

# 労働経済学

川田恵介

## Table of contents

1	比較研究	2
1.1	比較研究 . . . . .	2
1.2	単純差 . . . . .	2
1.3	差の理由 . . . . .	2
1.4	Example: CPS1985 . . . . .	2
1.5	Example Paper . . . . .	3
1.6	例: Dube et al. (2020) . . . . .	3
1.7	伝統的な方法 . . . . .	3
2	線形モデルによる比較	4
2.1	線形モデルによる記述 . . . . .	4
2.2	数値例: OLS . . . . .	4
2.3	例: OLS (独立の場合) . . . . .	5
2.4	例: OLS (関連している場合) . . . . .	5
2.5	まとめ . . . . .	6
3	OLS の別解釈	6
3.1	OLS Algorithm: 単回帰 . . . . .	6
3.2	OLS Algorithm: General case . . . . .	6
3.3	OLS Algorithm: General case . . . . .	6
3.4	OLS の解釈 . . . . .	7
4	Constant Difference モデルによる解釈	7
4.1	Constant Difference . . . . .	7
4.2	単回帰の解釈 . . . . .	7
4.3	単回帰の解釈 . . . . .	8
4.4	母集団への含意 (事例数無限大) . . . . .	8
4.5	重回帰の解釈 . . . . .	8
4.6	重回帰の解釈 . . . . .	8

4.7	母集団への含意 . . . . .	9
4.8	Mis-specification . . . . .	9
4.9	Overfit . . . . .	9
4.10	まとめ . . . . .	9
	Reference . . . . .	9

## 1 比較研究

- 研究の主たる関心とする変数を絞り込むことで、より頑強かつ透明性の高い分析が可能になる
  - 因果推論なども含み、労働経済学における主要な分析手法

### 1.1 比較研究

- 特定の  $Y$  と  $D$  の関係性について関心があるケースが多い
  - 例: 男女間賃金格差:  $Y = Wage, D = Gender$

### 1.2 単純差

- グループ  $D$  間における単純平均差を推定したいのであれば、母平均の差  $E[Y|D = 1] - E[Y|D = 0]$  ないし、単回帰  $Y \sim D$  を推定すれば十分な応用がほとんど
  - 研究として行う場合は、差が生じるの”理由”にも言及することが多い

### 1.3 差の理由

- なぜ差が生まれるのか?
  - 本スライドでは、データから観察可能な他の変数  $X$  に注目
    - \*  $X$  についての格差が、 $Y$  の差をもたらしている
    - \*  $X$  以外についての格差が、 $Y$  の差をもたらしている
- 例: 男女間格差: 男女間での職業分布の違いがどの程度賃金格差をもたらしているのか
  - $X = \text{職業}$

### 1.4 Example: CPS1985

Characteristic	manufacturing, N = 99	construction, N = 24	other, N = 411
wage	9.0 (5.7, 12.0)	9.8 (6.8, 11.5)	7.5 (5.2, 11.2)
gender			
male	60 (61%)	22 (92%)	207 (50%)
female	39 (39%)	2 (8.3%)	204 (50%)

## 1.5 Example Paper

- Part-Time Pay Penalty: Manning and Petrongolo (2008)
  - $D$  = Full/Part-Time,  $Y$  = Income
- Urban-Rural Income Gap: Sicular et al. (2007)
  - $D$  = Rural/Urban,  $Y$  = Income
- $X$  : 年齢、性別

## 1.6 例: Dube et al. (2020)

- Online 労働市場において、求人が提示する賃金水準 ( $= D$ ) と応募者数 ( $= Y$ ) はどのような関係にあるのか?
  - 賃金が高い仕事は、高い技能が要求される/きつい... ( $= X$ ) 可能性がある
  - 求人内容 ( $= X$ ) 以外の要因で、どの程度の差が生じているのか?
    - \* 労働市場の不完全性 (独占力) の指標 (Langella and Manning 2021)

## 1.7 伝統的な方法

- $X$  をコントロールする: 以下を推定

$$g_Y(D, X) = \underbrace{\beta_D}_{=X以外による格差} D + \underbrace{\beta_0 + \beta_1 X_1 + \dots}_{Xの影響を除去}$$

