

## OLS 推定量の導出

別所俊一郎

説明変数が  $k$  個，サンプルの大きさが  $n$  のモデルを考えてみましょう．さしあたって説明変数の選択の問題は解決されているとして，定数項を明示的に取り扱くと次の線形連立方程式で表現されます．

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \\ Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n \end{cases} \quad (1)$$

これを行列で表現したいので，式をじーっと見ると，右辺の第  $k+1$  項目までは係数が共通であることに気がつくでしょう．そこで，

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{U} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad (2)$$

とにおいて，

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \quad (3)$$

とすれば，もとの式になります．ここで，縦ベクトル  $\mathbf{y}$  は  $n \times 1$ ， $\mathbf{X}$  は  $n \times k$ ，縦ベクトル  $\boldsymbol{\beta}$  は  $k+1 \times 1$ ，縦ベクトル  $\mathbf{U}$  は  $n \times 1$ ，ですから，行列の積である  $\mathbf{X}\boldsymbol{\beta}$  の次元<sup>1</sup>は  $n \times 1$  となります．行列のふつうの積は「ヨコ×タテ」で定義されますから，1行目がちゃんと

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \quad (4)$$

となっていることを確認しましょう．

さてここで (3) についていくつか確認しておきましょう． $\mathbf{X}$  は確率変数， $\mathbf{U}$  のそれぞれの要素は実現値を観測できない確率変数です．観測できない  $\boldsymbol{\beta}$  は定数ですが， $\mathbf{U}$  が確率変数ベクトルなので  $\mathbf{y}$  のそれぞれの要素も確率変数となります．ここで (3) という式の形と， $\mathbf{u}$  についていくつかの仮定を置けば，それっぽい  $\boldsymbol{\beta}$  を決めることができる，というのが推定の手続きに他なりません．最小 2 乗法とは， $\mathbf{u}$  のそれぞれの要素はあまり大きくないと想定して，その 2 乗和を最小にするような  $\boldsymbol{\beta}$  をもって推定値にする，という方法で

<sup>1</sup>行列の縦にいくつ，横にいくつの要素が並んでいるか，ということ次元といいます（ルパン 3 世とは関係ない）．ランクとは違うものです．次元の違う行列は足したり引いたりできないですし，一般の積の場合は左の行列の横の次元と，右の行列の縦の次元が同じでないと計算できません．だから，行列のふつうの掛け算では交換則が成り立ちません．掛けられる行列のそれぞれの要素に，掛ける行列を掛けるという，積となる行列の次元がやたらと大きくなる掛け算も考えることがあって，そういうのをクロネッカー積といいます，さしあたって使わないので気にしないことにしましょう．行列では割り算は定義されません．

す。さてそこで (3) を用いて最小にされるべき 2 乗和を計算してみましょう。残差ベクトル  $\hat{U}$  を同様に定義して移項すると

$$\hat{U} = y - Xb \quad (5)$$

となります。2 乗和は

$$\hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2 = \hat{U}'\hat{U} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_n \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} \quad (6)$$

と書けます<sup>2</sup>から、右辺も同じように書いてみると、

$$\hat{U}'\hat{U} = (y - Xb)'(y - Xb) \quad (7)$$

和をとったものの転置行列は転置行列の和になるから<sup>3</sup>、

$$\hat{U}'\hat{U} = (y' - (Xb)')(y - Xb) \quad (8)$$

積の転置行列は転置したものの積になるから<sup>4</sup>、

$$\hat{U}'\hat{U} = (y' - b'X')(y - Xb) \quad (9)$$

カッコを外す分配則はふつうに成り立つから<sup>5</sup>、

$$\hat{U}'\hat{U} = y'y - y'Xb - b'X'y + b'X'Xb \quad (10)$$

3 項目を  $b'X'$  と  $y$  の積とみると、積の転置行列は転置したものの積だから、

$$\hat{U}'\hat{U} = y'y - y'Xb - y'Xb + b'X'Xb \quad (11)$$

2 項目と 3 項目をまとめると

$$\hat{U}'\hat{U} = y'y - 2y'Xb + b'X'Xb \quad (12)$$

となります。適当に  $\beta$  の推定値  $b$  を決めてやれば、それに応じた  $\hat{u} = y - Xb$  を計算することができるので、そこから計算された  $\hat{u}'\hat{u}$  のことを残差平方和とよびます。つまり、残差平方和を最小にする  $b$  を見つけるのが最小 2 乗法ということになります。残差平方和は推定値  $b$  の関数ですから、残差平方和を最小にする  $b$  を見つけるには、残差平方和を  $b$  で微分してゼロと置いて解けばよいわけです<sup>6</sup>。

