

# Nonparametric CATE: Estimation and Application

## 機械学習の経済学への応用

川田恵介

### 内容

前回:  $\tau(X) = E[Y|D = 1, X] - E[Y|D = 0, X]$  の Parametric な推定

1.  $\tau(X) = E[Y|D = 1, X] - E[Y|D = 0, X]$  の Nonparametric な推定
2. 予測への応用
3. “顕著な異質性を持つグループ” の発見と信頼区間計算

### ここまでのまとめ

- $\tau(X)$  の有限個の実数パラメータによる特徴づけ
  - $\tau(X) \simeq \beta_0$ : Partialling-out/AIPW
  - $\tau(X) \simeq \beta_0 + \beta_1 X_1 + \dots + \beta_L X_L$
- $\tau(X)$  を Nonparametric に推定する手法はあるが、有益？
  - 一般に信頼区間計算が難しい
  - 他の手法と同様に、分析目標をハッキリさせる必要がある

### 応用

- 個人因果効果の予測: できれば意思決定を強力に支援
  - 限界原理に基づく意思決定: 因果効果を知る必要がある
- 顕著な異質性を持つグループの”発見”
  - 平均効果は正だが、負の影響を受けるグループは存在しないか？
  - とてつもなく大きな正の影響を受けるグループはいるか？

## 条件付き平均差の推定

- 比較研究におけるゴールの一つ
- 条件付きランダム化が成り立っているのであれば、“個人効果”の最善の予測

## R Learner

- $\tau(X)$  を (擬似的な) Nonparametric (LASSO や Random Forest など) に推定する一つの手法
- Nie and Wager (2021)
- Causal Forest (Wager and Athey 2018; Athey, Tibshirani, and Wager 2019)
  - 入門 & 応用 (Athey and Wager 2019)

## Generalized Partial Linear Model

$$E[Y|D, X] = \beta_\tau(X) \times D + f(X)$$

- Partialing out

$$\underset{=Y-E[Y|X]}{\bar{Y}} = \beta_\tau(X) \times \underset{=D-E[D|X]}{\bar{D}} + u$$

- $E[u|D, X] = 0$

## Population Risk Minimization

$$\min_{\beta_\tau(X)} E[(\bar{Y} - \beta_\tau(X) \times \bar{D})^2]$$

- 無限大に大きなサンプルで丸暗記モデル推定 ( $X$  のあらゆる組み合わせについてのサブサンプル推定) をした結果得られる推定値

## Empirical Risk Minimization

$$\min_{\beta_\tau(X)} \sum_i [(\bar{Y}_i - \beta_\tau(X) \times \bar{D}_i)^2]$$

- まともにやると、ほぼほぼ過剰適合

- 何らかの単純化が必要
  - 罰則付き線形モデル  $\beta(\tau) \times \bar{D}$  を十分に多く複雑な交差項として、定式化
  - モデル集計: Random Forest など

## Causal Forest

- Population Risk を以下のように書き換えられる

$$\begin{aligned} \min_{\beta_\tau(X)} E[(\bar{Y} - \beta(\tau) \times \bar{D})^2] \\ = \min_{\beta_\tau(X)} E[(\frac{\bar{Y}}{\bar{D}} - \beta(\tau))^2 \times \bar{D}^2] \end{aligned}$$

- $\bar{Y}/\bar{D}$  を Target,  $\bar{D}^2$  を Sampling weight とした Random Forest や Stacking

## Inference

- 推計誤差を評価できるか?
  - Causal Forest は一応可能、 $X$  が多いと無理
- 一般には難しい (不可能?)
- 多くの研究課題において、“too much” な目標かも

