Estimation with Partial Linear Model: Asymptotics

川田恵介

Table of contents

1 大標本性質: Recap 1

1.1 例: 平均値の推定 . . . 1

1.2 例: 平均値の推定 . . . 2

1.3 大標本性質: 平均値 . . . 2

1.4 応用上の含意 . . . 2

1.5 平均値: $N = 2000$ . . . 3

1.6 平均値: $N = 200$ . . . 3

1.7 例: 速度の異なる収束 . . . 4

1.8 正規分布への収束 . . . 4

1.9 拡張: 合成指標 . . . 4

1.10 拡張: Implicit function . . . 5

2 大標本性質: with nuisance function 5

2.1 R-leaner . . . 5

2.2 Estimator . . . 5

2.3 分解 . . . 5

2.4 AI のミス: $N = 100$ . . . 6

2.5 DML: $N = 100$ . . . 6

2.6 DML: $N = 100$ . . . 7

2.7 AI のミス: $N = 500$ . . . 8

2.8 DML: $N = 500$ . . . 8

2.9 DML: $N = 5000$ . . . 9

2.10 Reference . . . 9

1 大標本性質: Recap

1.1 例: 平均値の推定

- Estimand: Y の母平均 $E[Y]$ の推定

- Estimator: サンプル平均 $\theta = \sum_i Y_i/N$
    - Moment 法 (“置き換え法”)

- Estimator は、データ上の Y の分布に依存するので、研究者によって異なる
  - 一般に $E[Y] \neq \theta$
  - 多くの実証研究では、点推定量と信頼区間 (ないし p 値) を報告し、対処する

1.2 例: 平均値の推定

```
readr::read_csv("Public/Data.csv") |>
  estimatr::lm_robust(
    Price ~ 1,
    data = _)
```

```
            Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper    DF
(Intercept) 39.00496  0.2015849 193.4915        0 38.60984 39.40008 22138
```

- どのような解釈ができるのか?
  - 何が根拠か?

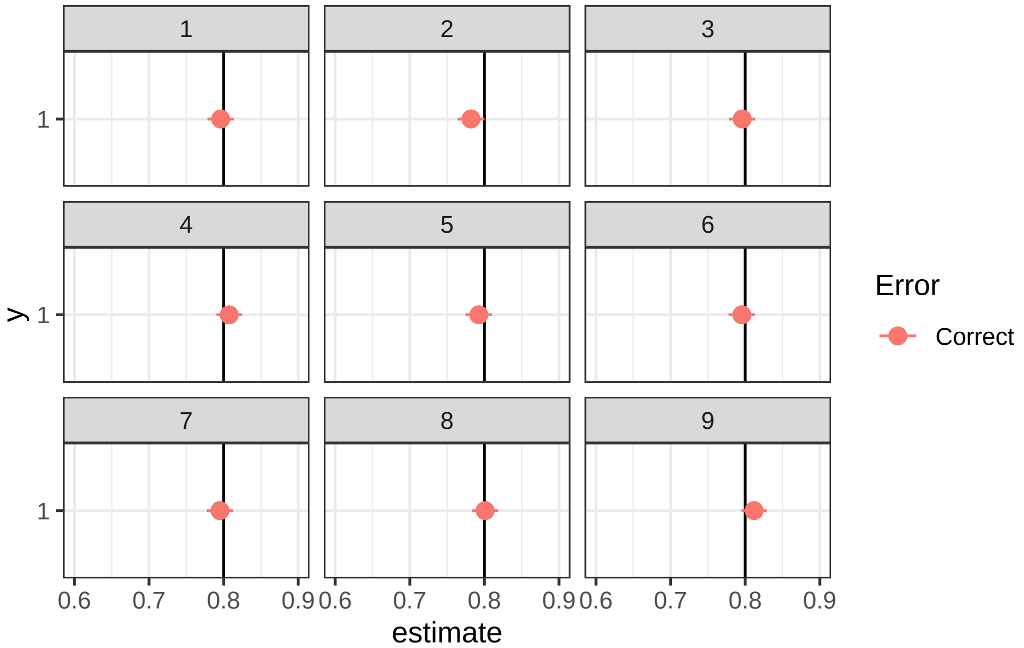
1.3 大標本性質: 平均値

- 事例数が無限大になると、Estimator の分布について、以下の性質が成り立つ
  - サンプル平均は、母平均 $E[Y]$ に収束する $\theta \to E[Y], N \to \infty$
  - θ の分布は、正規分布 $N(E[Y], \sigma^2/N)$ に収束する (中心極限定理)

