ワンステップアップ・セミナー
(不) 自然実験を用いた因果推論

鎌田 拓馬
大阪大学国際公共政策研究科

第 77 回数理社会学会大会 東北大学
2024/8/29

台風が懸念される中で開催された数理社会学会・・・研究の生産性に影響は？

The Role of Conferences on the Pathway to Academic Impact

Evidence from a Natural Experiment

Fernanda Leite Lopez de Leon
Ben McQuillin

ABSTRACT

We provide evidence for the effectiveness of conferences in promoting academic impact by exploiting the cancellation—due to Hurricane Isaac—of the 2012 American Political Science Association Annual Meeting. We assembled a data set of 29,142 papers and quantified conference effects, using difference-in-differences regressions. Within four years of being presented at the conference, a paper's likelihood of becoming cited increases by five percentage points. We decompose the effects by authorship and provide an account of the underlying mechanisms. Overall, our findings point to the role of short-term face-to-face interactions in the formation and dissemination of scientific knowledge.

Figure: de Leon and McQuillin (2020) *Journal of Human Resources*.

本セミナーの目的・範囲

（不）自然実験を用いた因果推論に関する概要

内容：

（不）自然実験を用いた因果推論を行うための研究志向
ノンテクニカルな説明と応用例の紹介
社会学者が実行可能な分析手法の提示

紹介する手法：

回帰不連続デザイン（Regression Discontinuity Design, RDD）（研究例を用いて簡単に説明）
差分の差分法（Difference-in-Differences Design, DiD）（DDD, continuous DiD 含め、重点的に説明）
操作変数差分の差分法（DDIV）・Shift-share（Bartik）操作変数法 （DiD の延長として説明）
その他の実用的な分析（例：プラセボテスト）
*Disclaimer: 時間の制約上、また応用に重点を置くため、手法に関する説明では重要度の高い仮定のみに言及。
* 扱わないテーマ：inference（仮説検定）については扱わない。適当なレベル（大体、処置レベル）でクラスタリング標準誤差・空間相関を許す Conley 標準誤差・ロバスト標準誤差、あるいはブートストラップをしておけば多くのケースでは大丈夫です 笑

因果効果識別の重要性

なぜ因果効果を識別する必要があるのか？

1. 理論・仮説の検証

   地域のソーシャル・キャピタル向上 ！ 犯罪率低下
   移民に対するネガティブな報道 ！ 移民への信頼低下
   貧困地域での教育 ！ 高校中退率減少
   これらは相関ではなく、因果関係の特定に関心がある

2. メカニズム解明の第一歩 (¸ 外的妥当性)

   因果の識別なしにメカニズムの識別は不可能（D ！ M ！ Y）
   ASR のエディターらによる計量分析のガイドライン (Mustillo et al., 2018)

   媒介分析については言及あるにも関わらず、**因果効果を識別するための適切なリサーチデザインの構築については言及なし・・・**

3. 政策評価

   相関と因果の区別が不可欠
   適切な政策提言のための基礎 （¸ 外的妥当性）

応用計量経済学における研究デザインとトレンド

Figure: Source: Currie et al. (2020)

応用計量経済学における研究デザインとトレンド

Figure: Source: Currie et al. (2020)

因果推論関連の最近のレビュー・実践論文: 社会学 vs. 他分野

社会学

DiD に関してはまだ書かれていない・・・？

IV (Felton and Stewart, 2022): 本セミナーで扱う Bartik 操作変数、DDIV などは扱われていない

(因果推論と機械学習 (Brand et al., 2023)：本セミナーで扱うテーマは扱われていない)

他分野

DiD

経済学: (De Chaisemartin and d'Haultfoeuille, 2023; Roth et al., 2023)

Goodman-Bacon (2021) 以来の「みんなやっている DiD って実は何をやっているかよくわかっていないよね」により「週刊 DiD」が出来上がり、最新の DiD 推定量の開発に関するサーベイ

イベントスタディに焦点を絞った実践論文 (Miller, 2023)

政治学: Liu et al. (2024); Li and Strezhnev (2024)

疫学: Wang et al. (2024)

IV: 操作変数そのものに関しては多くの分野で書かれているが、本セミナーで扱うトピックに関しては：

Shift-share IV (Borusyak et al., 2024; Goldsmith-Pinkham et al., 2020)

