主成分回帰に関する覚書

井　上　勝　雄

§0

単一構造方程式の回帰分析を行うとき，いくつかの説明変数の間に多重共線性が存在すると最小二乗推定値が不安定になる．関係式の計測のために，基本的には最小二乗推定法を用いる経済分析の場合，データはしばしば多重共線性をもつので，これは重大な問題を投げかけることになる．したがって，分析者は計測に際して様々な工夫をしなければならないが，その一つに，最小二乗推定法に替わる計測方法として主成分回帰推定は有効な方法と考えられる．

主成分回帰についてはケンドール〔6〕が提案し，その後，数は少ないものの文献〔8〕，〔9〕，〔2〕，〔7〕でとり挙げられている．筆者も〔5〕において，主成分回帰推定量について，リッジ推定量や最小二乗推定量と比較しつつ，その統計学的性質を考察し，〔4〕では，実験によって，リッジ推定や最小二乗推定との比較を試みた．また，〔3，3章〕においては，主成分回帰の採用可否を問うための統計的仮説検定の方式をも明らかにした．本稿は主成分回帰推定に関して幾何学的な説明を加えながら，その性質についての考察を若干補足的にとり扱うものである．

§1

1.1　最初にわれわれが考察の対象とする単一回帰モデルの定式化をしておこう．モデルはp個の説明変数によって被説明変数を説明するもので，標本数はT個あるとする．つまり，

(1.1)　$y=X\beta+u,\ u\sim N(0,\ \sigma^2 I_T)$

で表現できる標準的な回帰モデルである．(1.1) において，y は $T\times 1$ の被説明変数ベクトル，X は $T\times p$ の説明変数行列，β は $p\times 1$ の推定すべき回帰係数ベクトル，u は $T\times 1$ の攪乱項ベクトルである．また，確率変数ベクトル u は期待値が 0 で，互いに無相関，分散は一定 σ^2 であるとする．

われわれは，(1.1) の回帰係数 β を推定する際の説明変数間の多重共線性を問題にしている．したがって，考察にあたって，説明変数間の多重共線性の尺度を導入しなければならない．このためには，各変数が基準化されているのが適当である．つまり，p 個の説明変数の各々が基準化されていると，積和行列 $X'X$ は原データの相関係数行列となる．以下にみるように，この相関係数行列の固有値が多重共線性の尺度となる．

1.2　さて，以下に考察する主成分回帰のために，モデル (1.1) を，直交する説明変数への回帰モデルに変換する．

いま，説明変数間に完全な多重共線性がない場合，$X'X$ は正値定符号行列である．ゆえに，

(1.2)　$X'XP=P\Lambda$

となる正則行列 P が存在する．ここで P は，

(1.3)　$P'P=PP'=I_P$

を満たす直交行列であり，$p\times p$ 行列 Λ は $X'X$ の正の固有値を対角要素にもつ対角行列である．さらに，(1.2)，(1.3) より，

(1.4)　$P'X'XP=\Lambda$

が導け，P は相関係数行列 $X'X$ を対角化する行列であることがわかる．

さて，上で定義した P を用いて，回帰モデル (1.1) を次のように変換でき

る．

(1.5)　$y = XP\cdot P'\beta + u = Z\alpha + u$

ここで，$Z = XP,\ \ \alpha = P'\beta$

上に定義した変数変換 Z と (1.4) より，

(1.6)　$Z'Z = P'X'XP = \Lambda$

である．このことから，Z の各列は互いに直交する．つまり，(1.5) はモデル (1.1) を「被説明変数 y の，直交する説明変数 Z への回帰」に変換していることになる．あるいは，P の定義より明らかなように，(1.5) は「y の X の主成分 Z への回帰[1]」である．

