

# 計量経済学特論 B 不明点

前川 大空

2025 年 12 月 10 日

## 1 Introduction to Causal Inference

■p.7 Simpson's paradox あるデータを全体で見ると一方の傾向が現れるのに、グループに分けて詳しく見ると逆の傾向が出てしまう現象。

p.8 Ex. ワクチン 基礎免疫という交絡因子が邪魔

■p.10 DiD について DevEcon とかはこの修正に近年悩まされている？平行トレンドの不適切性？

p.11 Ex. Card and Krueger (1994) 労働経済学では、最低賃金引上げは企業退出により失業率を上昇させるとする。理論と NJ のデータあってなくて、原因は平行トレンドの不成立？

■p.13 Shigeoka (2014) 医療費負担率の設計によるモラルハザードの説明を RDD で。被説明変数がいろいろ変わっていて、Overall outpatient visits のジャンプだけが大きいことがその証左になっていることを確認せよ。『制度の記述』から始まるることも注目しよう。

■p.18 線形の特定化

$$Y_i = \alpha_i + \beta_i D_i + \varepsilon_i \quad \text{where } D_i \in \{0, 1\}$$

$\beta_i$  のインデックスからも分かるように、効果には異質性が許容される。 $Y_i(1)$ ,  $Y_i(0)$  は簡約表現に過ぎない。

■Exercise 1.1 セレクションバイアスとその原因を明らかにする練習問題だ。

- $\hat{\beta}$  の確率極限を求め、 $Y_i(1) - Y_i(0)$  と比較してこれを解釈せよ。
- $(Y_i(1), Y_i(0)) \not\perp D_i$  の下で、確率極限を処置効果とセレクションバイアスに分解せよ。

*Proof.*  $D_i \in \{0, 1\}$  は処置割当てを表す。 $Y_i = \alpha + \beta D_i + u_i$  で OLS 推定を行うと、 $\hat{\beta}$  の確率極限は：

$$\text{plim } \hat{\beta} = \frac{\text{Cov}(D_i, Y_i)}{\text{Var}(D_i)}.$$

ここで  $Y_i$  に潜在結果の表現  $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$  を代入すると：

$$\text{Cov}(D_i, Y_i) = \text{Cov}(D_i, Y_i(0)) + \text{Cov}(D_i, (Y_i(1) - Y_i(0)) D_i).$$

次に  $\tau_i := Y_i(1) - Y_i(0)$  と置くと：

$$\begin{aligned} \text{Cov}(D_i, \tau_i D_i) &= \mathbb{E}[D_i^2 \tau_i] - \mathbb{E}[D_i] \mathbb{E}[D_i \tau_i] = \mathbb{E}[D_i \tau_i] - \mathbb{E}[D_i] \mathbb{E}[D_i \tau_i] \\ &= \mathbf{P}(D_i = 1)(1 - \mathbf{P}(D_i = 1)) \mathbb{E}[\tau_i | D_i = 1] \end{aligned} \quad \because D_i^2 = D_i$$

最後の変形は  $\mathbb{E}[D_i \tau_i] = \mathbb{E}[\tau_i | D_i = 1] \mathbf{P}(D_i = 1)$ ,  $\mathbb{E}[D_i] = \mathbf{P}(D_i = 1)$  を利用した。同様に：

$$\text{Cov}(D_i, Y_i(0)) = \mathbb{E}[D_i Y_i(0)] - \mathbb{E}[D_i] \mathbb{E}[Y_i(0)] = \mathbf{P}(D_i = 1) (\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)])$$

一方：

$$\text{Var}(D_i) = \mathbf{P}(D_i = 1)(1 - \mathbf{P}(D_i = 1))$$

であるから、これを用いると：

$$\frac{\text{Cov}(D_i, Y_i(0)) + \text{Cov}(D_i, (Y_i(1) - Y_i(0))D_i)}{\text{Var}(D_i)} = \mathbb{E}[\tau_i | D_i = 1] + \frac{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)]}{1 - \mathbf{P}(D_i = 1)}.$$

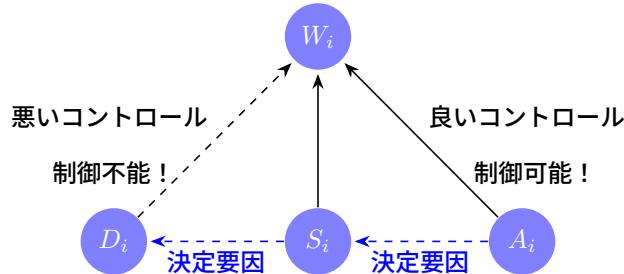
LIE による  $\mathbb{E}[Y_i(0)] = \mathbf{P}(D_i = 1)\mathbb{E}[Y_i(0) | D_i = 1] + (1 - \mathbf{P}(D_i = 1))\mathbb{E}[Y_i(0) | D_i = 0]$  から：

$$\frac{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)]}{1 - \mathbf{P}(D_i = 1)} = \mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]$$

$$\text{plim } \hat{\beta} = \underbrace{\mathbb{E}[\tau_i | D_i = 1]}_{\text{処置群の平均効果}} + \underbrace{(\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0])}_{\text{選択バイアス}} = \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]$$

従って確率極限は「処置群における平均処置効果 (ATT)」と「選択バイアス項」の和に分解できる。  
 $(Y_i(1), Y_i(0)) \not\perp\!\!\!\perp D_i$  の下では第二項は消えず、 $\hat{\beta}$  は  $\beta$  の一致推定量とはならない。□

■p.22 bad control 語の定義としては、一般的には 興味のある説明変数が決定要因になるようなコントロール変数のことなのでズレてる気がする。



## 2 Randomized Control Trials

■p.2 RCT の意義 セレクションバイアスは  $(Y_i(1), Y_i(0)) \not\perp\!\!\!\perp D_i$  の下で発生する

$$\begin{aligned} \text{plim } \hat{\beta} &= \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0] \\ &= \underbrace{\mathbb{E}[\tau_i | D_i = 1]}_{\text{処置群の平均効果}} + \underbrace{(\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0])}_{\text{選択バイアス}} \end{aligned}$$

これを消すための方法として第一に考えられることはランダムな割当：

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$$

$$\mathbf{P}(Y_i(t) \leq y | D_i = 0) = \mathbf{P}(Y_i(t) \leq y | D_i = 1) \quad \forall y \in \text{supp}(Y_i)$$

■Brook et al. (1983) より良い保険を提供することで、家計の健康状態は改善されるかだろうか？これに答えを与えるため、米国の家族に異なる保険プランをランダムに割り当てる。

p.5 結果の解釈 不明。

■p.8 ネイマン推定量 RCT のうち、特に完全無作為化実験で、処置・統制群の平均の差を推定量とする：

$$\tau_{\text{dif}} \equiv \frac{1}{N_1} \sum_{i=1}^N Y_i \mathbb{1}\{D_i = 1\} - \frac{1}{N_0} \sum_{i=1}^N Y_i \mathbb{1}\{D_i = 0\} \quad \text{where } N_d = \sum_{i=1}^N \mathbb{1}\{D_i = d\}$$

このとき、Neyman 推定量は次の性質をもつ：

$$\mathbb{E}_D[\tau_{\text{dif}}] = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)).$$

■p.9 標本平均となる理由 標本平均である理由は、抽出に関する不確実性を考慮していないため。

- 標本 ATE  $\frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$ : 固定された標本 での ATE
- 母集団 ATE  $\mathbb{E}[Y_i(1) - Y_i(0)]$ : 母集団全体 での ATE

したがって、Neyman 推定量の不偏性は、所与の標本での割当  $D_i$  のランダム性のもとで不偏。母集団で不偏とは限らない。i.i.d. 抽出の場合は次が成り立つ：

$$\mathbb{E}[\tau_{\text{dif}}] = \mathbb{E}[Y_i(1) - Y_i(0)],$$

ここで  $\mathbb{E}[\cdot]$  は 抽出・割当両方 で取られ、推定量は母集団 ATE でも不偏。

■p.11 線形回帰での代用 共変量の表現が容易で、かつネイマン推定量を構成することが可能であるため、一般的には線形回帰を利用する。現状は標本数は固定されているため、大標本理論（一致性）は考慮せず、推定量の 不偏性 のみを考える。

■p.12 回帰による RCT の推定 ランダム化実験において、単純な回帰モデル

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

を考える。ここで  $D_i \in \{0, 1\}$  は処置割当てであり、潜在結果表記で

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

が成り立つ。OLS 推定量  $\hat{\tau}$  は標本共分散／分散の比で表される：

$$\hat{\tau} = \frac{\widehat{\text{Cov}}(D_i, Y_i)}{\widehat{\text{Var}}(D_i)} = \frac{\frac{1}{n} \sum_{i=1}^n D_i Y_i - \left(\frac{1}{n} \sum_{i=1}^n D_i\right) \left(\frac{1}{n} \sum_{i=1}^n Y_i\right)}{\frac{1}{n} \sum_{i=1}^n D_i^2 - \left(\frac{1}{n} \sum_{i=1}^n D_i\right)^2}.$$

不偏性を示そう。

*Proof.* まず次の母集団における期待値を考える：

$$\mathbb{E}[D_i Y_i] = \mathbb{E}[D_i (D_i Y_i(1) + (1 - D_i) Y_i(0))] = \mathbb{E}[D_i Y_i(1)] \quad \because D_i(1 - D_i) = 0.$$

同様に全体の期待値は

$$\mathbb{E}[Y_i] = \mathbb{E}[D_i Y_i(1)] + \mathbb{E}[(1 - D_i) Y_i(0)] = \mathbb{E}[D_i] \mathbb{E}[Y_i(1)] + \mathbb{E}[1 - D_i] \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$$

これらを用いて分子 (の母集団における期待値) を計算する:

$$\begin{aligned} \mathbb{E}[D_i Y_i] - \mathbb{E}[D_i] \mathbb{E}[Y_i] &= \mathbb{E}[D_i Y_i(1)] - \mathbb{E}[D_i](\mathbb{E}[D_i Y_i(1)] + \mathbb{E}[(1 - D_i) Y_i(0)]) \\ &= \mathbb{E}[D_i] \mathbb{E}[Y_i(1)] - \mathbb{E}[D_i](\mathbb{E}[D_i] \mathbb{E}[Y_i(1)] + (1 - \mathbb{E}[D_i]) \mathbb{E}[Y_i(0)]) \\ &= \mathbb{E}[D_i](1 - \mathbb{E}[D_i])(\mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]). \end{aligned}$$

分母は

$$Var[D_i] = \mathbb{E}[D_i](1 - \mathbb{E}[D_i]).$$

$$\mathbb{E}[\hat{\tau}] = \frac{\mathbb{E}[D_i Y_i] - \mathbb{E}[D_i] \mathbb{E}[Y_i]}{Var[D_i]} = \mathbb{E}[Y_i(1) - Y_i(0)].$$

が成り立つ.  $\hat{\tau}$  は不偏推定量である.  $\square$

### p.13 別解: 条件期待値を使った直観

*Proof.* 条件付き期待値を用いるとより直観的に説明できる.

$$\varepsilon_i \equiv Y_i - \mathbb{E}[Y_i | D_i]$$

とおくと,

$$\begin{aligned} \varepsilon_i &= Y_i - \mathbb{E}[Y_i(1) | D_i = 1] D_i - \mathbb{E}[Y_i(0) | D_i = 0] (1 - D_i) \\ &= Y_i - \mathbb{E}[Y_i(1) - Y_i(0)] D_i - \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i. \end{aligned}$$

つまり条件付き期待値関数は線形の形を取り,

$$\mathbb{E}[Y_i | D_i] = \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0)] D_i,$$

これは回帰モデルの  $\alpha + \tau D_i$  に対応する. したがって OLS による推定は不偏性を満たす.  $\square$

### ■p.14 Q. ネイマン推定量と OLS は同等か? どちらも不偏推定量, 不均一分散下で同等.

p.14 Ex. 完全無作為化実験: ネイマン分散推定量は頑健 OLS 分散推定量と同等.

