慶應義塾大学ビジネス・スクール

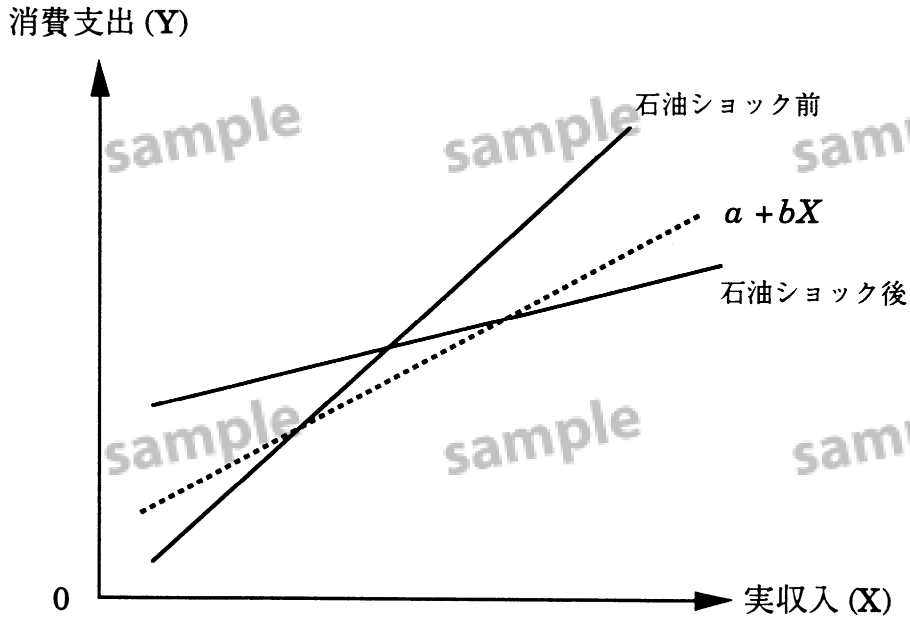
回帰分析シリーズ４

— 構造の同質性とダミー変数 —

回帰分析は、被説明変数（たとえば消費や利益）がどのような要因（所得や戦略）に依存するかを導き出す究めて有用な手法である。しかしこの手法は、すべてのデータを同質的なものと見なし、一つの回帰線に集約するものである。したがって、もし分析されるデータの構造が同質ではないとすれば、導き出された回帰線の信頼性は明らかに低くなる。

たとえば、男性と女性、低額所得者と高額所得者とでは、所得と消費の関係が異なることが考えられる。また企業の収益性と企業行動との関係についても、バブル経済のときとバブル崩壊後とでは違ったもになるかもしれない。図１は、構造の違いによる誤った推定線を概念化したものである。たとえば家計の収入と消費の関係において、石油ショック前と後では両者の関係が異なっているものとしよう。しかし、もし分析者が石油ショック前後のサンプルを区分せず、安易に同じ回帰モデルに組み込んだ場合には、点線で示された推定線を導き出してしまうのである。

（図１）構造の違いと推定線



回帰分析だけでなく、あらゆる統計的方法を用いることのできる条件の一つは、データの構造が異なっているという仮説を棄却することである。このノートでは、回帰分析における構造の同質性検定について基礎的な手法のいくつかを紹介する。

このノートは、慶應義塾大学ビジネス・スクールにおける補助教材として、同ビジネス・スクール教授矢作恒雄と博士課程磯辺剛彦が作成した。（1995年5月作成）

「回帰分析シリーズ３－行列－」において、回帰モデルの仮説検証を行うための便利な制約式である $\mathbf{R\beta = r}$ を紹介した。ここで $\mathbf{R}$ は $[(g-1)\times k]$、$\boldsymbol{\beta}$ は $(k\times 1)$、$\mathbf{r}$ は $[(g-1)\times 1]$ の行列である。k は説明変数の数を、また g は制約条件の数を示している。たとえば４つの説明変数をもつ回帰モデルで、

$$\beta_2 = \beta_4 \qquad \beta_3 + 2\beta_4 + \beta_5 = 1$$

という仮説を検証するには、$\mathbf{R}$ と $\mathbf{b}_*$ にそれぞれ次のような行列式が与えられる。ここで $\mathbf{b}_*$ とは制約条件が与えられた推定値であり、制約条件が与えられていない通常最小二乗法 (OLS) による推定値 $\mathbf{b}$ と区分している。

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 1 \end{pmatrix} \qquad \mathbf{r} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

最小二乗法は、このように与えられた制約条件の下で、予測誤差を最小にするような $\mathbf{b}_*$ を推定しなければならない。そのためにはスカラー関数を (1) 式のように定義し、$\mathbf{b}_*$ と $\boldsymbol{\lambda}$ でそれぞれ偏微分したものがゼロとなる $\mathbf{b}_*$ と $\boldsymbol{\lambda}$ を導き出せばよい。

$$\varphi = (\mathbf{y} - \mathbf{X}\mathbf{b}_*)'(\mathbf{y} - \mathbf{X}\mathbf{b}_*) - 2\boldsymbol{\lambda}'(\mathbf{R}\mathbf{b}_* - \mathbf{r}) \tag{1}$$