他方，母回帰係数については，

(1.7)　$\beta = P\alpha$

が成立する．

1.3　次に，モデル (1.5) における α の最小二乗推定量 $\hat{\alpha}$ を導出しておこう．(1.5) の正規方程式

$$Z'Z\hat{\alpha} = Z'y$$

より，(1.6) を用いて，

(1.8)　$\hat{\alpha} = \Lambda^{-1}Z'y$

が得られる．また，(1.5)，(1.8) より，

$$\hat{\alpha} = \alpha + \Lambda^{-1}Zu$$

が導かれ，これから直ちに，最小二乗推定量 $\hat{\alpha}$ の性質として，

1)　主成分の意味については〔1，11ch〕，〔3，3章〕参照．

$$(1.9)\quad E(\hat{\alpha})=\alpha$$
$$V(\hat{\alpha})=\sigma^2\Lambda^{-1}$$

が明らかとなる．最小二乗推定量$\hat{\alpha}$は不偏推定量であるが，その分散は説明変数の相関係数行列$X'X$の固有値の大きさに反比例する．したがって，その固有値が0に近いとき，対応する推定量の分散は非常に大きくなる．これが，最小二乗推定値を不安定にする原因なのである．

われわれが本来分析の対象としているのは，モデル（1.1）である．したがって，次に（1.1）の回帰係数βの最小二乗推定量$\hat{\beta}$と上述の$\hat{\alpha}$との関係を示しておこう．周知のように（1.1）の最小二乗推定量$\hat{\beta}$は，

$$\hat{\beta}=(X'X)^{-1}X'y$$

である．（1.4）より，

$$X'X=P\Lambda P'$$

だから，これとZの定義より，

$$(1.10)\quad \hat{\beta}=(P\Lambda P')^{-1}X'y=P\Lambda^{-1}P'X'y=P\Lambda^{-1}(XP)'y$$
$$=P\Lambda^{-1}Z'y=P\hat{\alpha}$$

が導ける．つまり，最小二乗推定量$\hat{\alpha}$，$\hat{\beta}$の間には，母回帰係数α，βの関係（1.7）と同様の関係がある．さらに，

$$(1.11)\quad E(\hat{\beta})=P\alpha=\beta$$
$$V(\hat{\beta})=PV(\hat{\alpha})P'=\sigma^2P\Lambda^{-1}P'=\sigma^2(X'X)^{-1}$$

が（1.9），（1.10）より得られるが，これらも周知に属する．

1.4　上に見た（1.2）より明らかなように，Λは相関係数行列$X'X$の固有値$\lambda_i(i=1,2,\cdots\cdots,p)$を対角要素にする．また，$P$の各列は，それぞれの固有

値に対応する固有ベクトルである．ここで，固有値λ_1について，

(1.12)　$\lambda_1 \geqq \lambda_2 \geqq \cdots\cdots \geqq \lambda_p$

としても一般性を失うことはない．いま，最小固有値λ_pについて，

(1.13)　$\lambda_p \fallingdotseq 0$

と想定する．このとき，モデル（1.5）の係数α_pの最小二乗推定量$\hat{\alpha}_p$の分散は$\dfrac{\sigma^2}{\lambda_p}$であるから非常に大きい値になる．また，モデル（1.1）の最小二乗推定量$\hat{\beta}$は，（1.10）に示したように$\hat{\alpha}$の一次結合であるから，相対的に大きい分散を持つことになる．

そこで，モデル（1.1）の精度の良い推定量を求めるため，まず，

(1.14)　$\tilde{\alpha} = [I - J_p]\,\hat{\alpha}$

を定義し，モデル（1.1）の係数βの推定量$\hat{\beta}$として，

(1.15)　$\tilde{\beta} = P\tilde{\alpha} = P[I - J_p]\,\hat{\alpha}$

とする．上で，J_pは第pp要素だけが1で，残りはすべて0である行列であり，$\hat{\alpha}$は（1.8）に示した最小二乗推定量である．つまり，推定量$\tilde{\alpha}$は，分散の非常に大きくなる$\tilde{\alpha}_p$を0として，係数α_i（$i=1, 2, \cdots\cdots, p-1$）に関しては最小二乗推定量$\hat{\alpha}_i$を採用することになる．換言すれば，モデル（1.5）において，yの主成分Zへの回帰を行う際，p番目の主成分Z_pを欠落させて残りの（$p-1$）個の主成分にだけyを回帰させることになる．

