

# 比較研究: 発展

## 機械学習

川田恵介  
東京大学  
keisukekawata@iss.u-tokyo.ac.jp

2025-11-26

## 1 恒常所得仮説の例

### 1.1 推奨される予測モデルの活用

! Important 1: 予測モデルを活用した比較

1. データの事例について、 $Y/D$  の予測値を算出する
2.  $Y/D$  の予測誤差:  $(Y - Y \text{ の予測値}) / (D - D \text{ の予測値})$  を計算する
3.  $Y$  の予測誤差を、 $D$  の予測誤差で OLS 推定する

- 具体的な例を用いて、直感を確認

### 1.2 消費の決定

- 教科書的なケインズ型消費関数

$$\text{消費} = \beta_0 + \beta_1 \times \text{所得}$$

- ▶ 所得は、消費と貯蓄 (含む借金の返済) に振り分ける
- 乗数効果: 所得の増加は、消費を増やし、他の人の所得を増やす
  - ▶ 給付金や減税等による”景気刺激政策”の論拠の一つ

### 1.3 現代的議論

- 所得変化の理由が論点
  - ▶ 「今、川田が一万円を皆さんに配った場合、今月の消費は増える?、それとも貯蓄に回る?」
- 恒常所得仮説: 生涯所得に大きな影響を与えない「一時的な所得」は、消費をほとんど増やさない

## 1.4 現代的議論

- 恒常所得仮説への反論: 「一時的な所得」であったとしても、消費を大きく増加させる可能性がある
  - ▶ 金融市場の不完全性、人間の思考の癖
- 個人差があるので、傾向把握は実証的な課題

## 1.5 想定

- データから観察できる変数: ある月の[消費額、所得額、個人属性 (職業、過去の賃金履歴、学歴、技能等)]
- 所得の”構造”: 所得の内訳

$$\text{所得} = \underbrace{\text{一時所得} + \text{恒常所得}}_{\text{観察できない}}$$

## 1.6 想定

- 平均消費の”構造”: 所得の内訳

$$E[\text{消費} \mid \text{一時、恒常所得}] = 10 + 0.5 \times \underbrace{\text{恒常所得}}_{\text{観察できない}}$$

- ▶ 恒常所得仮説が厳密に成り立っている
- ▶ 恒常所得が同じ回答者内で比較すると、消費と一時所得の間に関係性はない

## 1.7 悪例

- 「一時所得  $\simeq$  所得」と考えて、消費  $\sim$  所得 を OLS で推定
  - ▶ 「消費と一時所得の間に正の関係性がある」というミスリーディングな結論を得てしまう

## 1.8 自然実験の活用

- 一時所得のみを変動させるイベント (自然実験) があれば、消費  $\sim$  イベント ( $\simeq$  一時所得) を OLS で推定できる
  - ▶ 例: 宝くじ購入内で、当たった人と当たらなかった人の消費パターンを比べる
    - 米国における政府機関の一時閉鎖を利用し、閉鎖により一時的に所得が下がった人とそれ以外を比較 (Baker and Yannelis, 2017)

## 1.9 個人属性を活用したバランス後の比較

- 恒常所得 =  $X$  から予測可能な所得と仮定
- $X$  が全く同じ人の中で、バランス後の比較

- ・  $X$  が同じであれば、所得の個人差 = 一時所得の個人差なので、消費 ~ 所得 の関係性は、一時所得と消費の関係性を捉えている

## 1.10 予測モデルの活用

- ・ 所得  $\simeq$  一時所得 + 予測所得
  - ・ 一時所得  $\simeq$  所得 -  $\underbrace{\text{予測所得}}_{\text{所得の予測誤差}}$
- ・ 消費量 -  $X$  からの予測消費量 も同時に行うことで、 $X$  の影響を念入りに排除

## 2 予測モデルへの要求

### 2.1 予測精度

- ・  $Y/D$  について、理想的な予測モデルに近いモデルを推定する必要がある。
- ・ OLS や LASSO、Random Forest などの特定の推定方法だけでは、不十分な可能性があるので、Stacking 法が推奨される

### 2.2 過剰適合の回避

- ・ 予測モデルを比較的シンプルな OLS や LASSO など推定した場合、全データを用いて予測モデルを推定できる
- ・ Random Forest など推定する場合は、交差推定が必要
- ・ 交差推定を利用した Stacking を用いる場合、アルゴリズムの数が少数であれば、全データを用いて推定できる

## 3 差の異質性

### 3.1 Fukai, Ichimura and Kawata (2021)

- ・ コロナ前(2019 年)と後(2020 年)の 4-6 月の就業率を比較したい
  - ・ COVID とその対応策の影響を間接的に評価する
- ・  $Y$  = 就業状態、 $D$  = 2020/2019、 $X$  = 年齢等の基礎変数 + 前年の就業状態
- ・ 平均的な差はあまりない
  - ・ 就業状態が流動的であると考えられる、一部の層に大きな影響がある

### 3.2 個人因果効果の予測

- ・ 個人ごとに因果効果が予測できれば、
  - ・ 医療行為の個別化: “きき方”/副反応の深刻さ、に応じて医療行為を変える
    - － ある医療の行為の因果効果を、 $X$  から予測する

- ▶ 単位取得計画のサジェスト: 学生の志望進路や興味関心などに応じて、おすすめ単位取得を示す
  - 講義の受講の因果効果を、 $X$ から予測する
- 実務の例: EconML (MicroSoft), CausalML (Uber)