- 課題
  - コントロールとは?
  - 本当に  $X$  以外の理由による格差?
- より精緻かつ明確な議論が必要

## 2 線形モデルによる比較

- 伝統的に用いられてきた方法
  - モデルの定式化に強く依存することに注意

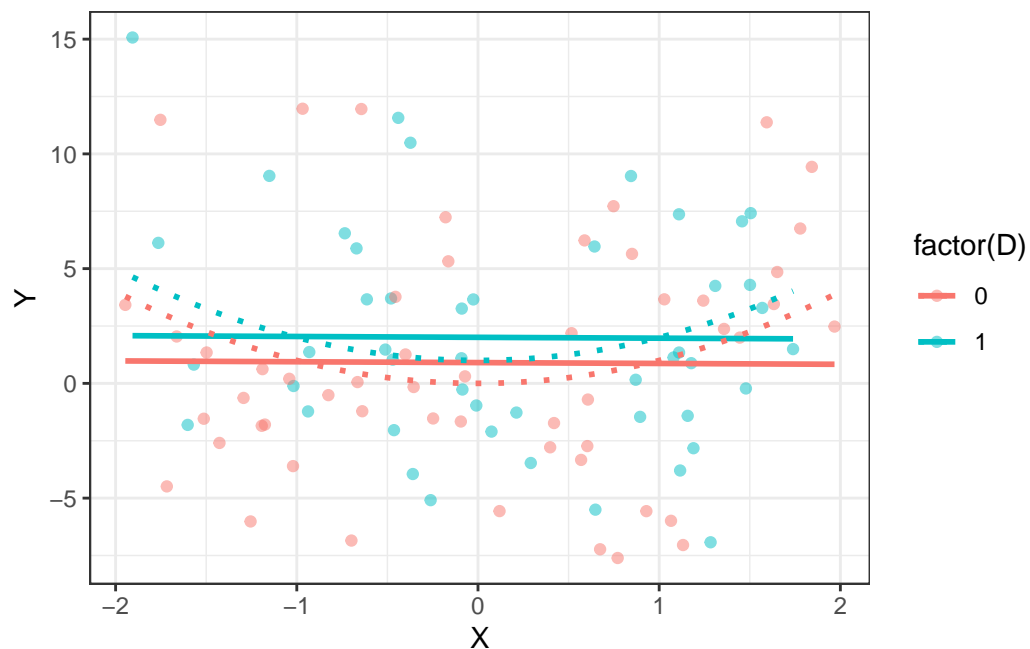
### 2.1 線形モデルによる記述

- $$g_Y(X) = \underbrace{\beta_D}_{Interest} \times D + \underbrace{\beta_0 + \beta_1 X_1 + \dots + \beta_L X_L}_{Nuisance\ term}$$
- $\beta_D$  の解釈:  $X$  を”一定にした上での”  $D$  間の格差
- Estimand = Best Linear Projection であることに注意
  - 致命的にミスリードな推定結果となりうる