DDIV (Hudson et al., 2017)

RDD

Cattaneo et al. (2019): RDD の実践をほぼ網羅している

マッチングに関して

社会学においてしばしば観測される因果推論の誤解：

- 「OLS は相関しかわからない」「マッチング（e.g., 傾向スコアマッチング）は反実仮想アプローチで因果効果わかる」という誤解
  - 両手法とも selection on observables の仮定に基づく
- それではマッチングの利点とは？
  - 共通サポート（common support）問題への注意を促す
  - 処置群・比較群の比較可能性を明示的に考慮

共通サポートに関して

応用計量経済学のマッチングに対する見方

マッチングのみでは一定水準の雑誌掲載は困難

研究デザインがあって、そこにマッチングを組み合わせて使用することは多々ある

例：Matched DiD (e.g., Aneja and Xu, 2022; Fenizia and Saggio, 2024)

因果効果を探る方法

因果効果を探る方法：

- 実験データ：ランダム化比較試験（RCT）
- 観察データ：共変量調整型
  - 共変量に条件づけた上で、処置変数が無作為であればマッチング、OLS などで推定
- 観察データ：準実験・（不）自然実験アプローチ
  - 単純に固定効果を入れるのと何が違う?
    - 準実験・（不）自然実験アプローチ: ターゲットパラメータがある・識別仮定がハッキリしている
  - 個人がコントロールできない「外生的」な要因を利用：イベント、天気の変動、あるいは制度変更など
    - 70 歳での医療額が減る ）比較：70 歳前後の個人 (Shigeoka, 2014)
  - 「何に対して」外生なのか？ 条件付けに外生なのか？ については後に詳しく議論
  - 処置前の処置群・比較群間に差異がない
    - 何に対する差（共変量のレベル・結果変数のトレンド）なのかはデザインによって異なる
  - ただし、政策の内生性 (Besley and Case, 2000)、割当メカニズムの不明確さ (Keele et al., 2020) （多くの場合 DiD）があると、不自然実験

潜在結果変数モデル

信頼性革命

「信頼性革命（credibility revolution）」(Angrist and Pischke, 2008)：

共変量による交絡要因の統制よりも、デザインによる交絡の統制

処置群と比較群の比較可能性：統計モデルよりリサーチデザインを重視

処置変数の変動がどこからきているか・割当メカニズムの理解

）ドメインスペシフィックな知識（制度的背景）の重要性

例：医療大麻合法化がメキシコ麻薬組織の活動に打撃を与えることを検討する際に、医療大麻合法化が犯罪率に対して内生的に導入されていない/医療的な目的で導入されていることを議論 (Gavrilova et al., 2019)

換言すれば、「それっぽい推定結果」が出ていても、制度的な説明が十分ではないと説得的ではない！

「原因」と「結果」を捉え方

「原因」と「結果」を捉える上で大きく 2 つの枠組みがある

原因の結果 (effects of causes):

処置（独立）変数の視点 ！ 結果変数を考える

* 処置変数以外にはあまり関心がない（その他の共変量はあくまで統制変数）

結果の原因 (causes of effects):

結果変数の視点 ！ 独立変数を考える

「結果の原因」と「原因の結果」の関係 (Gelman and Imbens, 2013) ：

「結果の原因」から問を立てるが、「結果の原因」から因果推論を行うのではない

「なぜ」（結果の原因）から問いを立て、「もし〜ならば」（原因の結果）の問いの動機づけをする

（不）自然実験アプローチを行う上で必要な研究志向

単一のデータセットのみを用いた分析からの脱却：

例：GSS データを用いて、個人の人種に対する態度を個人の属性やその他の態度や認識に回帰

$$\underset{\textit{Race attitude}}{y_{|\{z\}}} = V_0 \text{ , } V_1 \underset{\textit{Education}}{x_{1|\{z\}}} \text{ , } V_2 \underset{\textit{Income}}{x_{2|\{z\}}} \text{ , } V_3 \underset{\textit{Diversity attitude}}{x_{3|\{z\}}} \cdots \text{ , } n$$

外部要因への注目 ）複数のデータを組み合わせる：

例：Diversity attitude を**外部から変えるような要因**を探る

）既存の調査データを用いながらオリジナルな分析

人種的マイノリティが活躍するミュージカルが 2016 年にトニー賞を受賞したことを利用 (Ikeuchi, 2023)：

2016 年の GSS データを用いて、トニー賞受賞以前に調査を受けた人に比べて、以後に調査を受けた人は、黒人の不平等は差別によるものではないと考えるようになることを明らかにした

