Estimation with Partial Linear Model

川田恵介

Table of contents

1 Moment 推定への応用 2
- 1.1 学習への動機 . . . 2
- 1.2 一般的手順 . . . 2
- 1.3 R(obinson/siduals) learner . . . 3
- 1.4 直感 . . . 3

2 性質: 一致推定量 3
- 2.1 AI のミス . . . 3
- 2.2 重要な仮定 . . . 3
- 2.3 性質: 一致推定量 . . . 4
- 2.4 Estimand . . . 4
- 2.5 性質: 書き換え . . . 4
- 2.6 OLS/Double selection との接続 . . . 5
- 2.7 まとめ . . . 5

3 Local robustness 5
- 3.1 比較対象: Single Model . . . 5
- 3.2 現実の予測値 . . . 6
- 3.3 例 . . . 6
- 3.4 例 . . . 6
- 3.5 例 . . . 6
- 3.6 例: 事例数の影響 . . . 7
- 3.7 例: 事例数の影響 . . . 7
- 3.8 例: Single Model . . . 7
- 3.9 まとめ . . . 7
- 3.10 まとめ . . . 8

4 数値例 8
- 4.1 計算結果 . . . 8

4.2 計算結果: 推定結果 . . . 9
4.3 推定結果: AI のミス . . . 9
4.4 推定結果: AI のミス . . . 10
4.5 計算結果: 事例数の増加 . . . 10
4.6 計算結果: 低品質のアルゴリズム . . . 11

5 補論 11
5.1 補論: 分母への影響 . . . 11
5.2 補論: e_Y の影響 . . . 11
Reference . . . 12

1 Moment 推定への応用

- Chernozhukov et al. (2018); Chernozhukov et al. (2022)
  - gentile introduction (Ichimura and Newey 2022; Fisher and Kennedy 2021; Hines et al. 2022)
  - Chap 10 in CausalML
- 大枠: 大表本理論 + Neyman's orthogonal condition + Cross fitting

1.1 学習への動機

- 多くの応用: Panel data (see Sant'Anna and Zhao 2020; Roth et al. 2023), Instrumental variable (Chap 13 in CausalML), Causal (Heterogeneity) Learning (Chap 15 in CausalML), Cause of Effect (Cuellar and Kennedy 2020), Mediation Analysis (Opacic, Wei, and Zhou 2023)
- 多くの実装: DoubleML (R/Python), EconML (Python), causalml (Python), DDML (STATA/R)

1.2 一般的手順

- データを 2 分割 (auxiliary/main data) し、2 段階の推定を行う
  - 交差推定も可

1. Estimand を母集団上でモーメント条件として定義
2. Auxiliary data を用いて、"補助的な" 予測モデルを (機械学習で) 推定し、main data に適用する
3. Main data と予測モデルを用いて、OLS/平均値の推定を行う

1.3 R(obinson/siduals) learner

1. Y, D を予測するモデル $g_Y(X), g_D(X)$ を auxiliary data で推定し、main data に適用
2. main data で以下を OLS し、“通常” の信頼区間を報告

$$Y - g_Y(X) \sim D - g_D(X) | Auxiliary\ data$$

- 注: g_D, g_Y の Uncetainly を無視している

1.4 直感

- キャチーに言うと、AI に補助を受けながら、古典的な回帰をしている
  - $g_Y(X), g_D(X)$ を機械学習により推定している
- 大きな論点は、β_D の推定誤差を生み出す二つの要因
  - Main Data によって、Y, D の分布が異なる (古典的要因)
  - Auxiliary Data によって、$g_Y(X), g_D(X)$ が異なる (“AI のミス”)
    - 処理しにくく、扱いに工夫が必要

