



豊富な実証例から
計量経済学の
「生きた」知識を身につける！

東洋経済新報社

第9章 不均一分散

藪友良

『入門 実践する計量経済学』

(東洋経済新報社)

PPT

- 不均一分散
- 不偏性と一致性
- ロバスト標準誤差
- 加重最小2乗法
- どの推定法を用いるべきか

不均一分散

- 均一分散

$$V(u_i) = E[u_i^2] = \sigma^2$$

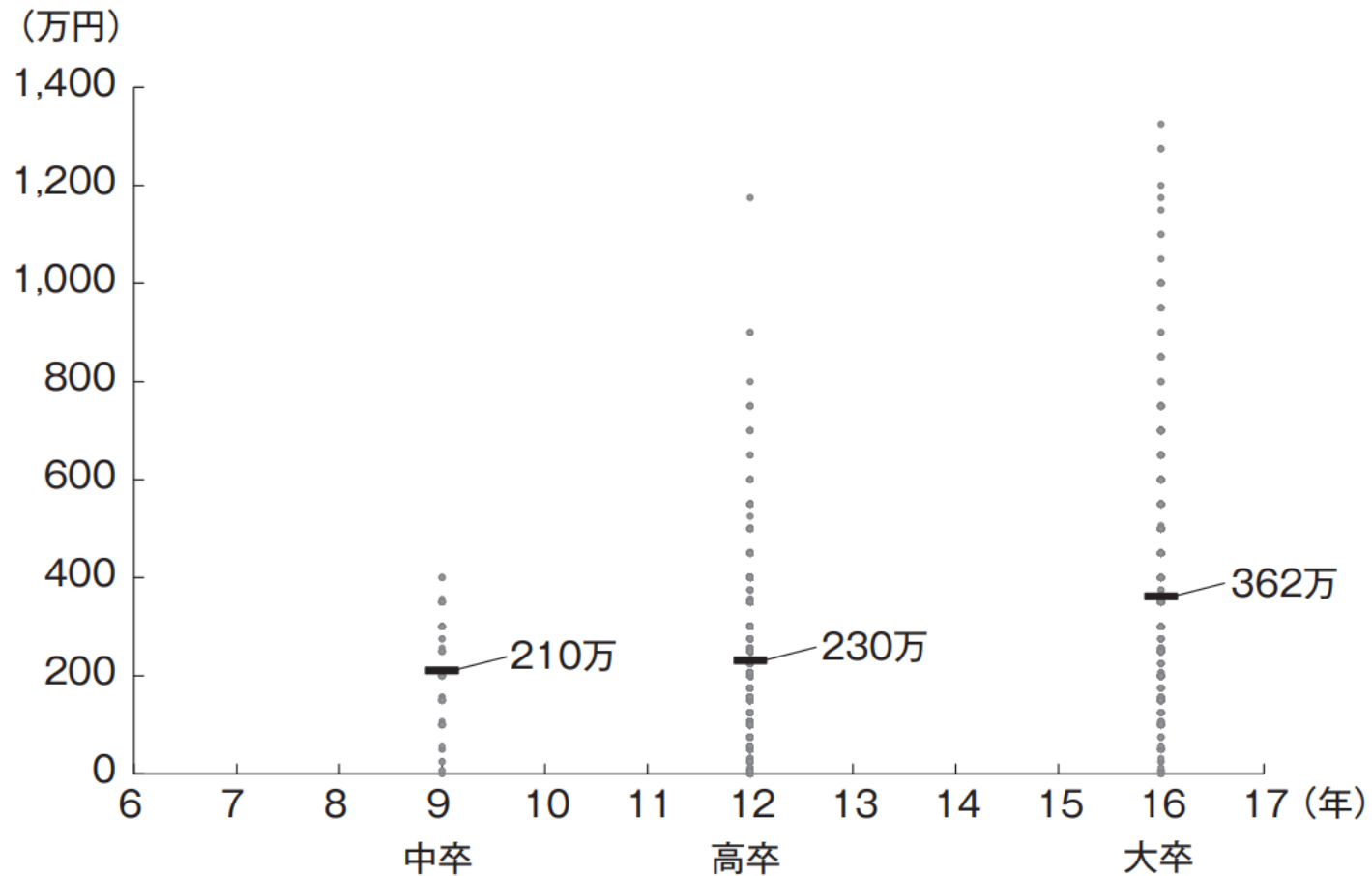
--- 均一分散は非現実的仮定

- 不均一分散

$$V(u_i) = E[u_i^2] = \sigma_i^2$$

--- 不均一分散は現実的仮定

図 8 - 6 教育年数と所得との関係



(注) 田中隆一(2015)『計量経済学の第一歩』有斐閣, の疑似マイクロデータを用いました。

例) グループ平均から生じる不均一分散

県別家計データ(i 県 j 世帯消費額 Y_{ij} 、所得額 X_{ij})