## Treatment Effect Risk

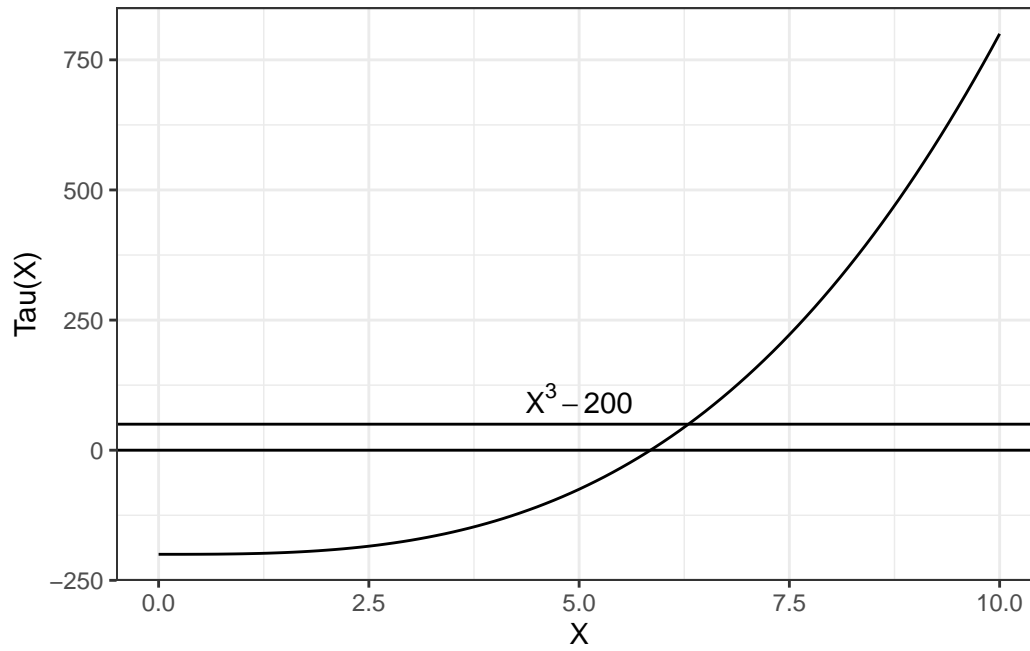
- Kallus (2022)
- 大きな異質性があるグループが存在するかどうかを探索する
  - $\tau(X)$  の予測値  $s(X)$  をシグナルとして用いる

## Estimand

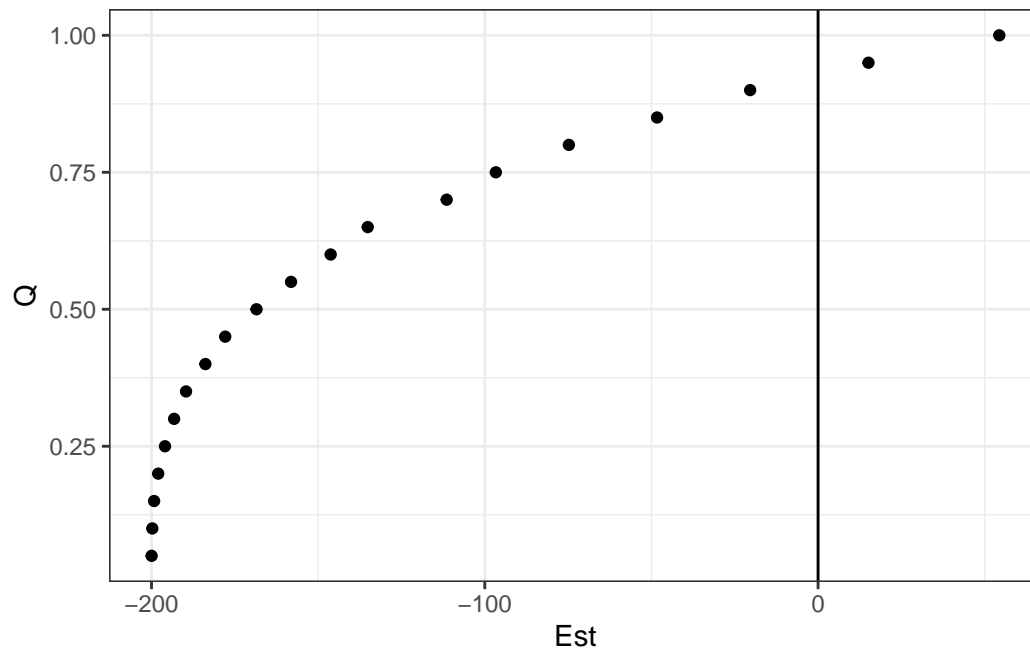
$$R(q) = E[\tau(X) | \tau(X) \leq Q(\tau(X), q)]$$