ここで $\boldsymbol{\lambda}$ は、ラグランジュ乗数の列ベクトルである。(1) 式を偏微分し、それぞれの方程式にゼロを与えれば、

$$\frac{\partial\varphi}{\partial\mathbf{b}_*} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}_* - 2\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0} \tag{2}$$

$$\frac{\partial\varphi}{\partial\boldsymbol{\lambda}} = -2\mathbf{R}\mathbf{b}_* + 2\mathbf{r} = \mathbf{0} \tag{3}$$

さらに (2) 式に $\mathbf{R(X'X)}^{-1}$ を掛ければ、

$$-\mathbf{R(X'X)}^{-1}\mathbf{X'y} + \mathbf{R}\mathbf{b}_* - \mathbf{R(X'X)}^{-1}\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0} \tag{4}$$

ここで通常最小二乗法を用いた制約のない推定値 $\mathbf{b}$ は (5) 式のように与えられる。

$$\mathbf{b} = \mathbf{(X'X)}^{-1}\mathbf{X'y} \tag{5}$$

(4) 式に (3) 式と (5) 式を代入すれば、

$$-\mathbf{R(X'X)}^{-1}\mathbf{X'Xb} + \mathbf{r} - \mathbf{R(X'X)}^{-1}\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0} \quad \text{より、}$$

$$\boldsymbol{\lambda} = \left[\mathbf{R(X'X)}^{-1}\mathbf{R}'\right]^{-1}(\mathbf{Rb} - \mathbf{r}) \tag{6}$$

これを (2) 式に代入し、両辺に $\mathbf{(X'X)}^{-1}$ を掛ければ、

$$(\mathbf{X'X})^{-1}\mathbf{X'y} - (\mathbf{X'X})^{-1}\mathbf{X'Xb}_{\bullet} + (\mathbf{X'X})^{-1}\mathbf{R'}\left[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1}(\mathbf{Rb} - \mathbf{r}) = \mathbf{0} \quad \text{より、}$$

$$\mathbf{b}_{\bullet} = \mathbf{b} + (\mathbf{X'X})^{-1}\mathbf{R'}\left[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1}(\mathbf{r} - \mathbf{Rb}) \qquad \because (5)\text{式} \tag{7}$$

次に制約モデルの残差 $\mathbf{e}_{\bullet}$ は、

$$\mathbf{e}_{\bullet} = \mathbf{y} - \mathbf{Xb}_{\bullet}$$

によって求めることができるが、この式は次のようにも表すことができる。

$$\begin{aligned}\mathbf{e}_{\bullet} &= \mathbf{y} - \mathbf{Xb} - \mathbf{X}(\mathbf{b}_{\bullet} - \mathbf{b}) \\ &= \mathbf{e} - \mathbf{X}(\mathbf{b}_{\bullet} - \mathbf{b})\end{aligned} \tag{8}$$

ここで (8) 式の両辺を二乗すると、

$$\mathbf{e}_{\bullet}'\mathbf{e}_{\bullet} = \mathbf{e'e} + (\mathbf{b}_{\bullet} - \mathbf{b})'\mathbf{X'X}(\mathbf{b}_{\bullet} - \mathbf{b}) \qquad \because \mathbf{X'e} = \mathbf{0} \tag{9}$$

(9) 式に (7) 式を代入すると、制約モデルと制約のないモデルとの残差平方和の差は (10) 式のように簡単に示すことができる。

$$\begin{aligned}\mathbf{e}_{\bullet}'\mathbf{e}_{\bullet} - \mathbf{e'e} &= (\mathbf{b}_{\bullet} - \mathbf{b})'\mathbf{X'X}(\mathbf{b}_{\bullet} - \mathbf{b}) \\ &= (\mathbf{r} - \mathbf{Rb})\left[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1}(\mathbf{r} - \mathbf{Rb})\end{aligned} \tag{10}$$

この (10) 式は、「回帰分析シリーズ 3 －行列－」での F 検定 (48) 式の分母に等しいことが分かる。したがって帰無仮説 $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ の検定は、(48) 式の代わりに (11) 式を用いてもよいことになる。(11) 式は、構造の同質性を検定する際の重要な分析手法となる（q と k については例題のところで述べる）。

$$F = \frac{(\mathbf{e}_{\bullet}'\mathbf{e}_{\bullet} - \mathbf{e'e}) \,/\, q}{\mathbf{e'e} \,/\, (n - k)} \tag{11}$$

構造の同質性

構造の同質性を検証するために、次のような仮説を構築したものとする。

帰無仮説：第一次石油ショック以前とそれ以降とでは、勤労者世帯の収入と消費支出との関係に変化はない。

この帰無仮説を検証するために、まずモデルの非制約型 (unrestricted form) を考えてみる。ここでは (12) 式の上半分が石油ショック以前、下半分が石油ショック以降とし、それぞれのサンプル数は $n1$ と $n2$ である。