### ■p.15 ネイマン推定量との違い 共変量が入れられること. 以下の (真の) 構造モデルを考えよう:

$$Y_i = \alpha + \tau D_i + X_i \beta + \varepsilon_i$$

$X_i$  は共変量で  $(Y_i(d), X_i) \perp\!\!\!\perp D_i$  とすると, 過小定式化により得られる  $\hat{\tau}$  の OLSE は  $\mathbb{E}[Y_i(1) - Y_i(0)]$  に対し不偏性を持たないが, 一致性は保たれる. これは, 欠落変数バイアス (確率収束先のバイアス) が  $Cov(X_i, D_i)$  の形で表現されるためである. 次の練習問題で確認しよう:

■p.16 Exercise 2.1 ネイマン推定量と同等な  $\hat{\tau}$  は, 共変量を落とした過小定式化に基づく推定量として得られて,  $(Y_i(d), X_i) \perp\!\!\!\perp D_i$  によって, 真の ATE との間に一致性は保持される. 真のデータ生成過程が:

$$Y_i = \alpha + \tau D_i + X_i \beta + \varepsilon_i, \quad (\varepsilon_i, X_i) \perp\!\!\!\perp D_i$$

を満たすとする。ここで  $D_i$  は二値の確率変数。誤った定式化（過小定式化）

$$Y_i = \tilde{\alpha} + \tilde{\tau} D_i + \tilde{\varepsilon}_i$$

で得られる OLS 推定量  $\hat{\tau}$  は  $\tau$  に一致性を持つことを示せ。

*Proof.* 真のモデル  $Y = \mathbf{1}\alpha + D\tau + X\beta + \varepsilon$  として、過小定式化に基づく OLSE  $\hat{\tau}$  は：

$$\begin{aligned} (\hat{\alpha}, \hat{\tau})^\top &= (Z^\top Z)^{-1} Z^\top Y = (Z^\top Z)^{-1} Z^\top (\mathbf{1}\alpha + D\tau + X\beta + \varepsilon) \\ &= (Z^\top Z)^{-1} Z^\top (\mathbf{1}\alpha + Z^\top D + X\beta) + (Z^\top Z)^{-1} Z^\top \varepsilon \\ &= (Z^\top Z)^{-1} (\mathbf{1}\alpha + Z^\top D + X\beta) + (Z^\top Z)^{-1} Z^\top \varepsilon \\ &= \left(\frac{Z^\top Z}{n}\right)^{-1} \left(\frac{Z^\top \mathbf{1}\alpha + Z^\top D\tau + Z^\top X\beta}{n}\right) + \left(\frac{Z^\top Z}{n}\right)^{-1} \frac{Z^\top \varepsilon}{n} \end{aligned}$$

ここで、 $Z = [\mathbf{1} \ D]$ ,  $\gamma = [\alpha, \tau]^\top$  は過小定式化で用いた説明変数行列で：

$$Z^\top Z = \begin{pmatrix} \mathbf{1}^\top \mathbf{1} & \mathbf{1}^\top D \\ D^\top \mathbf{1} & D^\top D \end{pmatrix}, Z^\top D = \begin{pmatrix} \mathbf{1}^\top D \\ D^\top D \end{pmatrix}, Z^\top \mathbf{1} = \begin{pmatrix} \mathbf{1}^\top \mathbf{1} \\ D^\top \mathbf{1} \end{pmatrix}, Z^\top X = \begin{pmatrix} \mathbf{1}^\top X \\ D^\top X \end{pmatrix}, Z^\top \varepsilon = \begin{pmatrix} \mathbf{1}^\top \varepsilon \\ D^\top \varepsilon \end{pmatrix}.$$

仮定  $(\varepsilon_i, X_i) \perp\!\!\!\perp D_i$  から  $\mathbb{E}[X_i D_i] = 0$ ,  $\mathbb{E}[D_i \varepsilon_i] = 0$  であって、

$$\begin{aligned} \mathbb{E}[Z^\top Z] &= \begin{pmatrix} n & np \\ np & np \end{pmatrix}, \mathbb{E}[Z^\top D] = \begin{pmatrix} np \\ np \end{pmatrix}, \mathbb{E}[Z^\top \mathbf{1}] = \begin{pmatrix} n \\ np \end{pmatrix}, \mathbb{E}[Z^\top X] = \begin{pmatrix} n\mathbb{E}[X_i] \\ \mathbf{0} \end{pmatrix}, \mathbb{E}[Z^\top \varepsilon] = \mathbf{0} \\ \mathbb{E}[Z^\top Z]^{-1} &= \frac{1}{n^2 p - n^2 p^2} \begin{pmatrix} np & -np \\ -np & n \end{pmatrix} = \frac{1}{n(1-p)} \begin{pmatrix} p & -p \\ -p & 1 \end{pmatrix} \quad \text{if } p \in (0, 1), \end{aligned}$$

ここで、 $p = \mathbb{E}[D_i]$ . LLN より：

$$\begin{aligned} \operatorname{plim} \begin{pmatrix} \hat{\alpha} \\ \hat{\tau} \end{pmatrix} &= (\mathbb{E}[Z^\top Z])^{-1} (\mathbb{E}[Z^\top \mathbf{1}]\alpha + \mathbb{E}[Z^\top D]\tau + \mathbb{E}[Z^\top X]\beta) + (\mathbb{E}[Z^\top Z])^{-1} \mathbb{E}[Z^\top \varepsilon] \\ &= (\mathbb{E}[Z^\top Z])^{-1} (\mathbb{E}[Z^\top \mathbf{1}]\alpha + \mathbb{E}[Z^\top D]\tau) \\ &= \frac{1}{n(1-p)} \begin{pmatrix} p & -p \\ -p & 1 \end{pmatrix} \begin{pmatrix} n\alpha + np\tau \\ np\alpha + np\tau \end{pmatrix} \\ &= \frac{1}{n(1-p)} \begin{pmatrix} \cdot \\ -p(n\alpha + np\tau) + (np\alpha + np\tau) \end{pmatrix} = \begin{pmatrix} \cdot \\ \tau \end{pmatrix} \end{aligned}$$

題意は示された。FWL 定理を用いるならば最初の  $\tilde{\alpha}$  に関して利用すればよかったわけだ。□

**■補足: FWL 定理** 真の構造モデルにおける  $D_i$  の係数  $\tau$  は、次の手順で得られる係数と一致する：

1.  $Y_i$  を  $1, X_i$  に回帰して残差  $\tilde{Y} = M_{1,X}Y = Y - P_{1,X}Y$  を得る ( $P_{1,X}$ : 射影行列)
2.  $D_i$  を  $1, X_i$  に回帰して残差  $\tilde{D}_i = M_{1,X}D_i$  を得る
3.  $\tilde{Y}_i$  を  $\tilde{D}_i$  に回帰したときの係数が、 $\tau$  と等しい

即ち以下が成り立つ：

$$\hat{\tau} = (D^\top M_{1,X} D)^{-1} D^\top M_{1,X} Y.$$

*Proof.* FWL 定理を用いて示そう (期待値が分からぬ):

$$\begin{aligned}\tilde{Y}_i &= Y_i - \bar{Y}, \quad \tilde{\epsilon}_i^* \equiv Y_i - X_i(X_i^\top X_i)^{-1}X_i^\top Y_i = M_X Y_i \\ \hat{\tau}_i^* &\xrightarrow{p} \frac{\mathbb{E}[(M_X \tilde{D}_i)^\top (M_X Y_i)]}{\mathbb{E}[(M_X \tilde{D}_i)^\top (M_X \tilde{D}_i)]} = \frac{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{Y}_i]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} \\ &= \frac{\mathbb{E}[\tilde{D}_i^\top M_X(\tau \tilde{D}_i + \tilde{X}_i \beta + \varepsilon_i)]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} = \tau + \frac{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{X}_i]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} = \tau \\ \because \mathbb{E}[\tilde{D}_i M_X \tilde{X}_i] &= \mathbb{E}[\tilde{D}_i(I - \tilde{X}_i(\tilde{X}_i^\top \tilde{X}_i)^{-1}\tilde{X}_i^\top)\tilde{X}_i] = 0 \\ \hat{\tau} &\xrightarrow{p} \frac{\mathbb{E}[\tilde{D}_i \tilde{Y}_i]}{\mathbb{E}[\tilde{D}_i^2]} = \frac{\mathbb{E}[\tilde{D}_i(\tau \tilde{D}_i + \tilde{X}_i^\top \beta + \varepsilon_i)]}{\mathbb{E}[\tilde{D}_i^2]} = \tau + \beta \frac{\mathbb{E}[\tilde{D}_i \tilde{X}_i^\top]}{\mathbb{E}[\tilde{D}_i^2]} = \tau (\because X \perp\!\!\!\perp D)\end{aligned}$$

□

■p.17 Estimations with RCT: Regression with Covariates 真のモデルが

$$Y_i = \alpha + \tau D_i + X_i \beta + D_i(X_i - \bar{X})\gamma + \varepsilon_i$$

で与えられるとする.  $(Y_i(d), X_i) \perp\!\!\!\perp D_i$  の下で, OLSE  $\hat{\tau}$  は ATE の一致推定量. ここで  $\mathbb{E}[Y_i(d) | X_i]$  の関数形は一致性には関係なく,  $\hat{\gamma}$  はこの線形射影と見なすことが出来る.

p.18 条件付期待関数 (CEF) と線形射影 処置効果は線形なら:

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \tau d_i + d_i(X_i - \bar{X})\gamma$$

では? OLS 推定は完全に正しい定式化である. しかし一般には,

$$\mathbb{E}[Y_i(1) - Y_i(0) | X_i = x] = \alpha(x) + \tau(x)D_i$$

のように非線形である. このとき OLS は  $\tau(x)$  の線形近似として

$$\tau(x) \approx \tau + x' \gamma$$

を推定しているとみなせる.

■p.20 共変量を追加したときの分散への影響 共変量を追加することによって推定量の分散は弱く減少する:

$$\text{Var}(Y_i - \mathbb{E}[Y_i | D_i = d]) \geq \text{Var}(Y_i - \mathbb{E}[Y_i | D_i = d, X_i = x]).$$

ただし, サンプルにおいてはトレードオフが存在する:

- 共変量を追加しても母集団分散は決して増加しない.
- 無関連な変数を追加すると標本分散はむしろ増加しうる.

証明が Exercise 2.2 で行われる.

*Proof.* 假定  $\mathbb{E}[Y_i^2] < \infty$  の下で, 次の 2 つの記法を用いる:

$$m(D_i) = \mathbb{E}[Y_i | D_i], \quad m(D_i, X_i) = \mathbb{E}[Y_i | D_i, X_i].$$

このとき示すべき不等式は

$$\text{Var}(Y_i - m(D_i)) \geq \text{Var}(Y_i - m(D_i, X_i)).$$

まず LIE より

$$m(D_i) = \mathbb{E}[m(D_i, X_i) | D_i]$$

さらに全分散の法則から:

$$\text{Var}(Y_i | D_i) = \mathbb{E}[\text{Var}(Y_i | D_i, X_i) | D_i] + \text{Var}(m(D_i, X_i) | D_i).$$

この式はすべての  $D_i$  に対して成り立つ。凸関数  $\kappa(x) = x^2$  に Jensen の不等式を適用すると:

$$\mathbb{E}[m(D_i, X_i)^2 | D_i] \geq (\mathbb{E}[m(D_i, X_i) | D_i])^2 = m(D_i)^2.$$

$$\text{Var}(m(D_i, X_i) | D_i) = \mathbb{E}[m(D_i, X_i)^2 | D_i] - m(D_i)^2 \geq 0.$$

$D_i$  について期待値を取ると:

$$\begin{aligned} \mathbb{E}[\text{Var}(Y_i | D_i)] &= \mathbb{E}[\mathbb{E}[\text{Var}(Y_i | D_i, X_i) | D_i]] + \mathbb{E}[\text{Var}(m(D_i, X_i) | D_i)] \\ &\geq \mathbb{E}[\mathbb{E}[\text{Var}(Y_i | D_i, X_i) | D_i]] = \mathbb{E}[\text{Var}(Y_i | D_i, X_i)]. \end{aligned}$$

一方,

$$\text{Var}(Y_i - m(D_i)) = \mathbb{E}[(Y_i - m(D_i))^2] = \mathbb{E}[\text{Var}(Y_i | D_i)],$$

$$\text{Var}(Y_i - m(D_i, X_i)) = \mathbb{E}[(Y_i - m(D_i, X_i))^2] = \mathbb{E}[\text{Var}(Y_i | D_i, X_i)].$$

よって,

$$\text{Var}(Y_i - m(D_i)) \geq \text{Var}(Y_i - m(D_i, X_i)).$$

$$\text{Var}(Y_i - \mathbb{E}[Y_i | D_i]) \geq \text{Var}(Y_i - \mathbb{E}[Y_i | D_i, X_i])$$