- $Q(s(X), q) = s(X)$  の  $q$  分位点
  - $q$  は研究者が事前に (複数) 指定
  - 多重検定の補正は現実的な選択肢

Estimand



Estimand



## Signal as signal

- Chernozhukov et al. (2020)
- $\tau(X) \neq s(X)$  を受け入れる
  - 異質性についてのシグナル (にすぎない) とみなす
- シグナルの応用: 例えば Group Average Difference
  - $E[\tau(X)|Q(s(x), q) \geq s(X)]$
- 注意:  $s(X)$  は母集団上で定義された (明確な解釈がある) ものとは見做さない

## 推定

- $s(X)$  は交差推定され、所与 (非確率変数) とみなすのであれば、推定は容易
- $s(X) \geq Q(s(X), q)$  を満たす事例について、AIPW 推定などを行えば OK

## AIPW 推定の応用

- 推定値  $= \sum_i \phi_i / N$

$$\phi_i = \frac{I(Q(\bar{s}, q))}{q} \times \phi_i^{AIPW}$$

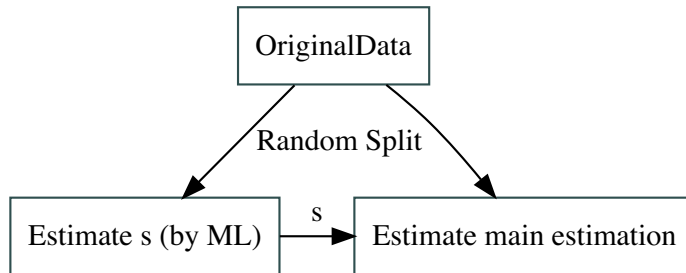
$$\phi_i^{AIPW} = f_Y(1, X_i) - f_Y(0, X_i)$$

$$+ \frac{D_i(Y_i - f_Y(1, X_i))}{f_D(X_i)} - \frac{(1 - D_i)(Y_i - f_Y(0, X_i))}{1 - f_D(X_i)}$$

## 交差推定の活用

- $s(X)$  を固定のルールとみなすには、交差推定が必要
  - (ランダムに分割された) 別データで行う
- $f_Y(D, X), f_D(X)$  ももちろん交差推定が必要

## 推定フロー: 例



## 具体的な推定手順

- $s$  の推定: 例えば R learner、あるいは  $\phi_i^{AIPW}$  を”予測する”
  - Training データ内で
    1.  $\phi_i^{AIPW}$  を算出 (Training データ内で交差検証)
    2.  $E[\phi^{AIPW}|X]$  を Stacking などで推定
- メインの推定は、Test データ内で AIPW 推定

## 解釈

$$E[\tau(X)|Q(s(X), q) \geq s(X)]$$

$$\geq E[\tau(X)|Q(\tau(X), q) \geq \tau(X)]$$

- 下位  $q$ th サブグループ内での平均差の上限 (の推定値)

## Signal as Nuisance

- やっぱ  $E[\tau(X)|Q(\tau(X), q) \geq \tau(X)]$  を推定したい
  - $s(X)$  という解釈困難な指数を用いた分析???
- 問題は一般に  $s(X) \neq \tau(X)$ 
  - 乖離している可能性を考慮する必要がある

## Neyman's orthogonality の応用

- Neyman's orthogonality を満たすスコアを設定し、漸近的に  $s(X)$  の推定誤差を無視できないか?
- Kallus (2022) の推定値:  $= \sum \phi_i / N$

$$\phi_i = \frac{I(Q(s, q) \geq s)}{q} \times \phi_i^{AIPW} + \underbrace{\frac{q - I(Q(s, q) \geq s)}{q} Q(s, q)}_{Adjusted}$$

## 推定手順

- Signal as signal とほぼ同様
- Training データで推定された  $s$  を用いて、 $Q(s, q)$  も計算
- 最終推定において  $s(X)$  の推定誤差を無視するには、 $s(X)$  が  $\tau(X)$  の一致推定量かつ  $n^{1/4}$  以上の速度で収束 (Kallus 2022)

