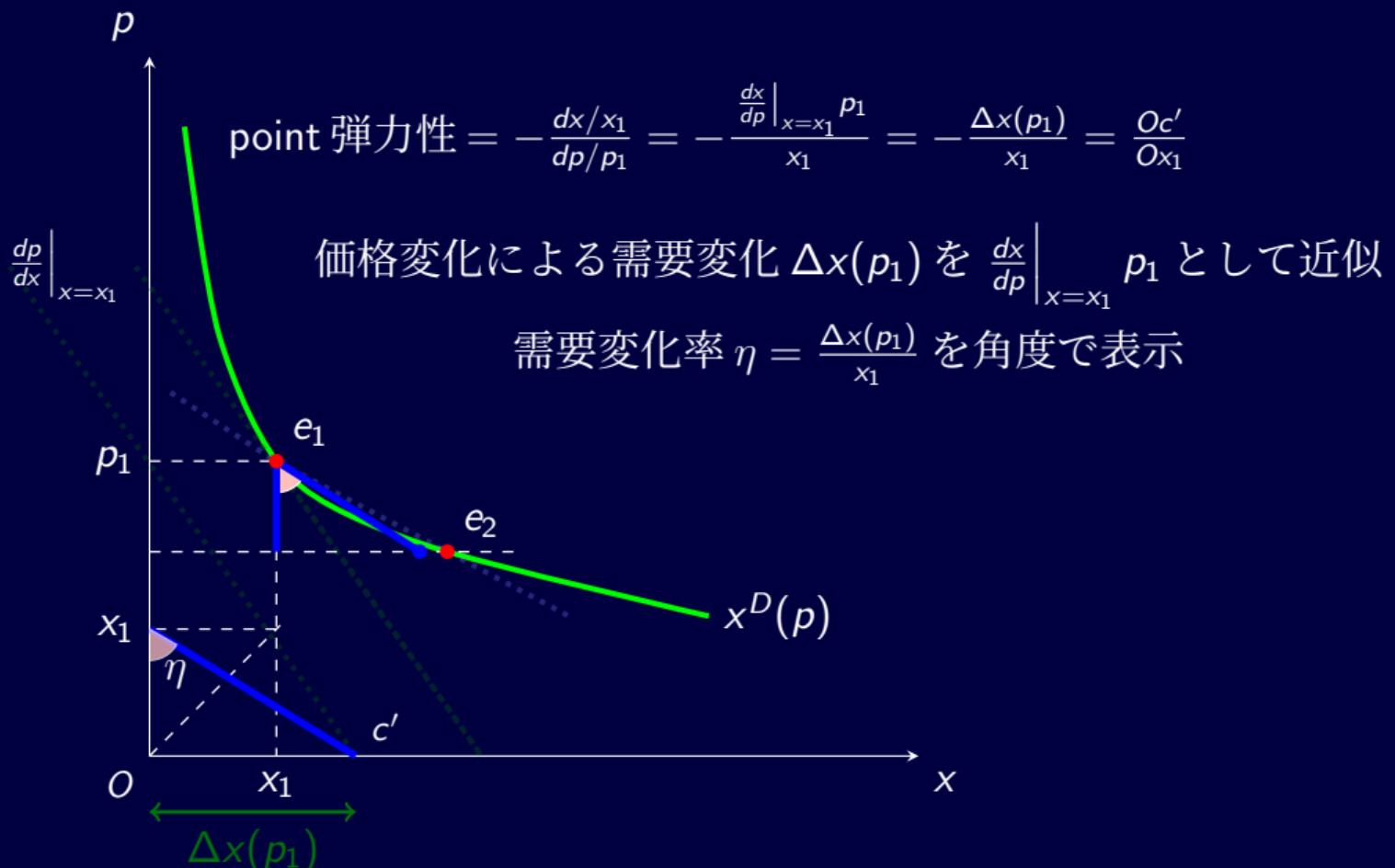


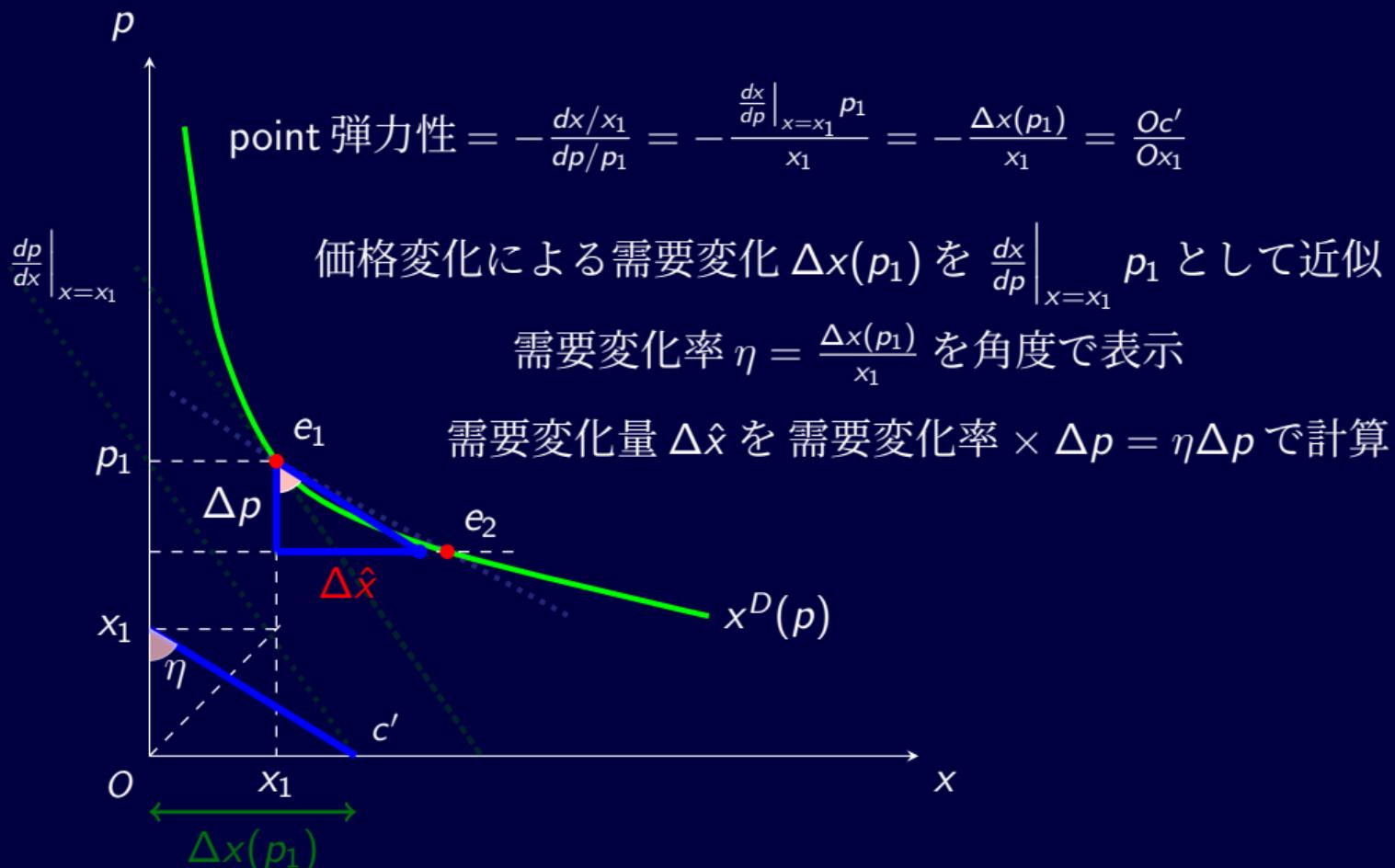