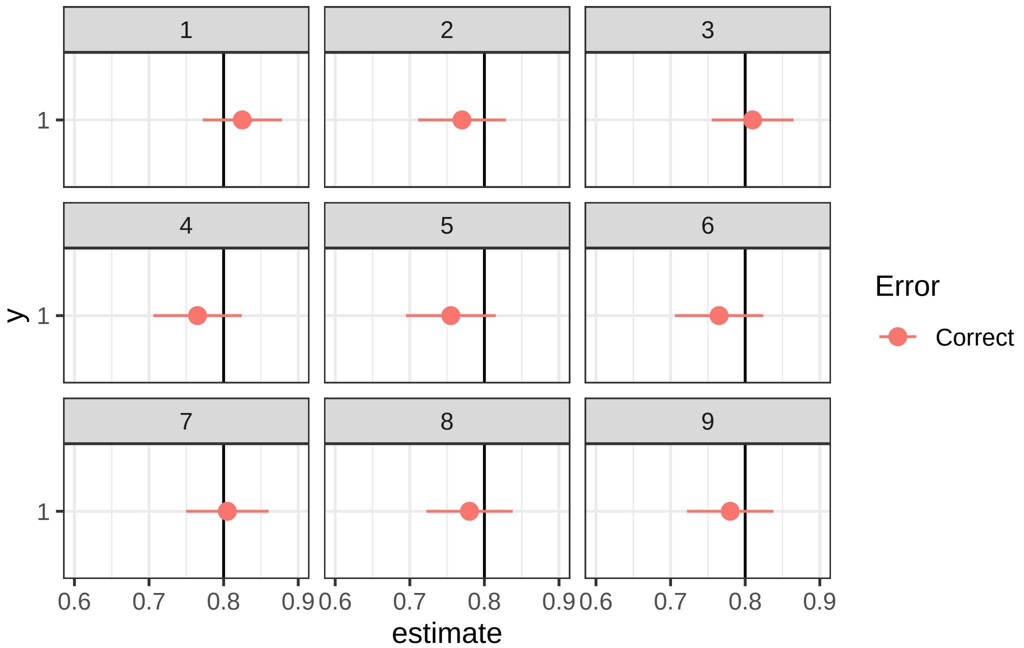
1.4 応用上の含意

- 事例数が**十分に大きい**と
  - 点推定量は、ほぼほぼ母平均と一致する
  - 信頼区間は、ほぼほぼ 95% の” 確率” で母平均を含む
- ただし、十分に大きい、の水準は違う

