

# 線型モデルによる異質性分析

## 機械学習

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

2025-12-15

### 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

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

- 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  $\approx$  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( $\approx$  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  $\simeq \frac{\text{所得} - \text{予測所得}}{\text{所得の予測誤差}}$
- 消費量 -  $X$  からの予測消費量 も同時にを行うことで、 $X$  の影響を念入りに排除

## 2 差の異質性

### 2.1 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

### 2.2 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)

### 2.3 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

## 3 異質性の推定

### 3.1 部分線型モデルの一般化

- 部分線型モデルを一般化する

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

+ $X$ についての複雑な関数

- ここまででは  $\beta_D(X)$  は一定の値

### 3.2 推定

- $\beta_D(X)$  を以下の手順で推定

- $Y, D$  の予測モデル  $g_Y(X), g_D(X)$  を交差推定で推定
- $\beta_D(X) \times (D - E[D | X])$  から、 $Y - E[Y | X]$  を極力予測するように  $\beta_D(X)$  を推定

### 3.3 「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

### 3.4 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

### 3.5 例

```
set.seed(111)

library(tidyverse)

data("CPS1985", package = "AER")

Y <- CPS1985$wage

D <- if_else(
  CPS1985$gender == "female",
  1,
  0)
```

```
X <- CPS1985 |>
  select(
    -wage,
    -gender) |>
  fastDummies::dummy_columns(
    remove_selected_columns = TRUE
  )
```

### 3.6 例: 部分線型モデルの推定

```
model <- ddml::ddml_plm(
  y = Y,
  D = D,
  X = X |> data.matrix(),
  learners = list(
    list(fun = ddml::ols),
    list(fun = ddml::mdl_ranger)
  ),
  shortstack = TRUE,
  silent = TRUE
)
```

### 3.7 例: 部分線型モデルの推定

```
model |> summary()
```

```
PLM estimation results:

, , nnls

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0357     0.189   0.189 8.50e-01
D_r         -1.7550     0.440  -3.988 6.65e-05
```

### 3.8 例: 部分線型モデルの推定

```
CPS1985$y_r <- model$ols_fit$model$y_r
CPS1985$D_r <- model$ols_fit$model$D_r

estimatr::lm_robust(y_r ~ D_r, CPS1985)
```

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper
(Intercept)	0.0357106	0.1884311	0.1895155	8.497611e-01	-0.3344496	0.4058708

```

D_r      -1.7550374  0.4389362 -3.9983882 7.279257e-05 -2.6172983
-0.8927765
DF
(Intercept) 532
D_r         532

```

### 3.9 例: 「差」の線型モデルの推定

```

estimatr::lm_robust(
  y_r ~ D_r +
  D_r:(scale(age) + scale(education) + occupation),
  CPS1985)

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.02437532	0.1895742	0.1285793	0.8977397579
D_r	-2.88612695	0.7561122	-3.8170616	0.0001511523
D_r:scale(age)	-0.52709066	0.5341472	-0.9867890	0.3242006180
D_r:scale(education)	0.12323118	0.4454943	0.2766167	0.7821832859
D_r:occupationtechnical	1.64652394	1.3588924	1.2116661	0.2261852761
D_r:occupationservices	1.90980964	0.9792718	1.9502344	0.0516799911
D_r:occupationoffice	2.12648512	1.0377261	2.0491777	0.0409412118
D_r:occupationsales	-0.50304856	1.2082911	-0.4163306	0.6773382379
D_r:occupationmanagement	0.84989994	2.6204724	0.3243308	0.7458167760
	CI Lower	CI Upper	DF	
(Intercept)	-0.34804183	0.3967925	525	
D_r	-4.37150404	-1.4007499	525	
D_r:scale(age)	-1.57641911	0.5222378	525	
D_r:scale(education)	-0.75193922	0.9984016	525	
D_r:occupationtechnical	-1.02301055	4.3160584	525	
D_r:occupationservices	-0.01396287	3.8335821	525	
D_r:occupationoffice	0.08787968	4.1650906	525	
D_r:occupationsales	-2.87672773	1.8706306	525	
D_r:occupationmanagement	-4.29799942	5.9977993	525	

### 3.10 Reference

#### Bibliography

Baker, S.R. and Yannelis, 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.