## Kallus (2022) の例

- フランスにおける失業者への支援サービス主体を民間から公営に移行させる施策が再就職率に与える影響 (Behaghel, Crépon, and Gurgand 2014) を再検証
  - RCT により正の平均効果を確認
- [Section 5: Figure 1-5](#)
  - 大きな異質性: 負の効果

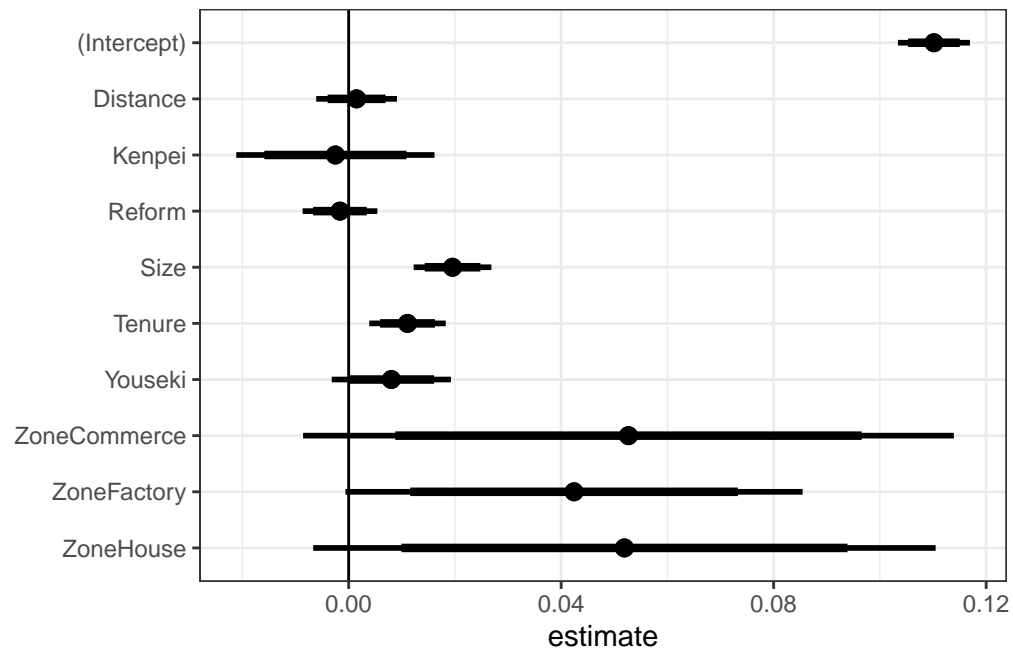
## 例を用いた復習

- Research Question: 2019 年と 2021 年において、“観察できる特徴が共通な” 中古マンションの取引価格はどのように変化したのか
- Data: 2019 年と 2021 年の全国の中古マンション取引データ
  - $Y$  : 取引価格
  - $D$  : 2021/2019
  - $X$  : 広さ、立地市区町村、駅からの距離、広さ、築年、建ぺい・容積率、都市区間、改装