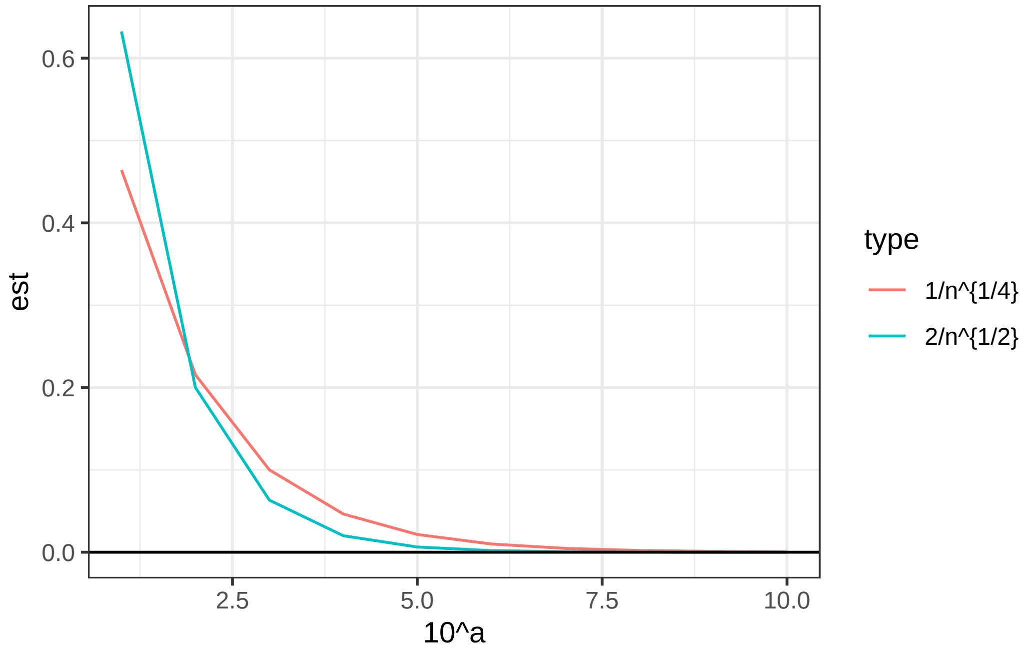
1.5 平均値: $N = 2000$



1.6 平均値: $N = 200$



1.7 例: 速度の異なる収束



1.8 正規分布への収束

- Berry-Esseen's Centraol Limit Theorem (see Chap 1 in CausalML)
- 任意の 標準化された X (平均 0, 分散 1) について

$$|\Pr[X \leq x] - \Pr[N(0,1) \leq x]| \leq KE[|X|^3]/\sqrt(n)$$

- $K =$ 何らかのパラメタ (< 0.5)

1.9 拡張: 合成指標

- Estimand: 複数の変数 $O = \{X_1, .., X_L\}$ によって、定義される指標 $g(O)$ の平均値 $E[g(O)]$
  - サンプル平均値 $\theta = \sum g(O)/N$ で置き換える
  - ただし関数 $g(O)$ は既知であり、全ての研究者が同じ式を用いる必要がある
- 例: 国語 X と算数 Y の合計点の平均値

$$g(O = \{X, Y\}) = X + Y$$

1.10 拡張: Implicit function

- Estimand $= \theta$ 、ただし以下の関数を満たす

$$E[m(\theta, O)] = 0$$

  - Estimator = サンプル平均 $0 = \sum m(\theta, O)$ を満たす θ

- 例: OLS
  - Estimand $= \min E[(Y - \theta X)^2]$ を達成する θ
  - $m(O, \theta) = X(Y - \theta X)$

2 大標本性質: with nuisance function

2.1 R-leaner

- Estimand = 以下を満たす τ

$$0 = E[m(O, g(X), \tau)]$$

  where

$$O = \{X, D, Y\}$$

$$\mu(X) = \{\mu_D(X), \mu_Y(X)\}$$

$$m = (D - \mu_D(X)) \times [Y - \mu_Y(X) - \tau(D - \mu_D(X))]$$

2.2 Estimator

- データ上で置き換えると、$m(O, g(X), \tau)$ 、ただし $g(X) = \{g_D(X), g_Y(X)\}$ は Auxiliary data を用いて推定された関数
- 一見すると Moment 法がそのまま適用できそうだが、AI のミスに注意
  - $\mu(X) \neq g(X)$
  - 研究者によって異なる
- AI のミスが推定結果に与える影響は、R-learber と Single model approach で異なる

