

2

2010 年度 学部 エコノメトリックス 第 8 回講義メモ (訂正増補版)

2010 年 10 月 27 日

2.7 推定した回帰式の当てはまりの良さ

第 2 回講義で、回帰式の当てはまりの良さを**決定係数**

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{\text{ESS}}{\text{TSS}}$$

という ESS(回帰式で説明される変動) と TSS(全変動) の比で表すことを説明した。

ここでは、決定係数 R^2 に関する二つの話題について触れることにしよう。

3

説明変数の数と決定係数

単回帰モデルで、標本の大きさが $n = 2$ のとき、 $R^2 = 1$ となる。というのも、異なる座標の 2 点があればそれを含む直線はユニークに決まるからである。同様に、重回帰モデルの場合にも、定数項を含む説明変数が $k = n$ 個であるときに、 $R^2 = 1$ となる。

このように、説明変数の数が増えれば、それに伴って R^2 は 1 に近づく。では、説明変数の数の影響を割り引いた回帰式の「当てはまりの良さ」を考えるにはどうしたらよいであろうか。

一つの考え方は、RSS(残差 2 乗和) や TSS(全変動) を自由度で割った

4

ものの比を使うというものである。すなわち、当てはまりの良さを

$$\bar{R}^2 = 1 - \frac{\frac{\sum_{i=1}^n e_i^2}{n-k}}{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

で定義する。ここで k は定数項を含む説明変数の数である。この $\bar{R}^2$ を自由度調整済決定係数と呼ぶ。実は、

$$1 - R^2 = \left(\frac{n-k}{n-1} \right) (1 - \bar{R}^2)$$

という関係があるので、決定係数 R^2 が $0 \leq R^2 \leq 1$ であったのに対し、自由度調整済決定係数 $\bar{R}^2$ は

$$-\frac{k-1}{n-k} \leq \bar{R}^2 \leq 1$$

であり、 $k > 2$ の時には負の値をとることもある。

5

定数項がない場合の決定係数

回帰モデルに定数項を含む場合、この回帰式は必ず座標 $(\bar{Y}, \bar{X})$ を通るため、全変動 (TSS) や説明変動 (ESS) を考える際に、 $\bar{Y}$ からの偏差を考えた。

しかし、定数項を含まない場合には、回帰式は原点 $(0, 0)$ を通るが、 $(\bar{Y}, \bar{X})$ は通らない。したがって、この場合には全変動を $\sum_{i=1}^n Y_i^2$ で定義する。そのため、決定係数 R^2 の定義を以下のように変更する。

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n Y_i^2}$$

6

2.8 予測

単回帰モデル $Y_i = \alpha + \beta X_i + \epsilon_i$ ($i = 1, \dots, n$) が安定的であって、 $i = n+1$ においても成立する状況を考えよう。このとき、 X_{n+1} さえ与えられれば、 α, β の OLS 推定量 a, b を用いて、 Y_{n+1} の予測が可能になる。 Y_{n+1} の予測量を

$$\hat{Y}_{n+1} = a + bX_{n+1} \quad (2.8.1)$$

とすると、その予測誤差は $Y_{n+1} - \hat{Y}_{n+1}$ 、予測の平均 2 乗誤差は $\text{MSE}(Y_{n+1} - \hat{Y}_{n+1}) = E[(Y_{n+1} - \hat{Y}_{n+1})^2]$ で与えられる。予測誤差の期待値をとると、

$$E(Y_{n+1} - \hat{Y}_{n+1}) = E[(\alpha - a) + (\beta - b)X_{n+1} + \epsilon_{n+1}] = 0$$

7

となり、不偏予測量 (unbiased predictor) になっていることが分かる。

また $\hat{Y}_{n+1}$ が不偏予測量であることより、予測の平均 2 乗誤差は予測誤差の分散に等しい。よって、

$$\begin{aligned} \text{MSE}(Y_{n+1} - \hat{Y}_{n+1}) &= E \left[((\alpha - a) + (b - \beta)X_{n+1} + \epsilon_{n+1})^2 \right] \\ &= \text{Var}(a) + X_{n+1}^2 \text{Var}(b) + \text{Var}(\epsilon_{n+1}^2) \\ &\quad + 2X_{n+1} \text{Cov}(a, b) \end{aligned}$$

第 4 回講義で得た結果より、

$$b - \beta = \sum_{i=1}^n w_i \epsilon_i, \quad a - \alpha = \sum_{i=1}^n \left(\frac{1}{n} - w_i \bar{X} \right) \epsilon_i$$

であるから、 $\text{Var}(b) = \sigma^2 \sum_{i=1}^n w_i^2$, $\text{Var}(a) = \sigma^2 \frac{\frac{1}{n} \sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$ となり、

8

また、

$$\begin{aligned} \text{Cov}(a, b) &= E[(a - \alpha)(b - \beta)] \\ &= E \left[\left(\sum_{i=1}^n w_i \epsilon_i \right) \left(\sum_{i=1}^n \left(\frac{1}{n} - w_i \bar{X} \right) \epsilon_i \right) \right] \\ &= -\sigma^2 \bar{X} \left(\sum_{i=1}^n w_i^2 \right) \end{aligned}$$

という関係があるので、予測誤差の分散は

$$\begin{aligned} \text{MSE}(Y_{n+1} - \hat{Y}_{n+1}) &= \text{Var}(Y_{n+1} - \hat{Y}_{n+1}) \\ &= \sigma^2 \frac{\frac{1}{n} \sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \sigma^2 X_{n+1}^2 \left(\sum_{i=1}^n w_i^2 \right) + \sigma^2 \\ &\quad - 2\sigma^2 \bar{X} X_{n+1} \left(\sum_{i=1}^n w_i^2 \right) \end{aligned}$$

9

$$= \sigma^2 \left(1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

として得られる。これより、 X_{n+1} が $\bar{X}$ に近いほど予測誤差が小さくなることが分かる。

さらに攪乱項に関する仮定 5 が成立する場合には、予測誤差の区間推定を厳密に求めることができる。

2.8 関数形と線形化変換

今まで Y が X の線形関数であるケースを扱ってきた。線形回帰モデルは、 Y が X の非線形関数であっても適当な変換を施すことによって線形化できるものについて幅広く適用できる。

通常、以下の変換がよく用いられる。

10

(a) **逆数** 説明変数 (独立変数) ないし被説明変数 (従属変数) に逆数をとるもの。

$$Y = \alpha + \beta X + \epsilon$$

(b) **ロジスティック変換** 成長曲線として用いられるロジスティック曲線に回帰モデルを組み合わせたもの。

$$Y = \frac{e^{\alpha + \beta X + \epsilon}}{1 + e^{\alpha + \beta X + \epsilon}} \Leftrightarrow \log \left(\frac{Y}{1 - Y} \right) = \alpha + \beta X + \epsilon$$

ここで、 $0 \leq Y \leq 1$ である。

(c) **対数線形変換 log-log** 両辺ともに対数線形化するもの。

$$Y = \alpha X^\beta e^\epsilon \Leftrightarrow \log Y = \log \alpha + \beta \log X + \epsilon$$

(d) **半対数線形変換 log-lin** 左辺を対数線形化するもの。

$$Y = e^{\alpha + \beta X + \epsilon} \Leftrightarrow \log Y = \alpha + \beta X + \epsilon$$

11

(e) **半対数線形変換 lin-log** 右辺を対数線形化するもの。

$$e^{Y - \alpha} = X^\beta e^\epsilon \Leftrightarrow Y - \alpha = \beta \log X + \epsilon$$