) 多様性を進める取り込みをすると、人種差別がなくなったと認識するようになる

調査日とイベントのタイミングを利用した Unexpected Event during Survey Design (UESD) (Muñoz et al., 2020, ある状況下では回帰不連続デザインと同じ)

（不）自然実験アプローチを行う上で必要な研究志向

外部要因への注目：複数のデータを組み合わせることについて

社会学研究において、複数のデータソースを統合する際にマルチレベルモデルが頻繁に使われる（例：個人を対象とする調査データと地域データの結合）。しかし、地域変数（例：経済状況、人種構成、移民比率など）そのものが内生的であることが多いため、単純に複数のデータソースを組み合わせるだけでは不十分である。

（不）自然実験アプローチと研究デザイン

因果推論で用いられる主要な研究デザイン：

回帰不連続デザイン (RDD)

処置変数の割り当てメカニズム = ランニング変数
閾値（カットオフ）付近での局所的な因果効果を推定

操作変数法 (IV)

処置変数の割り当てメカニズム = 操作変数
操作変数によって動かされた内生変数における局所的な因果効果を推定

RDD と IV の共通点：

ランニング変数・操作変数 = 内生変数を擬似的に無作為化する
RDD にはシャープとファジーがあり、ファジー RDD は IV と同等である

差分の差分法 (DiD)

処置変数の割り当てメカニズムを考えない (強いて言えば selection on unobservables)
識別仮定: 平行トレンド仮定 = **結果変数**に置かれる
内生変数を無作為するという発想に立っていない

*DiD は RDD と IV に比べて大きく異なる発想に基づいている。しかし、本セミナーでは DiD の延長として IV（Bartik IV / DDIV）を考えることにするため、順序は、RDD, DiD, IV（の一部）と進める。

各研究デザインとデータ生成過程

よく “Let data speak for itself” と言うが、全然違うやん！ 笑

各研究デザインと内的妥当性の程度具合

Figure: Source: Brodeur et al. (2020)

研究デザインによって恣意的な操作のしやすさが全く異なる・・・

RCT と RDD は同じくらいの内的妥当性

一方、DiD と IV は恣意的な操作がしやすい

因果推論を理解するのに最低限必要な概念

Estimand: 関心のあるパラメータ（ターゲットパラメータ）

- 平均処置効果 (Average Treatment Effect, ATE)
  - 母集団全てを対象とした効果
- 処置群における平均処置効果 (Average Treatment Effect on the Treated, ATT)
  - 処置を受けた母集団における効果
- 局所的平均処置効果 (Local Average Treatment Effect, LATE)
  - RDD: カットオフ周辺における処置効果
  - IV: 操作変数によって行動変容が促された個体（complier と呼ぶ）における処置効果
  - 内的妥当性: ATT・ATE・LATE あり
  - 外的妥当性: ATT は処置を受けてない母集団に関しては当てはめることは保証しない。従って ATT<ATE となることがある。この場合においては、ATT は外的妥当性は持たない。LATE に関しても外的妥当性は保証されない。

潜在結果変数：

- （観測できる）結果変数 Y_i
- 潜在的結果 (potential outcomes) : i さん（個人、地域、組織）の**全く同じ時間における 2 つの別々の世界の状態**
  - Y_i^1: i さんが処置を受けた世界の状態下での潜在結果
  - Y_i^0:（同じ）i さんが処置を受けなかった世界の状態下での潜在結果

処置変数 D_i: 処置を受ける $D_i = 1$、処置を受けない $D_i = 0$

（シャープ）回帰不連続デザイン（RDD）

処置の割り当てが連続変数のカットオフによって決定される状況を考える：

$$D_i = \begin{cases} 0 & \text{if } X_i \ddot{Y} c \\ 1 & \text{if } X_i \quad c \end{cases} \tag{1}$$