題意は示された。 □

■p.24 層化無作為化実験 等確率割当 等確率の割り当ての場合には,  $(Y_i(d), G_i) \perp\!\!\!\perp D_i$  だが  $G_i \not\perp\!\!\!\perp Y_i(d)$  で, コントロール変数と同じ扱いが出来ることを注意せよ。一致性は保たれるため, 入れても入れなくともよい。

■p.25 等確率割当でない場合 等確率割当で無ければ,  $G_i \not\perp\!\!\!\perp D_i$  であって, ダミーとして入れる必要がある。

1. ダミーとして追加する場合:  $\beta^W = \sum_{g \in \bar{G}} \mathbf{P}(G_i = g) e(g)(1 - e(g)) \beta_g$  への一致推定量
2. 交差項も追加する場合:  $\sum_{g \in \bar{G}} \beta_g \mathbf{P}(G_i = g)$  への一致推定量

後者の方が分散が大きくなるため, トレードオフがある。

■p.28 no spillover effect 適切な効果量の 範囲設定 が必要。

■p.29 Ex. Angrist and Lavy (2009) 成績優秀賞の効果分析。治療群と対照群の割り当ては 学校単位。個人に割り当てた場合, 足の引っ張り合いが起きうることが理由。

■p.32 SUTVA 以下のように定式化される:

**Assumption: SUTVA**

潜在的結果が以下のように表されることを指す。

$$Y_i = Y_i^*(Z_i) = \begin{cases} Y_i^*(1) & \text{if } D_i = 1 \\ Y_i^*(0) & \text{if } D_i = 0 \end{cases} \quad \text{where } D_i \in \{0, 1\}$$

つまり、以下の 2 条件がこの式で表現されている：

1. no spillover: 潜在的結果は他者の処置の影響を受けない (引数が  $D_i$  のみ)
2. consistency: 処置が施されたとき、観測されるのは潜在結果をその処置で評価した値
3. binary treatment: 処置が均一 (デジタルな処置)

■ p.33 Ex. Angrist and Lavy (2009) 以下のクラス内平均ならば他クラスの処置  $C'$  に依存しない：

$$\bar{Y}_C(d) := \frac{1}{n_C} \sum_{i=1}^{n_C} Y_{i,C}(d)$$

■ p.36 フィッシャーの正確確率検定 以下のように、すべての潜在結果が特定化できる帰無仮説を考える：

$$H_0^{\text{Fisher}} : Y_i(0) = Y_i(1) \quad \forall i \in \{1, \dots, N\}$$

この下では  $Y_i = Y_i(d)$  は固定されるため、以下は  $D_i$  に依存しない：

$$T(D, Y) = \left| \frac{\sum_i Y_i \mathbb{1}\{D_i = 1\}}{\sum_i \mathbb{1}\{D_i = 1\}} - \frac{\sum_i Y_i \mathbb{1}\{D_i = 0\}}{\sum_i \mathbb{1}\{D_i = 0\}} \right|$$

この  $D_i$  に関する全組合せが帰無分布を構成する。これは厳密分布であり、漸近理論に依存しない。これを用いて行われるのが **ランダム化検定** である。一方で Neyman の帰無仮説は平均処置効果についての仮説で：

$$H_0^{\text{Neyman}} : \mathbb{E}[Y_i(1) - Y_i(0)] = 0.$$

Fisher の検定はより強い仮定に基づく検定である。

■ p.43 Ex. Young (2019) AEA 掲載論文の RCT を Fisher 検定で再評価し、およそ 13 ~ 22% の結果が有意でなくなることを示した。これは Fisher 検定が有限標本により厳格であること、特に **unbalanced** な割当の影響が大きい場合に有意差が消えることを意味する (p.47)。

■ p.51 検定法の選択  $N < 250$  のとき、下方バイアスに対応するための頑健標準推定量としては **HC3** が適切だが、Stata のデフォルトは HC1。十分大きな標本数では逆に **HC1** を利用することが適切。ランダム化検定は、保守的になりすぎるところがあるため、あまりにも標本数が小さいときのみ利用すべき。

■ 教科書 p.79 多重検定問題 様々な種類の潜在結果を考える際に、ただ仮説検定を繰り返すだけでは、本来制御の対象である検定のサイズを制御することが出来ないという問題。

■ 教科書 p.83 ボンフェローニ検定 多重検定問題の対策として、和集合不等式に基づいて、それぞれの過誤率を  $5/M\%$  とすることで族過誤率 (FWER) を 5% に抑える検定。

■ 教科書 p.85 ボンフェローニ・ホルム検定 明らかに帰無仮説から生成されていない検定統計量を逐次的に消去し棄却水準を下げて、検出力をボンフェローニ検定よりも高める一方で、族過誤率 (FWER) は依然 5% に抑える検定。少なくとも一つの帰無仮説を棄却する確率自体はボンフェローニ検定と変わらない ( $5/M\%$ )。

■教科書 p.87 適切な制御対象と分析結果の利用目的 ある処置が企業における複数の業績指標に及ぼす効果を評価するような場合、どの項目に差があるかによって異なる対応をとるような場合は族過誤率を気にする必要がある。だがある処置が企業の業績に何らかの影響があるかを見るだけならば気にする必要がなく、偽検出比率 (FDP) の期待値である 傷検出率 (FDR) を考えれば十分である。

$$\text{FDP} := \frac{\#\{i \in \{1, \dots, n\} : H_i \text{ が真かつ } H_i \text{ が棄却される}\}}{\max\{\#\{i \in \{1, \dots, n\} : H_i \text{ を棄却する}\}, 1\}} \quad (3.5)$$

$$\text{FDR} = \mathbb{E}[\text{FDP}] \leq \alpha$$

■教科書 p.89 ベンジャミン・ホッシュバーグ検定 10% が水準として慣例となっている。手順としては、今までとは逆に、棄却しにくいものから順に  $p$  値の検証を行い、棄却された場合にはそれ以上に棄却しやすいものを全て棄却するような検定となっている。例えば  $j$  番目に小さい  $p_{(j)}$ 、すなわち  $j$  番目に棄却されやすい帰無仮説については、棄却の基準を少し厳しくして、 $p_{(j)} \leq j\alpha/M$  ならば  $H_{(1)} \sim H_{(j)}$  まで全て棄却する。検定統計量が独立である際には、論理的には怪しいところがあるが、相関構造に対してはある程度頑健なことを考慮して、よく用いられる検定といえる。

■p.52 多重検定問題 ステップアップ法 (Romano and Shaikh, 2006), (See List et al., 2023)あたりを使うのが検出力を保ったうえで、族過誤率の制御をするには良いと思われる。FDR に関しては Benjamini-Hochberg 検定を用いればいいだろう。

### 3 Instrumental Variables

■p.3 片側遵守者 処置受取  $D_i$  の、操作変数たる処置割当  $Z_i$  との関係性から判断して、今回の設定は片側遵守者のケースである。無作為化実験は  $Y_i(d) \perp\!\!\!\perp Z_i$  を保証する。まず 処置割当効果 (ITT 効果) について考察する: ITT 効果は、割当変数による平均の差

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]$$

として定義される。独立性  $(Y_i(1), Y_i(0)) \perp\!\!\!\perp Z_i$  の下で、

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbb{E}[D_i | Z_i = 1].$$

右辺第 2 項  $\mathbb{E}[D_i | Z_i = 1] = \mathbf{P}(D_i = 1 | Z_i = 1)$  は、割当群のうち実際に処置を受けた割合を意味する。したがって、ITT 効果は処置効果に、実際の処置受取率を掛け合わせた形で表される。

*Proof.* 以下のように、片側遵守者の仮定と、独立性によって確認ができる:

$$\begin{aligned} & \mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] \\ &= \mathbb{E}[Y_i(1)D_i + Y_i(0)(1 - D_i) | Z_i = 1] - \mathbb{E}[Y_i(1)D_i + Y_i(0)(1 - D_i) | Z_i = 0] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0))D_i | Z_i = 1] + \mathbb{E}[Y_i(0) | Z_i = 1] - \mathbb{E}[Y_i(0) | Z_i = 0] \quad \because D_i |_{Z_i=0} \equiv 0 \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbf{P}[D_i = 1 | Z_i = 1] \\ &+ \mathbb{E}[0 | D_i = 0, Z_i = 1] \mathbf{P}[D_i = 0 | Z_i = 1] - \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp Z_i \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbb{E}[D_i | Z_i = 1] \end{aligned}$$

これで題意は満たされた.

□

■p.7 Bloom (1984) ITT 効果を用いると, ATT を次のように推定できる:

$$\text{ATT} = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbf{P}(D_i = 1 | Z_i = 1)} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1].$$

これは Bloom (1984) により提案された推定式であり, 一般形として

$$\frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]}$$

なる Wald 推定量として知られる.

■p.8 Wald 推定量と IV 推定量の関係 Wald 推定量は, 二値の  $Z_i$  を操作変数とした IV 推定量と一致する:

$$\frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)}$$

背後にある構造方程式系は

$$\begin{cases} Y_i = \alpha + \beta D_i + \varepsilon_i, \\ D_i = \pi_0 + \pi_1 Z_i + \xi_i, \end{cases}$$

であり,  $Z_i$  が  $D_i$  に関連し ( $\pi_1 > 0$ ), かつ  $Z_i$  が  $Y_i$  に直接影響しない ( $Y_i(D_i, Z_i) = Y_i(D_i)$ ,  $(\varepsilon_i, \xi_i) \perp\!\!\!\perp Z_i$ , 排除制約) ならば, 一致性を持つ因果効果推定量となる.

■p.9 Exercise 3.1 Wald 推定量を導出しよう.  $0 < \text{Cov}(D, Z)$  のとき,

$$\frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} = \frac{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}$$

が成り立つことを示せ.

*Proof.* まず, 共分散の定義と  $Z \in \{0, 1\}$  を利用して:

$$\begin{aligned} \text{Cov}(Y, Z) &= \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]. \\ \mathbb{E}[YZ] &= \mathbb{E}[Y | Z = 1]\mathbf{P}(Z = 1) = \mathbb{E}[Y | Z = 1]\mathbb{E}[Z]. \\ \mathbb{E}[Y] &= \mathbb{E}[Y | Z = 1]\mathbb{E}[Z] + \mathbb{E}[Y | Z = 0]\mathbb{E}[1 - Z]. \end{aligned}$$

したがって,

$$\begin{aligned} \text{Cov}(Y, Z) &= \mathbb{E}[Z](\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y]) \\ &= \mathbb{E}[Z](\mathbb{E}[Y | Z = 1] - (\mathbb{E}[Y | Z = 1]\mathbb{E}[Z] + \mathbb{E}[Y | Z = 0]\mathbb{E}[1 - Z])) \\ &= \mathbb{E}[Z]\mathbb{E}[1 - Z]\{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]\} \\ &= \text{Var}(Z)\{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]\}. \end{aligned}$$

同様に,

$$\text{Cov}(D, Z) = \text{Var}(Z)\{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]\}.$$

これらの代入によって以下を得る:

$$\frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} = \frac{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}.$$

よって二値  $Z$  の場合における Wald 推定量を得た.  $\square$

■p.16 Exercise 3.2 両側遵守の場合, は追加の仮定を置かない限り ITT の解釈が出来ない. 独立性:

$$(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$$

の下で, 以下が成り立つことを示せ:

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) | \text{cp}] \mathbf{P}(\text{cp}) - \mathbb{E}[Y_i(1) - Y_i(0) | \text{df}] \mathbf{P}(\text{df}).$$

*Proof.* 独立性  $(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$  を用いる. タイプ集合を:

at (always-taker), nt (never-taker), cp (complier), df (defier)

とする. まず条件付き期待の差をタイプごとに分解する:

$$\begin{aligned} & \mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] \\ &= \mathbb{E}[D_i(1)Y_i(1) + (1 - D_i(1))Y_i(0)] - \mathbb{E}[D_i(0)Y_i(1) + (1 - D_i(0))Y_i(0)] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \end{aligned}$$

ここで  $D_i(1) - D_i(0)$  は各タイプで

$$D_i(1) - D_i(0) = \begin{cases} 1 & (\text{complier}), \\ 0 & (\text{always-taker} \text{ または } \text{never-taker}), \\ -1 & (\text{defier}) \end{cases}$$

を取る. 従って期待値はタイプの差として書ける:

$$\begin{aligned} \mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] &= \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{\text{cp}\}] - \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{\text{df}\}] \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | \text{cp}] \mathbf{P}(\text{cp}) - \mathbb{E}[Y_i(1) - Y_i(0) | \text{df}] \mathbf{P}(\text{df}). \end{aligned}$$

これがスライド記載の式に一致する.  $\square$

■p.21 LATE と単調性 単調性により defier の不存在を仮定することによって, ITT は LATE と complier の割合の積 に分解できる. これを変形することにより, Wald 推定量と LATE の一致が確認できる.

