

パネルデータ: Advanced

川田恵介

2025-07-07

1 推定: 2 × Many case

1.1 2 × Many case

- 2 × 2の議論は、多期間に拡張できる
 - ▶ ずっと $D = 0$ のグループ VS 途中で $D = 1$ に切り替わったグループ
 - ▶ Event study と呼ばれる推定方法が活用可能

1.2 Event study

- 4 期間パネルで、Treatment Group に対して、3 期目に介入が入るのであれば、

$$E[Y | Z_{it}] = \beta_1 Z_1 + \underbrace{\beta_2}_{=0 \text{ と基準化}} Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + f_i + f_t$$

- $Z_t = \text{Treatment Group}$ かつ t 期目であれば 1、それ以外であれば 0
 - ▶ Control Group であれば、常に 0

1.3 Event study

- 識別の仮定のもとで、以下の式を推定すれば、動学効果、および Parallel trends のチェックができる

$$E[Y | Z_{it}] = \underbrace{\beta_1}_{\text{Parallel Trend のもとで } = 0} Z_1 + \beta_3 Z_3 + \beta_4 Z_4 + f_i + f_t$$

1.4 別表現

- 別の表現もよく用いられる
- E_i : 個人 i に対して介入が行われる時期
 - ▶ Control group については、 $E_i = \infty$

1.5 別表現

- $\beta_{-1} = 0$ と基準化:

$$E[Y_{it} \mid E_i = e] = \sum_{l=l_-}^{l_+} \beta_l \mathbb{I}[t - e = l] + f_i + f_t$$

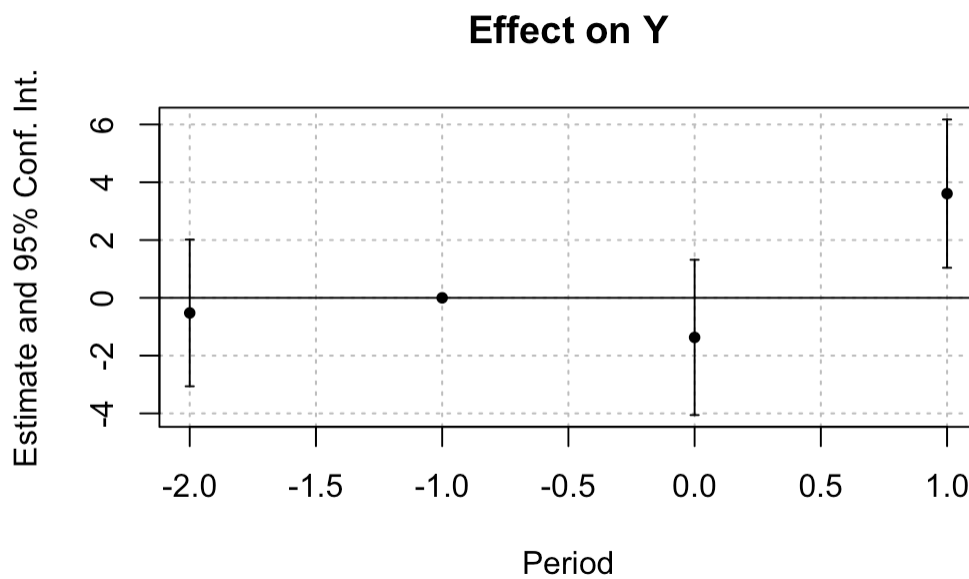
- $\mathbb{I}[t - e = l] = t - e$ が l であれば 1、それ以外は 0
- l_-, l_+ = 研究者が定める分析期間

1.6 R Example

```
# |
DataLong = mutate(
  DataLong,
  Z = case_when(
    TreatGroup == 1 ~ 3,
    .default = 100000
  )
)

Model = fixest::feols(
  Y ~ sunab(Z, Period) | Period + ID,
  DataLong,
  panel.id = "ID"
)
```

1.7 R Example



2 *Many* × *Many*

2.1 一般化の限界

- *Many* × *Many* case (介入が開始するタイミングが3つ以上存在) での推定方法は、まだよく分かっていない
 - ▶ 推定モデルの正しさに、結果が強く依存
- 例外は、介入が徐々に行われるケース (Staggered design)
 - ▶ 直近で研究が進む

2.2 応用例: Marriage premium/penalty

- 研究対象: 結婚”経験”が労働供給に与える影響を推定したい
- データ: 徐々に結婚していく
 - ▶ 離婚したとしても、結婚経験がある $D = 1$ と定義する
 - ▶ Control group は一度も結婚したことがないグループ