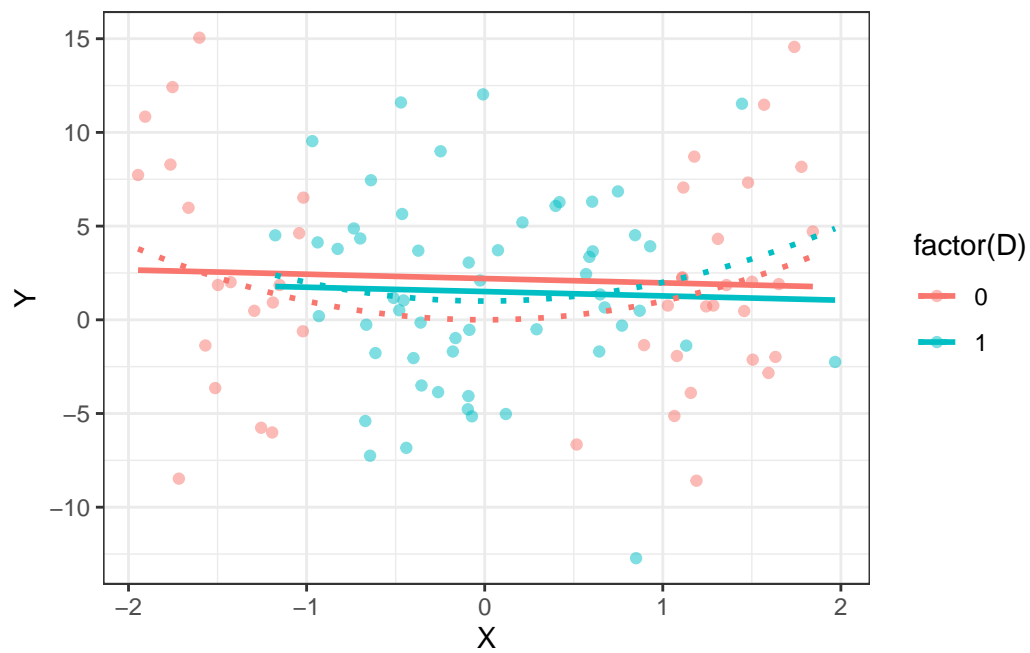
### 2.2 数値例: OLS

- $Y = \underbrace{D}_{\in \{0,1\}} + \underbrace{X^2}_{\sim Uniform(-2,2)} + \underbrace{u}_{\sim N(0,10)}$ 
  - 独立している場合:
    - \*  $\Pr(D = 1) = 0.5$
  - 相関している場合:
    - \*  $\Pr(D = 1 | 1 \geq X \geq -1) = 0.9$
    - \*  $\Pr(D = 1 | X \geq 1 | X \leq -1) = 0.1$

### 2.3 例: OLS (独立の場合)



### 2.4 例: OLS (相関しているの場合)



## 2.5 まとめ

- Linear Model を用いた比較研究は、現在でもよく行われるが、定式化に強く依存することが懸念されてきた。
- 様々な解決案が提案され、多様な方法論開発が進んでいる
  - 機械学習の導入も有力視されている
- とりあえず OLS が何をやっているのか、しっかり理解することが重要

## 3 OLS の別解釈

- $g_Y(D) = \beta_0 + \beta_D D + \beta_1 X_1 + \dots$  を OLS で推定する
- (BLP の推定ではなく) Weight 推定としても再解釈できる

### 3.1 OLS Algorithm: 単回帰

- $g_Y(D) = \beta_D D + \beta_0$  を OLS 推定すると

$$\beta_D = \sum_{i:D_i=1} \underbrace{\frac{1}{N_1}}_{=W_i} Y_i - \sum_{i:D_i=0} \underbrace{\frac{1}{N_0}}_{=W_i} Y_i$$

- 事例数の逆数を Weight とした比較と解釈できる

– 注:  $\sum_{i:D_i=1} W_i = \sum_{i:D_i=0} W_i = 1$

### 3.2 OLS Algorithm: General case

1. 全ての  $X = [X_1, \dots, X_L]$  について、

$$\begin{aligned} \sum_{i:D_i=1} W_i X_{i,l} &= \sum_{i:D_i=0} W_i X_{i,l}, \\ \sum_{i:D_i=1} W_i &= \sum_{i:D_i=0} W_i = 1 \end{aligned}$$

を満たす  $W$  から、分散が最も小さいものを選ぶ

### 3.3 OLS Algorithm: General case

- 2.

$$\beta_D = \sum_{i:D_i=1} W_i Y_i - \sum_{i:D_i=0} W_i Y_i$$

### 3.4 OLS の解釈

- $Y$  の Weight 付き平均差として解釈できる
  - データ上で、 $D$  間で  $X$  の平均値が”Balance”するように Weight は選ぶ
- $X^2$  もモデルに加えれば、 $X^2$  の平均値 (分散) も等しくなるように選ばれる
- $X_1 * X_2$  もモデルに加えれば、 $X_1, X_2$  の共分散も等しくなるように選ばれる

## 4 Constant Difference モデルによる解釈

- 母集団に対する、かなり強い仮定を用いて、OLS の推定結果を解釈
- 注記: 不必要に強い仮定であり、将来緩める

### 4.1 Constant Difference

- $E[Y|1, X] - E[Y|0, X] = \tau$  を母集団上で仮定
  - $\tau = X$  が全く同じ集団間での平均格差
  - \* ”  $X$  をコントロール/Ceteris paribus”

- 以下のモデルで表現できる

$$Y = \tau \times D + \underbrace{h(X)}_{\text{なんらかの関数}} + \underbrace{u}_{=E[u|X]}$$

- Semiparametric estimation では、 $h(X)$  は Nuisance function と呼ばれる

### 4.2 単回帰の解釈

- $Y$  を代入すると

$$\begin{aligned}\beta_D &= \frac{\sum_{i:D_i=1} Y_i}{N_1} - \frac{\sum_{i:D_i=0} Y_i}{N_0} \\ &= \frac{\sum_{i:D_i=1} (\tau_D + h(X_i) + u_i)}{N_1} \\ &\quad - \frac{\sum_{i:D_i=0} (h(X_i) + u_i)}{N_0}\end{aligned}$$