■p.22 LATE と共変量 Wald 推定量は共変量を考慮していない. 存在する場合, 仮定を再構成すべき:

1. 条件付き独立性:  $(Y_i(0), Y_i(1), D_i(0), D_i(1)) \perp\!\!\!\perp Z_i | X_i$
2. 排除制約:  $Y_i(d, z) = Y_i(d)$
3. 関連性:  $\mathbb{E}[D_i | Z_i = 1, X_i] \neq \mathbb{E}[D_i | Z_i = 0, X_i]$
4. 単調性:  $D_i(1) \geq D_i(0)$  a.s.

$X_i$  が離散変数 であり,  $Z_i, D_i$  が二値変数のとき, 2SLS は次の 完全飽和モデル で構成される:

$$\begin{aligned} \text{第 1 段階: } D_i &= \sum_{k=1}^K \mathbb{1}\{X_i = k\}(\pi_{0k} + \pi_{1k}Z_i) + \xi_i, \\ \text{第 2 段階: } Y_i &= \sum_{k=1}^K \mathbb{1}\{X_i = k\}(\alpha_{0k}) + \tau D_i + \varepsilon_i, \end{aligned}$$

この設定下では、各  $X_i = k$  ごとに LATE を考えられる:

$$\beta(X_i = k) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = k, D_i(1) > D_i(0)].$$

このとき、2SLS の推定量  $\hat{\tau}_{2SLS}$  は、LATE の加重平均に一致する:

■p.25 Exercise 3.3 2SLS 推定量  $\hat{\tau}_{2SLS}$  が以下のような加重平均で表せるこことを示す:

$$\hat{\tau}_{2SLS} \xrightarrow{p} \mathbb{E}[\omega(X)\tau(X)] \quad \text{where } \omega(X_i) = \frac{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)}{\mathbb{E}[\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)]}.$$

この重み  $\omega(X_i)$  は、 $X_i = x$  での  $Z_i$  による  $D_i$  の変動の大きさに比例する。したがって、操作変数がより強く効いている集団 (compliance のばらつきが大きい集団) ほど、推定量において重く評価される。

*Proof.* 完全飽和モデルのもとでは、各  $X_i = k$  ごとに独立した第1段階  $D_i = \pi_{0k} + \pi_{1k}Z_i + \xi_i$  と、第2段階  $Y_i = \alpha_{0k} + \tau D_i + \varepsilon_i$  を構築できる。このとき、各  $k$  ごとに Wald 推定量:

$$\tau(k) = \frac{\mathbb{E}[Y_i \mid Z_i = 1, X_i = k] - \mathbb{E}[Y_i \mid Z_i = 0, X_i = k]}{\mathbb{E}[D_i \mid Z_i = 1, X_i = k] - \mathbb{E}[D_i \mid Z_i = 0, X_i = k]}$$

を計算できる。完全飽和モデルでは、FWL 定理で、 $X_i$  ごとに偏差を取った 2SLS 推定量が

$$\hat{\tau}_{2SLS} = \frac{\text{Cov}(P_{Z,X}D_i, P_{Z,X}Y_i)}{\text{Var}(P_{Z,X}D_i)} \xrightarrow{p} \frac{\text{Cov}(\mathbb{E}[Y_i \mid X_i, Z_i], \mathbb{E}[D_i \mid X_i, Z_i])}{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i])}$$

と表される。ここで、確率収束先を整理すると:

$$(分子) = \text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z]) = \mathbb{E}[\text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z] \mid X)]$$

条件付き共分散の中身を代入すると:

$$\begin{aligned} \mathbb{E}[\text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z] \mid X)] &= \mathbb{E}[\text{Cov}(a(X) + \tau(X)m(X, Z), m(X, Z) \mid X)] \\ &= \mathbb{E}[\tau(X)\text{Var}(m(X, Z) \mid X)], \end{aligned}$$

となる ( $a(X)$  は  $X$  のみに依存する関数)。一方、分母は全分散の法則により:

$$\text{Var}(\mathbb{E}[D \mid X, Z]) = \mathbb{E}[\text{Var}(m(X, Z) \mid X)] + \text{Var}(\mathbb{E}[m(X, Z) \mid X]),$$

であるが、完全飽和モデルでの 2SLS は  $X$  による「間の変動」を除いた（すなわち  $M_X$  による偏差をとった）成分に基づくため、分母に使うのは内部変動  $\mathbb{E}[\text{Var}(m \mid X)]$  に対応する部分である（標準的な導出では  $M_X$  を掛けた後の経験分散の期待が  $\mathbb{E}[\text{Var}(m \mid X)]$  に収束する）。したがって、射影後の 2SLS の母集団極限は次の形に帰着する:

$$\hat{\tau}_{2SLS} \xrightarrow{p} \mathbb{E}\left[\frac{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)}{\mathbb{E}[\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)]} \cdot \tau(X_i)\right].$$

したがって、 $\hat{\tau}_{2SLS}$  は各  $X_i$  ごとの LATE  $\tau(X_i)$  の加重平均に一致する。□

*Proof.* 第1段階を  $D$  を  $(X, Z)$  に回帰した射影とし、

$$\hat{D} = P_{Z,X}D, \quad \hat{D} \xrightarrow{p} \mathbb{E}[D \mid X, Z].$$

第2段階は  $Y$  の  $(\widehat{D}, X)$  への回帰. FWL 定理より,  $X$  による偏差を取った回帰式は,

$$Y - \widehat{\mathbb{E}}[Y | X] \text{ を } \widehat{D} - \widehat{\mathbb{E}}[\widehat{D} | X] \text{ に回帰する OLS の係数}$$

で表される. したがって, 母集団形は次のようになる:

$$\widehat{\tau}^{2SLS} \xrightarrow{p} \tau^* = \frac{\text{Cov}(Y - \mathbb{E}[Y | X], \mathbb{E}[D | X, Z] - \mathbb{E}[\mathbb{E}[D | X, Z] | X])}{\text{Var}(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X])}. \quad (*)$$

分子を条件付き期待を用いて展開すると,

$$\text{Num} = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y | X])(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X]) | X]].$$

ここで,  $X$  で条件付けた内側の期待値を,  $Z$  の二値性を利用して書き換える:

$$\text{Num} = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y | X])(Z - \mathbb{E}[Z | X])\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\} | X]].$$

$Z$  が  $X$  から独立にランダム化されているため,  $\mathbb{E}[(Z - \mathbb{E}[Z | X])^2 | X] = \text{Var}(Z | X)$  であり, また,

$$\mathbb{E}[Y | X, Z = 1] - \mathbb{E}[Y | X, Z = 0] = \tau(X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}.$$

したがって,

$$\text{Num} = \mathbb{E}[\tau(X) \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2].$$

同様に, 分母は

$$\begin{aligned} \text{Den} &= \text{Var}(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X]) = \mathbb{E}[\text{Var}(\mathbb{E}[D | X, Z] | X)] \\ &= \mathbb{E}[\text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]. \end{aligned}$$

分子・分母を対応させると,

$$\tau^* = \frac{\mathbb{E}[\tau(X) \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]}{\mathbb{E}[\text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]}.$$

したがって,

$$\tau^* = \mathbb{E}[\omega(X)\tau(X)], \quad \omega(X) \propto \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2.$$

題意は示された. □

■p.27 Abadie の  $\kappa$  上記の推定量は,  $X_i$  が離散変数である場合にのみ「完全飽和モデル」として構成できる. 現実には  $X_i$  が連続あるいは高次元であるため, この方法を直接適用することは困難である. そこで Abadie (2003) により提案されたのが, 共変量を含む状況でも complier だけの平均を識別できる加重法 ( $\kappa$  重み) である. Abadie は, 任意の可積分関数  $g(Y_i, D_i, X_i)$  に対して次を示した:

$$\mathbb{E}[g(Y_i, D_i, X_i) | D_i(1) > D_i(0)] = \frac{\mathbb{E}[\kappa_i g(Y_i, D_i, X_i)]}{\mathbb{E}[\kappa_i]},$$

ただし, 重み  $\kappa_i$  は

$$\kappa_i = 1 - \frac{D_i(1 - Z_i)}{1 - \mathbf{P}(Z_i = 1 | X_i)} - \frac{(1 - D_i)Z_i}{\mathbf{P}(Z_i = 1 | X_i)}.$$

$D_i = Z_i$  の個体（すなわち complier）のみが非ゼロの重み  $\kappa_i$  を持つ。 $\kappa_i$  を用いることで、 $\mathbb{E}[Y | X, D, D(1) > D(0)]$  に対する関数  $h(D_i, X_i; \theta)$  の近似のため、次の最小二乗問題を考えればよい：

$$\theta_0 = \arg \min_{\theta} \mathbb{E}[\kappa_i \{Y_i - h(D_i, X_i; \theta)\}^2],$$

この方法により、 $X_i$  が連続変数であっても、complier 平均効果を整合的に推定できる。

■p.30 線形確率モデル (LPM) の利用 IV では、第 1 段階で  $\mathbb{E}[D_i | Z_i, X_i]$  を推定する必要があるが、 $D_i$  が二値変数の場合、その条件付き期待値は確率として  $[0, 1]$  に制約される。しかし、多くの実証研究では、第 1 段階を 線形確率モデル (Linear Probability Model, LPM) で近似しており、確率としての性質を満たさない予測を返しうる。本来線形射影によって推定され機械的にその残差は第 2 段階の誤差と無相関となるが、仮に第 1 段階を ロジットやプロビットなどの非線形モデルで推定した場合はこれが成り立たない。誤った定式化の懸念もある以上、線形モデルでの第一段階推定が無難だと結論できる。

## 4 Conditional Independence Approaches

■p.2 LaLonde (1986) 実験データで得られた真の因果効果を、観察データと回帰による補正でどこまで再現できるかについての研究。

■p.6 バイアス 従来のバイアス概念は無条件の差

$$\mathbb{E}[Y(0) | D = 1] - \mathbb{E}[Y(0) | D = 0]$$

で表されるが、これには少なくとも三種類の要因が含まれると説明される：

1.  $X | D = d$  のサポートの違い
2.  $X | D = d$  の分布の違い
3. 共通サポートの選択バイアス  $\mathbb{E}[Y(0) | D = 1, X = x] - \mathbb{E}[Y(0) | D = 0, X = x]$

■p.7 Heckman et al. Heckman らは LaLonde の問題点を整理し、バイアスの源泉を示すための表を提示している。概念的には、単純な平均差では原因を特定できないことを強調している。

■p.8 共変量分布のバランス 共変量間の差は主に二つに分かれれる：

- unbalanced: 同一サポート上で分布が異なる
- not overlapped: サポート自体が異なり共通領域が小さい（または無い）

共変量バランスやオーバーラップを評価するため、共変量ごとの平均や標準偏差の比較が確認できる。

■p.11 オーバーラップの評価基準 オーバーラップ評価の一つの方法として、ロジット（またはロジットの線形予測）に基づく近傍判定が提示されている。十分に小さい  $\ell^U$  について：

$$\xi_i = \begin{cases} 1 & \text{if } \sum_{i': D_{i'} \neq D_i} \mathbb{1}\{\hat{\ell}(X_{i'}) - \hat{\ell}(X_i) \leq \ell_u\} \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \ell(x) = \log \frac{p(x)}{1 - p(x)}$$

これは「反対群に十分近い傾向スコアを持つ観測が存在するか」を示す。オーバーラップの要約指標は：

$$q_c = \frac{1}{N_c} \sum_{i: D_i=0} \xi_i, \quad q_t = \frac{1}{N_t} \sum_{i: D_i=1} \xi_i$$

処置群・対照群がどれだけ重なっているかを数値化する。LaLonde では  $q_c = 0.21$ ,  $q_t = 0.97$  が報告された。ATE の推定は困難であり ATT しか見られず, overlapping のない処置グループに関しては、傾向スコアの低い群を除外することで観測を除外することが検討される。

■p.13 マッチング法 共変量がバランスしていない場合やオーバーラップが不十分な場合の解決策としてマッチング法が紹介される。理想は「同一の  $X$  を持つ対」を作ることであり、これにより疑似ランダム化の近似を行う。単純なマッチング差分の推定量は、ある観測  $i$  に対して同じ  $X$  を持つ  $m_i^c$  を見つけ,

$$\hat{\tau}_i^{\text{match}} = Y_i - Y_{m_i^c},$$

処置群に対する ATT 推定量は

$$\hat{\tau}^{\text{match}} = \frac{1}{N_t} \sum_{i:D_i=1} (Y_i(1) - Y_{m_i^c}(0))$$