2 性質: 一致推定量

- Auxiliary/main data の事例数 $\{N_A, N_M\}$ が無限大になると、どのような値が推定されるか？

2.1 AI のミス

- 重要概念:

$$e_Y = \underbrace{E[Y|X]}_{=\mu_Y(X)} - g_Y(X),$$

$$e_D = \underbrace{E[D|X]}_{=\mu_D(X)} - g_D(X)$$

  - 本講義では以降、“AI のミス”、と呼ぶ

2.2 重要な仮定

- auxiliary data の事例数 N_A が無限大になれば、$g_Y(X), g_D(X)$ は、$\mu_Y(X), \mu_D(X)$ に収束する:

$$\{e_Y(X), e_D(X)\} \to 0, n_A \to \infty$$

- 以下の議論で仮定
  - Simple な Linear model の推定では、満たされない.
  - 機械学習が有効

2.3 性質: 一致推定量

- $N_A, N_M \to \infty$ となると
- Step 1 で用いる機械学習アルゴリズムが一致性を満たすとすると、$\{g_Y(X), g_D(X)\} \to \{\mu_Y(X), \mu_D(X)\}$
- Step 2 は、以下の Population OLS と一致

$$\underbrace{Y - \mu_Y(X)}_{Irreducible\ error} \sim \underbrace{D - \mu_D(X)}_{Irreducible\ error}$$

- Irreducible error の Population OLS と一致

2.4 Estimand

- R-leanrer は、Partial linear model (Robinson 1988) 上で定義される Estimand τ の推定方法と見做せる
- $$E[Y|D, X] = \underbrace{\tau}_{Constant\ effect} \times D + f(X)$$
  - 次回、$\tau(X)$ に拡張

2.5 性質: 書き換え

- $$Y = \tau D + f(X) + \underbrace{U}_{E[U|D,X]=0}$$
- $$\underbrace{E[Y|X]}_{\mu_Y(X)} = \tau \underbrace{E[D|X]}_{\mu_D(X)} + f(X) + \underbrace{E[U|X]}_{=0}$$
- 片片を引くと
$$Y - \mu_Y(X) = \tau(D - \mu_D(X)) + U$$
  - τ = Irreducible error 同士を Population OLS した結果

2.6 OLS/Double selection との接続

- FWL 定理を用いれば、OLS 推定の一般化として解釈可能
- OLS 推定は、以下の Algorithm と一致

0. Double selection を用いて、$Z \subset X$ を選ぶ
1. **全データと OLS** を用いて、$g_Y(Z), g_D(Z)$ を推定
2. **全データと OLS** を用いて、$Y - g_Y(Z) \sim D - g_D(Z)$ を推定

- 違いは、データ分割と予測モデルの柔軟な推定

2.7 まとめ

- Estimand: Irreducible error の Population OLS の結果

$$\tau = \frac{E[(Y - \mu_Y(X)(D - \mu_D(X))]}{E[(D - \mu_D(X))^2]}$$

- Estimator:

$$\frac{\sum_i [(Y_i - g_Y(X_i))(D_i - g_D(X_i))]}{\sum_i (D_i - g_D X_i)^2}$$

- AI のミス ($g \neq \bar{\mu}$) を無視し、信頼区間を計算している

3 Local robustness

- R leaner は、Neyman の直行条件 (次回説明) を満たすので、1 段階目の推定誤差 (“AI のミス”) の影響は軽減される
  - 事例数が増えればより軽減される

3.1 比較対象: Single Model

- 繰り返し期待値の法則より、Estimand は書き換えられる

$$\begin{aligned} \tau &= \frac{E[(Y - \mu_Y(X))(D - \mu_D(X))]}{E[(D - \mu_D(X))^2]} \\ &= \underbrace{\frac{E[Y(D - \mu_D(X))]}{E[(D - \mu_D(X))^2)]}}_{D \ model \ based \ approach} \end{aligned}$$

- $Y \sim D - g_D(X)$ (Single Model) を回帰してもいいのは?

– $\beta_D \to \tau, \{N_A, N_M\} \to \infty$、は保証される

3.2 現実の予測値

- 有限のデータで推定する限り、AI のミス $e_Y(X), e_D(X)$ は常に発生する

$$\mu_Y(X) = g_Y(X) + e_Y(X),$$
$$\mu_D(X) = g_D(X) + e_D(X)$$

- 以下 D を予測する AI のミス $e_D(X)$ が β_D に与える影響を議論
  - $e_Y(X)$ についても同様の議論が適用できる

3.3 例

- $X =$ 立地 (23 区) のみとする
- **I**tabashi についての D の予測ミス $e_D(I)$ の影響

3.4 例

- R - learner:

$$\frac{\sum(Y_i - g_Y(X_i))(D_i - \mu_D(X_i) - e_D(X_i))/N_M}{\sum(D_i - \mu_D(X_i) - e_D(X_i))^2/N_M}$$