2.3 分解

-

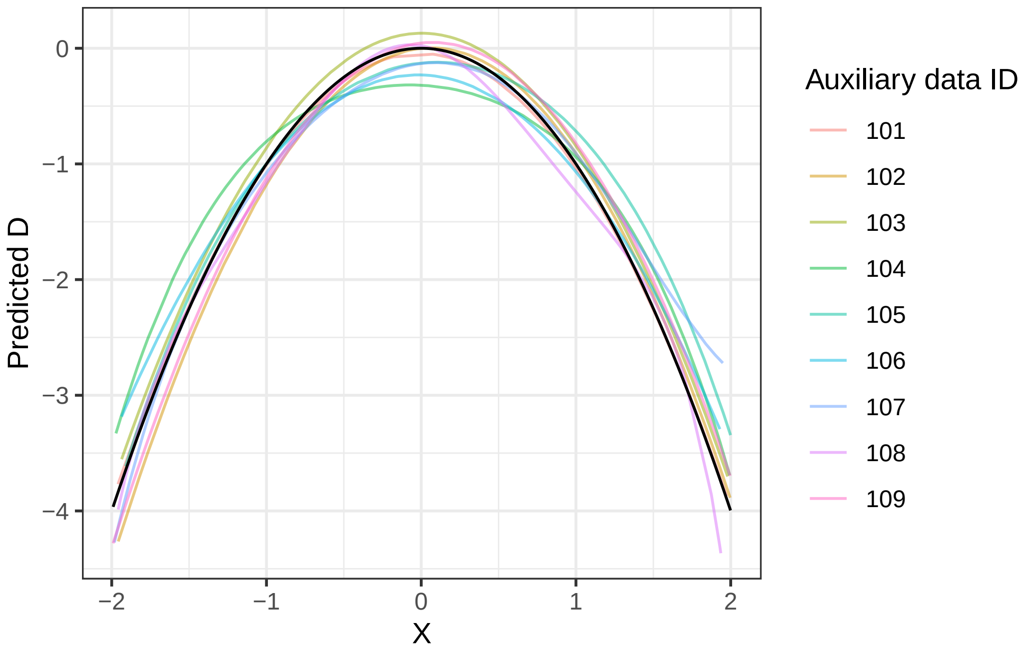
$$m(O, g(X), \tau)$$

- $$= \underbrace{m(O, \mu(X), \tau)}_{Oracle} + \underbrace{m(O, g(X), \tau) - m(O, \mu(X), \tau)}_{AIのミスの影響}$$

- μ は、母平均であり、すべての研究者にとって共通
  - 第 1 項については、通常の Moment 法が適用可能
- 第 2 項については、適用不可能だが、R-learner においては、AI のミスの影響は軽減されている