1.5　p_pを$X'X$の固有値λ_pに対する固有ベクトルとして，したがって（1.2）で定義する行列Pの第p列とすると，

(1.16)　$Z_p{}'Z_p = p_p{}'X'Xp_p = \lambda_p p_p{}'p_p = \lambda_p$

が成立する．上式において，Z_pはZの第p列を表わしており，

(1.17)　$Z_p = Xp_p$

である．(1.13) で想定したように，$\lambda_p \fallingdotseq 0$ ならば，(1.16) よりわかるように第 p 主成分 Z_p の絶対値は 0 に近似でき，したがって，

$$Z_p = Xp_p \fallingdotseq 0$$

を意味している．つまり，説明変数 X の一次結合 Xp_p が 0 ベクトルに近似できるのである．換言するならば，$\lambda_p \fallingdotseq 0$ は説明変数間に多重共線性が存在することを表わしていると言えるのである．

§2

2.1　ここで，前節で考察した説明変数 X の一次変換 $Z_p = Xp_p$ を幾何学的な視点でみてみよう．

いま，われわれは T 個の標本を想定しているが，これは，p 次元の説明変数空間内の T 個の点で表わせる．そして，この空間に ($p-1$) 次元平面を想定し，この平面と T 個の標本点との距離の二乗和が最小なるように決定する．$Z_p = Xp_p$ はこのように決定した平面と T 個の点との（符号付）距離を表わし，また，これらの距離の二乗和が λ_p であることを以下に示そう．

まず，一般的に p 次元空間内の ($p-1$) 次元平面を，

(2.1)　$x'c = \Sigma_i x_i c_i = 0$

で表わし，係数 c は基準化されているとする．したがって，

(2.2)　$c'c = 1$

を仮定する．次に，説明変数の標本点 (x_{t1}，x_{t2}．……．x_{tp}) と平面 (2.1) との距離は，

$$(2.3)\quad \frac{|\Sigma_i x_{ti} c_i|}{\sqrt{\Sigma_i c_i^2}} = |\Sigma_i x_{ti} c_i| \qquad (\because \quad c'c = 1)$$

である．したがって，T個の標本点と平面との距離の二乗和は，

$$(2.4)\quad \Sigma_t \,|\, \Sigma_i x_{ti} c_i \,|^2 = c'X'Xc$$

となり，これを最小にするためには，

$$(2.5)\quad S = c'X'Xc - \mu(c'c - 1)$$

を最小化する数学的問題を解くことになる．(2.5) でμはスカラーのラグランジェ未定乗数である．Sを最小化するため，

$$\frac{\partial S}{\partial c} = 2X'Xc - 2\mu c = 0$$

$$\frac{\partial S}{\partial \mu} = c'c - 1 = 0$$

を導く．これより，

$$(2.6)\quad X'Xc = \mu c$$

$$(2.7)\quad c'c = 1$$

が得られ，(2.6) によって，求める平面の係数ベクトルcは$X'X$の固有ベクトルであり，μはそれに対応する固有値でなければならないことが分かる．しかも，(2.6)，(2.7) より，

$$(2.8)\quad c'X'Xc = \mu$$

が得られ，μは最小にすべきものとなるから，μは$X'X$の最小固有値でなければならない．結局，c，及びμは，前節におけるp_p，λ_pに等しい．

2.2　他方，説明変数間の多重共線性を，説明変数間の回帰分析によって考慮することも自然な考え方であろう[1]．

一般に，説明変数間において

$$(2.9)\quad x_1 = X_2 a + v$$

の関係が想定できる．上式において，

$$x_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \cdot \\ \cdot \\ x_{T1} \end{pmatrix} \quad X_2 = \begin{pmatrix} x_{12} \cdots x_{1p} \\ x_{22} \cdots x_{2p} \\ \cdot \cdots \cdot \\ \cdot \cdots \cdot \\ x_{T2} \quad x_{Tp} \end{pmatrix} \quad a = \begin{pmatrix} a_2 \\ a_3 \\ \cdot \\ a_p \end{pmatrix}$$

である．つまり，x_1は説明変数行列Xの第1列を，X_2はXの第2列から第p列までの部分行列を表わしており，aは定係数ベクトル，vは計算上必然的なT次元の残差ベクトルである．

