回帰分析の基礎 5

crimsonbach

2004 年 10 月 8 日

特に時系列データを扱う場合、誤差項どうしが相関することがある。これを系列相関という。ここでは 1 次の自己回帰（auto-regression）を考える。1 次の自己回帰では誤差項間に以下のような関係がある。

$$u_t = \rho u_{t-1} + \epsilon_t, \qquad t = 1, \cdots, T. \tag{1}$$

ここで、 ρ は定常仮定であるための条件として $|\rho| < 1$ である。（1）式の ϵ_t は通常の仮定

$$E(\epsilon_t) = 0, \quad V(\epsilon_t) = {\sigma_\epsilon}^2, \quad Cov(\epsilon_t, \epsilon_s) = E(\epsilon_s \epsilon_t) = 0, \quad s \neq t.$$

を満たす。誤差項 u_t の分散および共分散は

$$V(u_t) = \frac{{\sigma_\epsilon}^2}{(1-\rho^2)}, \quad Cov(u_t, u_{t-s}) = \rho^s V(u_t).$$

誤差項間の自己相関係数[*1]は

$$\begin{aligned} r_s &\equiv \frac{Cov(u_t, u_{t-s})}{\sqrt{V(u_t)V(u_{t-s})}} \\ &= \frac{Cov(u_t, u_{t-s})}{V(u_t)} \\ &= \rho^s. \end{aligned} \tag{2}$$

（2）式から分かるように、 s が大きくなるほど、つまり 2 つの期間が離れるほど、自己相関係数 r_s は小さくなる。

では、誤差項間に系列相関がある単回帰モデル

$$Y_t = \beta_1 + \beta_2 X_t + u_t.$$

[*1] そもそも誤差項の生成メカニズム（ u_t が従う確率モデル）は $u_t = \rho u_{t-1} + \epsilon_t$ （ $E(\epsilon_t) = 0, E(\epsilon^2) = {\sigma_\epsilon}^2, E(\epsilon_t \epsilon_s) = 0$ ）である。u_t と u_{t-1} の相関係数は $u_t = \rho(\rho u_{t-2} + \epsilon_{t-1}) + \epsilon = \rho^2 u_{t-2} + \epsilon_t + \rho\epsilon_{t-1} = \cdots = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}$ 。定常性の条件から $E({u_{t-1}}^2) = \frac{{\sigma_\epsilon}^2}{(1-\rho^2)} = {\sigma_u}^2$ であり、 $\frac{1}{n}\sum_{t=2}^{T} {u_{t-1}}^2$ の確率収束は $\operatorname*{plim}_{t\to\infty} \frac{1}{n}\sum_{t=2}^{T} {u_{t-1}}^2 = E({u_{t-1}}^2) = {\sigma_u}^2$ 。同様に $E(u_t u_{t-1}) = E(\sum \rho^i \epsilon_{t-i})(\sum \rho^i \epsilon_{t-i-1}) = \rho \frac{{\sigma_e}^2}{1-\rho^2} = \rho {\sigma_u}^2$ であり、 $\frac{1}{n}\sum_{t=2}^{T} u_{t-1}u_t$ の確率収束は $\operatorname*{plim}_{t\to\infty} \frac{1}{n}\sum_{t=2}^{T} u_{t-1}u_t = E(u_{t-1}u_t) = \rho {\sigma_u}^2$ 。自己相関係数 $\hat{r} \equiv \frac{\sum_{t=2}^{T} \hat{u}_{t-1}\hat{u}_t}{\sum_{t=2}^{T} \hat{u}_{t-1}^2} \approx \frac{\sum_{t=2}^{T} u_{t-1}u_t}{\sum_{t=2}^{T} {u_{t-1}}^2}$ であるから、分母と分子を T で割ると自己相関係数は r に確率収束する（ $\operatorname*{plim}_{t\to\infty} \hat{r} = r$ ）。

