

## 第四章 多変数の回帰<sup>1</sup>

劉慶豊<sup>2</sup>

小樽商科大学

December 16, 2009

---

<sup>1</sup>第二章からの資料は森棟公夫先生著「基礎コース 計量経済学」をもとに作成したものである。

<sup>2</sup>E-mail:qliu@res.otaru-uc.ac.jp, URL:<http://www.otaru-uc.ac.jp/~qliu/> ◀ ▶ ≡ 🔍 ↺

# K変数回帰（重回帰）のモデル

## モデル式

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + u_i, i = 1, 2, \cdots, n-1, n \quad (1)$$

誤差項に関する仮定  $u_i$  が  $i$  に関して独立で、かつ

$$E(u_i) = 0, V(u_i) = \sigma^2, i = 1, 2, \cdots, n$$

である。

## データ

$$\{(y_1, x_{11}, \cdots, x_{K1}), (y_2, x_{12}, \cdots, x_{K2}), \cdots, (y_n, x_{1n}, \cdots, x_{Kn})\} \quad (2)$$

$$x_{1i} = 1, i = 1, 2, \cdots, n$$

# 回帰係数の推定

OLS推定量 OLS(最小二乗) 推定量は

$$\Phi = \sum_{i=1}^n \{y_i - (\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_K x_{Ki})\}^2 \quad (3)$$

を最小化するように決まる。

正規方程式

$$\sum_{i=1}^n \{y_i - (\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_K x_{Ki})\} x_{mi} = 0, m = 1, 2, \cdots, K \quad (4)$$

もう一つの求め方  $\hat{\beta}_K$  を求めるが、説明変数  $x$  の順番が特に意味がないため、他の係数の推定量、 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{K-1}$  も同じ方法で計算できる。方法は第3章で説明した説明変数が三つの場合の発展である。

1.  $x_K$  を他の説明変数に回帰し、残差  $\hat{u}_{Ki}$  を求める

$$x_{Ki} = c_1 x_{1i} + c_2 x_{2i} + \dots + c_{K-1} x_{K-1i} + u_{Ki} \quad (5)$$

$$\hat{u}_{Ki} = x_{Ki} - \hat{c}_1 x_{1i} - \hat{c}_2 x_{2i} - \dots - \hat{c}_{K-1} x_{K-1i}$$

2. 偏標本分散、偏標本共分散を求める

$$\begin{aligned} s_{x_K x_K | x_1 \dots x_{K-1}} &= \frac{1}{n - (K - 1)} \sum_{i=1}^n (\hat{u}_{Ki})^2 \\ &= \frac{1}{n - K + 1} \sum_{i=1}^n \hat{u}_{Ki} \cdot x_{Ki}, \end{aligned} \quad (6)$$

$$s_{y x_K | x_1 \dots x_{K-1}} = \frac{1}{n - (K - 1)} \sum_{i=1}^n \hat{u}_{Ki} \cdot y_i \quad (7)$$

### 3. 推定量は

$$\hat{\beta}_K = \frac{s_{yx_K|I_{K-1}}}{s_{x_K x_K|I_{K-1}}}$$
$$\hat{\beta}_K = r_{yx_K|I_{K-1}} \sqrt{\frac{s_{yy|I_{K-1}}}{s_{x_K x_K|I_{K-1}}}} \quad (8)$$

となる。ただし、 $\cdot|I_{K-1}$  は  $\cdot|x_1 \cdots x_{K-1}$  と同じことを意味する。

### 4. 正規方程式から

$$\sum_{i=1}^n y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum_{i=1}^n x_{2i} - \cdots - \hat{\beta}_K \sum_{i=1}^n x_{Ki} = 0$$
$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}_2 - \cdots - \hat{\beta}_K \bar{x}_K \quad (9)$$

### 5. 標本平均からなる座標点 $(\bar{y}, \bar{x}_2, \cdots, \bar{x}_K)$ は推定された回帰直線上に位置する。

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x}_2 + \cdots + \hat{\beta}_K \bar{x}_K \quad (10)$$