(2.9）を第1説明変数の第2説明変数以下への回帰式と見なすならば，a，vは最小二乗法によって求められる．また，これは説明変数の間に一次の関係が規定でき，したがって，共線関係の可能性を表わしている．もし，ベクトルvがゼロベクトルならば，説明変数の間には完全な多重共線性が存在することになる．しかしながら，説明変数間に完全な共線関係があることは実際には起こり得ないであろうし，このような場合はむしろ重大な問題ではないであろう[2]．したがって（2.9）のvがゼロベクトルに近いことが計量経済学的分析のときに問題になるのである．

1）〔3，1章〕参照．

2）完全な多重共線性が存在する場合は，モデル（1.1）を分析することが不可能である．たとえば，回帰係数β_iは他の変数の値を一定にして，変数X_iを微少変化させたときの被説明変数の変化を表わすが，完全な共線性が有るときは，他の変数の値を一定に維持できないのである．

さて，(2.9) における定係数ベクトル a，残差ベクトル v を求めるため最小二乗法を適用すると，

$$(2.10)\quad a=(X_2'X_2)^{-1}X_2'x_1$$
$$v=x_1-X_2a=X\begin{pmatrix}1\\-a\end{pmatrix}=Xq \quad ここで，q=\begin{pmatrix}1\\-a\end{pmatrix}$$

が容易に得られる．周知のように (2.10) は (2.9) における残差 v の二乗和 $v'v$ を最小化することによって，つまり，X の一次結合 Xq について，

$$(2.11)\quad v'v=q'X'Xq$$

を最小化することによって得られる．

先述のように，前節の多重共線性の尺度については，X の一次結合 $Z_p=Xp_p$ の二乗和であった．以下では (2.10) における v と (1.17) における Z_p との違いについて若干補足しておこう．

まず，両者に関連して，p 次元の説明変数空間内に ($p-1$) 次元の平面を想定することができる．いずれもその背景に p 次元空間内の ($p-1$) 次元平面を考えることが出来るが，その平面と T 個の標本点との関係は異なる．2.1 では T 個の標本点と平面との距離の二乗和を最小になるよう平面を決定する．そして，その平面と標本点との距離の二乗和は $X'X$ の最小固有値 λ_p に等しく，また，これが共線性の尺度であった．(2.11) は周知の通り，T 個の標本点より x_1 軸に平行に沿って計る平面までの距離を考え，それらの二乗和を最小にしている．このように決定した平面と標本点との距離の二乗和を共線性の尺度にしているのである．

§3

3.1　モデル (1.1) と，あるいはモデル (1.5) に対して，(1.14)，(1.15) に

示される主成分回帰推定による推定係数の幾何学的意味について考える.[1]

モデルを推定するということは，幾何学的には，標本空間内に回帰平面を決定することである．そして，(1.14) に示したように，$\tilde{\alpha}_p=0$ とすることは主成分回帰によって決定すべき回帰平面が Z_p 軸となす角度を 0 とすることである．これは説明変数空間内の標本点が，Z_p 軸に沿っては広がりを持たないため，回帰平面を決定する際に，この情報を無視するのである．というのは，この情報は推定を不安定にするからである．これに対して，Z_i 軸 $(i=1,2,\cdots,p-1)$ については最小二乗推定によって，その軸と回帰平面とがなす角を決定する．ここに，最小二乗推定によって決定される係数 $\hat{\alpha}_i$ は回帰平面が Z_i 軸となす角度の正接（タンジェント）である．このように決定した回帰平面と，本来分析の対象となっている各々の X_i 軸 $(i=1,2,\cdots,p)$ とがなす角度の正接が主正成分回帰による推定係数 $\tilde{\beta}_i$ である．

3.2　以上の幾何学的解釈を一つの標本例で示してみよう．表 1 はこの節で例示しようとする標本データである.[2] われわれがここで扱うモデルの説明変数は 3 個である．表 1 における10個の標本からそれぞれのデータの平均，標準偏差を計算すると，

$$\bar{w}_1=781.63 \qquad \bar{w}_2=73.2 \qquad \bar{w}_3=8.02 \qquad \bar{g}=75.56$$
$$s_1=174.38 \qquad s_2=24.71 \qquad s_3=0.66 \qquad s_g=23.18$$

が得られる．これらをもとに，原データを基準化すると表 2 が得られる．次に，表 2 のデータより説明変数の積和行列，及び説明変数と被説明変数との積和ベクトルを，つまり，原データの相関係数は，