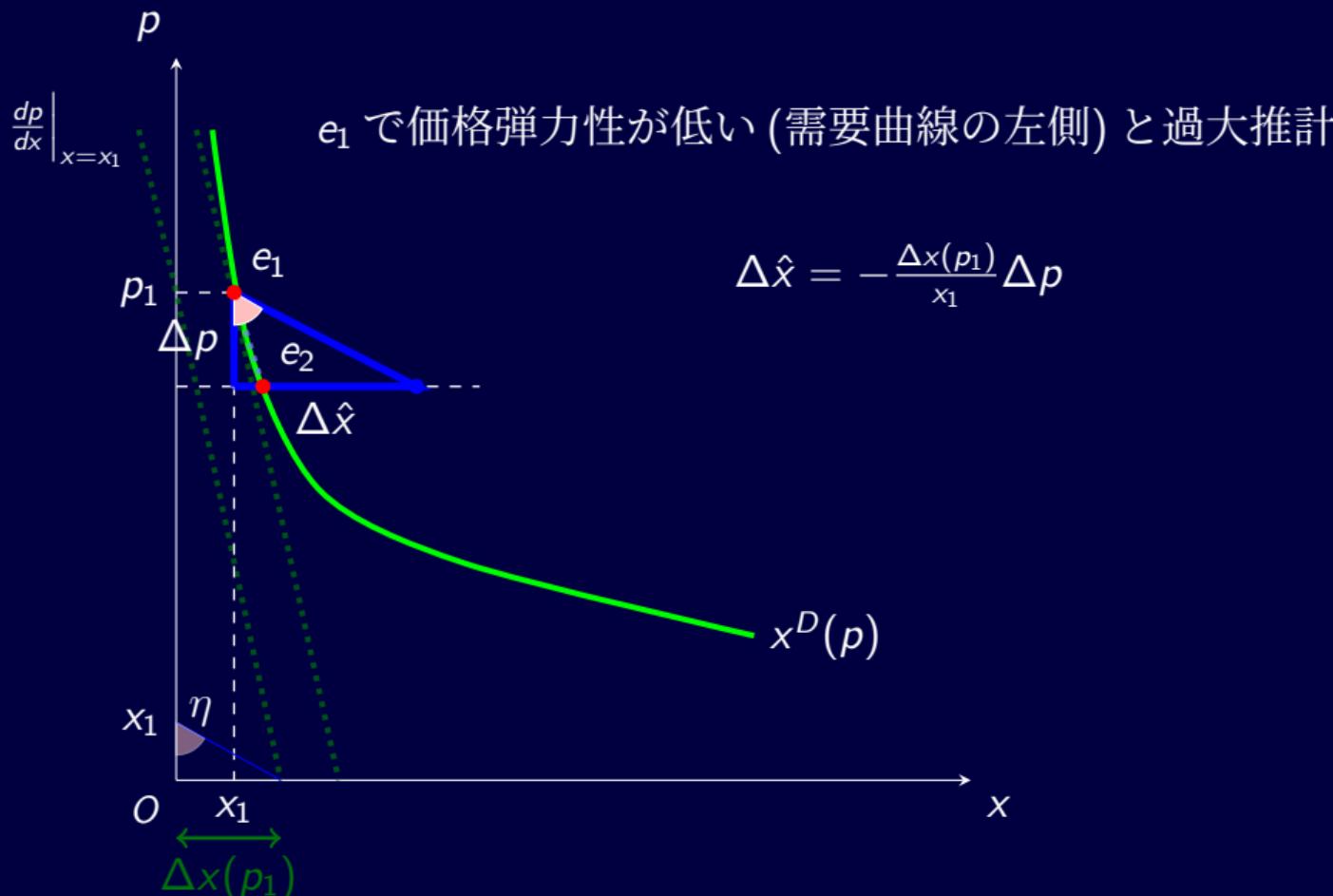












*p*

$e_1$  で価格弾力性が高い(需要曲線の右側)と過小推計

$$\Delta \hat{x} = -\frac{\Delta x(p_1)}{x_1} \Delta p$$