## Linear Model

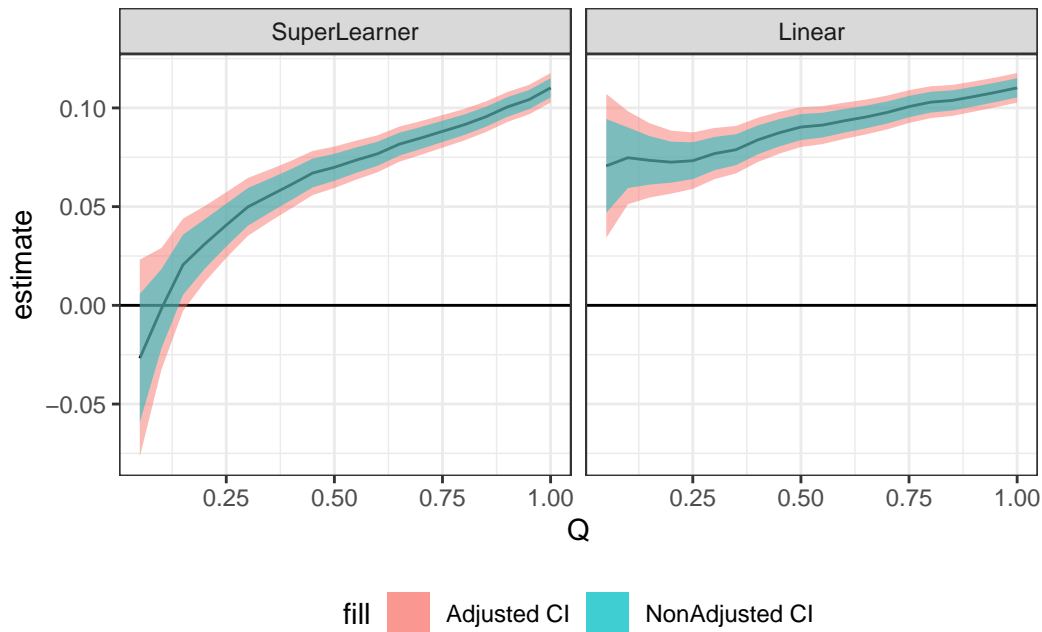
- 解釈が容易 & 信頼区間形成が容易なモデル
  - Semenova and Chernozhukov (2021)
- Effect Risk の推定としての解釈:  $E[s(X)|X]$  を線形モデルで推定