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n1} \\ \hdashline Y_{n1+1} \\ Y_{n1+2} \\ \vdots \\ Y_{n1+n2} \end{bmatrix}
=
\begin{bmatrix} 1 & X_1 & 0 & 0 \\ 1 & X_2 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{n1} & 0 & 0 \\ \hdashline 0 & 0 & 1 & X_{n1+1} \\ 0 & 0 & 1 & X_{n1+2} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & X_{n1+n2} \end{bmatrix}
\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \hdashline \alpha_2 \\ \beta_2 \end{bmatrix}
=
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n1} \\ \hdashline u_{n1+1} \\ u_{n1+2} \\ \vdots \\ u_{n1+n2} \end{bmatrix}
\tag{12}
$$

(12) 式を行列型で示すと、

$$
\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \tag{13}
$$

そしてＯＬＳを用いることによって、(13) 式から各時期の切片と傾きの推定値を導き出すことができる。

$$
\mathbf{b} = \begin{bmatrix} a_1 \\ b_1 \\ \hdashline a_2 \\ b_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2'\mathbf{X}_2)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 \end{bmatrix} \tag{14}
$$

ここでの帰無仮説は、

$$
\mathrm{H}_0 : \ \mathbf{b}_1 = \mathbf{b}_2
$$

または、

$$
\mathrm{H}_0 : \ \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \tag{15}
$$

ここで (15) 式で表した仮説検定に、$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ の制約式を用いると、$\mathbf{R}$ と $\mathbf{r}$ はそれぞれ次のように表すことができる。

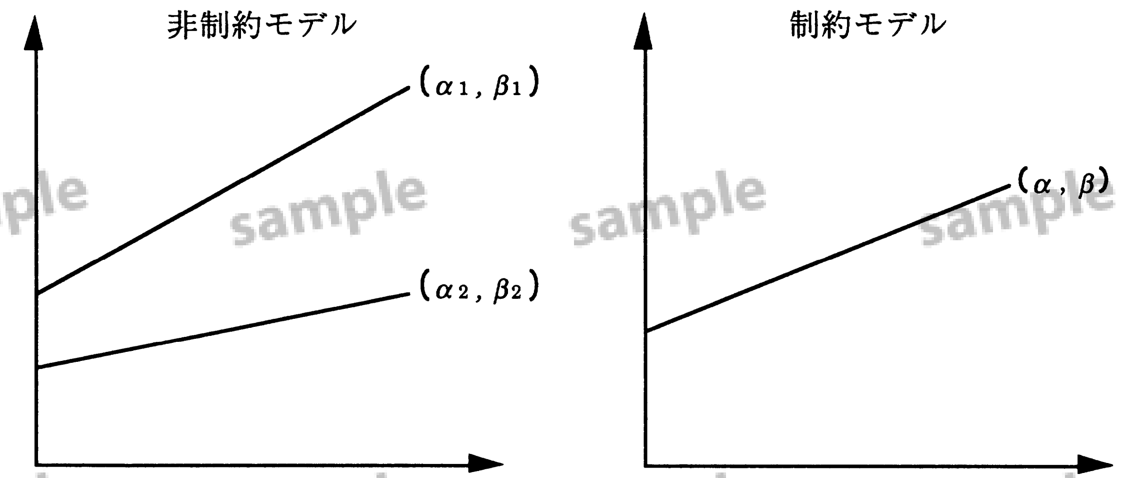
$$
\mathbf{R} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \end{bmatrix}
$$

$$
\mathbf{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}
$$

次にモデルの制約型 (restricted form) とは、石油ショック前と後のデータを区分せずに同じモデルに組み込むものである。したがって、推定される回帰モデルは (16) 式のように示される。また制約モデルと非制約モデルとの相違は、図２に示している。

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \tag{16}$$

（図２）制約モデルと非制約モデルとの相違



これまでの分析を実際の数値例に適用してみよう。表１は、総務庁統計局による勤労者世帯の一ヶ月当たりの収入と消費支出の推移である。第一次石油ショックが日

（表１）実収入と消費支出の推移

(単位：千円)

年 (n)	実収入 (X)	消費支出 (Y)
1963	53.3	41.1
1964	59.7	45.5
1965	65.1	49.3
1966	71.3	53.6
1967	78.7	58.8
1968	87.6	65.5
1969	97.7	72.6
1970	112.9	82.6
1971	124.6	91.3
1972	138.6	99.3
1973	165.9	117.0
1974	205.8	142.2
1975	236.2	166.0
1976	258.2	180.7
1977	286.0	197.9
1978	304.6	208.2

（出所：総務庁統計局）

本を襲ったのは 1973 年であり、ここでは1963 年から 1972 年までの１０サンプル ($n1$) と 1973 年から1978 年までの６サンプル ($n2$) を採用した（1979 年にも第二次石油ショックが起こっているが、ここではその影響はないものと仮定している）。

表１から、石油ショック前と後の **y** と **X** の行列はそれぞれ次のようになる。

石油ショック前

$$\mathbf{y}_1 = \begin{bmatrix} 41.1 \\ 45.5 \\ 49.3 \\ 53.6 \\ 58.8 \\ 65.5 \\ 72.6 \\ 82.6 \\ 91.3 \\ 99.3 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} 1 & 53.3 \\ 1 & 59.7 \\ 1 & 65.1 \\ 1 & 71.3 \\ 1 & 78.7 \\ 1 & 87.6 \\ 1 & 97.7 \\ 1 & 112.9 \\ 1 & 124.6 \\ 1 & 138.6 \end{bmatrix}$$