において最小二乗法を行う。β_2 の最小二乗推定量は

$$\begin{aligned}\hat{\beta_2} &= \frac{\sum(X_t-\overline{X})Y_t}{\sum(X_t-\overline{X})^2} \\ &= \beta_2 + \frac{\sum(X_t-\overline{X})u_t}{\sum(X_t-\overline{X})^2}.\end{aligned}$$

$E(u_t)=0$ だから、

$$E(\hat{\beta_2})=\beta_2.$$

$\hat{\beta_2}$ は不偏推定量および一致推定量である。ただし、

$$\begin{aligned}V(\hat{\beta_2}) &= E(\hat{\beta_2}-\beta_2)^2 \\ &= E\left[\left\{\tfrac{\sum(X_t-\overline{X})u_t}{\sum(X_t-\overline{X})^2}\right\}^2\right]\end{aligned}$$

$t\neq\tau$ では $E(u_tu_\tau)\neq 0$ だから、分散は $V(\hat{\beta_2})\neq\frac{\sigma^2}{\sum(X_t-\overline{X})^2}$ となって、通常の最小二乗法では計算できない。また、推定量 $\hat{\beta_2}$ はもはや最良線型不偏推定量（BLUE）ではない。

誤差項間に系列相関があるかどうかの検定については、帰無仮説

$$H_0:\rho=0.$$

を設定する。検定の方法として、ここではダービン・ワトソンの d 統計量（Durbin - Watson d statistic）を使う（以下、 DW とする）。

$$DW=\frac{\sum\limits_{t=2}^{T}(e_t-e_{t-1})^2}{\sum\limits_{t=1}^{T}{e_t}^2}. \tag{3}$$

また、DW は 0 から 4 の間の値しかとらない ($0\leq DW\leq 4$)。T が十分に大きいならば、 $\sum\limits_{t=1}^{T}{e_t}^2$ 、 $\sum\limits_{t=2}^{T}{e_t}^2$ 、 $\sum\limits_{t=2}^{T}{e_{t-1}}^2$ はほぼ等しい。したがって、

$$DW\approx\frac{\sum{e_t}^2+\sum{e_{t-1}}^2-2\sum e_t\sum e_{t-1}}{\sum{e_t}^2}\approx 2(1-\rho). \tag{4}$$

$\hat{\rho}$ には、 $e_t=\rho e_{t-1}+\epsilon^*$ での最小二乗推定量を利用して、 $u_t\approx e_t$ と考えられる。よって、 $\rho\approx\rho$ となり、$DW\approx 2(1-\rho)$ である。

$$DW\approx\begin{cases}0 & \rho\approx 1\\ 2 & \rho\approx 0\\ 4 & \rho\approx -1\end{cases}.$$

DW の値が 2 に近い場合は H_0 を採択し、 0 あるいは 4 に近い場合は H_0 を棄却する。$\rho=0$ のとき

$$P(DW>DW_\alpha)=\alpha,\qquad P(DW>DW_{1-\alpha})=1-\alpha.$$

となるためのパーセント点を DW_α および $DW_{1-\alpha}$ とする。DW の分布は説明変数に依存するため、パーセント点は容易に求められない。しかし、 DW は説明変数に依存しない上限値 DW_U および下限値 DW_L が存在し、

$$DW_L<DW_{1-\alpha}<DW_U,\qquad 4-DW_U<DW_\alpha<DW_L.$$

となることが知られている。したがって、

$$\begin{cases} 0 \leq DW \leq DW_L & \text{棄却 } (\rho > 0) \\ DW_L \leq DW \leq DW_U & \text{留保} \\ DW_U \leq DW \leq 4 - DW_U & \text{採択} \\ 4 - DW_U \leq DW \leq 4 - DW_L & \text{留保} \\ 4 - DW_L \leq DW \leq 4 & \text{棄却 } (\rho < 0) \end{cases}$$

因みに、系列相関が生じた場合は回帰モデルが誤っている可能性が高い。たとえば、真の回帰式に含まれる説明変数を見落として回帰モデルを設定すると、誤差項にその部分が含まれ相関が強くなるのがそれである。