- 分子への影響は、

$$\sum_{i|X_i=I} (Y_i - g_Y(I)) \times e_D(I)/N$$

  - 分母への影響についても、以下と同じような議論が適用可能

3.5 例

- 分子への影響は、

$$\underbrace{\underbrace{\frac{N_M(I)}{N_M}}_{X=I\text{の割合}} \times \Big(\underbrace{\sum_{i|X_i=I} Y_i/N_M(I)}_{=\bar{\mu}_Y(I)} - g_Y(I) \Big)}_{AI\text{のミスの影響を緩和}} \times e_D(I)$$

- サンプル平均 $\bar{\mu}_Y(I)$ と $g_Y(I)$ が近づくと、$e_D(I)$ の影響は伝わりにくくなる

3.6 例: 事例数の影響

- $\{N_A, N_M\}$ が増えると、1 段階目の悪影響が、2 重に削減される
  - $$\{\bar{\mu}_Y(X), g_Y(X)\} \to \mu_Y(X)$$
  - $$e_D(X) \to 0$$
- D を予測する AI と Y を予測する AI がダブルチェックしている

3.7 例: 事例数の影響

- よって
$$-N_M(I)/N_M \times \underbrace{(\bar{\mu}_Y(I) - g_Y(I))}_{\to 0} \times \underbrace{e_D(I)}_{\to 0},$$
$$\{N_A, N_M\} \to \infty$$
- 事例数の増加は、D の予測モデルを改善しつつ、**同時に**、ダブルチェック機能も改善する

3.8 例: Single Model

- Single Model:
$$\frac{\sum Y_i(D_i - g_D(X_i) - e_D(X_i))}{\sum(D_i - g_D(X_i) - e_D(X_i))^2}$$
- 分子への影響は、
$$\sum_{i|X_i=Itabashi} Y_i \times e_D(I)$$
- AI のミス e_D の影響をもろに受ける

3.9 まとめ

- R-Learner は、AI のミスを緩和する仕組みが内蔵されている
$$(\bar{\mu}_Y(I) - g_Y(I)) \times e_D(I)$$
  - Y の予測モデルの性能が良く、Main Data の事例数が十分であれば、D の予測モデルのミス $e_D(X)$ の悪影響を削減できる

3.10 まとめ

- 性能の良いアルゴリズム (Stacking など) や Auxiliary data の増加は、2 重の利点をもたらす
  - 注意: 予測性能が悪く $\bar{\mu}(X) - g(X)$ が 1 を超える場合が多ければ、予測誤差を" 増幅" する可能性がある

4 数値例

- $X \in \{-1, 0, 1\}$
- $D = X^2 + \underbrace{U_D}_{\sim N(0,1)}$
- $Y = 2D - X^2 + \underbrace{U_D}_{\sim N(0,1)}$
- とりあえず $N_{Main} = N_{Auxiliary} = 200$