石油ショック後

$$\mathbf{y}_2 = \begin{bmatrix} 117.0 \\ 142.2 \\ 166.0 \\ 180.7 \\ 197.9 \\ 208.2 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 1 & 165.9 \\ 1 & 205.8 \\ 1 & 236.2 \\ 1 & 258.2 \\ 1 & 286.0 \\ 1 & 304.6 \end{bmatrix}$$

$$\mathbf{X}_1'\mathbf{X}_1 = \begin{bmatrix} 10 & 889.5 \\ & 86{,}621.0 \end{bmatrix}, \quad \mathbf{X}_1'\mathbf{y}_1 = \begin{bmatrix} 659.6 \\ 63{,}861.0 \end{bmatrix}$$

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \begin{bmatrix} 1.155 & -0.112 \\ & 1.33-\mathrm{E}4 \end{bmatrix}, \quad (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 = \begin{bmatrix} 4.411 \\ 0.692 \end{bmatrix}$$

$\mathbf{y}_1'\mathbf{y}_1 = 47{,}100.3$

また、

$$\mathbf{X}_2'\mathbf{X}_2 = \begin{bmatrix} 6 & 1{,}456.7 \\ & 366{,}911.3 \end{bmatrix}, \quad \mathbf{X}_2'\mathbf{y}_2 = \begin{bmatrix} 1{,}012.0 \\ 254{,}558.0 \end{bmatrix}$$

$$(\mathbf{X}_2'\mathbf{X}_2)^{-1} = \begin{bmatrix} 4.616 & -0.018 \\ & 0.75-\mathrm{E}5 \end{bmatrix}, \quad (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 = \begin{bmatrix} 6.282 \\ 0.669 \end{bmatrix}$$

$\mathbf{y}_2'\mathbf{y}_2 = 176{,}630.0$

これらを (14) 式に代入すると、

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \\ \hdashline a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} 4.411 \\ 0.692 \\ \hdashline 6.282 \\ 0.669 \end{bmatrix}$$

したがって、石油ショック前と後の推定線はそれぞれ次のようになる。

$$Y_P = 4.411 + 0.692\ X_P$$
$$Y_A = 6.282 + 0.669\ X_A$$

ここで残差平方和は、$\mathbf{e'e = y'y - b'x'y}$ によって求めることができる。したがってそれぞれの残差平方和は、

$$\mathbf{e_1'e_1} = 47{,}100.3 - \begin{bmatrix} 4.411 & 0.692 \end{bmatrix} \begin{bmatrix} 659.6 \\ 63{,}861.0 \end{bmatrix} = 2.191$$

$$\mathbf{e_2'e_2} = 176{,}630.0 - \begin{bmatrix} 6.282 & 0.669 \end{bmatrix} \begin{bmatrix} 1{,}012.0 \\ 254{,}558.1 \end{bmatrix} = 12.428$$

以上の計算から、非制約モデルでの残差平方和は、

$$\mathbf{e'e = e_1'e_1 + e_2'e_2} = 2.191 + 12.428 = 14.619$$

になる。また制約モデルについては、

$$\mathbf{X_*'X_*} = \begin{bmatrix} 16 & 2346.2 \\ & 453{,}532.2 \end{bmatrix}, \quad \mathbf{X_*'y_*} = \begin{bmatrix} 1{,}671.6 \\ 318{,}419.1 \end{bmatrix}$$

$$(\mathbf{X_*'X_*})^{-1} = \begin{bmatrix} 0.259 & -1.3-\mathrm{E}3 \\ & 9.1-\mathrm{E}6 \end{bmatrix}, \quad (\mathbf{X_*'X_*})^{-1}\mathbf{X_*'y_*} = \begin{bmatrix} 6.307 \\ 0.669 \end{bmatrix}$$

$$\mathbf{y_*'y_*} = 223{,}730.3$$

により、次のような推定線が導かれる。

$$Y_* = 6.307 + 0.669\ X_*$$

また制約モデルの残差平方和は、

$$\mathbf{e_*'e_*} = 18.707$$

制約モデルと非制約モデルとの残差平方和を用いて、石油ショック前と後との構造の違いを分析する。これらの残差平方和を (11) 式に代入すると、

$$F = \frac{(\mathbf{e_*'e_* - e'e})\ /\ q}{\mathbf{e'e}\ /\ (\mathrm{n} - k)} = \frac{(18.707 - 14.619)\ /\ 2}{14.619\ /\ (16 - 4)} = 1.678$$

ここで q は $\alpha_1 = \alpha_2$、$\beta_1 = \beta_2$ という２つの制約があることを示し、また k とは非制約変数の未知数の数である。２と１２の自由度をもつ F 値の 5% レベルは 3.88 であり、したがって両者の構造が同じであるという帰無仮説が棄却されないことが示されている。