$$Y_{ij} = \alpha + \beta X_{ij} + u_{ij}$$

誤差項 u_{ij} は、標準的仮定をすべて満たす

$$V(u_{ij}) = E[u_{ij}^2] = \sigma^2$$

- i 県の世帯 j に関して和をとって、

$$\sum_{j=1}^{N_i} Y_{ij} = N_i \alpha + \beta \sum_{j=1}^{N_i} X_{ij} + \sum_{j=1}^{N_i} u_{ij}$$

集計世帯数 N_i で割る

$$\underbrace{\frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}}_{Y_i} = \alpha + \beta \underbrace{\frac{\sum_{j=1}^{N_i} X_{ij}}{N_i}}_{X_i} + \underbrace{\frac{\sum_{j=1}^{N_i} u_{ij}}{N_i}}_{u_i}$$

- 平均消費 Y_i と平均所得 X_i に線形関係がある

$$Y_i = \alpha + \beta X_i + u_i$$

ただし、 $E[u_i^2] = \frac{\sigma^2}{N_i}$ となる

- 誤差項には不均一分散がある

$$E[u_i^2] = \frac{\sigma^2}{N_i}$$

[証明]

$$\begin{aligned}
 E[u_i^2] &= E \left[\left(\frac{\sum_{j=1}^{N_i} u_{ij}}{N_i} \right)^2 \right] \\
 &= \frac{\sum_{j=1}^{N_i} E[u_{ij}^2]}{N_i^2} \\
 &= \frac{N_i \sigma^2}{N_i^2} = \frac{\sigma^2}{N_i}
 \end{aligned}$$

$$\begin{aligned}
 &N_i = 2 \text{ のケースでは、} \\
 &E[(u_{i1} + u_{i2})^2] \\
 &= E[u_{i1}^2] + E[u_{i2}^2] + 2E[u_{i1}u_{i2}] \\
 &= 2\sigma^2
 \end{aligned}$$

[終]

- 集計世帯数 N_i が多ければ分散は小さくなり、
集計世帯数 N_i が少なければ分散は大きくなる

例) 線形確率モデルから生じる不均一分散
被説明変数 Y_i はダミー変数とする

$$Y_i = \alpha + \beta X_i + u_i$$

- $Y_i = 1$ となる確率を P_i とすると、

$$P_i = \alpha + \beta X_i$$

[証明]

$$E[Y_i] = \alpha + \beta X_i$$

$$E[Y_i] = 1 \times P_i + 0 \times (1 - P_i) = P_i \quad [\text{終}]$$

- β は X が1単位増加したとき、 $Y = 1$ となる確率がどれぐらい変化するか

・誤差項は不均一分散

[証明]

$Y_i = \alpha + \beta X_i + u_i$ から、

$$\begin{aligned} u_i &= Y_i - (\alpha + \beta X_i) \\ &= Y_i - P_i \end{aligned}$$

$$P_i = \alpha + \beta X_i$$

したがって、

$$u_i = 1 - P_i \quad (Y_i = 1 \text{ なら})$$

$$u_i = -P_i \quad (Y_i = 0 \text{ なら})$$

また、 $V(u_i) = E[(u_i - E[u_i])^2] = E[u_i^2]$ から、

$$\begin{aligned} E[u_i^2] &= (1 - P_i)^2 P_i + (-P_i)^2 (1 - P_i) \\ &= (1 - P_i) [(1 - P_i) P_i + P_i^2] \\ &= (1 - P_i) P_i \end{aligned}$$

[終]

不偏性と一致性

- 単回帰モデル $Y_i = \alpha + \beta X_i + u_i$ のOLS推定量は、

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

- 仮定4に関係なく、OLS推定量は不偏性を満たす

$$E[\hat{\beta}] = \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) E[u_i]}{\sum_{i=1}^n (X_i - \bar{X})^2} = \beta$$

