単回帰 (2)

別所俊一郎

2006 年 5 月 12 日

OLS 推定の前提

OLS 推定量が「望ましい」性質を持つためにはいくつかの前提が必要。

- どのようなときに用いればよいか、がわかる
- どのようなときに用いにくいか、がわかる
- 以下の条件を満たしていないときでも OLS 推定量は計算は可能

ここでの条件は以下の 3 つ

1. 説明変数を所与としたときの誤差項の条件付分布の平均がゼロ
2. 各観測値の分布は i.i.d.
3. 説明変数と誤差項の 4 次モーメントは有限

仮定 1：$E[u_i|X_i] = 0$

X_i を所与としたときの u_i の条件付分布の平均がゼロ。

- u_i が表している「その他の要素」についての仮定
- 説明変数を所与としたときに誤差項の平均がゼロ、という意味において、誤差項と説明変数のあいだに関係はない
- 観測値は（真の）回帰線の周囲に均等に分布
- 説明変数 X_i を所与としたとき、Y_i の条件付期待値は真の回帰直線上に並ぶ

仮定1：$E[u_i|X_i] = 0$

説明変数と誤差項の相関との関係

- 条件付き期待値がゼロなら共分散もゼロだから、

$$E[u_i|X_i] = 0 \Longrightarrow \text{corr}(X_i, u_i) = 0$$

- ただし、逆は必ずしも成り立たないから、説明変数と誤差項の相関がゼロだからといって誤差項の条件付き期待値がゼロになるとは限らない
- 対偶は成り立つから、説明変数と誤差項に相関があれば誤差項の条件付き期待値はゼロにならない
- それゆえ、説明変数と誤差項の相関で考えてもよい。

仮定 2：(X_i, Y_i) は i.i.d.

標本抽出の方法についての仮定

- 個人の無作為抽出のケースには成り立つ

成り立たないケース

- X_i が実験の一部として設定されているケース（稀）
- 時系列データ：同じ主体の通時的変化を追っているケース
  - 時点が近ければ相関を持つ可能性が高い（独立でない）
  - 特殊な扱いが必要
- （Oversampling のケース）

仮定 3：$0 < E[X_i], E[u_i] < \infty$

X_i と u_i が有限の 4 次モーメントを持つ

- 極端な外れ値を持たない
- OLS の検定統計量の大標本近似を正当化する仮定：中心極限定理の応用のため（標本分散の一致性の証明にも用いたことを想起せよ）
- 確認するのは困難だが、成立しているものとして扱うことがほとんど（観測される値は有限個）。
- 正規分布の 4 次モーメントは有限

OLS 推定の仮定とは…

数学的なもの

- これらの仮定が成り立てば、OLS 推定量の標本分布は漸近的に正規分布に従う
- 仮説検定や信頼区間の形成が可能になる

OLS が使いやすい／使いにくい状況の特定化

- 実際にはさまざまな事情でこれらの仮定は厳密には成り立たない
- とくに時系列データのばあい
- それらへの対処法はまた後ほど。

OLS 推定量の標本分布

OLS 推定量は確率変数で標本分布を持つ

- OLS 推定量は標本によって決まるから、母集団が同じであっても標本が変われば値は変わる
- 小標本の分布は複雑だが、大標本ならば中心極限定理によって漸近的に正規分布に従う
- 仮説検定などを行うためには標本分布の性質を知っておくことが必要
- 前述の仮定のもとで、OLS 推定量は一致性と不偏性を持ち、漸近的に正規分布に従う

推定量の分布

標本平均の分布

- 小標本のときに分布の形状を特定化するのは困難だが、大標本（$n \to \infty$）のとき、無作為標本であれば

$$（不偏性）\quad E[\overline{Y}] = \mu_y$$

$$（一致性）\quad \overline{Y} \xrightarrow{d} N\left(\mu_y, \sigma^2_{\overline{Y}}\right)$$

OLS 推定量の分布

- 小標本のときに分布の形状を特定化するのは困難だが、サンプルの大きさによらず

$$（不偏性）\quad E[\hat{\beta}_0] = \beta_0, \qquad E[\hat{\beta}_1] = \beta_1$$

OLS 推定量の不偏性の証明

OLS 推定量は

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

いま、$Y_i = \beta_0 + \beta_1 X_i + u_i$ だから

$$Y_i - \overline{Y} = \beta_1 (X_i - \overline{X}) + u_i - \overline{u}$$

$\hat{\beta}_1$ の式の分子に代入して整理すると

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(\beta_1 (X_i - \overline{X}) + u_i - \overline{u})}{\sum_{i=1}^n (X_i - \overline{X})^2} = \beta_1 + \frac{\sum_{i=1}^n (X_i - \overline{X}) u_i}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

両辺期待値をとると、

$$\begin{aligned} E[\hat{\beta}_1] &= \beta_1 + E\left[E\left[\left. \frac{\sum_{i=1}^n (X_i - \overline{X}) u_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \right| X_i \right] \right] \\ &= \beta_1 + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X}) E[u_i | X_i]}{\sum_{i=1}^n (X_i - \overline{X})^2} \right] = \beta_1 \qquad Q.E.D. \end{aligned}$$