この結果を別の手法で検証してみよう。残差平方和を用いた (11) 式の検証方法は、「回帰分析シリーズ３－行列－」での (48) 式と同じであることをすでに述べた。したがって、これまでの計算結果と (48) 式の結果は同じになるはずである。(5.48) 式の分母は、

$$(\mathbf{Rb}-\mathbf{r})'\left[\mathbf{R}\,(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1}(\mathbf{Rb}-\mathbf{r}) \tag{17}$$

ここで、

$$\begin{aligned}\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'} &= \begin{bmatrix}\mathbf{I} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}(\mathbf{X_1'X_1})^{-1} & \mathbf{0}\\ \mathbf{0} & (\mathbf{X_2'X_2})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ -\mathbf{I}\end{bmatrix}\\ &= (\mathbf{X_1'X_1})^{-1}+(\mathbf{X_2'X_2})^{-1}\\ &= \begin{bmatrix}5.771 & -0.030\\ & 2.1+\mathrm{E}4\end{bmatrix}\end{aligned}$$

$$\left[\mathbf{R}(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1} = \begin{bmatrix}0.711 & 102.7\\ & 19{,}638.4\end{bmatrix}$$

$$\mathbf{Rb} = \begin{bmatrix}\mathbf{I} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{b_1}\\ \mathbf{b_2}\end{bmatrix} = \begin{bmatrix}1 & 0 & -1 & 0\\ 0 & 1 & 0 & -1\end{bmatrix}\begin{bmatrix}4.411\\ 0.692\\ \hdashline 6.282\\ 0.669\end{bmatrix} = \begin{bmatrix}-1.871\\ 0.023\end{bmatrix}$$

また $\mathbf{r}=\mathbf{0}$ なので、(17) 式は、

$$(\mathbf{Rb}-\mathbf{r})'\left[\mathbf{R}\,(\mathbf{X'X})^{-1}\mathbf{R'}\right]^{-1}(\mathbf{Rb}-\mathbf{r}) = 4.088$$

この結果は、$\mathbf{e}_*'\mathbf{e}_* - \mathbf{e}'\mathbf{e} = 18.707 - 14.619 = 4.088$ と同じであることがわかる。

傾きの変化

これまで回帰モデル全体の構造の違いについて分析を進めてきた。しかし分析の目的のよっては、切片や傾きの違いだけに注目しなければならないことも考えられる。まず最初に、ここでは傾きが同じかどうかを検証する手法を例題を用いて紹介しよう。

切片を無視して傾きが同じかどうかを検証するには、制約条件を $\beta_1=\beta_2$ に設定し、非制約モデルと比較すればよい。制約モデルは次のように表すことができる（ただし、$\mathbf{X}=\begin{bmatrix}\mathbf{i} & \mathbf{x}\end{bmatrix}$ である）。

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix} = \begin{bmatrix}\mathbf{i_1} & \mathbf{0} & \mathbf{x_1}\\ \mathbf{0} & \mathbf{i_2} & \mathbf{x_2}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \beta\end{bmatrix} + \begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix} \tag{18}$$

ここで例題の数値を (18) 式に代入すれば、

sample

$$\mathbf{X}_\bullet'\mathbf{X}_\bullet = \begin{bmatrix} n_1 & 0 & \mathbf{i}_1'\mathbf{x}_1 \\ 0 & n_2 & \mathbf{i}_2'\mathbf{x}_2 \\ \mathbf{x}_1'\mathbf{i}_1 & \mathbf{x}_2'\mathbf{i}_2 & \mathbf{x}_1'\mathbf{x}_1 + \mathbf{x}_2'\mathbf{x}_2 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 0 & 890 \\ 0 & 6 & 1{,}457 \\ & & 453{,}532 \end{bmatrix}$$

$$(\mathbf{X}_\bullet'\mathbf{X}_\bullet)^{-1} = \begin{bmatrix} 0.481 & 1.041 & -0.004 \\ 0 & 3.008 & -0.012 \\ 0 & 0 & 4.8-\mathrm{E}5 \end{bmatrix}, \quad (\mathbf{X}_\bullet'\mathbf{X}_\bullet)^{-1}\mathbf{X}_\bullet'\mathbf{y} = \begin{bmatrix} 5.724 \\ 4.255 \\ 0.677 \end{bmatrix}$$

$$\mathbf{X}_\bullet'\mathbf{y} = \begin{bmatrix} \mathbf{i}_1'\mathbf{y} \\ \mathbf{i}_2'\mathbf{y} \\ \mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 660 \\ 1{,}012 \\ 318{,}419 \end{bmatrix} \qquad \mathbf{y}'\mathbf{y} = 223{,}730$$