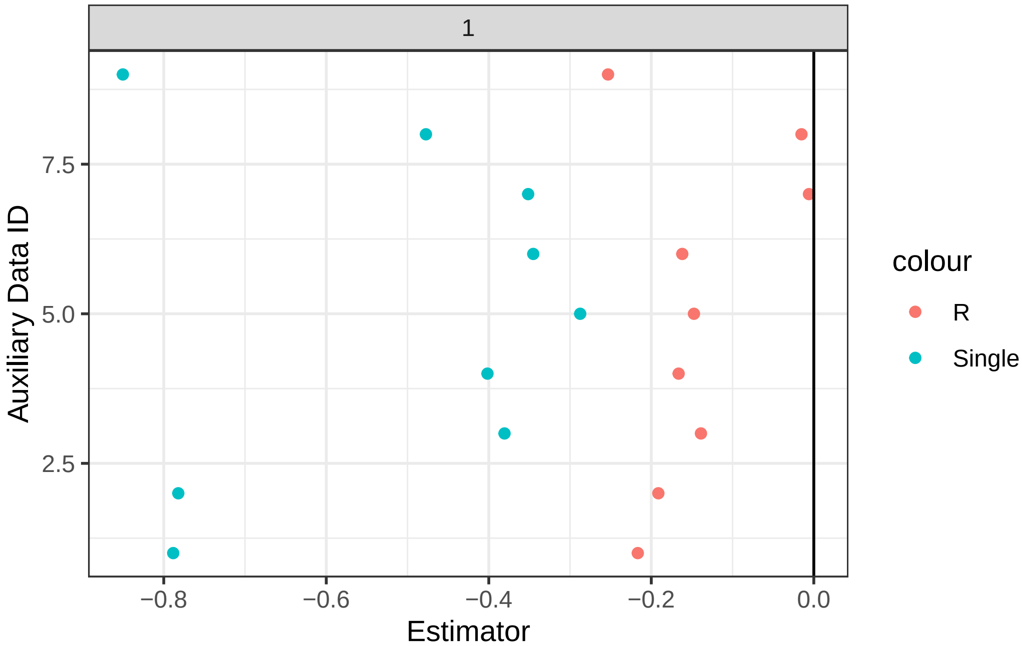
2.4 AI のミス: $N = 100$

- Main Data は共通、Auxiliary data は 101-109 まで存在
  - 黒線 = 母平均



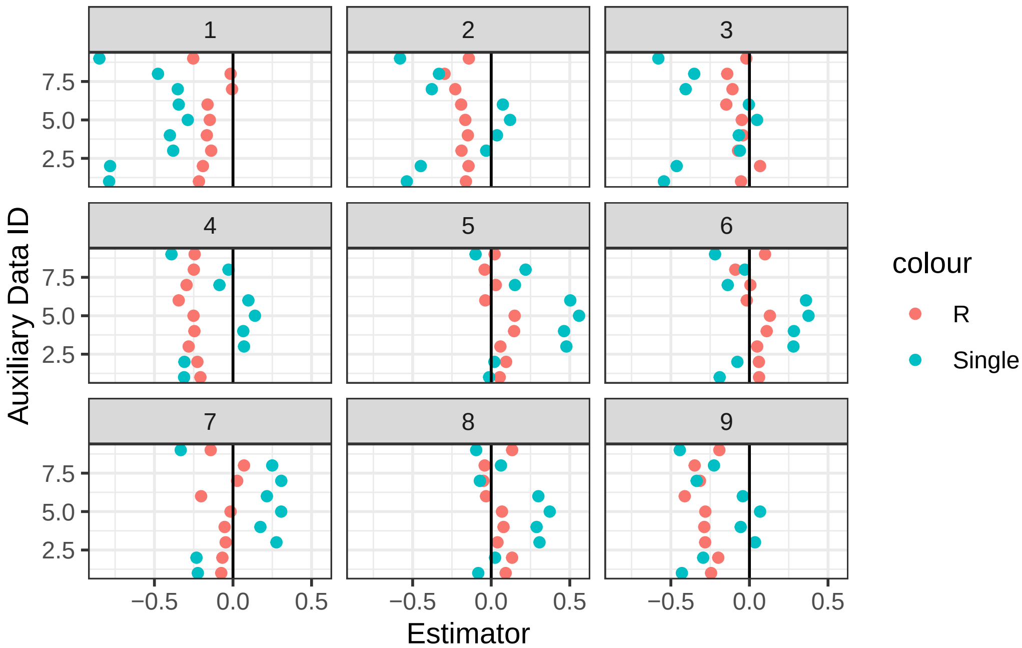
2.5 DML: $N = 100$

- Main data が共通でも、Auxiliary data の違いにより、推定値の分布が生まれる



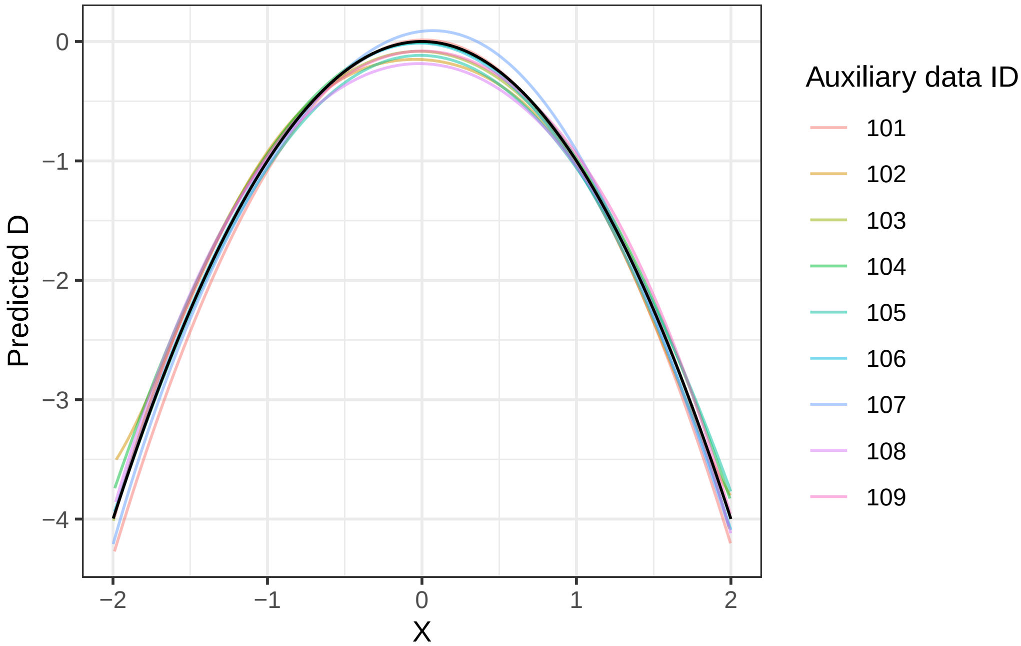
2.6 DML: $N = 100$