1）前節では，被説明変数を除ぞいた，p 個の説明変数だけで構成される p 次元空間内での考察であった．これに対して，この節では被説明変数も含めた $(p+1)$ 次元空間内での議論である．

2）表 1 の標本データは，

$$g=34.0+0.06\cdot W_1+0.5\cdot W_2-5.4\cdot W_3+u,\quad u\sim N(0,2.4^2)$$

を母集団の真の構造として生成したものである．

表 1

	W_1	W_2	W_3	g
1	470.47	24.33	7.84	30.01
2	836.89	93.10	7.43	93.14
3	786.88	68.99	8.13	72.59
4	900.31	77.64	8.75	83.37
5	893.65	90.52	8.55	86.53
6	987.37	105.39	7.86	108.51
7	470.15	46.40	6.44	49.20
8	762.15	59.45	8.44	66.12
9	970.98	103.72	8.00	103.26
10	737.43	62.66	8.75	62.84

表 2

	X_1	X_2	X_3	y
1	−1.78441	−1.97823	−0.26389	−1.96500
2	0.31692	0.80428	−0.88881	0.75835
3	0.03011	−0.17105	0.16209	−0.12800
4	0.68062	0.17883	1.09394	0.33706
5	0.64240	0.69996	0.79486	0.47346
6	1.17988	1.30193	−0.23711	1.42182
7	−1.78623	−1.08520	−2.36962	−1.13693
8	−0.11170	−0.55704	0.62782	−0.40712
9	1.08589	1.23403	−0.02357	1.19517
10	−0.25348	−0.42750	1.10429	−0.54880

$$X'X=\begin{pmatrix} 1.0 & 0.93362 & 0.50265 \\ & 1.0 & 0.19434 \\ & & 1.0 \end{pmatrix} \quad X'y=\begin{pmatrix} 0.94672 \\ 0.99319 \\ 0.20360 \end{pmatrix}$$

となる．更に，上の相関係数行列 $X'X$ の固有ベクトル，つまり§1における（1.2）の主成分変換行列 P の各列，

$$\begin{pmatrix} 0.67688 \\ 0.61760 \\ 0.40048 \end{pmatrix} \begin{pmatrix} 0.10517 \\ 0.45734 \\ -0.88304 \end{pmatrix} \begin{pmatrix} 0.72853 \\ -0.63984 \\ -0.24461 \end{pmatrix}$$

が計算でき，これらの固有ベクトルに対応する固有値

$$2.149 \qquad 0.839 \qquad 0.011$$

を得る．また，これらをもとに主成分変換データを表3に掲げている．

表3の主成分 Z_1，Z_2，Z_3 は互いに独立となり，これらの主成分と被説明変数 y との標本点を図示すると図1のように描ける．

次に，被説明変数の主成分への回帰モデルにおける最小二乗推定が，

表 3

	Z_1	Z_2	Z_3
1	−2.53529	−0.85938	0.03031
2	0.35529	1.18602	−0.06632
3	−0.02035	−0.21820	0.09173
4	1.00926	−0.81263	0.11384
5	1.18546	−0.31422	−0.17429
6	1.50776	0.92890	0.08455
7	−2.82830	1.40831	−0.02734
8	−0.16821	−0.82090	0.12147
9	1.48773	0.69940	0.00729
10	0.00665	−1.19732	−0.18125

$$\hat{\alpha}=\begin{pmatrix}0.62149\\0.44552\\0.39320\end{pmatrix}$$

と得られる．これらの係数は標本空間内に決定した回帰平面が主成分の軸 Z_1，Z_2，Z_3 となす角の正接を表わしている．さらに，(1.10) にしたがって，β に関する最小二乗推定は，