## Example: Linear Model





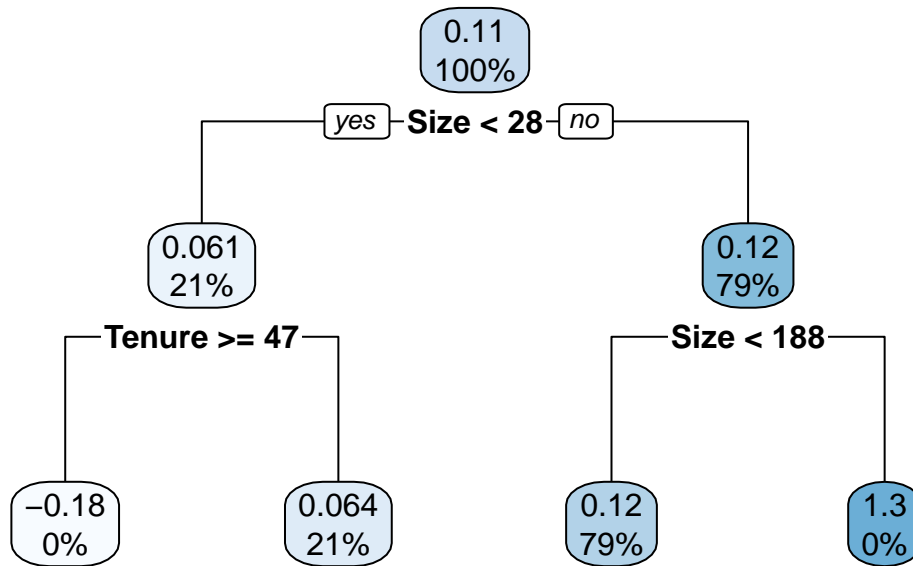
## Example: Stacking VS Linear



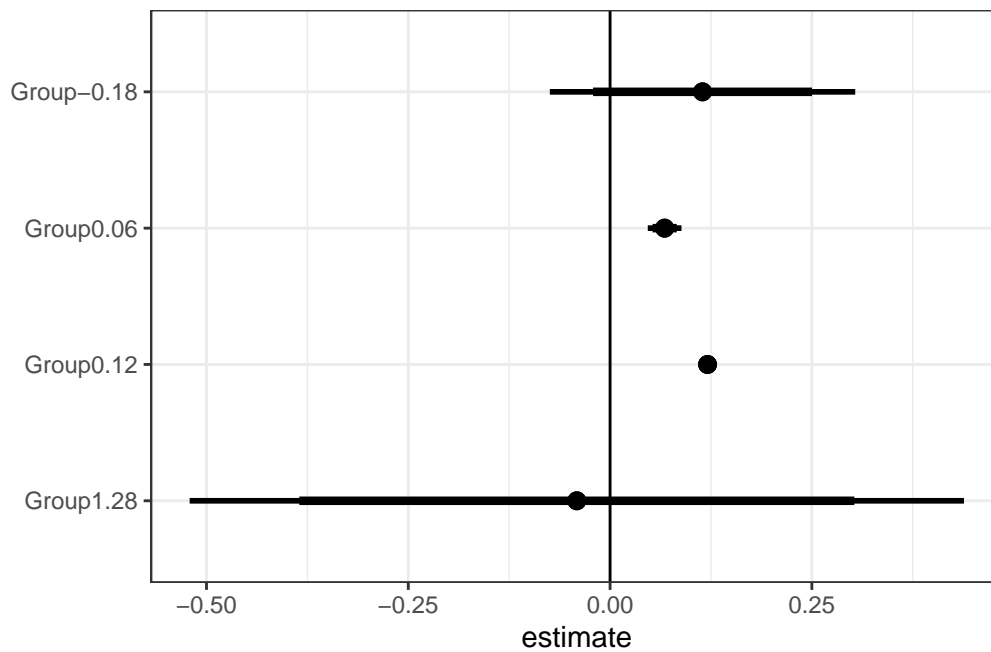
## Tree Model

- データ主導かつ解釈モデリングモデル: Tree, LASSO など
  - Honest Tree (Athey and Imbens 2016) : Tree の形成を行うデータとサブグループ平均の推定を行うデータを分割する
- 一般に推定するパラメータの選択 (Tree 構築, 変数選択) を行うデータとパラメータ推定を行うデータを分ければ、**選択されたモデル**を前提に、信用できる信頼区間計算が可能 (Kuchibhotla, Kolassa, and Kuffner 2022)
  - 選択されたモデルと母集団の関係性は、一般に不明瞭
- Effect Risk の推定としての解釈:  $E[s(X)|X]$  を予測木モデルで推定