X_i はランニング変数、c はカットオフ（処置割当のルール）

$$\mathrm{E}^1 Y_i \mathrm{j} X_i{}^{\mathrm{o}} = \begin{cases} \mathrm{E}^1 Y_i^0 \mathrm{j} X_i{}^{\mathrm{o}} & \text{if } X_i \ddot{Y} c \\ \mathrm{E}^1 Y_i^1 \mathrm{j} X_i{}^{\mathrm{o}} & \text{if } X_i \quad c \end{cases} \tag{2}$$

局所平均処置効果 (LATE) / 条件付き平均処置効果 (CATE) = カットオフ付近での処置効果

RDD を回帰分析で捉える

RDD を回帰の枠組みで示すと：

$$Y_i = U \text{ ¸ } gD_i \text{ ¸ } V_1{}^1X_i \quad c^{\text{o}} \text{ ¸ } V_2D_i{}^1X_i \quad c^{\text{o}} \text{ ¸ } n_i \tag{3}$$

関心のあるパラメータ： g

RDD の重要な仮定：

連続性の仮定：カットオフにおいて潜在結果 $\mathrm{E}^1 Y_i^1 \mathrm{j} X_i = x^{\text{o}}$ と $\mathrm{E}^1 Y_i^0 \mathrm{j} X_i = x^{\text{o}}$ は連続であり、処置以外の要因では不連続にならない

) 偽のカットオフで検証してジャンプが起きないことを確認（プラセボテスト）

実践的な点として：

カットオフ付近のデータのみを使用) バンド幅を変えることでの頑健性チェック

関数形の柔軟性) セッティングによるが、ランニング変数を線形と二乗項でフィット

カットオフ周辺で共変量のバランスチェック = カットオフ付近で処置群と比較群が比較可能

カットオフ周辺で操作（manipulation）がない) Density test を行う

社会学者が応用可能な RDD セッティング：人口の分布を利用

人口規模が少数集団間で転換することを利用 (Ikeuchi and Kamada, 2024)：

黒人と他のマイノリティ集団の人口規模の転換周辺において白人と黒人の住居の棲み分けが観測

）関心のある集団（この場合、黒人）の規模のみならず、他の集団との相対的な規模（i.e., ランク = 参照依存）も、住居の棲み分けに対して重要

社会学者が応用可能な RDD セッティング：調査データ内で分析可能

男女間賃金差を利用した RDD (Bertrand et al., 2015)：

夫婦において妻の収入が 50%を少し超えたところで、妻の収入シェアの分布が急激に減少

) 性別役割分業意識を捉えている

RDD の実行可能性

利点：RCT と同様の内的妥当性。p-hacking などの恣意的な分析が困難。
不利点：カットオフ付近でたくさんの観測数が必要) 実行可能性が低い

実行可能性が高いデザインは・・・

ベストではないが、実行可能性が高い分析は・・・

) DiD

DiD を行うのに必要なデータ

DiD を行うのに必要なデータ

1. 処置群と比較群
2. 処置が導入された前後

ではどのようなデータを使うことができるのか？

パネルデータ: 同一の個体に対して複数の時点に渡って変化を記録したデータ

ある調査の年以降にある集団で処置が入る

繰り返しクロスセクショナル・データ: 異なる時点のクロス セクショナルデータを集計したデータ (同じ母集団を対象とし、異なる時点で異なるサンプルのデータ)：

ある調査の年以降にある集団で処置が入る
仮定：処置前後でサンプル構成が変わらないこと

クロスセクショナル・データ:

あるコーホート以降にある集団で処置が入る

差分の差分法（DiD）

標準的な DiD のセットアップとして 2 2 の DiD を考える

2 グループ:

$D = 1$: 処置グループ

$D = 0$: 比較グループ

2 期間:

$t = 0$: 処置前

$t = 1$: 処置後

DiD: 結果変数の

1. 差の (Difference-in): **処置群内・比較群内**における処置が導入される**前後比較**
2. 差 (Differences): **処置群と比較群間の比較** により因果効果を識別

$$\overbrace{\overbrace{\mathrm{E}[Y_{i1}|D_i = 1] - \mathrm{E}[Y_{i0}|D_i = 1]}^{\text{処置群内における 1 期・0 期の差}} - \overbrace{\mathrm{E}[Y_{i1}|D_i = 0] - \mathrm{E}[Y_{i0}|D_i = 0]}^{\text{比較群内における 1 期・0 期の差}}}^{\text{処置群・比較群間の差}} \tag{4}$$