と表される。マッチング推定量の推論は非標準的で、ブートストラップが無効になりうる。マッチング推定の分散推定やブートストラップ改良の研究が続いている (Ex. Abadie & Imbens (2006): 分散推定量, Otsu & Rai (2017): 実現可能なブートストラップ法・同対照群を流用すると失敗しうる)。

■p.17 非完全一致の場合 完全一致が困難な場合、 $X$  が最も近い観測を探る最近傍マッチングを用いる:

$$m_i^c = \arg \min_{i': D_{i'} \neq D_i} \|X_i - X_{i'}\|$$

距離はユークリッドや傾向スコア差  $|p(X_i) - p(X_j)|$  を使う。傾向スコアは **balancing** の性質を持つ:

$$X \perp\!\!\!\perp D \mid b(X)$$

を満たす  $b(X)$  を balancing score と呼ぶ。つまり  $b(X)$  を固定すれば、処置・対照群の  $X$  分布は同一。

■p.20 Rosenbaum & Rubin (1983)

1. 傾向スコアはバランススコア:

$$X \perp\!\!\!\perp D \mid p(X) \quad p(X) = \Pr(D = 1 \mid X)$$

2. 傾向スコアは最小次元のバランススコア:

$b(X)$  がバランススコアなのは  $b(X)$  が  $p(X)$  より細かい ( $p(x_1) \neq p(x_2) \Rightarrow b(x_1) \neq b(x_2)$ ) とき

■p.22 傾向スコアの因果的含意 もし潜在結果が条件付き独立

$$(Y(1), Y(0)) \perp\!\!\!\perp D \mid X$$

を満たすなら、任意のバランススコア  $b(X)$  について

$$(Y(1), Y(0)) \perp\!\!\!\perp D \mid b(X)$$

も成立する。 $p(X)$  は (1) 共変量の balancing に有効で (2) 条件付き独立の利用において最も効率的。

■p.23 傾向スコアの利用: Heckman et al. (1998) 条件付き独立の下で

$$\mathbb{E}[Y(0) \mid D = 1, X \in S] = \mathbb{E}[\mathbb{E}[Y \mid D = 0, p(X) = p(X_i)] \mid D = 1, X \in S]$$

が成り立つ。ここで  $S$  は共通サポート  $S = \text{Supp}(X \mid D = 1) \cap \text{Supp}(X \mid D = 0)$  である。

■p.24 傾向スコアによる推定式 推定量はカーネル法を用いて

$$\hat{\tau} = \frac{1}{N_t} \sum_{i:D_i=1} [Y_i - \hat{E}(Y | D=0, p(X) = p(X_i))] I\{X_i \in \hat{S}\}$$

で与えられる。傾向スコア推定は一段階として扱われ、標準的な二段階推定の漸近理論を適用できる。

■p.25 傾向スコアマッチング (Abadie & Imbens, 2016) 非完全一致 ( $M$  個の最近傍) マッチングを定式化：

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (2D_i - 1) \left( Y_i - \frac{1}{M} \sum_{j \in J_M(i)} Y_j \right),$$

$J_M(i)$  は  $D_j = 1 - D_i$  かつ  $|p(X_i) - p(X_j)|$  が最も近い  $M$  個の集合。Wang & Zubizarreta (2023) は、バランス条件を保ちつつ最大の  $M$  を選ぶマッチング法を提案し、その推定量が  $\sqrt{n}$ -一致することを示した。

■p.26 Exercise 4.1 仮定：

- 潜在結果の条件付き独立  $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
- すべての  $x$  で  $0 < p(x) < 1$

このとき以下が成立する：

$$\mathbb{E}\left[\left(\frac{D}{p(X)} - \frac{1-D}{1-p(X)}\right) Y\right] = \mathbb{E}[Y(1) - Y(0)].$$

つまり、傾向スコアを用いた逆確率重み付けにより母平均の因果効果を識別できる。

*Proof.*  $Y = DY(1) + (1-D)Y(0)$  から、

$$\begin{aligned} & \mathbb{E}\left[\left(\frac{D}{p(X)} - \frac{1-D}{1-p(X)}\right) Y\right] \\ &= \mathbb{E}\left[\frac{D}{p(X)}(DY(1) + (1-D)Y(0))\right] - \mathbb{E}\left[\frac{1-D}{1-p(X)}(DY(1) + (1-D)Y(0))\right] \\ &= \mathbb{E}\left[\frac{D}{p(X)}Y(1)\right] - \mathbb{E}\left[\frac{1-D}{1-p(X)}Y(0)\right] \because D^2 = D, D(1-D) = 0 \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{D}{p(X)}Y(1) | X\right]\right] - \mathbb{E}\left[\mathbb{E}\left[\frac{1-D}{1-p(X)}Y(0) | X\right]\right] \because \text{LIE} \\ &= \mathbb{E}\left[\frac{\mathbb{E}[D | X]}{p(x)} \mathbb{E}[Y(1) | X]\right] + \mathbb{E}\left[\frac{\mathbb{E}[1-D | X]}{1-p(x)} \mathbb{E}[Y(0) | X]\right] \because (Y(1), Y(0)) \perp\!\!\!\perp D | X \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}[Y(1) - Y(0)] \because \mathbb{E}[D | X] = p(X), \text{LIE} \end{aligned}$$

以上で証明された。 □

■p.27 Inverse Probability Weighting (IPW) IPW 推定量には二形式がある：

1. Horvitz–Thompson (1952) 型：

$$\hat{\tau}_{HT} = \frac{1}{n} \sum_i \left( \frac{D_i Y_i}{p(X_i)} - \frac{(1-D_i) Y_i}{1-p(X_i)} \right)$$

2. Hájek (1971) 型：

$$\hat{\tau}_H = \frac{1}{n} \sum_i h(D_i, X_i) Y_i,$$

ここで  $h(D_i, X_i)$  は処置・統制群の重みをそれぞれの標本全体で標準化したもの.

■p.29 傾向スコア推定の誤った定式化 傾向スコア推定の誤った定式化は深刻な影響を与える. Hirano et al. (2003), Imai & Ratkovic (2014) などが推定法を提案している.

■p.30 二重頑健推定 Robins et al. (1994) の Augmented IPW (AIPW), 二重頑健推定 は,  $p(X; \gamma)$  または  $\mathbb{E}[Y | X] = X^\top \beta$  の, いずれか一方の定式化が正しければ一致推定量が得られる. したがって AIPW は, モデルの不確実性に対して頑健な推定法として広く用いられる. 毎度のごとく, 末石先生の ミクロ計量分析 第七回 p.14 ~ 16 の方が分かりやすい. アイデアは以下の通り: 第一項の  $Y(1)$  について言えば,

$$\mathbb{E}[Y_i | D_i = 1, X_i] = X_i^\top \beta_1$$

を仮定して OLS 推定量  $\hat{\beta}_1$  を得る. この回帰残差  $\hat{\varepsilon}_i = Y_i - X_i^\top \hat{\beta}_1$  を用い, バイアス補正を行うことで,

$$\hat{\mu}_1 = \frac{1}{n} \sum_i X_i^\top \hat{\beta}_1 + \frac{1}{n} \sum_i \frac{D_i}{\hat{p}(X_i)} \hat{\varepsilon}_i$$

を構築する. ここで  $\hat{p}(X_i)$  は傾向スコアの推定量. 第 1 項は「回帰による予測値」, 第 2 項は「逆確率重み付けによる残差補正」であり, いずれかの定式化が正しければ一致推定が得られる.

■p.32 Exercise 4.2 上述の頑健性を確認するために, 以下のケースを考える.

- 傾向スコアは 誤って定式化 されており,  $\hat{p}(x) \xrightarrow{p} p^*(x) \neq \mathbb{E}[D | X = x] > 0$ .
- 回帰関数は 正しく定式化 されており,  $\mathbb{E}[Y_i | D_i = 1, X_i] = X_i^\top \beta_1$ .

このとき以下が成り立つことを示せ :

$$\frac{1}{n} \sum_i X_i^\top \hat{\beta}_1 + \frac{1}{n} \sum_i \frac{D_i}{\hat{p}(X_i)} \hat{\varepsilon}_i \xrightarrow{p} \mathbb{E}[Y(1)] \quad (\text{スライドはタイポ}).$$

*Proof.* 顕在結果の観測手段  $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$  と, Exercise 4.1 で識別対象を得た際に置いた仮定である  $(Y(1), Y(0)) \perp\!\!\!\perp D | X$  から, 以下が成り立つ:

$$\begin{aligned} \mathbb{E}[Y_i | D_i = 1, X_i] &= \mathbb{E}[D_i Y_i(1) + (1 - D_i) Y_i(0) | D_i = 1, X_i] \\ &= \mathbb{E}[Y_i(1) | D_i = 1, X_i] = \mathbb{E}[Y_i(1) | X_i] = X_i^\top \beta_1. \end{aligned}$$

誤差項を  $\varepsilon_i := Y_i - X_i^\top \beta_1$  と定義すれば, 以下は回帰モデルとなる:

$$Y_i = X_i^\top \beta_1 + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | D_i = 1, X_i] = 0.$$

まず回帰予測値の項について, OLS の一致性により  $\hat{\beta}_1 \xrightarrow{p} \beta_1$  なので:

$$\frac{1}{n} \sum_{i=1}^n X_i^\top \hat{\beta}_1 \xrightarrow{p} \mathbb{E}[X_i^\top \beta_1] = \mathbb{E}[\mathbb{E}[Y_i(1) | X_i]] = \mathbb{E}[Y_i(1)].$$

次に補正項を考える。 $D_i Y_i = D_i Y_i(1)$  を用いると、補正項は残差  $\hat{\varepsilon}_i = Y_i - X_i^\top \hat{\beta}_1$  を用いて

$$\frac{1}{n} \sum_{i=1}^n \frac{D_i}{\hat{p}(X_i)} \hat{\varepsilon}_i = \frac{1}{n} \sum_{i=1}^n \frac{D_i}{\hat{p}(X_i)} (Y_i(1) - X_i^\top \hat{\beta}_1).$$

$\hat{\beta}_1 \xrightarrow{p} \beta_1$  と LLN より、

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{D_i}{\hat{p}(X_i)} (Y_i(1) - X_i^\top \hat{\beta}_1) &\xrightarrow{p} \mathbb{E}\left[\frac{D_i}{\hat{p}(X_i)} (Y_i(1) - X_i^\top \beta_1)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{D}{\hat{p}(X)} (Y(1) - X^\top \beta_1) \mid X\right]\right] \quad \because \text{LIE} \\ &= \mathbb{E}\left[\frac{\mathbb{E}[D \mid X]}{\hat{p}(X)} (\mathbb{E}[Y(1) \mid X, D=1] - X^\top \beta_1)\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[D \mid X]}{\hat{p}(X)} (X^\top \beta_1 - X^\top \beta_1)\right] = 0. \end{aligned}$$

ゆえに

$$\frac{1}{n} \sum_i X_i^\top \hat{\beta}_1 + \frac{1}{n} \sum_i \frac{D_i}{\hat{p}(X_i)} \hat{\varepsilon}_i \xrightarrow{p} \mathbb{E}[Y(1)].$$

題意は示された。  $\square$

### ■p.33 Exercise 4.3 次は先ほどとは逆のケースを考える:

- 傾向スコアは 正しく定式化 されており、 $\hat{p}(x) \xrightarrow{p} p(x) = \mathbb{E}[D \mid X = x]$ .
- 平均関数は 誤って定式化 されており、 $\mathbb{E}[Y_i \mid D_i = 1, X_i] \neq X_i^\top \beta_1$ .