# 残差 2 乗和と総変動の分解

第3章で説明した3変数の回帰分析の拡張であると言える。

回帰値  $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_K x_{Ki}$

残差  $\hat{u}_i = y_i - \hat{y}_i$

直行性  $\sum_{i=1}^n \hat{u}_i x_{mi} = 0, m = 1, 2, \dots, K$

誤差分散  $n - K$ で割っていることを注意しよう（自由度が  $n - K$ ）。

$$s^2 = \frac{\sum_{i=1, n} \hat{u}_i^2}{n - K} \quad (11)$$

$$s^2 = s_{yy|I_K} \quad (12)$$

# 残差 2 乗和と総変動の分解

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{RSS}{TSS} \quad (13)$$

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (14)$$

$$TSS = ESS + RSS \quad (15)$$

$$R^2 = \frac{ESS}{TSS} \quad (16)$$

# 最小2乗法の性質

単回帰 ( $K = 2$ ) や3変数の回帰分析 ( $K = 3$ ) を特殊ケースとして含む。

性質1. 残差和は0である。

性質2. 残差と説明変数は，直交する： $\sum_{i=1}^n \hat{u}_i x_{mi} = 0, m = 1, 2, \dots, K.$

性質3. 観測値の和は，回帰値の和に等しい： $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i.$

性質4. 観測値の平均は，回帰値の平均に等しい： $\bar{y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i.$

性質5. 残差と回帰値は，直交する： $\sum_{i=1}^n \hat{u}_i \hat{y}_i = 0.$

性質6.  $x_1$  は定数だから，標本平均からなる座標  $(\bar{y}, \bar{x}_2, \dots, \bar{x}_K)$  は，回帰直線上に位置する。

性質7. 観測値と回帰値の積和は，回帰値の平方和に等しい：

$$\sum_{i=1}^n y_i \hat{y}_i = \sum_{i=1}^n \hat{y}_i^2.$$

性質8. 決定係数  $R^2$  は，回帰値  $\hat{y}_i$  と観測値  $y_i$  の相関係数（重相関係数）の2乗に等しい。



# 残差変動の性質

性質1. 説明変数を増やせば、 $RSS$  が必ず同じになるか減少する。

$$\Phi(y|I_K) = \sum_{i=1}^n \{y_i - (\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{K-1} x_{K-1i} + \hat{\beta}_K x_{Ki})\}^2, \quad (17)$$

$$\Phi(y|I_{K-1}) = \sum_{i=1}^n \{y_i - (\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{K-1} x_{K-1i})\}^2 \quad (18)$$

$$RSS(y|I_K) = \min_{\hat{\beta}_1, \dots, \hat{\beta}_K} \Phi(y|I_K) \leq \min_{\hat{\beta}_1, \dots, \hat{\beta}_{K-1}} \Phi(y|I_{K-1}) = RSS(y|I_{K-1})$$

証明：説明変数が  $K-1$  個のときの推定量を  $\tilde{\beta}_1, \dots, \tilde{\beta}_{K-1}$  とする。説明変数が  $K$  個のときの推定値を  $\hat{\beta}_1, \dots, \hat{\beta}_{K-1}, \hat{\beta}_K$  とする。

$\hat{\beta}_1 = \tilde{\beta}_1, \dots, \hat{\beta}_{K-1} = \tilde{\beta}_{K-1}, \hat{\beta}_K = 0$  とすれば、 $\Phi(y|I_K) = \Phi(y|I_{K-1})$ 。  
ゆえに、 $\Phi(y|I_K)$  の最小値は必ず  $\Phi(y|I_{K-1})$  に到達することができ、大きくならない。