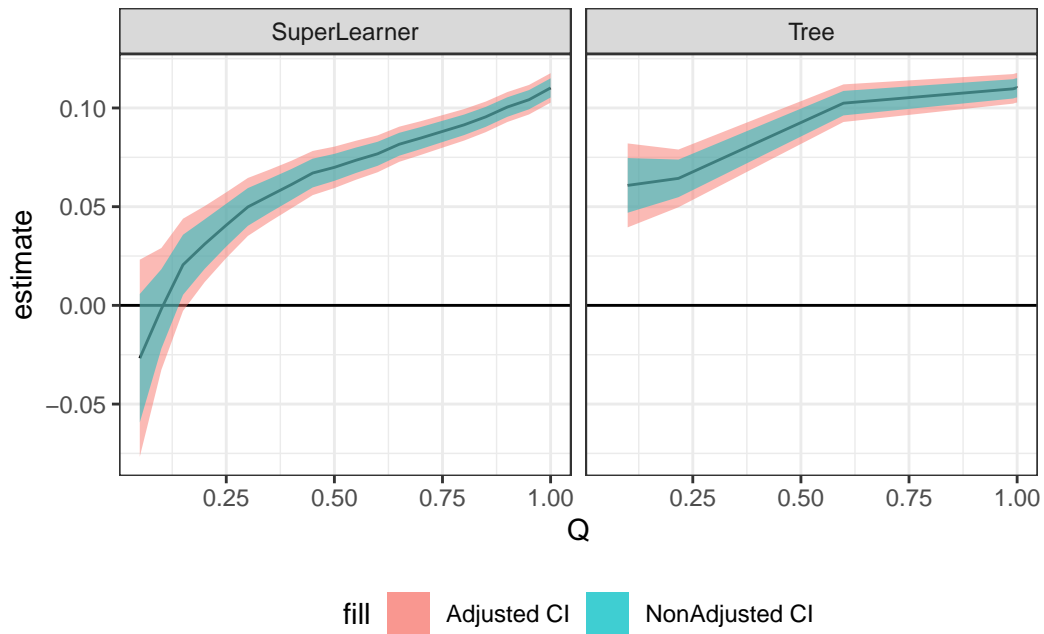
# Example: Tree Model



# Example: Tree Model



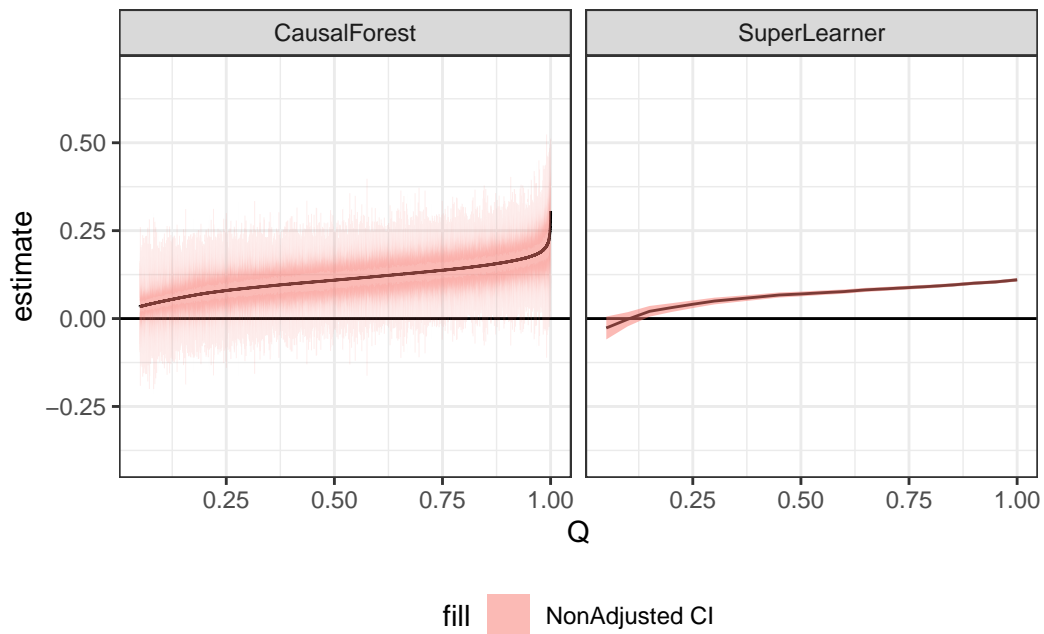
### Example: Stacking VS Tree



### Condiitonal Aaverage Difference

- Causal Forest などは、条件付き平均差  $\tau(X)$  それ自体を推定
  - Effect Risk よりも野心的な試み
  - 一般に信頼区間が広くなる/計算できない

### Example: Estimated CATE VS Risk



### Classification Analysis

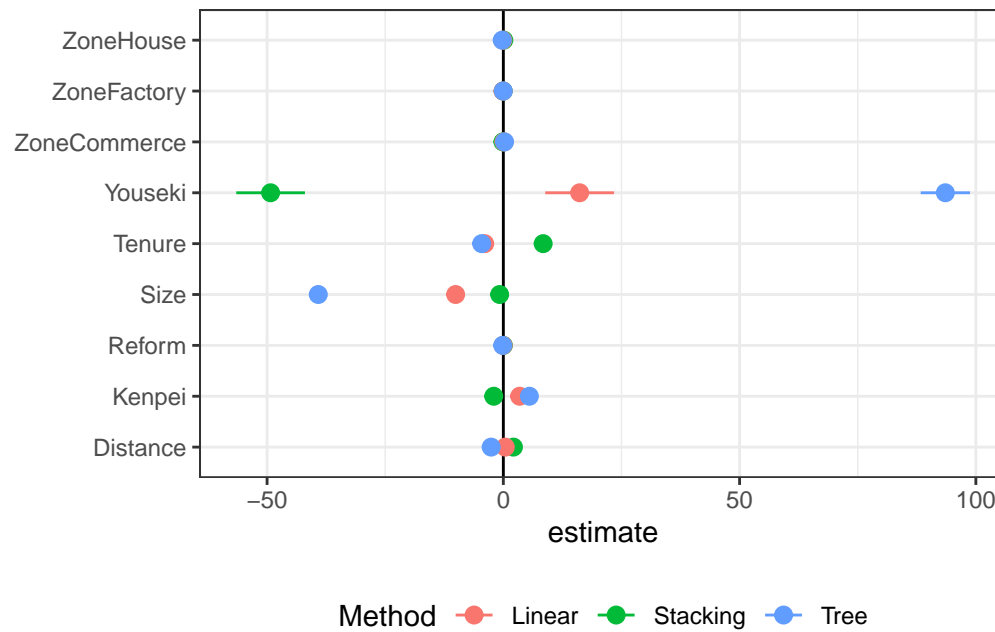
- 顕著な異質性を持つ集団はどのようなものか？
- Chernozhukov et al. (2020)