これまでの計算から制約モデルの残差平方和は、

$$\mathbf{e}_\bullet'\mathbf{e}_\bullet = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_\bullet(\mathbf{X}_\bullet'\mathbf{X}_\bullet)^{-1}\mathbf{X}_\bullet'\mathbf{y}$$

$$= (47{,}100 + 176{,}630) - \begin{bmatrix} 660 & 1{,}012 & 318{,}419 \end{bmatrix} \begin{bmatrix} 5.724 \\ 4.255 \\ 0.677 \end{bmatrix}$$

$$= 17.175$$

また非制約モデルはすでに計算されており、その残差平方和 $\mathbf{e}'\mathbf{e}$ は 14.619 であった。そして両者の残差平方和を用いて F 検定を行えば、

$$F = \frac{(17.715 - 14.619)/1}{14.619/(16-4)} = 2.541$$

ただし制約条件は $\alpha_1 = \alpha_2$ だけなので 1 になる。この F 値の水準は、$F_{0.95} = 4.75$ よりも小さいことが分かる。したがって、石油ショック前と後の回帰モデルの傾きが変化していないという帰無仮説が棄却できない。

切片の変化

次に、構造の同質性について切片だけが異なるかどうかについて分析したいものとする。このとき注意しなければならないことは、非制約モデルが「傾きの変化」での制約モデルに変わることである。その理由は、切片の変化が傾きの変化にも影響されるからである。したがって制約モデルについても、傾き β に変化がないことを仮定する。この場合の制約モデルと非制約モデルとはそれぞれ次のようになる。

sample

制約モデル　　　　　　　　　　非制約モデル

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_1\end{bmatrix}=\begin{bmatrix}\mathbf{i}_1 & \mathbf{x}_1\\ \mathbf{i}_2 & \mathbf{x}_2\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}+\mathbf{u} \qquad \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_1\end{bmatrix}=\begin{bmatrix}\mathbf{i}_1 & \mathbf{0} & \mathbf{x}_1\\ \mathbf{0} & \mathbf{i}_2 & \mathbf{x}_2\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \beta\end{bmatrix}+\mathbf{u}$$

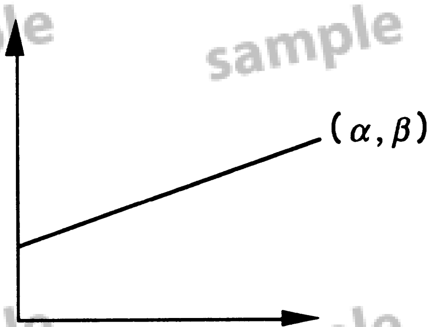
そしてそれぞれの残差平方和は、$\mathbf{e}_*'\mathbf{e}_* = 18.707$　および　$\mathbf{e}'\mathbf{e} = 17.715$　より、

$$F=\frac{(18.707-17.715)/1}{17.715/(16-3)}=0.728$$

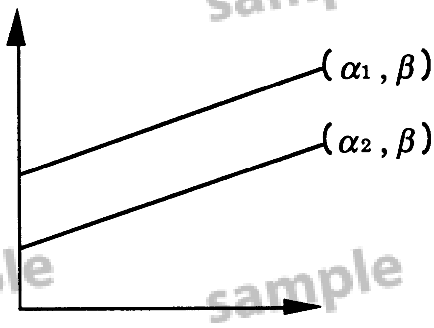
となり、この場合にも傾きの変化と同様に $F_{0.95}=4.67$ よりも小さい。

今までの議論をまとめると、構築される構造モデルは以下の (I) から (III) のいずれかのモデルに該当する。(I) は切片と傾きがともに共通であることを仮定したモデルであり、(I) は傾きだけが共通で切片が異なることを仮定したモデル、そして (III) は切片と傾きがともに異なることを仮定したモデルである。

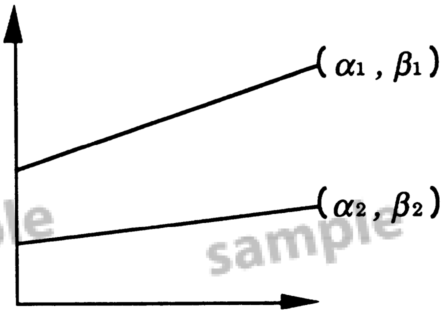
(I) $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_1\end{bmatrix}=\begin{bmatrix}\mathbf{i}_1 & \mathbf{X}_1\\ \mathbf{i}_2 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}+\mathbf{u}$



(II) $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\begin{bmatrix}\mathbf{i}_1 & \mathbf{0} & \mathbf{X}_1\\ \mathbf{0} & \mathbf{i}_2 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \beta\end{bmatrix}+\mathbf{u}$



(III) $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\begin{bmatrix}\mathbf{i}_1 & \mathbf{0} & \mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{i}_2 & \mathbf{0} & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \beta_1\\ \beta_2\end{bmatrix}+\mathbf{u}$



そして切片、傾き、モデル全体という３つの構造の同質性を検定するには、表２のような制約モデルと非制約モデルとを比較すればよい。

（表２）構造の同質性の検定

	制約モデル	非制約モデル
切片の検定	(I)	(II)
傾きの検定	(II)	(III)
モデル全体の検定	(I)	(III)

これまで、単回帰における構造の同質性だけを紹介してきたが、たとえ重回帰モデルであってもその手続きは同じである。たとえばモデルに含まれる説明変数の数が１から$(k-1)$に増加したものとしよう。このとき、モデル (I) からモデル (III) の行列記号 $\mathbf{X}$ を$\mathbf{X}_{\mathbf{M}}$ に変えると、その行列型は$(n\times 1)$ から$[n\times(k-1)]$になる。

$$\mathbf{X}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbf{X}_{\mathbf{M}}=\begin{pmatrix} x_{12} & x_{13} & \cdots & x_{1k} \\ x_{22} & x_{24} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

