2010 年度 学部 エコノメトリックス
第 6 回講義メモ

2010 年 10 月 20 日

2.5 攪乱項の分散の推定

前回まで見てきたように、α, β の最小 2 乗推定量 a, b の分散は σ^2 の関数になっている。ところが、実際には σ^2 は未知であることが多い。それでは、どのように σ^2 を推定したらよいだろうか。

σ^2 は攪乱項 ϵ_i の分散であるから、ϵ_i の推定量である残差 e_i を用いて推定することを考えよう。いま、残差 e_i は、

$$e_i \equiv Y_i - \hat{Y}_i = (\alpha - a) + (\beta - b)X_i + \epsilon_i \tag{2.5.1}$$

と書ける。ここで、以前得た結果より、

$$b - \beta = \sum_{j=1}^{n} w_j \epsilon_j$$

$$a - \alpha = \sum_{j=1}^{n} \left(\frac{1}{n} - w_j \bar{X} \right) \epsilon_j$$

であるから、(2.5.1) 式は、

$$e_i = -\left[\sum_{j=1}^{n} \left(\frac{1}{n} - w_j \bar{X} \right) \epsilon_j \right] - \left(\sum_{j=1}^{n} w_j \epsilon_j \right) X_i + \epsilon_i \tag{2.5.2}$$

となる。

仮定 1 〜仮定 4 より、e_i^2 の期待値は、

$$\begin{aligned}
\mathrm{E}(e_i^2) = \mathrm{E}&\left[\left(\sum_{j=1}^{n} \left(\frac{1}{n} - w_j \bar{X} \right) \epsilon_j \right)^2 \right] + X_i^2 \mathrm{E}\left[\left(\sum_{j=1}^{n} w_j \epsilon_j \right)^2 \right] + \mathrm{E}(\epsilon_i^2) \\
&+2X_i \mathrm{E}\left[\left(\sum_{j=1}^{n} \left(\frac{1}{n} - w_j \bar{X} \right) \epsilon_j \right) \left(\sum_{j=1}^{n} w_j \epsilon_j \right) \right] \\
&-2\mathrm{E}\left[\left(\frac{1}{n} - w_i \bar{X} \right) \epsilon_i^2 \right] - 2X_i \mathrm{E}\left(w_i \epsilon_i^2 \right) \\
= \sigma^2 &\left[\frac{1}{n} + \left(\sum_{j=1}^{n} w_j^2 \right) \bar{X}^2 + \left(\sum_{j=1}^{n} w_j^2 \right) X_i^2 + 1 - 2 \left(\sum_{j=1}^{n} w_j^2 \right) \bar{X} X_i \right. \\
&\left. -\frac{2}{n} + 2w_i \bar{X} - 2w_i X_i \right]
\end{aligned}$$

$$= \sigma^2 \left[1 - \frac{1}{n} + \left(\sum_{j=1}^{n} w_j^2 \right) \left(X_i - \bar{X} \right)^2 - 2w_i \left(X_i - \bar{X} \right) \right]$$

になる。

ゆえに、残差 2 乗和 $\sum_{i=1}^{n} e_i^2$ の期待値は、

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{n} e_i^2 \right) &= \sigma^2 \left[n - 1 + \left(\sum_{j=1}^{n} w_j^2 \right) \left(\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 \right) - 2 \sum_{i=1}^{n} w_i \left(X_i - \bar{X} \right) \right] \\
&= \sigma^2 (n-2)
\end{aligned}$$

となる。

この結果より、

$$\hat{\sigma}^2 \equiv \frac{1}{n-2} \sum_{i=1}^{n} e_i^2 \tag{2.5.3}$$

は σ^2 の不偏推定量 (不偏分散) である。一方、

$$\tilde{\sigma}_n^2 \equiv \frac{1}{n} \sum_{i=1}^{n} e_i^2 \tag{2.5.4}$$

は不偏ではないが、一致推定量になっている。というのも、

$$\operatorname*{plim}_{n\to\infty} \tilde{\sigma}_n^2 = \operatorname*{plim}_{n\to\infty} \frac{1}{n-2} \sum_{i=1}^{n} e_i^2 \lim_{n\to\infty} \frac{n-2}{n} = \sigma^2 \times 1 = \sigma^2$$

であるからである。

加えて、仮定 5 ($\epsilon_i \sim iid\, N(0, \sigma^2)$) が満たされるときには、$\hat{\sigma}^2$ の厳密な分布を求めることができる。仮定 5 より、

$$\sum_{i=1}^{n} \frac{\epsilon_i^2}{\sigma^2} \sim \chi^2(n)$$

が導かれる。線形代数の結果を用いると、

$$v^2 \equiv \frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \frac{e_i^2}{\sigma^2} \sim \chi^2(n-2) \tag{2.5.5}$$

が得られ、残差 2 乗和を σ^2 で割ったものが自由度 $n-2$ のカイ 2 乗分布に従うことがわかる。これは、最小 2 乗推定量 a, b の分布を求める際に重要な結果であり、α, β の仮説検定のところで再論する。

2.6 回帰係数についての仮説検定

A) 検定統計量の分布

回帰係数 α, β についての仮説検定は、最小 2 乗推定量 a, b から導出される検定統計量の分布に基づいて行われる。まず、検定統計量の分布を求めることから始めよう。

撹乱項に関して仮定 5 が満たされるときには、a, b の厳密な分布が得られる。第 4 回講義の結果 ((2.3.2) 式) より、

$$b \sim N\left(\beta, \sigma^2 \sum_{i=1}^{n} w_i^2\right)$$

が得られるので、変形した τ_0 は

$$\tau_0 = \frac{b-\beta}{\sqrt{\sigma^2 \sum_{i=1}^{n} w_i^2}} \sim N(0,1)$$

標準正規分布に従う。

実際には撹乱項の分散 σ^2 は未知のことが多いが、これを推定量 $\hat{\sigma}^2$ で置き換えた τ_1 の分布も厳密に求めることができる。

$$\tau_1 = \frac{b-\beta}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^{n} w_i^2}} = \frac{b-\beta}{\sqrt{\sigma^2 \left(\sum_{i=1}^{n} w_i^2\right)}} \div \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{\tau_0}{\sqrt{\frac{v^2}{n-2}}} \sim t(n-2)$$

τ_1 は標準正規分布に従う τ_0 と自由度 $n-2$ のカイ 2 乗分布に従う v^2 を自由度で割ったものの平方根の比になっているので、自由度 $n-2$ の t 分布に従うことが知られている。

ところが、撹乱項に関して仮定 2' しか満たされないときは、a, b の厳密な分布が得られず、中心極限定理に基づく近似分布 (漸近分布という) しか得られない。したがって、統計量 τ_2 は

$$\tau_2 = \frac{b-\beta}{\sqrt{\hat{\sigma}^2 \sum_{i=1}^{n} w_i^2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

n が十分大きいときに (漸近的に) 標準正規分布に従う。当然 n は十分大きいので大数の弱法則より

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2 \quad \text{as } n \to \infty$$

になっている。