処置群における平均処置効果 (ATT): 処置後における処置群の結果とその処置群が仮にも受けなかったときの潜在的な結果の期待値の差

$$ATT = \mathrm{E}[Y_{i1}^1 - Y_{i1}^0|D_i = 1] \tag{5}$$

DiD: 平行トレンドの仮定

$$ATT = \mathrm{E}^1 Y_{i1}^1 \quad Y_{i1}^0 \mathrm{j} D_i = 1^{\mathrm{o}} \underbrace{, \mathrm{E}^1 Y_{i1}^0 \quad Y_{i0}^0 \mathrm{j} D_i = 1^{\mathrm{o}} \quad \mathrm{E}^1 Y_{i1}^0 \quad Y_{i0}^0 \mathrm{j} D_i = 0^{\mathrm{o}}}_{\text{平行トレンドの仮定が満たされれば 0}} \tag{6}$$

平行トレンドの仮定が満たされる: 処置が介入されなければ処置グループの結果と比較グループの結果にトレンドの差異がない

$\mathrm{E}^1 Y_{i1}^0 \quad Y_{i0}^0 \mathrm{j} D_i = 1^{\mathrm{o}}$: 処置グループに仮にも処置が割り当てられなかったときの 1 期と 0 期における潜在的結果の期待値の差 (観察できない)

処置効果 = (処置グループにおける処置前後の結果変数の平均差)　(比較グループにおける処置前後の結果変数の平均差)

処置前: 処置グループと比較グループの結果変数の平均的なトレンド（傾き）が等しい (ただしレベルは異なる)

処置後: 処置グループにおいて結果変数のトレンド（傾き）は比較グループのそれよりも大きい

処置後: 処置グループの傾きは比較グループのそれよりも大きい

処置前: 処置グループの傾きはすでに大きい = **処置以前の結果変数のトレンドが処置グループと比較グループで異なっている**

処置の因果効果とはいえず、**既存のトレンドによって引き起こされる**

DiD を回帰分析で捉える

標準的な DiD: 2 2（2 期間　2 グループ）

$$Y_{it} = U \text{ ¸ } XD_i \quad Post_t \text{ ¸ } \grave{}D_i \text{ ¸ } WPost_t \text{ ¸ } n_{it}. \tag{7}$$

関心のあるパラメータ： X

D_i : 処置を受ければ 1、受けなければであれば 0 を取るグループダミー

$Post_t$: 処置後であれば 1、処置前であれば 0 を取る時間ダミー

多くの応用: 複数の期間・複数のグループ

（ある一定の条件下で）個体固定効果と時間固定効果を含めた固定効果モデル (two-way fixed effects, TWFE) が使用可能

$$Y_{it} = XD_i \quad Post_t \text{ ¸ } \grave{}_i \text{ ¸ } W_t \text{ ¸ } n_{it} \tag{8}$$

$\grave{}_i$ (個別固定効果): 時間 t に依存しない個体 i 特有の効果 = 個体の切片を固定

W_t (時間固定効果): サンプル全てに共通するその時期 t の効果を捉える (e.g., 景気などのマクロショック)

DiD を回帰分析で捉える

式 7 が複数期間・複数グループから成り立っているとしよう。そのとき：

Q: 式 7 で $D_i(\grave{})$, $Post_t(W)$ をそれぞれ加えること、式 8 で $\grave{}_i$ と W_t を加えることの違いとは？

A: 式 7: $D_i(\grave{})$ は、処置群・比較群間の違いを拾っているが、処置群内・比較群内 の異質性がないことを仮定している。同様に、$Post_t(W)$ で処置以前・以後間の違いを拾っているが、処置以前内・処置以後内 の時間効果は異質性がないことを仮定している。

式 8 ： $\grave{}_i$ は、ユニット間の恣意的な異質性を拾い、W_t は、ピリオド間の恣意的な異質性を拾う。

平行トレンドの解釈

平行トレンド仮定：

処置がなかった場合、結果変数の**トレンド**は処置群と比較群で潜在的に同じように挙動する。

Q: 処置がなかった場合、結果変数の**レベル**が処置群と比較群が潜在的に同じであることは必要とする？

A: レベルが同じである必要はない。なぜなら、D_i を含めることで元々の集団間のレベルの違いをコントロールしているため。より一般的に言えば、個体の固定効果（$\grave{}_i$） を含め、各ユニットの切片を固定することで、ユニット間の元々のレベルを一定にしている。