### 4.3 単回帰の解釈

$$\beta_D = \tau_D + \underbrace{\left[ \frac{\sum_{i:D_i=1} h(X_i)}{N_1} - \frac{\sum_{i:D_i=0} h(X_i)}{N_0} \right]}_{\text{属性のずれ}} + \underbrace{\frac{\sum_{i:D_i=1} u_i}{N_1} - \frac{\sum_{i:D_i=0} u_i}{N_0}}_{\text{観察できない属性のずれ}}$$

### 4.4 母集団への含意 (事例数無限大)

$$\beta_D = \tau_D + \beta_X \underbrace{\left[ \frac{\sum_{i:D_i=1} h(X_i)}{N_1} - \frac{\sum_{i:D_i=0} h(X_i)}{N_0} \right]}_{\substack{\xrightarrow[N_1, N_0 \rightarrow \infty]{} E_X[h(X)|D=1] - E_X[h(X)|D=0]}} + \underbrace{\frac{\sum_{i:D_i=1} u_i}{N_1} - \frac{\sum_{i:D_i=0} u_i}{N_0}}_{\rightarrow 0}$$

- 観察できる属性のずれの影響が残る

### 4.5 重回帰の解釈

- Y を代入すると

$$\begin{aligned} \beta_D &= \sum_{i:D_i=1} W_i Y_i - \sum_{i:D_i=0} W_i Y_i \\ &= \sum_{i:D_i=1} W_i (\tau_D + h(X_i) + u_i) \\ &\quad - \sum_{i:D_i=0} W_i (h(X_i) + u_i) \end{aligned}$$

### 4.6 重回帰の解釈

- Y を代入すると

$$\begin{aligned} \beta_D &= \tau_D \\ &+ \underbrace{\left[ \sum_{i:D_i=1} W_i h(X_i) - \sum_{i:D_i=0} W_i h(X_i) \right]}_{h(X)=\beta_0+\beta_1 X \text{ であれ } u_i \neq 0} \\ &\quad + \sum_{i:D_i=1} W_i u_i - \sum_{i:D_i=0} W_i u_i \end{aligned}$$



#### 4.7 母集団への含意

$$\beta_D = \tau_D + \beta_X \left[ \underbrace{\sum_{i:D_i=1} W_i h(X) - \sum_{i:D_i=0} W_i h(X)}_{h(X)=\beta_0+\beta_1 X \text{であれば}=0} \right] + \underbrace{\sum_{i:D_i=1} W_i u_i - \sum_{i:D_i=0} W_i u_i}_{\rightarrow 0}$$

#### 4.8 Mis-specification

- $g_Y(D, X) = \beta_0 + \beta_D D + \beta_1 X$  を OLS 推定するが、 $h(X) = \beta_0 + \beta_1 X + \beta_2 X^2$   
 –  $X$  の分散 ( $X^2$ ) は Balance しないので、 $\beta_D$  は  $\tau_D$  に (事例数が無限大でも) 収束しない

#### 4.9 Overfit

- Mis-specification を避けるためには、 $X$  を十分に複雑にしてモデルに導入する必要がある
- より多くの変数の平均値を揃える必要があるので、Weight  $W_i$  の分散が大きくなる
- 特定の個人 ( $u_i$ ) の影響が非常に強くなり、推定精度が悪化

#### 4.10 まとめ

- OLS =  $X$  の平均値を Balance させる Algorithm  
 – 高次項 ( $X_1^2, X_1^3, X_1 \times X_2, \dots$ ) を導入すると、 $X$  の分布を Balance させられる  
 – 弊害: Weight の分散が大きくなり、推定精度が悪化する
- 課題: “重要な” $X$  のみ Balance させたい

#### Reference

- Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri. 2020. “Monopsony in Online Labor Markets.” *American Economic Review: Insights* 2 (1): 33–46.
- Langella, Monica, and Alan Manning. 2021. “Marshall Lecture 2020: The Measure of Monopsony.” *Journal of the European Economic Association* 19 (6): 2929–57.
- Manning, Alan, and Barbara Petrongolo. 2008. “The Part-Time Pay Penalty for Women in Britain.” *The Economic Journal* 118 (526): F28–51.

Sicular, Terry, Yue Ximing, Björn Gustafsson, and Li Shi. 2007. "The Urban–Rural Income Gap and Inequality in China." *Review of Income and Wealth* 53 (1): 93–126.