2.3 問題点

- 2×2 であれば、識別の仮定(Pallael trends 等)の下で、Treatment Group 内平均効果の信頼区間を提供
- *many* × *many*以外では?
 - ▶ 識別のための仮定が成り立っていても、推定のための単純化が不適切な比較を生み出す
 - ▶ 個別因果効果が全て正でも、負の平均効果が推定されてしまう可能性も存在

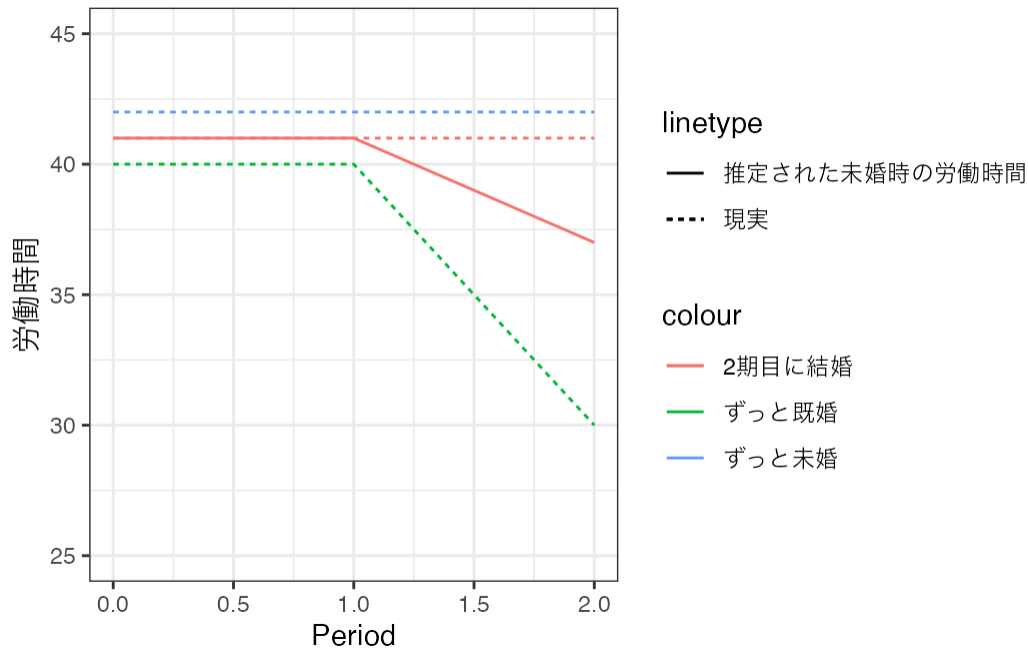
2.4 Simple Example

- 2 期間モデル: 結婚が就業状態に与える因果効果を推定するために、以下を比較
 - ▶ Treatment Group: 2 期目に結婚
 - ▶ Control Groups: ずっと未婚 ずっと既婚
- Two-ways Fixed Effect Model を推定すると、何らかの値が算出されるが、基本的に不適切

2.5 Simple Example

- シナリオ:
 - ▶ 結婚した期には、因果効果がほとんどない
 - ▶ 2 期目以降に労働時間を低下させる

2.6 例



- 2期目に結婚したグループの因果効果を推定する際に、**ずっと既婚グループ**も Control group として使用してしまう

2.7 問題点

- 識別の仮定ではなく、推定のために導入された仮定 (Two way fixed effect model)が問題
- D が変化していない(介入が生じていない 生じた後)期間を全て Control group として使ってしまう
- 2×2 では、最初から介入が生じているの事例を削除すれば良い
- 多期間の場合については、一般的な方法はよくわかっていない
 - ▶ 例外ケース: Staggered design

2.8 異質性を考慮した推定

- Sun & Abraham (2021) のアプローチを紹介
 - ▶ コホート別動学効果を集計する
- 他の手法としては、Roth et al. (2023), Miller (2023), De Chaisemartin & d'Haultfoeuille (2023) を参照

2.9 コホート別動学効果

- $$E[Y_{i,t} | e, l, f_i] = \sum_l \beta_{e,l} \mathbb{I}(e, t - e = l) + f_i + f_t$$
 - ▶ $\beta_{e,l}$ e 期目に介入を受けるグループにおける、介入発生から l 期経過した場合の因果効果
 - ▶ $\beta_{e,-1} = 0$ と基準化
- 介入発生からの期間 l と介入を受ける時期 e (コホート) に応じて、異なる平均効果を便宜的に推定