<sup>2</sup> $U'$  は行列のタテとヨコをひっくりかえしたもので転置行列といいます。数学の人たちは  $U^T$  とも書くことがあるようです。どっちにしても微分や累乗とまぎらわしい表現ですね。

<sup>3</sup>ただの足し算で位置が違うだけだからです。

<sup>4</sup>積を取るときの縦横をひっくりかえさないといけないからです。確認してみましょう。

<sup>5</sup> $(a + b)(c + d) = ac + ad + bc + bd$  が成り立つということ。

<sup>6</sup>最大値を見つかることになるかもしれないので、ほんとうはもうちょっとチェックが必要。

「残差平方和を  $\mathbf{b}$  で微分して」といっても、 $\mathbf{b}$  はベクトルですからそのまま微分を考えるとできるわけではありません。ベクトルで微分するというと難しそうですが、ベクトルを構成するそれぞれの要素でそれぞれ微分して、それを重ねて書き直したもののことを、ベクトルで微分するといっているだけのことです。つまり、関数  $f(\mathbf{x})$  が  $(x_1, x_2, \dots, x_k)$  の関数であるとき<sup>7</sup>、ベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_k)$  で微分したものは、

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_k} \end{bmatrix} \quad (13)$$

です。もし、関数  $f(\mathbf{x})$  が  $(x_1, x_2, \dots, x_k)$  の線形和であれば、

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x} = \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad (14)$$

と表現できて、その微分は

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \mathbf{a} \quad (15)$$

となるので、まとめて書けば、

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a} \quad (16)$$

となります。2次関数のばあいも同じようにこちゃこちゃと計算すれば、行列  $\mathbf{A}$  が対称行列<sup>8</sup>なら、

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \quad (17)$$

となることが分かっています。さて、これだけの準備をして

$$\hat{\mathbf{U}}'\hat{\mathbf{U}} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \quad (18)$$

を  $\mathbf{b}$  で微分することを考えましょう。最小2乗法で求められる推定値  $\mathbf{b} = \hat{\boldsymbol{\beta}}$  は微分をゼロにする<sup>9</sup>から、

$$\frac{\partial(\hat{\mathbf{U}}'\hat{\mathbf{U}})}{\partial \mathbf{b}} = -2\mathbf{y}'\mathbf{X} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0} \quad (19)$$

<sup>7</sup>関数の値がベクトルではなくてただの数になるケースのみを考えます。ベクトル値関数の場合は横にも並べればよいだけです。

<sup>8</sup>タテとヨコの次元が同じ行列（正方行列）のうち、転置しても同じ行列のことを対称行列といいます。分散共分散行列は対称行列ですし、任意の行列  $\mathbf{X}$  に対して、 $\mathbf{X}'\mathbf{X}$  は対称行列になります。

<sup>9</sup>各要素がすべてゼロになるということ。

が成り立ちます．転置を書き直して移項して2で割れば，正規方程式を得ます．

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \quad (20)$$

さて，行列では割り算が定義されていませんから，正規方程式から  $\mathbf{b}$  を求めるには，

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y} \quad (21)$$

の両辺に左から  $\mathbf{X}'\mathbf{X}$  の逆行列を掛ければよいことになります．ところが，逆行列というのはどんな行列に対しても定義されているわけではないことに注意しましょう．行列を構成するベクトルが1次独立でなければ逆行列は存在せず，その場合にその行列はランク落ちしている...のですが，さしあたっては，データを入れた行列  $\mathbf{X}$  が多重共線性 (perfect multicollinearity) をもっていれば逆行列が定義できず，あるいは多重共線に近い状況であれば逆行列の値が大きくなる，ということだけをおさえておきましょう．つまり，完全な多重共線が発生していれば，OLS 推定量は定義できないことになります．

さて，多重共線が発生していなければ，両辺に左から  $\mathbf{X}'\mathbf{X}$  の逆行列を掛けて，OLS 推定量を計算することができます．すなわち，

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (22)$$

であり，この  $\mathbf{b}$  をとくに  $\hat{\beta}$  で表します．