このとき、以下の同じ関係式が成り立つことを証明せよ：

$$\frac{1}{n} \sum_i X_i^\top \hat{\beta}_1 + \frac{1}{n} \sum_i \frac{D_i}{\hat{p}(X_i)} \hat{\varepsilon}_i \xrightarrow{p} \mathbb{E}[Y(1) \mid X].$$

*Proof.* まず残差の定義より、 $\hat{\varepsilon}_i = Y_i - X_i^\top \hat{\beta}_1$ . 式全体を展開すると、

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{n} \sum_i X_i^\top \hat{\beta}_1 + \frac{1}{n} \sum_i \frac{D_i}{\hat{p}(X_i)} (Y_i - X_i^\top \hat{\beta}_1) \\ &= \frac{1}{n} \sum_i \frac{D_i Y_i}{\hat{p}(X_i)} - \frac{1}{n} \sum_i \frac{D_i X_i^\top \hat{\beta}_1}{\hat{p}(X_i)} + \frac{1}{n} \sum_i X_i^\top \hat{\beta}_1. \end{aligned}$$

$\hat{p}(X_i) \xrightarrow{p} p(X_i)$ かつ  $\hat{\beta}_1 \xrightarrow{p} \beta^* \neq \beta$  より、

$$\hat{\mu}_1 \xrightarrow{p} \mathbb{E}\left[\frac{DY}{p(X)}\right] - \mathbb{E}\left[\frac{DX^\top \beta^*}{p(X)}\right] + \mathbb{E}[X^\top \beta^*].$$

第一項について、

$$\mathbb{E}\left[\frac{DY}{p(X)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{DY}{p(X)} \mid X\right]\right] = \mathbb{E}\left[\frac{\mathbb{E}[D \mid X]}{p(X)} \mathbb{E}[Y(1) \mid X]\right] = \mathbb{E}[\mathbb{E}[Y(1) \mid X]] = \mathbb{E}[Y(1)].$$

第二項と第三項をまとめると,

$$-\mathbb{E}\left[\frac{DX^\top \beta^*}{p(X)}\right] + \mathbb{E}[X^\top \beta^*] = -\mathbb{E}[X^\top \beta^*] + \mathbb{E}[X^\top \beta^*] = 0.$$

よって全体として,

$$\hat{\mu}_1 \xrightarrow{P} \mathbb{E}[Y(1)],$$

すなわち, アウトカムモデルが誤っていても傾向スコアが正しければ一致推定が得られる.  $\square$

これらの手順を  $\mu_0$  に関しても繰り返すことによって二重頑健性は確認できる.

■p.34 交絡因子 Ch.1 で見た Simpson (1951) のパラドックスは, 集計レベルと条件付き (層別) で推定される因果効果が符号や大きさを変える事象を示す. Ex. ワクチン:

- 無条件では  $\mathbf{P}(Y = 1 | D = 1) > \mathbf{P}(Y = 1 | D = 0)$  となり, 一見ワクチンが有害
- 年齢で条件付けすると  $\mathbf{P}(Y = 1 | D = 1, X = x) < \mathbf{P}(Y = 1 | D = 0, X = x)$  と有効性が明らかに

この齟齬は  $X$  が  $D$  (優先接種) と  $Y$  (高罹患リスク) 両方に影響を与える (交絡因子) ため.

■p.36 媒介因子, mediator  $X$  が  $D$  で変化し,  $X$  が  $Y$  に影響する場合 (Ex. 抗体レベル)

$$\mathbf{P}(Y = 1 | D = 1, \text{Age}, X) = \mathbf{P}(Y = 1 | D = 0, \text{Age}, X)$$

$$\mathbf{P}(Y = 1 | D = 1, \text{Age}) > \mathbf{P}(Y = 1 | D = 0, \text{Age})$$

となり,  $X$  は媒介因子と呼ばれる.  $X$  を制御するとトータル効果を消してしまうので注意.

■p.37 因果グラフ (DAG) ノードが変数を, 有向辺が因果方向を示す. DAG (Directed Acyclic Graph) は自己ループのない有向グラフで, 制御すべき変数を判断できる. Pearl (1995) は以下の基準を示した:

- Back-door criterion :  $X, Y$  間の バックドアパス をブロックできる変数集合  $Z$  を制御すべき
- Front-door criterion : もし中間点  $Z$  が  $X$  から  $Y$  のすべての因果経路を担い, かつ back-door path がブロックされるなら,  $Z$  を経由して効果を識別できる.

先ほど議論した交絡因子と媒介因子は, 三変数のみの関係で完結するような定義だったが, DAG に関する議論はそれに限らず, 経路には 3 つ以上の因子が入りうることに注意せよ.

■p.41 DAG の基本用語 親, 子, パスなどが定義される. 相変わらず説明不足なので ネット資料 を参照しよう. 以下の議論では Markovian parent property なる性質を仮定する:

**Assumption: Markovian parent property** —————

$(X_1, \dots, X_{i-1})$  を全ての先行要素,  $PA_i$  を直接の親として DAG では以下が成り立つ:

$$\mathbf{P}(X_i | PA_i) = \mathbf{P}(X_i | X_1, \dots, X_{i-1})$$

すなわち, DAG はすべての決定要素を捉えており,  $X_i$  は親以外の先行変数と条件付き独立といえる.

■p.42 Back-door 条件 (Pearl の定理) 集合  $Z$  が  $(X, Y)$  ( $X$  は  $Y$  の非子孫) に対し back-door 条件:

**Condition: back-door 条件**

1.  $Z$  に含まれるノードは  $X$  の子ノードではない ( $\Rightarrow$  媒介因子でない)
2.  $Z$  が  $X$  に入るすべての  $X, Y$  間のパスをブロックする ( $\Rightarrow$  交絡因子でない)

を満たすとき,  $X$  の  $Y$  への因果効果は以下で識別される (Theorem 3.3.2, Pearl, 2009).

$$\mathbf{P}(Y(x)) = \mathbf{P}(Y \mid \text{do}(X = x)) = \sum_z \mathbf{P}(Y \mid X = x, Z = z) \mathbf{P}(Z = z)$$

直感的には,  $Z$  を固定 (制御, バックドアパスを閉じる!) すれば  $X$  の操作が擬似的にランダム化される. すなわち,  $Z$  を条件付けることで  $X$  と  $Y(x)$  の独立性

$$Y(x) \perp\!\!\!\perp X \mid Z$$

が成立するため, 介入分布が観測分布から再構成できる.

**■p.43 Front-door 条件** 集合  $Z$  が  $(X, Y)$  ( $X$  は  $Y$  の非子孫) に対し front-door 条件:**Condition: front-door 条件**

1.  $Z$  は  $X$  から  $Y$  への全経路によって経由される (中間点)
2.  $X$  から  $Z$  への未遮断の back-door path がない
3.  $Z$  から  $Y$  への back-door path は  $X$  によりブロックされる

を満たすとき,  $X$  の  $Y$  への因果効果は以下で識別される (意味不明)

$$\mathbf{P}(Y(x)) = \sum_z \mathbf{P}(Z = z \mid X = x) \sum_{x'} \mathbf{P}(Y \mid X = x', Z = z) \mathbf{P}(X = x'),$$

p.45 Ex. たばこ 喫煙  $X$  と肺がん  $Y$ , 肺内部のタール量  $Z$  を例に取ろう.  $X \rightarrow Z \rightarrow Y$  で, 遺伝要因や環境曝露などが潜在的な back-door path として考えられるが, 条件が満たされれば  $Z$  経由で識別可能.

**■p.46 Back-door と Front-door の選択** 実務的には DAG を十分に知ることが重要. どちらの基準が使えるかは, 得られる変数とその因果関係の知識に依る. 文献で提案される変数選択の主流的アプローチは以下:

1. **Pre-treatment approach:** 全ての事前変数を調整
2. **Common cause approach:**  $X, Y$  の共通原因を中心に調整
3. **Disjunctive cause approach:**  $X$  または  $Y$  の原因となる変数を調整

**■p.48 Pre-treatment principle** 長所: シンプルで分かりやすい. 「介入前の全てを条件付けすればよい」という直感. 短所: 条件付けて collider を導入してバイアスを生む可能性 (collider bias). すべての事前変数を盲目的に調整するのは誤りになりうる. (というか 合流点 だからじゃない?)

p.50 Ex. Collider bias Figure は, ある選択  $L$  に条件付けると  $A$  と  $Y$  の間に偽の相関が出る例を示す (選択に入るか否かが  $A$  と  $Y$  の双方に影響するため). 例: 血圧降下薬  $A$  と Covid 感染  $Y$ , サブサンプル  $L$  (自己申告参加) を考えると,  $L$  を条件付けすることで  $A \rightarrow Y$  の偽相関が生じ得る.

**■p.51 Common cause approach** Glymour らは, 共通原因を見つけて条件付けする戦略を提案した. 共通原因是 back-door path を遮断しがち, だがそれだけではなく, 実際 p.53 の例なら  $C$  が back-door path をふさぐ.

■p.53 Disjunctive cause approach  $X$  または  $Y$  の原因である変数を調整する。理論的には、もしバックドアを遮断する集合が存在するなら、**disjunctive cause** はその集合を含む。ただし disjunctive cause でも害が出る場合がある (Figure 4)。特に変数が強い IV に近い場合、調整はバイアスを増幅 (減衰?) しうる。Pearl (1995) の基準に則るならば、 $U$  のみをブロックすべき。

#### ■p.55 結語：実務上のガイドライン

- 因果グラフができるだけ明確にすることが最重要
- path がブロックによってどうなるかを考えるべき
- 単純なルール（すべての事前変数を入れる、原因を入れる等）は便利だが無批判には使えない  
(そもそも、ここでいう『原因』って定義されてない直感的な言葉に過ぎないし……)
- 複数の手法比較が推奨される

## 5 Difference-in-Differences

### 5.1 Part 1

■p.25 共変量のある DiD 共変量がある場合には  $X \perp\!\!\!\perp D$  が保証されないと、バイアスが生じる可能性があり、 $X$  はコントロールの対象となる。しかし  $\mathbb{E}[Y | T = t, G = g, X = x]$  を推定するためには、誤った定式化の危険性もあり、折衷案として **セミパラメトリックな IPW method** が利用される。具体的には、傾向スコアが以下のように一部定式化される：

$$q(\mathbf{x}) = \Pr(\mathbf{x}^\top \gamma \geq \varepsilon) \quad \text{where } \varepsilon | X \sim N(0, 1)$$

$\mathbf{x}^\top \gamma$  の部分がパラメトリックであることに注目せよ。

### 5.2 Part 2

■p.2 Bertrand et al. (2004) *How much should we trust difference-in-differences estimates?* 系列相関を適切に考慮しないと、第一種の過誤につながる。クラスタリングや時系列依存を考慮した分散推定が必要。

■p.4 クラスター頑健標準誤差 Cluster-Robust Variance (CRVE) は一般的だが、クラスター数と期間どちらも大きくないときには使えない。<sup>\*1</sup> Wild bootstrap はクラスタサイズの不均衡や非 i.i.d. の設定下でも有用な代替法である (Cameron, Gelbach & Miller 2008)。図は、クラスタ数が極端に少ない場合の振る舞いを示しているのだろうが意図が不明。

■p.12 少数クラスタのみへの処置に関する対処法 Donald and Lang (2007) ではクラスタ内サンプルサイズが大きい場合を考慮し、Conley and Taber (2010) では時系列相関・均一分散のケースでの、統制群の残差分布を利用した処置群の残差分布の推定を対処法として提示している。Ferman & Pinto (2018) は、クラスタごとの不均一分散下で、ブートストラップ標本の分布を再スケールすることを提示した。具体的には：

1. クラスタ内分散の推定：各クラスタ  $j$  に対して、クラスタサイズを  $M_j$  とすると、クラスタごとの FE

<sup>\*1</sup> Carter et al. (2017) はクラスターサイズが不均一な場合の CRVE を提示している。

推定量の残差  $W_j$  の分散が  $M_j$  に依存するとして、これを回帰モデルで推定する：

$$\widehat{\text{Var}}(W_j | M_j) : M_j \text{ に回帰して推定する.}$$

2. 等分散仮定下のブートストラップ生成：通常のブートストラップで、各  $j$  の統計量  $W_j^b$  を生成する.
3. 再スケーリング： $W_j^b$  を、推定された不均一分散に基づいて再スケールする：

$$W_j^{(b,\text{adj})} = W_j^{(b)} \times \sqrt{\frac{\widehat{\text{Var}}(W_j | M_j)}{\widehat{\text{Var}}(W_j^b | M_j)}}.$$

■p.17 合成コントロール法 (Synthetic Control) 単一処理単位のケースで有力な選択肢となる。

- 目的：処置のない反事実  $Y_{1t}(0)$  を、他ユニットの重み付き和で再現すること
- 重みの選択：事前アウトカムや共変量を使った最適化問題解 (Ex.  $\arg \min(X_B - XW)^\top V(X_B - XW)$ )
- 推論は placebo ベースの検定や再サンプリングにより行う (Abadie et al. のアプローチ)