OLS 推定量の一致性

$n \to \infty$ のとき、OLS 推定量は中心極限定理によって 2 変量正規分布に漸近的に従う

（証明） OLS 推定量は

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

だから、まず分子に着目すると、$\overline{X}$ は μ_X の一致推定量だから分子は $v_i \equiv (X_i - \overline{X})u_i$ の標本平均で近似できる。ここで、$E[u_i|X_i] = 0$ だから $E[v_i] = 0$ であり、また標本は i.i.d.、$\mathrm{var}(v_i) = \mathrm{var}[(X_i - \overline{X})u_i] < \infty$ だから中心極限定理が成り立ち、

$$v \xrightarrow{d} N(0, \sigma_v^2/n)$$

分母は $\mathrm{var}(X)$ の一致推定量だから、$\hat{\beta}_1 - \beta_1 \cong \overline{v}/\mathrm{var}(X)$。それゆえ

$$\hat{\beta}_1 \xrightarrow{d} N\left(\beta_1, \frac{\mathrm{var}((X - \mu_X)u)}{n(\mathrm{var}(X))^2}\right) \qquad Q.E.D.$$

大標本理論の適用可能性

- $n > 100$ もあれば十分。今後の他の推定量についても同様。
- OLS 推定量は一致性を持ち、その標準誤差はサンプルサイズが大きいほど小さくなる
- X_i の（標本）分散が大きいほど OLS 推定量（$\hat{\beta}_1$）の分散は小さい：散らばっているほうが正確な線を引きやすい（Fig 4.5）
- 正規分布に漸近的に従うという性質を使うと仮説検定や信頼区間の設定が容易。

係数についての仮説検定

「クラスあたり児童数はテストの点数に影響しない」を調べたい

- 仮説を数字で表現する：$\beta_{\text{児童数}} = 0$
- 仮説検定を行う：どうやって？

復習：母平均についての仮説検定

1. 帰無仮説・対立仮説の設定：
   $H_0 : E[Y] = \mu_{Y,0}$ v.s. $H_1 : E[Y] \neq \mu_{Y,0}$
2. 標本平均 $\overline{Y}$ の標準誤差（$\mathrm{SE}(\overline{Y})$）の推定
3. t 値の算出：$t = (\overline{Y} - \mu_{Y,0})/\mathrm{SE}(\overline{Y})$
4. p 値の算出：H_0 を棄却できる有意水準の最小値
   - H_0 が正しいとしたときに、得られた値よりも「離れた」値が得られる確率
   - $n \to \infty$ では $t \xrightarrow{d} N(0,1)$ だから $p = 2\Phi(-|t^{\text{act}}|)$

OLS 推定量の仮説検定

$\hat{\beta} \xrightarrow{d} N$ だから、基本的な手続きは母平均の仮説検定と同じ

[1] 帰無仮説・対立仮説の設定

$$H_0 : \beta_1 = \beta_{1,0} \quad \text{v.s.} \quad H_1 : \beta_1 \neq \beta_{1,0}$$

[2] OLS 推定量 $\hat{\beta}_1$ の標準誤差（SE($\hat{\beta}_1$)）の推定

$$\mathrm{SE}(\hat{\beta}_1) = \sqrt{\sigma^2_{\hat{\beta}_1}} = \frac{1}{n}\frac{\frac{1}{n-2}\sum_{i=1}^n (X_i - \overline{X})^2 \hat{u}_i^2}{[\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2]^2}$$

(4.14) を対応する標本統計量で置き換えたもの

[3] t 値の算出：

$$t = \frac{\text{推定量} - \text{仮説の値}}{\text{推定量の標準誤差}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\mathrm{SE}(\hat{\beta}_1)}$$

OLS 推定量の仮説検定

[4] p 値の算出：H_0 を棄却できる有意水準の最小値

$$p = \Pr_{H_0}\left[|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{\mathrm{act}} - \beta_{1,0}|\right] = \Pr_{H_0}(|t| > |t|^{\mathrm{act}})$$

$\hat{\beta} \xrightarrow{d} N$ だから、

$$p = \Pr_{H_0}(|Z| > |t|^{\mathrm{act}}) = 2\Phi(-|t^{\mathrm{act}}|)$$

- H_0 が正しいとしたときに、得られた値よりも「離れた」値が得られる確率

OLS 推定量の片側検定

- 帰無仮説の設定

$$H_0 : \beta_1 = \beta_{1,0} \quad \text{v.s.} \quad H_1 : \beta_1 < \beta_{1,0}$$

- t 値の解釈、p 値の算出

$$p = \Pr_{H_0}(Z < t^{\text{act}}) = \Phi(t^{\text{act}})$$

片側検定を使うとき

- 仮説の値より大きな（小さな）値を取ることが理論的・実証的に自明なとき
- ただし、そのようなケースは多くないので、両側検定を使うケースが多い
  - 価格効果の符号条件？
  - 他の要因の混在の可能性