# 残差変動の性質

性質2. 説明変数の数  $K$  を標本数  $n$  と同じなるまで増やせば、 $RSS$  が0になる。ただし、説明変数の間に一次の関係がないことが条件である。一次の関係の例：すべての  $i$  に関して、 $x_{1i} = x_{2i}$ ,  $x_{1i} = 5x_{3i}$  など。  $K = n$  のとき、

$$RSS(y|I_n) = \sum_{i=1}^n \{y_i - (\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{n-1} x_{n-1i} + \hat{\beta}_n x_{ni})\}^2 = 0$$

$$y_1 = \hat{\beta}_1 x_{11} + \cdots + \hat{\beta}_{n-1} x_{n-11} + \hat{\beta}_n x_{n1}$$

$$y_2 = \hat{\beta}_1 x_{12} + \cdots + \hat{\beta}_{n-1} x_{n-12} + \hat{\beta}_n x_{n2}$$

...

$$y_n = \hat{\beta}_1 x_{1n} + \cdots + \hat{\beta}_{n-1} x_{n-1n} + \hat{\beta}_n x_{nn}$$

は  $n$  個の未知数を持つ  $n$  本の方程式となる。説明変数の間に一次の関係がなければ  $RSS(y|I_n) = 0$  が解ける。

**モデル選択** 異なった説明変数で異なったモデルを構成することができる。その中のどのモデルを選ぶかを決めることが一種のモデル選択の問題である。

**決定係数  $R^2$  の問題点** 以上のスライドで分かるのは、 $RSS$  を小さく（決定係数  $R^2$  を大きく、 $R^2 = 1 - RSS / TSS$ ）したいなら、説明変数を増やせばいい。 $R^2$  を回帰式の選択基準とすると、追加した説明変数は説明力がないとしても、説明変数を追加した回帰式が必ず選ばれる。

# 自由度修正済み決定係数(続き)

より良いモデル選択の基準：自由度修正済み決定係数  $\bar{R}^2$  説明変数を追加した場合、 $R^2$  が必ず大きくなるが、 $\bar{R}^2$  は  $\frac{n-1}{n-K}$  で調整したため、説明力があまりない説明変数を増やしても、 $\bar{R}^2$  は大きくなるしない。 $\bar{R}^2$  が大きいほど、回帰モデルが良い。

$$\bar{R}^2 = 1 - \frac{RSS/(n-K)}{TSS/(n-1)} \quad (19)$$

$$\bar{R}^2 = 1 - \frac{s_{yy}|I_K}{s_{yy}} \quad (20)$$

$$\bar{R}^2 = R^2 - \frac{K-1}{n-K}(1-R^2) \quad (21)$$

# 多変数回帰（重回帰）の例

## Example

人的資本を含むクロスセクション生産関数（変数の定義はテキストにある）

$$\begin{aligned}\widehat{\log Y} = & 0.37(0.40) + 0.61(5.8) \log L + 0.33(5.0) \log K \\ & + 0.12(0.56) \log H1 + 0.067(1.1) \log H2 + 0.014(0.09) \log H3 \\ & + 0.17(2.4) \log H4 + 0.18(1.7) \log H5 + 0.12(1.4) \log H6.\end{aligned}\quad (22)$$

$R^2 = 0.99648$ ,  $\bar{R}^2 = 0.9957$ 。t 値が小さい  $H1, H2, H3$  を除いて再推定

$$\begin{aligned}\widehat{\log Y} = & 0.77(1.3) + 0.57(5.8) \log L + 0.36(6.2) \log K \\ & + 0.20(4.5) \log H4 + 0.15(1.5) \log H5 + 0.13(1.7) \log H6,\end{aligned}\quad (23)$$

$R^2 = 0.99628$  上のモデルの  $R^2$  より減少した、 $\bar{R}^2 = 0.9958$  上のより大きくなった。

# 偏回帰係数に関する t 検定

$$V(\hat{\beta}_K) = \frac{\sigma^2}{(n - K + 1)s_{x_K x_K | I_{K-1}}}$$
$$t_{\beta_K} = \frac{\hat{\beta}_K}{\sqrt{s^2 / \{(n - K + 1)s_{x_K x_K | I_{K-1}}\}}} \quad (24)$$