■p.23 Abadie, Diamond, and Hainmueller (2010) 観測されるアウトカムは次のように表される：

$$Y_{it} = Y_{it}(0) + \alpha_{it} D_{it},$$

ここで  $D_{it} = 1$  は、第 1 ユニット ( $i = 1$ ) のみで、かつ処置時点以降 ( $t > T_0$ ) である。処置効果は

$$\alpha_{1t} = Y_{1t}(1) - Y_{1t}(0) = Y_{1t} - Y_{1t}(0)$$

で定義される。統制群の潜在結果  $Y_{it}(0)$  は次のような因子構造を持つと仮定する：

$$Y_{it}(0) = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it},$$

ここで、 $\delta_t$  は時間固有効果、 $Z_i$  は観測可能な共変量、 $\mu_i$  は観測されない個体固有の因子で、(二元) 固定効果モデル (DiD モデル) であるから、 $\lambda_t = \lambda$  で、 $\lambda \mu_i$  が個体固定効果である。理想的な  $W_j^*$  は次を満たす：

$$\sum_{j=2}^J W_j^* Z_j = Z_1, \quad \sum_{j=2}^J W_j^* \mu_j = \mu_1.$$

すなわち、観測可能な共変量  $Z_i$  だけでなく、観測されない因子  $\mu_i$  に関しても処置ユニットと完全に整合する重み付けを実現することが理想である。 $\mu_i$  は観測できないため、現実的には不可能。ただし、次の条件が成り立つとき、合成コントロール法は「理想的な重み付け」に近い性質を持つことが示される：

- 誤差  $\varepsilon_{it}$  が個体・時点間で独立、
- $\mathbb{E}[\varepsilon_{it} | \{Z_i, \mu_i\}] = 0$ ,
- $\sum_{t=1}^{T_0} \lambda_t^\top \lambda_t$  が正則行列。

このとき、理想的な重み  $W_j^*$  が存在すれば次が成り立つ：

$$Y_{1t} = \sum_{j=2}^J W_j^* Y_{jt}, \quad (\forall 1 \leq t \leq T_0), \quad Z_1 = \sum_{j=2}^J W_j^* Z_j.$$

すなわち、処置前期間ににおけるアウトカムと共変量が一致するような合成ユニットが構築できる。ただし注意点として、このような  $W_j^*$  がそもそも存在しない場合もある。

■p.26 プラセボ (placebo) テストによる推論 Abadie et al. (2010) では、統計的推論を行うためにプラセボ合成コントロール (placebo synthetic control) を構築する手法を提案している。具体的には：

1. 実際の処置個体 ( $i = 1$ ) に対して合成コントロールを作成する。
2. 他の統制個体 ( $i = 2, \dots, J$ ) についても、同様の手順で「もしこのユニットが処置されたとしたら」というプラセボ合成コントロールを構築する。
3. 処置後の期間  $t' > T_0$  において、実測値  $Y_{it'}$  と合成値  $\hat{Y}_{it'}$  の乖離の大きさ (MSE) を比較する。

帰無仮説（全てのユニットが未処置）の下では、処置群とプラセボ群の乖離分布は同様であるはずである。したがって、実際の処置ユニットの MSE がこの分布の中で極端に大きければ、その効果は偶然ではない（=処置効果が有意）と判断できる。この発想は 無作為化検定 に類似しており、有意水準を経験的に評価する「擬似実験的」な推論手法。

■p.28 Arkhangelsky et al. (2020) Arkhangelsky らは、Abadie らの合成コントロールと、DiD の平行トレンド（加法的固定効果）構造を統合した Synthetic Difference-in-Differences (SDiD) を提案した。すなわち、モデルの基礎に柔軟な因子構造を置き、因子そのものを一致推定することは目的とせず、処置効果の推定に集中する。以下のような制約が課される：

- サンプルサイズが十分大きいこと ( $N \rightarrow \infty, T \rightarrow \infty$  の両方が必要)
- 推定が不安定になるため、インタラクティブ FE より少し柔軟な程度が複雑性の限界

■p.30 複数期間・staggered adoption の混乱 2018 年頃に DiD の文献で問題のある分析が増えた。

■p.30 Callaway & Sant'Anna (2020) 1999 – 2007 の間で、連邦最低賃金が一定 (\$5.15) な一方で、各州が異なる時期に州別最低賃金を引き上げた。このように処置が群ごとに異なる時期に与えられる設定を staggered design (時期ずらしデザイン) という。

■p.32 staggered design の問題点 典型的な TWFE 回帰：

$$Y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \varepsilon_{it},$$

$T, N > 2$  の一般の場合、推定量  $\hat{\beta}$  は単純に ATT を表すとは限らない。 $2 \times 2$  の場合は：

$$\hat{\beta}_{1,0}^{2 \times 2} \equiv (\bar{Y}_{\text{post},j} - \bar{Y}_{\text{pre},j}) - (\bar{Y}_{\text{post},k} - \bar{Y}_{\text{pre},k}),$$

が ATT に収束するが、複数時点・複数コホートが混ざると話は複雑になる。naive な TWFE は、時期の早い処置群 (early treated), 時期の遅い処置群 (late treated), 一度も処置されない群 (never treated) など異なる比較を混ぜて一つの係数にして、 $\hat{\beta}$  は複雑な加重平均になり得る。

■p.35 Goodman–Bacon (2021) Goodman–Bacon は  $\hat{\beta}$  を「全ての  $2 \times 2$  比較の加重和」として分解できることを示した。記号として各  $2 \times 2$  比較の係数を次のように定義する：

- $\hat{\beta}_{k,U}^{2 \times 2}$  : 処置群  $k$  の pre-post と、常に未処置の群  $U$  の pre-post の比較
- $\hat{\beta}_{k,l}^{2 \times 2,k}$  : 早期処置  $k$  を treated, 遅い処置  $l$  を control として比較（遅い処置がまだ処置される前）
- $\hat{\beta}_{k,l}^{2 \times 2,l}$  : 逆に早期  $k$  を control, 遅い  $l$  を treated として比較（早期が処置された後）

**Theorem: Goodman-Bacon (2021)**

$$\hat{\beta} = \sum_{k \neq U} s_{kU} \hat{\beta}_{2 \times 2}^{k,U} + \sum_{k \neq U} \sum_{l > k} \left( s_{kl}^k \hat{\beta}_{2 \times 2}^{k,kl} + s_{kl}^l \hat{\beta}_{2 \times 2}^{l,kl} \right),$$

かつ重みは次を満たす：

$$\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{l > k} (s_{kl}^k + s_{kl}^l) = 1.$$

重要な点：第2項（早期 vs 遅延の比較）には  $\hat{\beta}_{2 \times 2}^{l,kl}$  のように「処置後の早期処置を control に用いる」比較が含まれ、これが *ill-behaved*（**負の重みや解釈困難な比較？**）になりやすい。

■p.37 Sun & Abraham (2020) イベントスタディ型回帰で次を推定する：

$$Y_{it} = \alpha_i + \gamma_t + \sum_{\ell} \mu_{\ell} \mathbb{1}\{t - E_i = \ell\} + \nu_{it},$$

ここで  $E_i$  はユニット  $i$  の処置開始時点、指標は相対時刻  $\ell$  を表す。Sun & Abraham は次を指摘した：

- 各時刻係数  $\mu_{\ell}$  は、複数コホートの ATT の複雑な線形結合（負の重みを含みうる）になり得る
- 通常行われる 事前トレンド検定（処置前の  $\mu_{\ell} = 0$  の確認）は不適切 **なぜ？**

■p.38 Wooldridge (2021, 2025) バランスパネルを仮定し、まず時間不变な処置効果  $\beta$  モデルを考える：

$$Y_{it} = \beta W_{it} + c_i + g_t + u_{it}, \quad W_{it} = D_i \times P_t,$$

$T \geq 2$ ,  $P_t = \mathbb{1}\{q \leq t\}$  は処置時刻以降かのダミー,  $D_i \in \{0, 1\}$  は処置群ダミー,  $W_{it}$  が処置ダミー。

**Corollary 5.1: Wooldridge (2021, 2025)**

全ての時点／個体ダミーを含む回帰における  $W_{it}$  の回帰係数  $\hat{\beta}_{FE}$  と、定数・ $W_{it} \cdot D_i \cdot P_t$  のみを説明変数とする回帰係数  $\hat{\beta}_{DD}$  は等しい、すなわち  $\hat{\beta}_{FE} = \hat{\beta}_{DD}$ .

この設定でも、DD 回帰の係数は正しく推定できている。つまり TWFE が誤っているのではなく、**単一係数仕様（ $\beta$  の時間不变性）が誤っている**、との指摘である。

■p.39 訂正 DiD では我々の推定対象とは何だろうか。係数が時刻ごとに異なる場合は：

$$Y_{it} = \beta_q (W_{it} \cdot f_{qt}) + \cdots + \beta_T (W_{it} \cdot f_{Tt}) + c_i + g_t + u_{it}, \quad f_{rt} = \mathbb{1}\{r = t\}.$$

ここで、時点毎の TWFE 推定量  $\hat{\beta}_r^{FE}$  は、拡張 DiD 回帰 ( $Y_{it}$  を 1,  $W_{it}f_{qt}, \dots, W_{it}f_{Tt}, D_i, f_{qt}, \dots, f_{Tt}$  に回帰) と同値。 $P_t = f_{qt} + \cdots + f_{Tt}$  と完全な多重共線性があるため、 $P_t$  は説明変数から落とされている。

■p.39 同一タイミング処置の TE の識別 先ず同一タイミングについて、以下の仮定を置こう：

**Assumption: NA (No Anticipation):**

任意の  $t < q$ 、すなわち処置前の任意の期間に対して：

$$\mathbb{E}[Y_t(1) - Y_t(0) | D = 1] = 0.$$

**Assumption: CT (Common Trend):** —————

処置を受けない潜在結果について処置/統制群は一定で:

$$\mathbb{E}[Y_t(0) - Y_1(0) \mid D] = \mathbb{E}[Y_t(0) - Y_1(0)] \equiv \theta_t, \quad t = 2, \dots, T.$$

NA は **主体が将来の処置を予期して行動を変えられない** ことを意味する。仮定の下、以下の表現を導ける:

$$\mathbb{E}[Y_{it} \mid D_i] = \lambda + \xi D_i + \theta_2 f_{2t} + \dots + \theta_T f_{Tt} + \tau_q (W_{it} \cdot f_{qt}) + \dots + \tau_T (W_{it} \cdot f_{Tt}),$$

$\tau_t \equiv \mathbb{E}[Y_t(1) - Y_t(0) \mid D = 1]$  ( $t \geq q$ )、つまり目標である **時間ごとの 処置効果** について考えられる。

■p.41 Exercise 5.3 (同時割当) 上記の内容を示そう。同時に時点  $q$  で処置  $D_i = 1$  が割り当てられる。群ごとの潜在結果は  $Y_t(d)$  ( $d \in \{0, 1\}$ )、 $Y_{it} = D_i Y_t(1) + (1 - D_i) Y_t(0)$ 。仮定 NA と CT の下で、以下を示せ:

$$\mathbb{E}[Y_{it} \mid D_i] = \lambda + \xi D_i + \theta_t + \theta_2 f_{2t} + \dots + \theta_T f_{Tt} + \tau_q (W_{it} \cdot f_{qt}) + \dots + \tau_T (W_{it} \cdot f_{Tt})$$

*Proof. (未精査)* 条件付き期待値を分解する:

$$\mathbb{E}[Y_{it} \mid D_i] = \mathbb{E}[D_i Y_t(1) + (1 - D_i) Y_t(0) \mid D_i] = D_i \mathbb{E}[Y_t(1) \mid D_i] + (1 - D_i) \mathbb{E}[Y_t(0) \mid D_i].$$

まず CT より、任意  $t$  について

$$\mathbb{E}[Y_t(0) \mid D_i] = \mathbb{E}[Y_1(0) \mid D_i] + \theta_t.$$

(ここで  $\theta_1 = 0$  を便宜上とる。) また NA により  $t < q$  のとき  $\mathbb{E}[Y_t(1) \mid D_i = 1] = \mathbb{E}[Y_t(0) \mid D_i = 1]$ 。一方  $t \geq q$  に対しては

$$\tau_t \equiv \mathbb{E}[Y_t(1) - Y_t(0) \mid D = 1]$$

を定義する。したがって一般に

$$\mathbb{E}[Y_t(1) \mid D_i = 1] = \mathbb{E}[Y_t(0) \mid D_i = 1] + \tau_t.$$

以上を用いると、

$$\begin{aligned} \mathbb{E}[Y_{it} \mid D_i] &= (1 - D_i)(\mathbb{E}[Y_1(0) \mid D_i] + \theta_t) + D_i(\mathbb{E}[Y_1(0) \mid D_i] + \theta_t + \tau_t) \\ &= \mathbb{E}[Y_1(0) \mid D_i] + \theta_t + D_i \cdot \tau_t. \end{aligned}$$