Q: 平行トレンド仮定を欠落変数で考えるとどういう意味であるか？

A: ユニットの固定効果を条件づけた上で、処置変数は結果変数のトレンドに対して外生的である。**平行トレンド仮定は、トレンドに関する欠落変数バイアスがない**ことを意味する。

トレンドに関するバイアス

考えてみよう

分析者は、新幹線の開通が経済発展にもたらす影響を差分の差分法を用いて検討しているとしている。新幹線をどこに建てるかは外生的ではなく内生的に決まっていると考えたとき、分析者は次の記述のうちどちらを気にするべきだろうか：

1. 新幹線は経済成長すると見込まれるところに作られるため内生性に対応しなければならない。
2. 新幹線は経済状況が良いところに作られるため内生性に対応しなければならない。

イベントスタディ分析

平行トレンド仮定は、反事実を含むため、実際に観察することは不可能。ただし、仮定の含意を間接的に検証することは可能。回帰分析を通じて示すためにイベントスタディ分析を見てみよう

イベントスタディ分析: 動学的な DiD 効果を推定

- 平行トレンドの仮定の間接的検証: 「未来」の処置変数を用いて処置が入る前の結果変数の挙動を見る
  - 処置が割り当てられる前なので、処置効果があるのあれば結果変数には影響し得ない。従って、係数は小さく有意ではない
  - 注意: 平行トレンドの仮定を間接的に検証するためには少なくとも 3 期分のパネルデータが必要
- 処置の長期的効果の検証: 「過去」の処置変数を用いて処置が入った後の結果変数の挙動を見る

イベントスタディ分析

$$Y_{it} = \sum_{k,g<1}^{\tilde{\mathcal{O}}} U_g D_i \quad \mathbf{1}^1 g = t \quad treat\ time^{o} \ , \ `_i \ , X_t \ , Y_{it'} \tag{9}$$

処置開始のタイミングで時間の標準化 ($g = calendar\ time \quad treat\ time$): 処置時期を 0、その一期後を 1, その一期前を　1... とする

$D_i \quad \mathbf{1}^1 g = t \quad treat\ time^{o}$: 処置以前　k 期より処置以後 l までのイベントスタディダミー変数
該当する期に対して 1、それ以外は 0 を取るダミー変数

ユニットの固定効果を含めた上で、全ての期の処置変数を放り込むと完全な多重共線性が生じるため、1 つ処置変数を抜く (通常処置 1 期前)

イベントスタディ変数

処置グループと比較グループの結果変数のトレンドを raw data でプロットしたもの

イベントスタディ分析による推定値と 95%信頼区間

たまにある誤解

「DiD・イベントスタディ分析で、同時性バイアス（例：警察の増員が犯罪を減少させる一方、犯罪の増加が警察の増員を招く）に対応できる」という主張は誤り。

ただし、処置前のトレンドの挙動が処置群と比較群で平行であることを確認することで、内生性の懸念が比較的小さいことを示唆することは可能（あくまでも suggestive であることを強調！）。

何と何を比較するのかを明確に！

Le Moglie and Sorrenti (2022) は、マフィアが合法的ビジネスに関与しているかを検証した。しかし、マフィアの行動そのものは内生的である。これを検証するために、マフィアの行動を外生的に変えるショックとしてサブプライム住宅ローンによる経済ショックを利用し、マフィアが活発な地域とそうではない地域での、新規企業数の変化を分析した。

Q: このイベントスタディ図はどう解釈するべきだろうか？　不景気後に（マフィアが活発な地域で）新規企業数が増えた？

サブプライム住宅ローン危機後に (ii) マフィアが多い地域はマフィアが少ない地域と比べて、**企業数の減り方が緩やか**だった（**危機の負の影響が小さい**）ことを明らかにした。

企業数が増えたではない！

別の角度からこの研究を見てみよう：

横断面のトリートメント：「マフィアの活動度合い」

縦断面トリートメント：「経済危機」

どちらの視点から考えている？

識別戦略 vs. 行政データ・・・！？　Massenkoff and Rose (ming)

個人の行動に対して外的な要素を探ると述べたが、個人レベルの処置変数と個人レベルの結果変数をイベントスタディで見ることもしばしばある。ただし、その場合、行政データなど質の高いデータが求められる可能性がある。

