大学院計量経済学講義ノート

劉　慶豊 (Qingfeng Liu)*

2009 年 2 月 8 日

小樽商科大学商学部

*qliu@res.otaru-uc.ac.jp

1　最小二乗法の統計学的性質

1.1　線形回帰モデルに関する条件 (強いバージョン)

線形回帰モデルの最小二乗（ＯＬＳ）推定量がいい性質を持つために様々な前提条件を必要とする。一般的なテキストでは条件ではなく仮定と言うが、上記の意味合いの文脈では条件と呼ぶほうが適切である。このような条件は強さによって様々なバージョンがあるが、ここではまずその中の一つ強いバージョンを紹介する。ここでの強いというのは普通の経済データはなかなかこのような条件を満たさないことを意味する。だが、データが以下の条件を満たされれば、単純な最小二乗法で分析できる。

A1. 線形性 y と x が以下の線形関係を持つ

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i \tag{1}$$

y : 非説明変数、$x_{i1}, x_{i2}, \cdots$: 定数項を含む説明変数、ε_i 攪乱項または誤差項。行列表現

$$y = X\beta + \varepsilon \tag{2}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{Kn} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

A2. フルランク (Full rank) $rank\,(X) = K$。ただし、K は説明変数の数で X の列数である。

A3. $E\,(\varepsilon|X) = 0$、X と ε が無相関であることを意味する。

A4. $Var\,(\varepsilon|X) = \sigma^2 I$、ただし、$Var\,(\varepsilon|X) = E\left[(\varepsilon - E\,(\varepsilon|X))\,(\varepsilon - E\,(\varepsilon|X))'\,|X\right]$。

A5. X が非確率変数または外生変数。すなわち、X は ε とまったく無関係である。

A.6 正規性 $\varepsilon \sim N\,(0, \sigma^2 I)$。

1.2　各条件の意味合い

A1. 線形性 $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i$。この線形性必ずしも成り立たないときが少なくない。あくまでも経済指標などの間の関係の線形近似である。たとえば、生産関数の規模と収穫の関係に関しては、収穫逓増や逓減のとき線形で近似しきれない場合がある。その時非線形のモデルを考える。これについては後の章で説明する。

A2. フルランク (Full rank) $rank(X) = K$。ただし、K は説明変数の数で X の列数である。この条件が満たされなければ最小二乗推定量 $\hat{\beta}$ が一意に決まらず、推定できない。(??) の導出を参照せよ。

練習 1 *Example* 2.5 を説明しなさい。

A3. $E(\varepsilon|X) = 0$、X と ε が無相関であることを意味する。$E(\varepsilon) = 0$、$Cov(\varepsilon_i, X) = 0$ を意味する。

問題 2 $E(\varepsilon) = 0$、$Cov(\varepsilon_i, X) = 0$ を証明しなさい。*Conditioning Theorem* を参照。

A4. $Var(\varepsilon|X) = \sigma^2 I$、ただし、$Var(\varepsilon|X) = E\left[(\varepsilon - E(\varepsilon|X))(\varepsilon - E(\varepsilon|X))' |X\right]$。異なった i と j に関しては $Cov(\varepsilon_i, \varepsilon_j) = 0$、$Var(\varepsilon) = \sigma^2 I$ を意味する。すなわち、分散均一で系列相関がないのである。講義中グラフを持って説明する。

練習 3 $Var(\varepsilon) = \sigma^2 I$ を証明しなさい。

A5. X が非確率変数または外生変数。すなわち、X は ε とまったく無関係である。$A5$ は仮定 $A3$ 及び $A4$ と意味合い上、重複するところがある。

A.6 正規性 $\varepsilon \sim N(0, \sigma^2 I)$。$A5$ と $A6$ により、最小二乗法の推定量の性質をより簡単に導出することができる。$\hat{\beta}$ の推定量としての性質が優れていることを保障してくれる。その性質を利用して $\hat{\beta}$ の信頼性を検定できる。以下からこれらの条件の下で最小二推定量の性質を導出していく。

1.3 不偏推定量

定義 4 β の推定量 b が以下の性質を満たすとき、b が β の不偏推定量という。

$$E(b|X) = E(b) = \beta \tag{3}$$

性質 1 条件 $A1 - A3$ が満たされるとき、最小二乗推定量 b が β の不偏推定量である。

証明.

$$b = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X'\varepsilon \tag{4}$$

$$E(b) = E\left[\beta + (X'X)^{-1} X'\varepsilon\right] = \beta + E\left(E\left((X'X)^{-1} X'\varepsilon|X\right)\right) = \beta \tag{5}$$

$$E(b|X) = E\left[\beta + (X'X)^{-1} X'\varepsilon|X\right] = \beta + E\left((X'X)^{-1} X'\varepsilon|X\right) = \beta \tag{6}$$