したがって回帰モデル (I) からモデル (III) の自由度は、それぞれ (I) $\mathrm{RSS}_1 : n-k$ (II) $\mathrm{RSS}_2 : n-k-1$ (III) $\mathrm{RSS}_3 : n-2k$ になる。(12) 式を用いて、重回帰モデルの切片の変化、傾きの変化、モデル全体の変化を検証するには、それぞれ次のような F 検定が必要とされる。

切片の検定：
$$F=\frac{\mathrm{RSS}_1-\mathrm{RSS}_2}{\mathrm{RSS}_2/(n-k-1)} \sim F(1,\ n-k-1)$$

傾きの検定：
$$F=\frac{(\mathrm{RSS}_2-\mathrm{RSS}_3)/(k-1)}{\mathrm{RSS}_3/(n-2k)} \sim F(k-1,\ n-2k)$$

モデル全体の検定：
$$F=\frac{(\mathrm{RSS}_1-\mathrm{RSS}_3)/k}{\mathrm{RSS}_3/(n-2k)} \sim F(k,\ n-2k)$$

さらにこれまでの分析方法以外にも、構造の同質性の検定にダミー変数を用いることもできる。この手法については、次の「ダミー変数」の項目において述べてゆくことにする。

ダミー変数

構造の同質性検定でのモデル (II) にダミー変数を適用すると、(19) 式のようにな

る。

$$Y_i = \alpha_1 D_{1,i} + \alpha_2 D_{2,i} + \beta X_i + u_i \qquad (19)$$

$D_{1,i}$ と $D_{2,i}$ は、比較する２つのサンプルに付与されるダミー変数である。例題でのダミー変数は、

$$D_{1,i} \left\langle \begin{array}{l} =1: i \text{ が石油ショック前のとき} \\ =0: i \text{ が石油ショック後のとき} \end{array} \right.$$

$$D_{2,i} \left\langle \begin{array}{l} =0: i \text{ が石油ショック前のとき} \\ =1: i \text{ が石油ショック後のとき} \end{array} \right.$$

になるが、ここで注意することは (19) 式に切片（定数項）を含めないことである。つまり $D_{1,i}$ と $D_{2,i}$ 、X を Y に回帰させるとき、コンピュータは自動的に切片を測定してしまうため、切片のダミー変数である $D_{1,i}$ と $D_{2,i}$ の係数の意味がなくなってしまうからである。したがって定数項を持つような回帰モデルは、ダミー変数の数が常にグループ内のカテゴリーの数よりも１つ少なくなければならない。したがって (19) 式の場合、カテゴリーの数は石油ショック前と後の２つなので、ダミー変数は１つになる。(19) 式は次のような定数項をもつ式に変換できる。

$$Y_i = \gamma_1 + \gamma_2 D_{2,i} + \beta X_i + u_i \qquad (20)$$

ここで、(19) 式と (20) 式との関係は、

$$\gamma_1 = \alpha_1 \ , \quad \gamma_2 = \alpha_2 - \alpha_1$$

になる。つまり γ_2 自体が切片の変化についての検定となりうるのである。このようにダミー変数を回帰モデルに組み込む際には、その係数の解釈に十分注意することが必要とされるのである。たとえばモデル (III) にダミー変数を使うと、

$$Y = \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 (D_1 X) + \beta_2 (D_2 X) + u \qquad (21)$$

によって計算できるが、この式を (20) 式と同様に定数項をもつモデルに変換すると、

$$Y = \alpha_1 + (\alpha_2 - \alpha_1) D_2 + \beta_1 X + (\beta_2 - \beta_1) D_2 X + u \qquad (22)$$

この場合には、切片だけでなく傾きの変化をも検証することができる。

またダミー変数は、自由度の許す限り、グループやカテゴリーをいくつでも増やすことができる。たとえば２つのカグループがそれぞれ４つのカテゴリーをもつような回帰モデルは (23) 式のようになる。

sample

$$Y = \alpha_1 + (\alpha_2 - \alpha_1)D_2 + (\alpha_3 - \alpha_1)D_3 + (\alpha_4 - \alpha_1)D_4 + \beta_1 X + (\beta_2 - \beta_1)D_2 X + (\beta_3 - \beta_1)D_3 X + (\beta_4 - \beta_1)D_4 X + u \qquad (23)$$

ダミー変数を用いて構造の同質性を検証する場合のメリットは、各回帰係数を t 検定することによって、どの変数が構造の違いをもっとも反映するかを認識できることにある。しかしダミー変数は分析の目的によっては強力な手法であるが、ダミー変数が２つ以上に増えるといくつかの複雑な問題が生じる。