4.1 計算結果

- $g_D(X) = X^2$ と推定できたとすると
- $g_Y(X)$ は $Y \sim poly(X, 2)$ を LASSO 推定

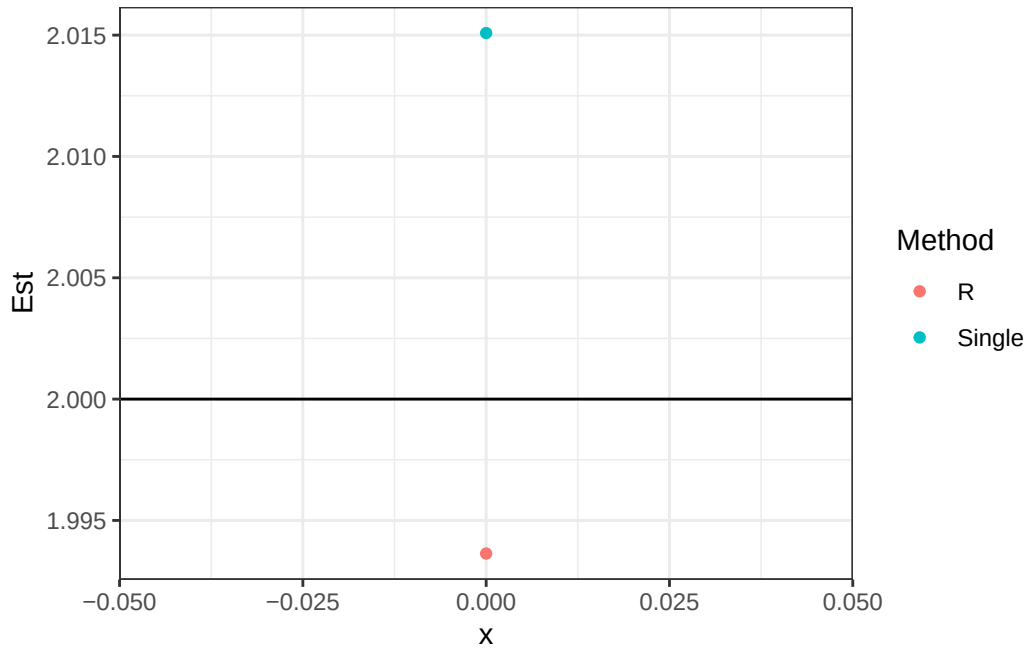
```
HatY = hdm::rlasso(
  Y ~ poly(X,2),
  ArxiliaryData,
  post = FALSE) |>
  predict(MainData) |>
  as.numeric()

MainData$ResY = MainData$Y - HatY
MainData$ResD = MainData$D - (MainData$X)^2

ModelR = lm(ResY ~ ResD, MainData)

ModelS = lm(Y ~ ResD, MainData)
```

4.2 計算結果: 推定結果

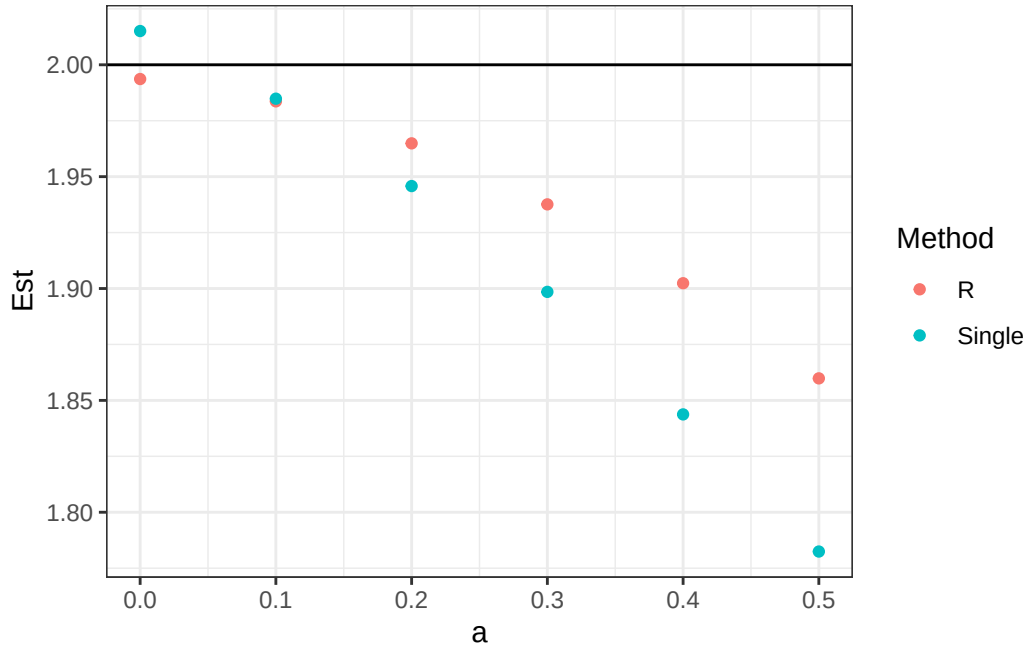


4.3 推定結果: AI のミス

- $g_D(X) = X^2 + \underbrace{a \times X^2}_{e_D(X)}$ と誤って推定された場合、AI のミスの影響は、$e_D(X) \times (\bar{\mu}_Y(X) - g_Y(X))$
- $a = 0.5$ の場合は、

```
# A tibble: 3 x 4
      X `mu_Y - g_Y` `e|a=0.5` `(e|a=0.5)*(mu_Y - g_Y)`
  <int>        <dbl>     <dbl>                    <dbl>
1    -1       -0.196       0.5                  -0.0979
2     1       -0.134       0.5                  -0.0672
3     0       -0.838       0                     0
```