### 3.3 職業訓練の効果

- フランスにおける実験: 職業訓練の提供主体を公的機関/民間企業にランダムに割り付ける
- Behaghel, Crépon and Gurgand (2014): 平均的には公的機関の方が効果的
  - ▶ Kallus (2023): 異質性が大きい
  - ▶ 一部の労働者によって公的機関による訓練の影響が強いだけで、平均的には民間企業の訓練の方が効果的

## 4 異質性の推定

### 4.1 部分線型モデルの一般化

- 部分線型モデルを一般化する

$$E[Y \mid D, X] = \underbrace{\beta_D(X)}_{\substack{X \text{ についての非常に複雑な関数}} \times D + X \text{ についての複雑な関数}$$

- ここまでは  $\beta_D(X)$  は一定の値

### 4.2 推定

- $\beta_D(X)$  を以下の手順で推定
  1.  $Y, D$  の予測モデル  $g_Y(X), g_D(X)$  を交差推定で推定
  2.  $\beta_D(X) \times (D - E[D \mid X])$  から、 $Y - E[Y \mid X]$  を極力予測するように  $\beta_D(X)$  を推定

### 4.3 「差」の線型モデル

- $\beta_D(X) = \beta_0 + \beta_1 X_1 + \dots$  と定式化する
- どのような属性が、大きな”平均差”をもたらすのか、推定できる

### 4.4 補論: 標準化

- 線型モデルにおいて、通常  $\beta_0$  の解釈は難しい
  - ▶ 全ての  $X$  が”0”であった場合の”値”
- $X$  を標準化  $(X - X \text{の平均値})/X \text{の標準偏差}$  すれば、

- ▶ 全ての $X$  が平均値であった場合の”値”となり、より解釈しやすい

## 4.5 例

```
set.seed(111)

library(tidyverse)

Data = read_csv("Public/Example.csv")

Y <- Data$Price

D <- Data$Reform

X <- select(
  Data,
  Size,
  Tenure,
  Distance)
```

## 4.6 例: 予測モデル

```
model <- grf::causal_forest(
  X = scale(X),
  Y = Y,
  W = D
) # Causal Forest
```

## 4.7 例: 「差」の線型モデル

```
model_linear <- grf::best_linear_projection(
  model,
  scale(X))

confint(model_linear)
```

	2.5 %	97.5 %
(Intercept)	2.623854226	5.5148527
Size	0.003509711	5.7935595
Tenure	2.061855409	4.6909830
Distance	-1.826427966	0.9313599

## 4.8 「差」の回帰木モデル

- $\beta_D(X)$  を回帰木モデルとして推定する
  - ▶ Random Forest が応用できる

– Causal Forest

## 4.9 例: 「差」の線型モデル

```
# A tibble: 6 × 4
  Size Tenure Distance pred[,1]
<dbl> <dbl>   <dbl>   <dbl>
1    40      3       3    -1.37
2    65     12       4     1.54
3    65     21       4     7.60
4    65     16       4     6.31
5    40      7       3    -1.08
6    95      5       3    11.1
```

## 4.10 Stacking の活用

- grf package のみでは、random forest を用いた  $Y/D$  の予測モデル推定が行われる
- Stacking に変更することも可能

## 4.11 例: Stacking の活用

```
library(ddml)

X <- select(
  Data,
  Size,
  Tenure,
  Distance)

model_Y <- shortstacking(
  y = Y,
  X = scale(X),
  learners = list(
    list(fun = ddml::ols),
    list(fun = ddml::mdl_ranger)),
  ensemble_type = "nnls1")
```

```
sample fold 1/2
sample fold 2/2
```

```
model_D <- shortstacking(
  y = D,
  X = scale(X),
  learners = list(
    list(fun = ddml::ols),
```

```
list(fun = ddml::mdl_ranger)),
ensemble_type = "nnls1")
```

```
sample fold 1/2
sample fold 2/2
```

## 4.12 例: Stacking の活用

```
model <- grf::causal_forest(
  X = scale(X),
  Y = Y,
  W = D,
  Y.hat = model_Y$soos_fitted,
  W.hat = model_D$soos_fitted) # Causal Forest

confint(grf::best_linear_projection(model, scale(X)))
```

	2.5 %	97.5 %
(Intercept)	2.7160864	5.2428268
Size	0.1077153	5.3172439
Tenure	2.4055694	4.5313925
Distance	-1.5042824	0.8121925

## 4.13 発展

- 異質性の推定については、多くの手法が提案され、現在でも研究が続いている
  - ▶ 参照: CausalML : Chap. 14-15

## 4.14 Reference

### Bibliography

Baker, S.R. and Yannellis, C. (2017) “Income changes and consumption: Evidence from the 2013 federal government shutdown,” *Review of Economic Dynamics*, 23, pp. 99–124.

Behaghel, L., Crépon, B. and Gurgand, M. (2014) “Private and public provision of counseling to job seekers: Evidence from a large controlled experiment,” *American economic journal: applied economics*, 6(4), pp. 142–174.

Fukai, T., Ichimura, H. and Kawata, K. (2021) “Describing the impacts of COVID-19 on the labor market in Japan until June 2020,” *The Japanese Economic Review*, 72(3), pp. 439–470.

Kallus, N. (2023) “Treatment effect risk: Bounds and inference,” *Management Science*, 69(8), pp. 4579–4590.