□

1.4　最小二乗推定量の分散

$$
\begin{aligned}
V(b) &= E(b-\beta)(b-\beta)' \\
&= E\left(\left(X'X\right)^{-1}X'\varepsilon\right)\left(\left(X'X\right)^{-1}X'\varepsilon\right)' \\
&= E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}\right) \\
&= E\left[E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}|X\right)\right] \\
&= E\left[\left(X'X\right)^{-1}X'E\left(\varepsilon\varepsilon'|X\right)X\left(X'X\right)^{-1}\right] \\
&= E\left[\left(X'X\right)^{-1}X'\sigma^2IX\left(X'X\right)^{-1}\right] \\
&= \sigma^2E\left[\left(X'X\right)^{-1}\right] \qquad (7)
\end{aligned}
$$

第 6 番の等式を導出するために性質 A4 が使われた。

性質 2 従って、X が非確率変数の時

$$V(b) = \sigma^2\left(X'X\right)^{-1} \qquad (8)$$

X が確率変数の時

$$
\begin{aligned}
V(b|X) &= E\left[(b-\beta)(b-\beta)'|X\right] \\
&= E\left(\left(X'X\right)^{-1}X'\varepsilon\varepsilon'X\left(X'X\right)^{-1}|X\right) \\
&= \left(X'X\right)^{-1}X'E\left(\varepsilon\varepsilon'|X\right)X\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}X'\sigma^2IX\left(X'X\right)^{-1} \\
&= \sigma^2\left(X'X\right)^{-1} \qquad (9)
\end{aligned}
$$

1.5　最小二乗推定量の分布

$b = \beta + \left(X'X\right)^{-1}X'\varepsilon$ であるのと $\varepsilon \sim N\left(0, \sigma^2I\right)$ であることに注意して、条件 A1-A6 の下で、以上の結果と正規分布の線形結合の性質を利用すれば、以下の結果が得られる

性質 3 X が非確率変数の場合、

$$b \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$$

X が確率変数の場合

$$b|X \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$$

1.6 Gauss-Markov Theorem(BLUE, Best Linear Unbiased Estimator)

条件 A1-A6 が満たされれば、最小二乗法は最小分散線形不偏推定量（ＢＬＵＥ）となる。

すなわち、最小二乗推定量が b とする、それと異なる線形不偏推定量を $b_0 = Cy$ とする。

必ず

$$V(b) \leq V(b_0)$$

となる。

ここでの線形の意味は推定量は撹乱項 ε の線形係数であることです。

証明. b_0 が不偏推定量から

$$\begin{aligned} E(b_0|X) &= \beta \\ E(CX'\beta + C\varepsilon|X) &= \beta \\ CX'\beta &= \beta \\ CX' &= I \end{aligned}$$

次に

$$\begin{aligned} V(b_0|X) &= E\left[(Cy-\beta)(Cy-\beta)'|X\right] \\ &= E\left[C\varepsilon(C\varepsilon)'|X\right] \\ &= \sigma^2 CC' \\ &= \sigma^2\left[\left(\left[C-(X'X)^{-1}X\right]+(X'X)^{-1}X\right)\left(\left[C-(X'X)^{-1}X\right]+(X'X)^{-1}X\right)'\right] \end{aligned}$$

$CX' = I$ から $\left[C-(X'X)^{-1}X\right]\times\left[(X'X)^{-1}X\right]' = 0$ なので

$$V(b_0|X) = \sigma^2(X'X)^{-1} + \sigma^2\left(\left[C-(X'X)^{-1}X\right]\left[C-(X'X)^{-1}X\right]'\right)$$

さらに任意の次元が合うベクトル q に関して $q'\left[C-(X'X)^{-1}X\right]\left[C-(X'X)^{-1}X\right]'q = z'z \geq 0$ なので $\left[C-(X'X)^{-1}X\right]\left[C-(X'X)^{-1}X\right]$ は非負値定符号行列となる。証明完了。 □

1.7 Minimum Mean Squared Error（最小ＭＳＥ予測値）

定義 5 $MSE = E\left(\hat{\theta}-\theta\right)^2$

y の X による線形予測を $\hat{y} = X'\beta$ とする。予測 MSE は

$$MSE = E(y - X'\beta)^2$$

この MSE の最小化条件は

$$\frac{\partial MSE}{\partial \beta} = 0$$
$$E\left(Xy\right) = E\left(XX'\right)\beta$$

となる。これの標本版は最小二乗法の正規方程式

$$\frac{1}{n}\sum_{i=1}^{n} X_i y_i = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right) b$$

になっている。ゆえに、大数の法則が満たされる場合、b は予測 MSE を最小にするような予測値 $X'\beta$ の中の係数の推定量であると考えられる。

1.8　分散 σ^2 の推定

推定量として

$$\hat{\sigma}^2 = \frac{1}{n}e'e$$

が考えられるが、しかし $\hat{\sigma}^2$ は不偏推定量ではない。確認していく。残差メーカー $M = I - X\left(X'X\right)^{-1}X'$ として

$$\begin{aligned} E\left(\hat{\sigma}^2\right) &= E\left(\frac{1}{n}e'e\right) \\ &= E\left(\frac{1}{n}\left(My\right)'\left(My\right)\right) \\ &= E\left(\frac{1}{n}\left[M\left(X\beta+\varepsilon\right)\right]'\left[M\left(X\beta+\varepsilon\right)\right]\right) \end{aligned}$$