- 不均一分散が正しいもとで、

$$\sigma_{\hat{\beta}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2}$$

均一分散が正しいもとで

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

--- サンプルサイズが増えると、仮定2から $\sigma_{\hat{\beta}}^2$ は0に収束する

--- OLS推定量は**一致性**も満たす

- OLS推定量はBLUEではないが、不偏性と一致性を備えた良い推定量である

$$\sigma_{\hat{\beta}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)^2} \text{ の証明}$$

[証明] 確率的表現 $\hat{\beta} = \beta + \frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$ を用いると、 $\sigma_{\hat{\beta}}^2$ は以下となる。

$$\begin{aligned} E[(\hat{\beta} - \beta)^2] &= E\left[\left(\frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^2\right] \\ &= \frac{E\left[\left(\sum_{i=1}^n (X_i - \bar{X}) u_i\right)^2\right]}{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)^2} \end{aligned}$$

* $n = 2$ として考える

$$\begin{aligned} &E\left[\left((X_1 - \bar{X})u_1 + (X_2 - \bar{X})u_2\right)^2\right] \\ &= E\left[(X_1 - \bar{X})^2 u_1^2 + (X_2 - \bar{X})^2 u_2^2 + 2(X_1 - \bar{X})(X_2 - \bar{X})u_1 u_2\right] \\ &= (X_1 - \bar{X})^2 E[u_1^2] + (X_2 - \bar{X})^2 E[u_2^2] + 2(X_1 - \bar{X})(X_2 - \bar{X})E[u_1 u_2] \\ &= (X_1 - \bar{X})^2 \sigma_1^2 + (X_2 - \bar{X})^2 \sigma_2^2 \end{aligned}$$

ロバスト標準誤差

$$\sigma_{\hat{\beta}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2} \text{の推定方法}$$

- ・ $\sigma_i^2 = E[u_i^2]$ から、 σ_i^2 の推定量として u_i^2 が適当である

$$Y_i = \alpha + \beta X_i + u_i$$

- ・ 誤差項 $u_i = Y_i - \alpha - \beta X_i$ を残差 $\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$ で代用する

$$s_{\hat{\beta}}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2}$$

---この平方根は、不均一分散に頑健な標準誤差、
ロバスト標準誤差もしくはホワイト標準誤差と呼ばれる

--- ロバスト標準誤差を使えば、通常の信頼区間、仮説検定を行える

例) 家賃と敷地面積の関係

- ・ H駅周辺の賃料Yと専有面積Xの関係
- ・ 724物件
- ・ OLS推定すると

$$\hat{Y} = 2.69 + 0.160X$$

通常の標準誤差	(0.101)	(0.0034)
ロバスト標準誤差	(0.117)	(0.0049)

--- ロバスト標準誤差は、通常の標準誤差と大きくは異なる

--- ロバスト標準誤差が、大きくなることも、小さくなることもある

--- Stataでは、rを最後につけるとロバスト標準誤差となる

```
reg Y X, r
```

例) アフリカ系アメリカ人への差別

- ・ 求人広告に架空の履歴書を送る
- ・ 履歴書には黒人風の名前(ラキーシャ)、白人風の名前(エミリー)をランダムに割り振る
- ・ 被説明変数 Y は企業から連絡があれば1となるダミー変数、説明変数 X は黒人風の名前なら1となるダミー変数

$$\hat{Y} = 0.0965 - 0.032X$$

	(0.0055)	(0.0078)
通常標準誤差		
ロバスト標準誤差	(0.0060)	(0.0078)

- 説明変数はランダムなので、バイアスは生じない
- 白人なら、9.65%の確率で連絡がある
- 黒人なら確率は3.2%ポイントも下がる(6.33%=0.0965-0.032)

- ・ 説明変数に実務年数Wを加える

$$\hat{Y} = 0.071 - 0.032X + 0.0033W$$

通常標準誤差 (0.0082) (0.0078) (0.00077)

ロバスト標準誤差 (0.0085) (0.0078) (0.00085)

- 白人なら、7.1%の確率で連絡がある
- 黒人なら確率は3.2%も下がる
- 勤続年数が1年増えると0.33%確率は上がる
- 勤続年数が10年増えると3.33%確率は上がる、つまり、黒人であると、実務年数が10年分少ないことと同じデメリットがある