2.10 推定

- 各コホート e と control group を使って、コホート単位で平均動学効果を推定する。

2.11 例. コホート期間別平均効果

```
Model = feols(  
  Y ~ sunab(Group, Period) | ID + Period,  
  DataLong,  
  cluster = ~ID  
)
```

```
summary(Model, agg = FALSE)
```

```
OLS estimation, Dep. Var.: Y  
Observations: 4,000  
Fixed-effects: ID: 500, Period: 4  
Standard-errors: Clustered (ID)  
  
              Estimate Std. Error  t value  Pr(>|t|)  
Period::-3:cohort::4  0.900886    0.817159  1.102461 2.7079e-01  
Period::-2:cohort::3  0.659769    0.870363  0.758038 4.4879e-01  
Period::-2:cohort::4  0.392347    0.805612  0.487017 6.2646e-01  
Period::0:cohort::3   1.780149    0.776116  2.293664 2.2225e-02 *  
Period::0:cohort::4   1.440264    0.836474  1.721829 8.5720e-02 .  
Period::1:cohort::3   5.230464    0.834848  6.265167 8.0526e-10 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
RMSE: 9.33386      Adj. R2: 0.001936  
Within R2: 0.01243
```

2.12 期間別平均効果

- $\beta_{e,l}$ の e についての平均値

```
summary(Model, agg = "period")
```

```

OLS estimation, Dep. Var.: Y
Observations: 4,000
Fixed-effects: ID: 500, Period: 4
Standard-errors: Clustered (ID)
      Estimate Std. Error t value Pr(>|t|)
Period::-3 0.900886    0.817159 1.102461 2.7079e-01
Period::-2 0.527643    0.595945 0.885389 3.7637e-01
Period::0  1.612222    0.581585 2.772119 5.7774e-03 **
Period::1  5.230464    0.834848 6.265167 8.0526e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 9.33386      Adj. R2: 0.001936
                Within R2: 0.01243

```

2.13 コホート別平均効果

- $\beta_{e,l}$ の l についての平均値

```
summary(Model, agg = "cohort")
```

```

OLS estimation, Dep. Var.: Y
Observations: 4,000
Fixed-effects: ID: 500, Period: 4
Standard-errors: Clustered (ID)
      Estimate Std. Error t value Pr(>|t|)
cohort::3  3.50531    0.619815 5.65541 2.6211e-08 ***
cohort::4  1.44026    0.836474 1.72183 8.5720e-02 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 9.33386      Adj. R2: 0.001936
                Within R2: 0.01243

```

2.14 平均効果

- $\beta_{e,l}$ の平均値

```
summary(Model, agg = "ATT")
```

```

OLS estimation, Dep. Var.: Y
Observations: 4,000
Fixed-effects: ID: 500, Period: 4
Standard-errors: Clustered (ID)
      Estimate Std. Error t value Pr(>|t|)
ATT    2.8278    0.544532 5.19308 3.0176e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

RMSE: 9.33386 Adj. R2: 0.001936
Within R2: 0.01243

2.15 コントロール変数との併用

- 個人内で変化する変数 X は導入可能 $\beta_1 X_{1,it} + \dots \beta_L X_{L,it}$
- 問題点: Cross section と同様に、介入の影響を受けない変数のみ導入すべき
 - ▶ 過去のイベントの影響を受けない変数とは???
- Callaway & Sant'Anna (2021) : 時間を通じて変化しない変数 (生まれ年、性別等) について、“マッチング”し推定
 - ▶ did

2.16 まとめ

- 介入変数 D が2値の場合、推奨は
 - ▶ 2 期間パネル: ずっと介入を受け続けているグループを排除し、通常の Two-way fixed effect model で推定
 - ▶ 多期間パネル: Staggered design になっているか確認し、問題なければ Sun & Abraham (2021) (他の手法は、Roth et al. (2023), Miller (2023), De Chaisemartin & d'Haultfoeuille (2023) などを参照) などの手法を用いて推定
- D が連続変数のケースなどの一般化は可能? (Callaway et al., 2024)

3 Reference

Bibliography

- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). Difference-in-differences with a continuous treatment.
- De Chaisemartin, C., & d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26(3), C1–C30.
- Miller, D. L. (2023). An introductory guide to event study models. *Journal of Economic Perspectives*, 37(2), 203–230.
- Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.