Massenkoff and Rose (ming)：

ワシントン州の行政データを用いて、犯罪リスクの高い個人が子供を出産すると犯罪をしなくなることを明らかにした。

ここで考えられる懸念は：

犯罪行動の指標として逮捕データを用いる：逮捕データは、個人の実際の犯罪行動 ¸ 警察の行動

仮に出産したことにより治安の良い近隣に引っ越す場合、取り締まられる確率も低くなるため、警察の行動を含んだ推定値となる

プレ・ポストの期間をどうするか問題

処置前の期間：反実仮想（比較群）として適切か？

- 処置直前：処置直後に対して良い比較対象
- 十分に長い処置前期間：
  - 利点：十分に長いプレトレンド期間がある方が一般的には良い
  - 欠点：長すぎる場合、反実仮想として適切でない可能性
  - 例：1980 年代に人種関連の処置があり、結果変数も人種に関連するもの。ただし、1960 年代に公民権運動がある場合は、1960 年以前は使えない。

処置後の期間：長期効果を考慮する場合

- 経済学では半世紀・数世紀にわたる長期効果の研究が盛んに行われている
- 欠点：平行トレンド仮定が成立しない可能性がある
  - 特に処置を受けるユニット数が少ない場合に顕著
  - 理由：長期間になればなるほど、別のショックによって結果変数の軌道が変わる可能性が高まる。処置を受けるユニット数が少ない場合、このような軌道の変化がより顕著になる。サンプルサイズが小さいため、個々のユニットの特異な変化が全体の傾向に大きな影響を与える。
- 長期効果を測る代替的なデザイン

RDD (Dell, 2010), zero-stage を利用した（Bartik 型）IV (Sequeira et al., 2020; Kamada, 2024), 一般的な IV (Acemoglu et al., 2001; Ang, 2023)

プレ・ポストの期間をどうするか問題

Q: プレトレンドのチェックが平行トレンド仮定の間接的な検証となると述べた一方で、処置後の期間が長いと平行トレンドが成り立たない可能性も指摘した。これらの記述（一方は処置前、もう一方は処置後に関する）は矛盾するか？

A: これらの記述は矛盾しない。

プレトレンドのチェック：処置前のトレンドを見ることで、処置群と比較群の結果変数が同じように変化しているかを確認。ここで言えることは、あくまでも「処置群と比較群の結果変数が処置前に同じように変化」していれば、**仮にも処置がなければ、潜在的には同じように変化し続けるでだろう**を意味する。したがって、プレトレンドのチェックは、平行トレンドを間接的に検証するための有効なツールではあるが、平行トレンドを保証するものではない。

時間を通じて変動する交絡要因

ユニットの固定効果を含めることで、時間を通じて変動しない観測されない異質性を統制することができた。ただし、DiD は**時間を通じて変動する交絡要因はない**ことを仮定している。時間を通じて変動する交絡要因に対する対策法をいくつか紹介しよう。

共変量を含める：

欠点： 固定効果が入っている以上、共変量は時間と共に変動するものである必要がある ($Z_{i,t}$)。しかし、共変量そのものが処置の影響を受けてしまう可能性があると、bad control (post-treatment バイアス）となる。

対策： 処置が導入される前の共変量の値に固定し（時間で変動しない共変量とする)、時間固定効果と交差項を取る (e.g., $Z_i \quad Time\ FE$)。かなりゴリゴリに統制される（demanding である)。

時間を通じて変動する交絡要因

ユニット別のトレンド（unit-specific time trends）を含める（$\sum_i \mathbb{1}_i \times t$、ただし、$t$ は線形トレンド）

利点：観測されない（線形）トレンドを統制する。平行トレンドが成り立たない時に有効。

欠点：処置後のトレンド（DiD で関心のあるトレンド！）さえも統制してしまい、過剰統制バイアスを引き起こす。数年前まではよく使われていたが、最近では以前ほど推奨されない。

解決法：Goodman-Bacon (2021) の detrend 方法：処置以前のパネルデータを用いてユニット固有のトレンドを推定し、全期間のパネルからフィットさせた分を引く。しかし、detrend された変数は、generated regressor となるので、仮説検定（inference）に関しては注意が必要。