まず回帰モデルに２つのグループが含まれる場合について考察する。ここでは新規に上場される株式を例として、野村証券や山一証券のような株式引き受け幹事会社が設定する「公開価格(OP_i)」と、市場で初めて取引が成立する「初値(IP_i)」との乖離を説明するモデルを構築したものとする。非説明変数は次のような式によって求め、これを「初期収益率 (IR_i)」と呼ぶことにする。ただし i は上場される株式を示している。

$$IR_i = \frac{IP_i - OP_i}{OP_i} \times 100$$

株式の初期収益率を説明するのに、上場される年と引き受け幹事会社という２つの変数をダミー変数に置き換えるものとする。上場される年を 1983 年と 1987 年とに、また引き受け幹事会社を野村証券、山一証券、大和証券とに区分する。このような分析目的においては、一般に次のような回帰モデルが想定されよう。なおサンプルは表３に記載している。

$$Y = \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 + \beta_1 Y_1 + \beta_2 Y_2 + u \qquad (24)$$

$$S_2 \begin{cases} = 1 : \text{幹事会社が野村証券のとき} \\ = 0 : \text{それ以外のとき} \end{cases}$$

$$S_3 \begin{cases} = 1 : \text{幹事会社が大和証券のとき} \\ = 0 : \text{それ以外のとき} \end{cases}$$

$$P_2 \begin{cases} = 1 : \text{上場年が 1986 年のとき} \\ = 0 : \text{上場年が 1983 年のとき} \end{cases}$$

ここで (24) 式を定数項をもつ式に変換すると、

$$Y = \gamma + \alpha_2 S_2 + \alpha_3 S_3 + \beta_2 P_2 + u \qquad (25)$$

(25) 式の係数の意味は、次の表のようにまとめることができる。たとえば (S_1, P_1) のカテゴリーに該当するサンプルの収益性は、係数 γ によって測定される。

（表３）新規上場株式の初期収益率

年	n	証券コード	幹事会社	初期収益率 (%)
83	1	6349	山一	88.2
	2	7999	野村	340.5
	3	8140	大和	141.0
	4	8141	大和	150.0
	5	6858	山一	224.8
	6	9764	山一	43.1
	7	1929	山一	20.5
	8	8841	野村	60.0
	9	1872	山一	90.0
	10	1973	大和	650.0
	11	6859	野村	57.1
	12	6963	野村	138.3
	13	6457	山一	116.7
	14	8592	大和	83.3
	15	8161	野村	14.2
	16	8199	山一	42.3
	17	1966	山一	112.9
86	1	7967	大和	62.3
	2	8158	大和	26.8
	3	6949	山一	70.5
	4	6143	野村	50.0
	5	4186	野村	42.7
	6	8037	山一	28.0
	7	8765	山一	8.8
	8	8159	野村	22.5
	9	5816	野村	25.8
	10	7965	野村	13.6
	11	8206	野村	10.0
	12	8209	大和	24.3
	13	7897	山一	18.5
	14	6860	大和	48.8
	15	8374	大和	12.2
	16	2902	野村	24.0
	17	6948	大和	20.2

またカテゴリー(S_2, P_2)の収益性に対する効果を測定するには、$\gamma + \alpha_2 + \beta_2$を計算すればよい。

		幹事会社		
		S_1	S_2	S_2
年	P_1	γ	$\gamma + \alpha_2$	$\gamma + \alpha_3$
	P_2	$\gamma + \beta_2$	$\gamma + \alpha_2 + \beta_2$	$\gamma + \alpha_3 + \beta_2$

α_2とα_3は、それぞれS_1と比較したS_2とS_3の効果を測定するものであり、β_2についてもS_1と比較したS_2の効果を測定する。各パラメータの効果は、次のようにまとめることができる。

$$
\begin{aligned}
&\alpha_2 && \rightarrow S_2 - S_1 \\
&\alpha_3 && \rightarrow S_3 - S_1 \\
&\beta_2 && \rightarrow P_2 - P_1 \\
&\alpha_3 - \alpha_2 && \rightarrow S_3 - S_2
\end{aligned}
$$

表４は(25)式の回帰分析の結果である。この表から、上場年が1986年の場合には、1983年とを比較すると初期収益率が約１２７％ほど高いことを示している。また幹事会社が野村証券の場合、1983年に上場した株式は約１１０％であるが、大和証券が幹事会社になった株式は、同じ年で約９０％程度高くなっている。このようにダミー変数の解釈には十分な配慮が必要である。

（表４）統計パッケージのアウトプット

```
DEP VAR:    IR      N:      34
MULTIPLE R: 0.542
SQUARED MULTIPLE R: 0.294
ADJUSTED SQUARED MULTIPLE R:  0.224
STANDARD ERROR OF ESTIMATE: 107.063
```

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	110.060	31.975	0.000	.	3.442	0.002
YAMAICHI	37.670	45.654	0.147	0.739	0.825	0.416
DAIWA	87.802	46.421	0.334	0.754	1.891	0.068
YEAR	-126.618	38.546	-0.529	0.908	-3.285	0.003

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	143305.289	3	47768.430	4.167	0.014
RESIDUAL	343874.011	30	11462.467		

sample

sample

sample

不　許　複　製

慶應義塾大学ビジネス・スクール

Contents Works Inc.