$MX = \left(I - X\left(X'X\right)^{-1}X'\right)X = 0$ なので、上式は

$$
\begin{aligned}
&= E\left(\frac{1}{n}\left(M\varepsilon\right)' M\varepsilon\right) \\
&= E\left(\frac{1}{n}\varepsilon' M\varepsilon\right) \\
&= E\left(\frac{1}{n} tr\left(\varepsilon' M\varepsilon\right)\right) \\
&= E\left(\frac{1}{n} tr\left(\varepsilon\varepsilon' M\right)\right) \\
&= E\left\{E\left(\frac{1}{n} tr\left(\varepsilon\varepsilon' M\right)|X\right)\right\} \\
&= E\left\{\frac{1}{n} tr\left(E\left(\varepsilon\varepsilon'|X\right)M\right)\right\} \\
&= E\left\{\frac{1}{n} tr\left(M\right)\right\}\sigma^2 \\
&= E\left\{\frac{1}{n} tr\left(I - X\left(X'X\right)^{-1}X'\right)\right\}\sigma^2 \\
&= E\left\{\frac{1}{n} tr\left(I\right) - tr\left(X'X\left(X'X\right)^{-1}\right)\right\}\sigma^2 \\
&= \frac{n-k}{n}\sigma^2
\end{aligned}
$$

以上の結果を踏まえて

$$
s^2 = \frac{n}{n-k}\frac{1}{n}e'e = \frac{1}{n-k}e'e
$$

が σ^2 の不偏推定量であることがわかる。

そして、b の分散 $V\left(b|X\right) = \sigma^2\left(X'X\right)^{-1}$ の推定量は $s^2\left(X'X\right)^{-1}$ となる。s は回帰分析の標準誤差（standard error）で、$s^2\left(X'X\right)^{-1}$ の第 (k,k) 要素は b_k の標準誤差と呼ぶ。

1.9 検定

以上で導出した推定量 b の分布を利用して幾つかの検定統計量を構成できる。

1.9.1 t 統計量

$$
t_k = \frac{b_k - \beta_k}{s_{b_k}} = \frac{b_k - \beta_k}{s\sqrt{\left[\left(X'X\right)^{-1}\right]_{kk}}}
$$

ただし、$\left[\left(X'X\right)^{-1}\right]_{kk}$ は $\left(X'X\right)^{-1}$ の第 (k,k) 要素。t_k は自由度が $(n-K)$ の t 分布に従う。

有意水準が $a\%$ として、帰無仮説 $H_0 : \beta_k = 0$ 対立仮説 $H_a : \beta_k > 0$ の片側検定を行うとき、まず

$$t_k = \frac{b_k - 0}{s_{b_k}}$$

を計算して、そして、t_k の値を自由度 $(n-K)$ の $1-a\%$ の分位点の値と比べる。t_k の方が大きい場合、帰無仮説を棄却する、逆の場合、帰無仮説を採択する。

1.9.2 t 統計量の導出

ベクトル z が標準正規分布に従うとする。M がべき等行列の場合、$z'Mz$ が自由度が $rank\,(M)$ の χ^2 分布に従う。

互いに独立な標準正規分布と χ^2_k 分布に従う二つの確率変数 x, y の比 $x/\sqrt{y/k}$ が t 分布に従う。

$z_k = (b_k - \beta_k) \Big/ \sqrt{\left[\sigma^2 \left(X'X\right)^{-1}\right]_{kk}}$ とする。b の分布から $z_k \sim N\,(0,1)$。さらに $v = (n-K)\,s^2\,/\sigma^2$ とする。

$$\begin{aligned} v &= e'e/\sigma^2 \\ &= \left(\frac{\varepsilon}{\sigma}\right)' M \left(\frac{\varepsilon}{\sigma}\right) \end{aligned}$$

$\varepsilon/\sigma \sim N\,(0,1)$ と M がべき等行列なので、v が自由度が $rank\,(M) = n-K$ の χ^2 分布に従う。

さらに

$$t_k = \frac{z_k}{v}$$

となることがわかる。t_k が t 分布に従うことを証明するためには、z_k と v が独立であることを証明する必要がある。定理 $B.12$ から、$(X'X)^{-1}\,XM = 0$ であるので、独立が証明される。

1.10 F 統計量

定数項以外すべての係数が 0 である、すなわち回帰モデルはまったく説明力がないという仮説を検定するとき、

$$F = \frac{R^2/\,(K-1)}{(1-R^2)\,/\,(n-K)}$$

は自由度 $(K-1, n-K)$ の F 分布に従う。