$$CLAN = E[X_l | s(X) \leq Q(s(X), q)]$$

$$-E[X_l | s(X) > Q(s(X), q)]$$

- $s(X)$  は所与（よくわからんルール）

## Example: CLAN



## 個人効果への含意

- 前提条件
  - 個人因果効果  $\tau_i = Y_i(1) - Y_i(0)$  が定義できる
  - $Y_i(d) \perp D_i | X_i = x$
- $\tau(X)$  は条件付き平均効果 + 個人効果の”最善”の予測値

## 個人効果への含意

- 一般に

$$E[\tau(X) | \tau_i \leq Q(\tau_i, q)] \leq E[\tau(X) | \tau(X) \leq Q(\tau(X), q)]$$

- 個人効果が下位  $q\%$  の事例内での平均効果の上限
- $E[\tau(X) | \tau(X) \leq Q(\tau(X), q)]$  が負であれば、個人効果の平均値も負

## 補論

- Kallus (2022) では個人効果の上限についての bound も議論
  - 下振れではなく上振れについても容易に応用可能
  - 一番シンプルなのは、 $D = 1$  をコントロールとする
- Nonparametric に推定された平均差の活用例
  - 個人効果の予測精度: Yadlowsky et al. (2021)
  - grf に実装済み

## まとめ

- “効果の異質性分析” はふんわりワード
  - もう少しシャープに関心を絞り込む
- Conditional Average Difference を推定し
  - 予測に使いたい? (grf)
  - 記述モデルを推定したい? (BLP)
  - 顕著な異質性を持つグループを探索したい? (Treatment Effect Risk)
- 野心的な目標 (予測) にはそれ相応のコスト (大きな信頼区間、信頼できない区間)

## Reference

- Athey, Susan, and Guido Imbens. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *Proceedings of the National Academy of Sciences* 113 (27): 7353–60. <https://doi.org/10.1073/pnas.1510489113>.
- Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. “Generalized Random Forests.” *The Annals of Statistics* 47 (2): 1148–1178. <https://doi.org/10.1214/18-AOS1709>.
- Athey, Susan, and Stefan Wager. 2019. “Estimating Treatment Effects with Causal Forests: An Application.” *Observational Studies* 5 (2): 37–51.
- Behaghel, Luc, Bruno Crépon, and Marc Gurgand. 2014. “Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment.” *American Economic Journal: Applied Economics* 6 (4): 142–74.
- Chernozhukov, Victor, Mert Demirer, Esther Duflo, et al. 2020. “Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments.” *Econometrica* Forthcoming.

- Kallus, Nathan. 2022. “Treatment Effect Risk: Bounds and Inference.” *arXiv Preprint arXiv:2201.05893*, (MR by Management Science).
- Kuchibhotla, Arun K, John E Kolassa, and Todd A Kuffner. 2022. “Post-Selection Inference.” *Annual Review of Statistics and Its Application* 9: 505–27.
- Nie, Xinkun, and Stefan Wager. 2021. “Quasi-Oracle Estimation of Heterogeneous Treatment Effects.” *Biometrika* 108 (2): 299–319.
- Semenova, Vira, and Victor Chernozhukov. 2021. “Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions.” *The Econometrics Journal* 24 (2): 264289. <https://doi.org/10.1093/ectj/utaa027>.
- Wager, Stefan, and Susan Athey. 2018. “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests.” *Journal of the American Statistical Association* 113 (523): 1228–42. <https://doi.org/10.1080/01621459.2017.1319839>.
- Yadlowsky, Steve, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. 2021. “Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects.” *arXiv Preprint arXiv:2111.07966*.