ここで  $\mathbb{E}[Y_1(0) \mid D_i]$  を定数分解して  $\lambda + \xi D_i$  と書けば ( $\lambda$  は基準群の定数、 $\xi$  は群差の定数)、

$$\mathbb{E}[Y_{it} \mid D_i] = \lambda + \xi D_i + \theta_t + D_i \cdot \tau_t.$$

最後に  $D_i \cdot \tau_t = \tau_q (W_{it} \cdot f_{qt}) + \dots + \tau_T (W_{it} \cdot f_{Tt})$  の形で展開可能であり ( $W_{it} = D_i P_t$  により  $D_i \cdot \tau_t = \tau_t (W_{it} \cdot f_{tt})$  等と表せる)。□

■p.42 コホート別 TE 本来の焦点である Staggered design にも議論を拡張する。 $q$  期から処置が開始されることを考え、コホート  $r$  (時点  $r \geq q$  に処置される群) で、**コホート・時点別** 効果を以下で定義する:

$$\tau_{rt} \equiv \mathbb{E}[Y_t(r) - Y_t(\infty) \mid D^r = 1], \quad r \geq q, t \geq r$$

$D^r$  は時点  $r$  に処置されるコホートのダミー、 $Y_t(\infty)$  は常に未処置 (never-treated) の潜在結果。もし never-treated が存在しない場合、最終的なユニットが処置を受ける直前まで解析を切り詰める (truncate)。

■p.43 Staggered での仮定 (NAS, CTS) 仮定を Staggered design においても適応させよう:

**Assumption: NAS (No Anticipation, Staggered):** —————

任意の処置開始後期間  $r \geq q$ , コホート  $\mathbf{D} = \{D^q, D^{q+1}, \dots, D^T\}$  において:

$$\mathbb{E}[Y_t(r) - Y_t(\infty) | \mathbf{D}] = 0, \quad t < r$$

**Assumption: CTS (Common Trend, Staggered):** —————

全コホート共通の never-treated のトレンドが存在し,

$$\mathbb{E}[Y_t(\infty) - Y_1(\infty) | \mathbf{D}] = \mathbb{E}[Y_t(\infty) - Y_1(\infty)] \equiv \theta_t, \quad t \geq 2.$$

**D = 0**, つまり全コホートが未処置な状況においても NAS は成り立つし, 条件もそれを排除しているようには見えないが, それは NAS が NA よりも条件として緩くなっていることを意味するのでは?

■p.44 Exercise 5.4 (Staggered) Staggered design. コホートダミー  $D^r$  ( $q \leq r \leq T$ ) は互いに排他で,  $\mathbf{D}$  は全コホートを網羅する. コホート  $r$  の潜在結果を  $Y_t(r)$  ( $r \in \{q, \dots, T, \infty\}$ ) とし, 観測手段は:

$$Y_{it} = Y_t(\infty) \left( 1 - \sum_r D_i^r \right) + \sum_r D_i^r Y_t(r).$$

仮定 NAS と CTS の下で, 次が成立することを示せ:

$$\mathbb{E}[Y_{it} | \mathbf{D}_i] = \eta + \lambda_q D_i^q + \dots + \lambda_T D_i^T + \sum_{s=2}^T \theta_s f_{st} + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (\mathbf{W}_{it} D_i^r f_{st}).$$

*Proof.* まず  $Y_{it}$  の条件付期待値をコホートで分解する:

$$\begin{aligned} \mathbb{E}[Y_{it} | \mathbf{D}_i] &= \mathbb{E}\left[Y_t(\infty) \left( 1 - \sum_r D_{ri} \right) + \sum_r D_{ri} Y_t(r) \mid \mathbf{D}_i\right] \\ &= \left( 1 - \sum_r D_i^r \right) \mathbb{E}[Y_t(\infty) | \mathbf{D}_i] + \sum_r D_i^r \mathbb{E}[Y_t(r) | \mathbf{D}_i] \quad \because \sigma\text{-加法性}. \end{aligned}$$

CTS より:

$$\mathbb{E}[Y_t(\infty) | \mathbf{D}_i] = \mathbb{E}[Y_1(\infty) | \mathbf{D}_i] + \theta_t$$

NAS により, 各コホート  $r$  について:

$$\mathbb{E}[Y_t(r) | \mathbf{D}_i] = \mathbb{E}[Y_t(\infty) | \mathbf{D}_i] \quad \text{for } t < r.$$

一方で  $t \geq r$  に対しては  $\tau_{rt} \equiv \mathbb{E}[Y_t(r) - Y_t(\infty) | D^r = 1]$  を定義でき, 一般に:

$$\mathbb{E}[Y_t(r) | \mathbf{D}_i] = \mathbb{E}[Y_t(\infty) | \mathbf{D}_i] + \tau_{rt} \quad \text{for } t \geq r,$$

これらを用いて条件付期待値を書き下すと, ( $\mathbf{D}_i \neq \mathbf{0}$  であることを前提として)

$$\begin{aligned} \mathbb{E}[Y_{it} | \mathbf{D}_i] &= \left( 1 - \sum_r D_i^r \right) (\mathbb{E}[Y_1(\infty) | \mathbf{D}_i] + \theta_t) + \sum_r D_i^r (\mathbb{E}[Y_1(\infty) | \mathbf{D}_i] + \theta_t + \mathbb{1}\{t \geq r\} \tau_{rt}) \\ &= \mathbb{E}[Y_1(\infty) | \mathbf{D}_i] + \theta_t + \sum_{r=q}^T D_i^r \mathbb{1}\{t \geq r\} \tau_{rt}. \end{aligned}$$

ここで  $\mathbb{E}[Y_1(\infty) | \mathbf{D}_i] = \eta + \sum_r \lambda_r D_i^r$  と コホートごとの定常差を取り出し 分解すると:

$$\mathbb{E}[Y_{it} | D_i] = \eta + \sum_{r=q}^T \lambda_r D_i^r + \theta_t + \sum_{r=q}^T \mathbb{1}\{t \geq r\} D_i^r \tau_{rt}.$$

次に, 指示関数  $\mathbb{1}\{t \geq r\}$  を時刻ダミー  $f_{st} \equiv \mathbb{1}\{s = t\}$  を用いて書き換える。任意  $t$  に対して

$$\mathbb{1}\{t \geq r\} = \sum_{s=r}^T f_{st},$$

であるから,

$$\begin{aligned} \sum_{r=q}^T \mathbb{1}\{t \geq r\} D_i^r \tau_{rt} &= \sum_{r=q}^T \left( \sum_{s=r}^T f_{st} \right) D_i^r \tau_{rt} = \sum_{r=q}^T \left( \sum_{s=r}^T \mathbb{1}\{s = t\} \tau_{rt} \right) D_i^r \\ &= \sum_{r=q}^T \left( \sum_{s=r}^T f_{st} \tau_{rs} \right) D_i^r = \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} D_i^r f_{st}. \end{aligned}$$

従って:

$$\mathbb{E}[Y_{it} | \mathbf{D}_i] = \eta + \lambda_q D_i^q + \dots + \lambda_T D_i^T + \sum_{s=2}^T \theta_s f_{st} + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (D_i^r f_{st}),$$

が得られる。(ここで  $D_i^r f_{st}$  は「コホート  $r$  のユニットで時刻  $s$  の観測か」を示す指標である。) □

**■要点** Staggered design では, 処置時期が群ごとに異なるため, 単純な TWFE の単一係数推定は誤りというより単一係数仕様が不適切になり得る。適切にコホート（処置開始時期）と時期の交差項を入れれば, コホート・時期ごとに異なる ATT ( $\tau_{rt}$ ) を回収できる。

**■p.45 Staggered design: 実務上の処方箋** 実務上の「正しい」やり方の一つは, 以下の TWFE 回帰の実行:

$$Y_{it} \text{ on } \{D_{iq} \cdot f_{qt}, \dots, D_{iq} \cdot f_{Tt}, \dots, D_{iT} \cdot f_{Tt}\},$$

すなわち全コホートと全処置期間の交差項を説明変数に入れること。これにより  $\tau_{rs}$  を直接推定できる。この回帰は,  $1, D_{ir}, f_{st}$  を完全に入れた pooled OLS と同値であり, 結果としてコホート別の ATT を識別可能にする(ただし識別のための仮定が必要)。

**■p.46 事前トレンドによる CTS の検定** CTS の妥当性を検討するために, 次のような回帰式を考える:

$$Y_{it} = \eta + \sum_{r=q}^T \lambda_r D_i^r + \sum_{s=2}^T \theta_s f_{st} + \underbrace{\sum_{r=q}^T \sum_{s=2}^{r-1} \omega_{rs} (D_{ir} \cdot f_{st})}_{\text{(placebo terms)}} + \sum_{r=q}^T \sum_{s=r}^T \delta_{rs} (D_{ir} \cdot f_{st}) + u_{it}.$$

ここで第三項は 処置前のプラセボ項<sup>\*2</sup> であり, この部分が検定される。NAS の下では:

$$\mathbb{E}[Y_t(r) - Y_1(\infty) | D^r = 1] - \mathbb{E}[Y_t(\infty) - Y_1(\infty) | D^\infty = 1] = 0 \quad t < r$$

が検定される。つまり, 各コホートの処置前が, never-treated のトレンドと異ならないかが確認される。

<sup>\*2</sup> 一般に  $\delta_{rs} \neq \tau_{rs}$  (真の処置効果) であり,  $\delta_{rs} = \tau_{rs}$  となるためには  $\omega_{rs} = 0 \forall r, s$  が必要。

### ■p.47 方法選択の指針

- 共変量がない（または扱わない）なら Wooldridge (2021) が有用
- Goodman-Bacon (2021) の分解プロットは視覚的に有用
- 共変量あり（離散）Wooldridge (2025) が有用
- 共変量が多岐に渡る場合は Callaway & Sant'Anna (2020) のセミパラ推定が推奨される

■p.47 非 Staggered Design の課題 Wooldridge (2021) では、処置が「一方向に進む (staggered)」ケースを想定していたが、実際にはより一般的な状況一すなわち、

- 時間が経つにつれて新たに処置を受ける群が現れる
- 一部の群では処置を中止する

という 非 Staggered なデザインが存在する。このようなより柔軟な設定を扱うのが de Chaisemartin & D' Haultfoeuille (2020) の枠組みである。ここでは、各グループ  $g$  が任意の時点  $t$  において部分的に処置されることを許容する。すなわち、平均的な処置状態の指標  $D_{g,t}$  に対して、

$$0 < \mathbb{E}[D_{g,t}] < 1$$

が成り立つ。このように、グループ内の一部のみが処置を受ける構造を **fuzzy design** と呼ぶ。識別には、少なくとも一部のグループが処置割合を変えない (**stable group**) 必要がある。即ち任意の時点  $t$  で：

$$\exists g' : \mathbb{E}[D_{g',t}] = \mathbb{E}[D_{g',t-1}] > 0$$

が成り立たねばならない。もあるグループ  $g$  で処置率が増加している ( $\mathbb{E}[D_{g,t}] > \mathbb{E}[D_{g,t-1}]$ ) なら、他方で変化しないグループ  $g'$  が存在する必要がある。これは、その時点  $t$  における「テイクアップ率変化」に対して、比較対象となるコントロール群（安定群）が必要であることを意味する。

- **Fair case:** コントロール群では処置率に変化がない場合。
- **Difficult case:** 全ての群でテイクアップ率が変化しており、適切なコントロールが存在しない場合。

**単調性** 各グループ  $g$  に対して、時間を通じて処置率が単調に変化することを仮定する：

$$D_{g,t} \geq D_{g,t-1} \quad \text{or} \quad D_{g,t} \leq D_{g,t-1} \quad (\text{a.s.}).$$

この非 Staggered デザインにおいて識別されるのは、平均処置効果 (ATT) ではなく、**スイッチャーに対する局所平均処置効果 (LATE)** である。すなわち、

$$\sum_{(g,t)} \omega_{g,t} \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0) \mid D_{g,t}(t) \neq D_{g,t}(t-1)],$$

ここで、 $(g,t)$  は「グループ  $g$  が時点  $t$  において処置状態を切り替える (switch する)」観測を示し、 $\omega_{g,t}$  はそのスイッチが多いほど大きくなる重みを表す。したがって、識別される効果は「処置状態が変化した群（スイッチャー）」に関する加重平均効果となる。