- Main data が異なれば、estimator はさらに異なる



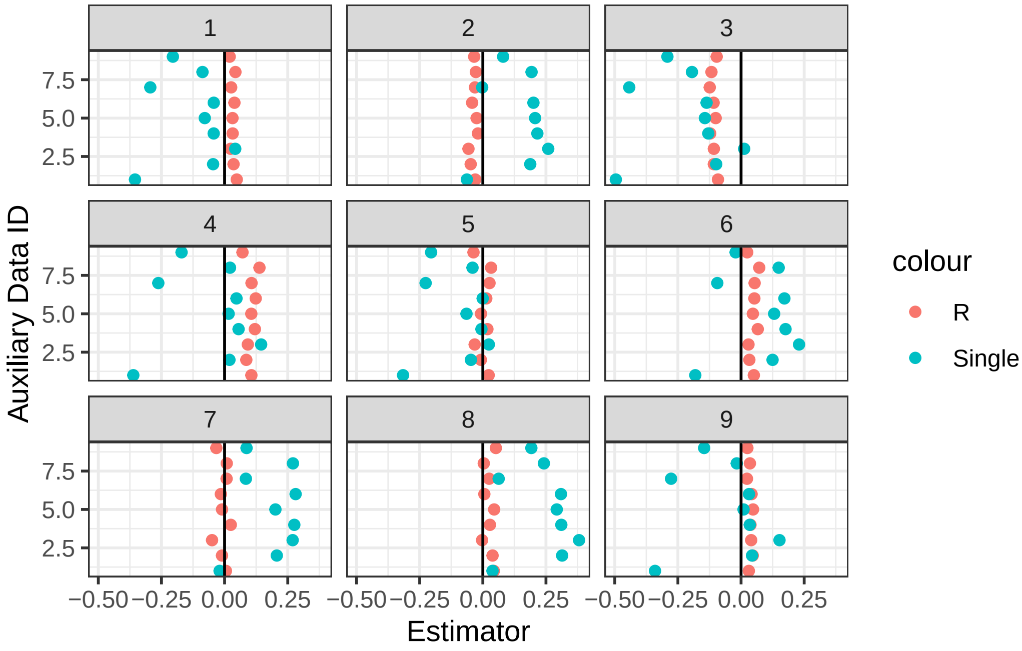
2.7 AI のミス: $N = 500$

- AI のミスは減る



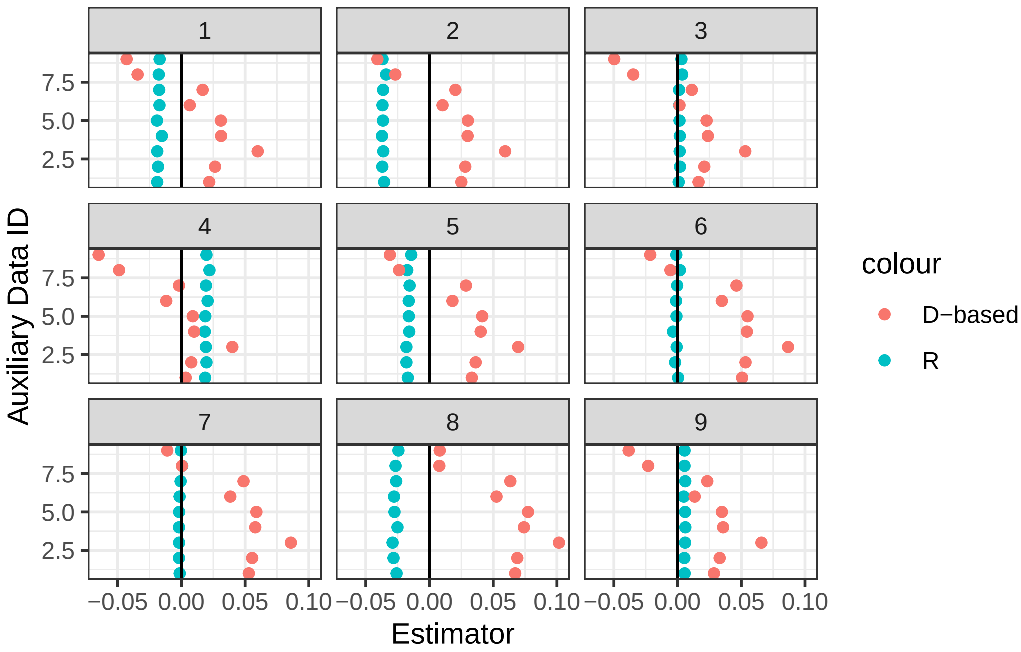
2.8 DML: $N = 500$

- R learner において、Auxiliary data がもたす分布が大きく減る



2.9 DML: $N = 5000$

- R learner において、Auxiliary data がもたす分布が大きく減る



2.10 Reference