DDD：次のページで詳しく見ていこう。

DDD: 時間を通じて変動する交絡要因への対処法

差分の差分の差分法 (Difference-in-Differences-in-Differences, DDD, Triple-D):

処置が起きた年に別のショックがある場合、前のページの対応策では対処不可能。ただし、関心のある処置は、ある集団にのみ導入されているが、もう片方のショックは全ての集団に影響がある場合) ある条件下で DDD が使える

（トレンドの）交絡がある = 平行トレンドが成り立たない) ある条件下で DDD が使える

その条件とは： 処置内（例えば地域）である集団（グループ A）は処置を受け、別の集団（グループ B）は処置を受けないとする。この場合、グループ B に対して zero effect assumption を仮定すれば、DDD が使える。

) 二集団のバイアスした DD 推定量の差は、不偏推定量となる (Olden and Møen, 2022)

DDD: 時間を通じて変動する交絡要因への対処法

DDD: 時間を通じて変動する交絡要因への対処法

（誤った DDD の概念図）Hoffmann and Velasco (2023). "Policy Effects on Mixed-Citizenship, Same-Sex Unions: A Triple-Difference Analysis." Social Forces.

DDD: 時間を通じて変動する交絡要因への対処法

シンプルなケースを考えよう。地域 i, 集団 j、時点 t とする：

$$y_{i,j,t} = U_0 \text{ , } U_1 D_i \text{ , } U_2 Post_t \text{ , } U_3 G_j \text{ , } U_4 D_i \quad G_j \text{ , } U_5 G_j \quad Post_t \text{ , } U_6 D_i \quad Post_t \text{ , } U_7 D_i \quad G_j \quad Post_t \text{ , } n_{i,j,t}. \tag{10}$$

DD with treatment effect heterogeneity：U_6 と U_7 をどう解釈する？

U_6: $G_j = 0$ における処置効果

U_7: $G_j = 0$ における処置効果が、$G_j = 1$ では U_7 だけ異なる（G による処置効果の異質性: treatment effect heterogeneity）。$G_j = 1$ の処置効果：U_6 , U_7

DDD：U_6 と U_7 をどう解釈する？

U_6: （zero effect assumption より）処置群を受けた地域 i における時間で変動する交絡要因 (関心のある処置以外に発生した別のショック)

U_7: $G_j = 1$ における処置効果！

U_6 が交絡要因である以上、U_6 , U_7 をする意味はない

（誤った DDD の解釈） Goldstein and Hamilton. (2022). “Dog Whistles and Work Hours: The Political Activation of Labor Market Discrimination.” Sociological Science.

DDD: 時間を通じて変動する交絡要因への対処法

DDD における平行トレンドとは？

集団間（G_j）の結果変数のトレンドが、処置地域と比較地域で潜在的に同じ

満たされなければならない平行トレンドは 1 つ！

それでは $G_j = 0$ が処置を受けていないことを強く正当化できない場合は？

DD with treatment effect heterogeneity

$G_i = 0$ の DD 推定値 (U_6) と $G_j = 1$ の DD 推定値 (U_6 , U_7) を求める。

Q: 満たすべき平行トレンドは？

A: $G_j = 0$ における平行トレンド、$G_j = 1$ における平行トレンドの 2 つ。

DDD：追加的な交絡要因を消す目的の場合

DDD が使える条件：$G_j = 0$ の処置効果がゼロ（プラセボグループ）であることを事前に正当化する必要がある

計量経済学的に U_6 が (i) $G_j = 0$ の処置効果なのか、(ii) 処置を受けた地域特有の時間で変動する交絡なのかを識別することは不可能

）解釈をするために、制度的な背景（ドメインスペシフィックな知識）や理論が必要！

識別仮定：

DDD の方が、DD with treatment effect heterogeneity よりも仮定が弱い

制度的背景：

DDD の方が、DD with treatment effect heterogeneity よりも仮定が強い

実は IV の exclusion restriction と同じくらい強い仮定かもしれない (Besley and Case, 2000)。

例： Kamada (2023) は、クラックコカインという、黒人コミュニティにて問題が顕在化した（白人コミュニティでは顕在化がほとんどされなかった）薬物が 1980 年代にアメリカの都市部に蔓延した直後、黒人の郊外化が 20 年以上にわたって長期的に進行し、約 100 万人の黒人の大規模な移動が起きたことを、DDD を用いて示した。