自由度  $n - K$  (テキストにミスがある) の  $t$  分布に従う。

## Example

### Per Capita 式と収穫不変性

$$\log Y_t = \alpha + \beta \log K_t + \gamma \log L_t + \dots \quad (25)$$

**収穫の不変性**  $\beta + \gamma = 1$  であれば、 $K$  と  $L$  を  $c$  倍に増やしたら、 $Y$  も  $c$  倍まで増える。 $\delta = \beta + \gamma - 1$  とする。

$$\log Y = \alpha + \beta \log K + (1 + \delta - \beta) \log L + \dots \quad (26)$$

$y = Y/L, k = K/L$  として

$$\log y = \alpha + \beta \log k + \delta \log L + \dots \quad (27)$$

$$H_0 : \beta + \gamma = 1, H_1 : \beta + \gamma < 1,$$

は  $H_0 : \delta = 0$  を等価である。

結果： $\widehat{\log y} = -0.07(-1.0) \log L + 0.36(6.2) \log k + \text{他の項は同じ}$ 。帰無仮説が棄却できない。

# 複数の偏回帰係数に関するF検定

検定には誤差項が正規分布に従うという仮定が必要。

## $H_A$ モデル

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i, i = 1, 2, \cdots, n \quad (28)$$

$H_A$  モデルの  $RSS$  を  $RSS(H_A)$  と表記する。

$$H_0 : \beta_{K-m+1} = 0, \cdots, \beta_{K-1} = 0, \beta_K = 0 \quad (29)$$

## $H_0$ モデルを立てる

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{K-m} x_{K-mi} + \varepsilon_i, i = 1, 2, \cdots, n \quad (30)$$

$H_0$  モデルの  $RSS$  を  $RSS(H_0)$  と表記する。



# 複数の偏回帰係数に関するF検定

$$\begin{aligned} f &= \frac{n - K}{m} \frac{RSS(H_0) - RSS(H_A)}{RSS(H_A)} \\ &= \frac{1}{m} \frac{RSS(H_0) - RSS(H_A)}{s^2} \end{aligned} \quad (31)$$

$$f = \frac{n - K}{m} \frac{R^2(H_A) - R^2(H_0)}{1 - R^2(H_A)} \quad (32)$$

自由度  $(m, n - K)$  の  $F$  分布に従う。 $m$  は制約の数 ( $0$  と仮定された説明変数の数)。

$\chi^2$

複数の偏回帰係数に関する $\chi^2$ 検定 誤差項が正規分布に従うという仮定が不要である。代わりに標本の数たくさんある ( $n$ が大きい) ことが必要。

$$C = \frac{RSS(H_0) - RSS(H_A)}{s^2} = (n - K) \frac{R^2(H_A) - R^2(H_0)}{1 - R^2(H_A)} \quad (33)$$

自由度  $m$  の  $\chi^2$  分布 (カイ自乗分布) に従う。

## Example

人的資本変数のF検定：帰無仮説は $H_1, H_2, H_3$ の係数が0である。

$$f = \frac{46 - 9}{3} \frac{0.116488 - 0.11016}{0.11016} = 0.71$$

$$f = \frac{46 - 9}{3} \frac{0.99648 - 0.99628}{1 - 0.99648} = 0.70$$

四捨五入しなければ同じ値になる。

# 回帰式全体に関するF検定

定数項以外のすべての説明変数の係数が0であるかどうか、すなわち、定数項以外の説明変数はまったく説明力がないかどうかを検定する。

$$H_0 : \beta_2 = 0, \dots, \beta_K = 0 \quad (34)$$

$$f = \frac{n-K}{K-1} \frac{TSS - RSS(H_A)}{RSS(H_A)} = \frac{n-K}{K-1} \frac{R^2}{1-R^2} \quad (35)$$

テキストの (3.50) 式の一般化である。