加重最小2乘法

$$Y_i = \alpha + \beta X_i + u_i$$

$$V(u_i) = E[u_i^2] = \sigma_i^2 = \sigma^2 h_i$$

- 分析者は h_i の値を知っている
- $\sqrt{h_i}$ で両辺を割ると(加重 $\frac{1}{\sqrt{h_i}}$ をつける)

$$\frac{Y_i}{\sqrt{h_i}} = \alpha \frac{1}{\sqrt{h_i}} + \beta \frac{X_i}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

このとき、 $\frac{u_i}{\sqrt{h_i}}$ は期待値0、分散 σ^2 となる

$$E \left[\left(\frac{u_i}{\sqrt{h_i}} \right)^2 \right] = \frac{E[u_i^2]}{h_i} = \frac{\sigma^2 h_i}{h_i} = \sigma^2$$

- 被説明変数 $\frac{Y_i}{\sqrt{h_i}}$ 、説明変数 $\frac{1}{\sqrt{h_i}}$ 、 $\frac{X_i}{\sqrt{h_i}}$ としてOLS推定する
 - 誤差項は標準的仮定を満たすため、OLS推定量はBLUE
 - **加重最小2乗(WLS)推定量**とも呼ばれる
 - h_i が小さいと情報量が大きいため、高い加重 $1/\sqrt{h_i}$ をつける 19

例) 県別家計データ(i 県 j 世帯消費額 Y_{ij} 、所得額 X_{ij})

$$Y_{ij} = \alpha + \beta X_{ij} + u_{ij}$$
$$V(u_{ij}) = E[u_{ij}^2] = \sigma^2$$

- 平均消費 Y_i と平均所得 X_i に線形関係がある

$$Y_i = \alpha + \beta X_i + u_i$$

ただし、 $E[u_i^2] = \sigma^2 / N_i$ となる

- $h_i = 1/N_i$ から、上式の両辺を $\sqrt{h_i} = 1/\sqrt{N_i}$ で割ると

$$\sqrt{N_i} Y_i = \alpha \sqrt{N_i} + \beta \sqrt{N_i} X_i + \sqrt{N_i} u_i$$

- 新しい誤差項の分散は一定となる

$$E[(\sqrt{N_i} u_i)^2] = N_i E[u_i^2] = N_i \frac{\sigma^2}{N_i} = \sigma^2$$

- 被説明変数を $\sqrt{N_i} Y_i$ とし、説明変数を $\sqrt{N_i}$ 、 $\sqrt{N_i} X_i$ として

OLS推定すればWLS推定量が得られる

---集計世帯数 N_i が多い県は情報量が多いため、高い加重をつける

実行可能な加重最小2乗法(FWLS)

$$Y_i = \alpha + \beta X_i + u_i$$
$$V(u_i) = E[u_i^2] = \sigma_i^2 = \sigma^2 h_i$$

- 分析者は h_i の値を知らない
- データから推定された $\hat{h}_i$ を用いて

$$\frac{Y_i}{\sqrt{\hat{h}_i}} = \alpha \frac{1}{\sqrt{\hat{h}_i}} + \beta \frac{X_i}{\sqrt{\hat{h}_i}} + \frac{u_i}{\sqrt{\hat{h}_i}}$$

- h_i の推定には、分散の構造を知る必要がある

例) 線形確率モデルから生じる不均一分散
被説明変数 Y_i はダミー変数とする

$$Y_i = \alpha + \beta X_i + u_i$$

- $Y_i = 1$ となる確率を P_i とすると、

$$P_i = \alpha + \beta X_i$$

- 誤差項は不均一分散

$$V(u_i) = (1 - P_i)P_i$$

- モデル $Y_i = \alpha + \beta X_i + u_i$ をOLS推定し、予測値として確率を推定

$$\hat{P}_i = \hat{\alpha} + \hat{\beta} X_i$$

$$\hat{h}_i = (1 - \hat{P}_i)\hat{P}_i$$

- 黒人への差別

OLS推定	$\hat{Y} = 0.071 - 0.032X + 0.0033W$
	(0.0085) (0.0078) (0.00085)

FWLS推定	$\hat{Y} = 0.070 - 0.031X + 0.0033W$
	(0.0085) (0.0077) (0.00085)

どの推定法を用いるべきか

均一分散

OLS推定

不均一分散

h_i の値は分からないし、
推定もできない

OLS推定(BLUEではない、
ロバスト標準誤差)

h_i の値を知っている

WLS推定(BLUE)

h_i の値を知らないが、
 h_i を推定できる

FWLS推定(OLSより
効率的)