4.4 推定結果: AI のミス



4.5 計算結果: 事例数の増加

- g_Y, g_D を LASSO で推定
- 事例数が、200 から 1000 まで増加

```
# A tibble: 6 x 6
      X     N     E_D `mu_Y - g_Y` `E_D*(mu_Y - g_Y)` Average
  <int> <dbl>   <dbl>        <dbl>              <dbl>   <dbl>
1    -1   200 -0.415        -0.196             0.0813   0.185
2    -1   500  0.0774       -0.271            -0.0210   0.0505
3     0   200 -0.500        -0.838             0.419    0.185
4     0   500 -0.417        -0.400             0.167    0.0505
5     1   200 -0.415        -0.134             0.0558   0.185
6     1   500  0.0494        0.118             0.00584  0.0505
```

- $e_D(X)$ の悪影響をより緩和
  - 個別の X で見れば、悪化しうるが、全体平均としては削減されている

4.6 計算結果: 低品質のアルゴリズム

- $g_Y(X)$ を $Y \sim X$ を OLS 推定して獲得

```
# A tibble: 6 x 7
      N     X Method     E_D `mu_Y - g_Y` `E_D*(mu_Y - g_Y)` Average
  <dbl> <int> <chr>    <dbl>        <dbl>              <dbl>   <dbl>
1  1000    -1 LASSO   0.0952       0.0229            0.00218  0.0412
2  1000    -1 OLS     0.357        0.285             0.102    0.237
3  1000     0 LASSO  -0.306       -0.368             0.113    0.0412
4  1000     0 OLS    -0.712       -0.773             0.550    0.237
5  1000     1 LASSO   0.0694       0.125             0.00868  0.0412
6  1000     1 OLS     0.220        0.276             0.0607   0.237
```

- $e_D(X)$ の悪影響が緩和されない

5 補論

5.1 補論: 分母への影響

$$\sum_{i|X_i=I} (D_i - \mu_D(I) - e_D(I))^2$$

$$= \sum_{i|X_i=I} (D_i - \mu_D(I))^2$$

$$-2 \underbrace{\sum_{i|X_i=I} e_D(D_i - \mu_D(X_i))}_{N_M(I) \times e_D \times \underbrace{[\bar{\mu}_D(I) - \mu_D(I)]}_{\to 0 \ , \ N_M \to \infty}} + \underbrace{e_D^2}_{<e_D \ if \ e_D<1}$$

5.2 補論: e_Y の影響

$$\frac{\sum_{i|X_i=I}(D_i - g_D(I))(Y_i - \mu_Y(I) - e_Y(I))}{\sum(D - g_D(I))^2}$$

$$= \frac{1}{\sum(D - g_D(I))^2} \times \underbrace{\sum_{i|X_i=I} (D_i - g_D(I))}_{\to 0 \ , \ \{N_M, N_A\} \to \infty} \times e_Y(I)$$

Reference

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." Oxford University Press Oxford, UK.

Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins. 2022. "Locally Robust Semiparametric Estimation." *Econometrica* 90 (4): 1501–35.

Cuellar, Maria, and Edward H Kennedy. 2020. "A Non-Parametric Projection-Based Estimator for the Probability of Causation, with Application to Water Sanitation in Kenya." *Journal of the Royal Statistical Society Series A: Statistics in Society* 183 (4): 1793–1818.

Fisher, Aaron, and Edward H Kennedy. 2021. "Visually Communicating and Teaching Intuition for Influence Functions." *The American Statistician* 75 (2): 162–72.

Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. "Demystifying Statistical Learning Based on Efficient Influence Functions." *The American Statistician* 76 (3): 292–304.

Ichimura, Hidehiko, and Whitney K Newey. 2022. "The Influence Function of Semiparametric Estimators." *Quantitative Economics* 13 (1): 29–61.

Opacic, Aleksei, Lai Wei, and Xiang Zhou. 2023. "Disparity Analysis: A Tale of Two Approaches."

Robinson, Peter M. 1988. "Root-n-Consistent Semiparametric Regression." *Econometrica*, 931–54.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2): 2218–44.

Sant'Anna, Pedro HC, and Jun Zhao. 2020. "Doubly Robust Difference-in-Differences Estimators." *Journal of Econometrics* 219 (1): 101–22.