$$\hat{\beta}=\begin{pmatrix}0.75401\\0.33601\\-0.24070\end{pmatrix}$$

と得られる．係数 $\hat{\beta}$ の各々の要素は，回帰平面と説明変数の軸 X_1，X_2，X_3 となす角の正接である．

他方，α の推定係数 $\hat{\alpha}$ の分散は (1.9) に示したように，それぞれに対応する固有値の逆数に比例する．つまり，

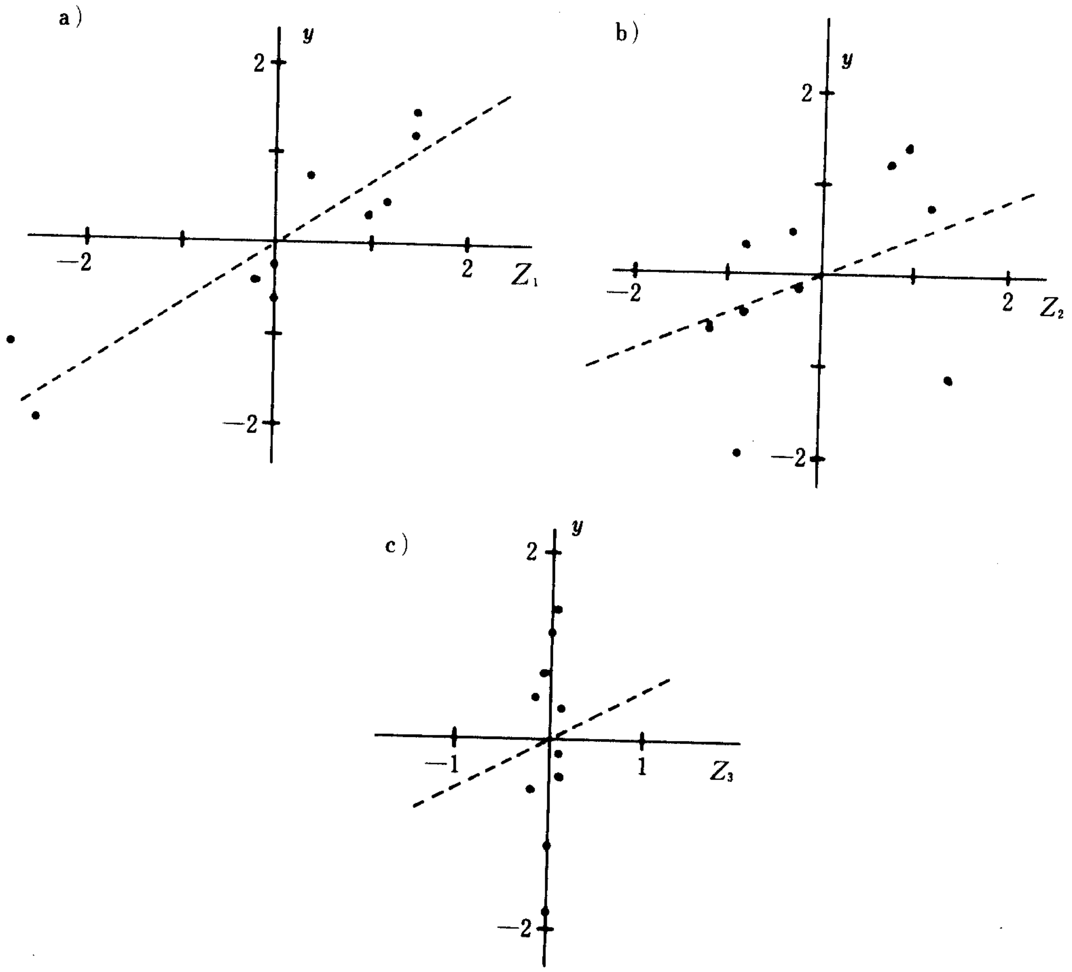
$$\frac{\sigma^2}{2.149}=0.465\cdot\sigma^2,\quad \frac{\sigma^2}{0.839}=1.191\cdot\sigma^2,\quad \frac{\sigma^2}{0.011}=90.9\cdot\sigma^2$$

である．上の結果でわかるように第３主成分に対する推定係数の分散は非常に大きくなる．このことは，図１（ｃ）を見ても明らかなように第３主成分のデータの広がりが小さく回帰平面の第３主成分の軸に対する傾きを決定するには情報不足であって，安定的な推定が得られないと考えられるのである．したがって，分散の大きい第３主成分への回帰推定係数を無視すると，

$$\tilde{\alpha}=[I-J_3]\,\hat{\alpha}=\begin{pmatrix}0.62149\\0.44553\\0.0\end{pmatrix}$$

となる．つまり，決定すべき回帰平面と第３主成分となす傾きをゼロとし，他

図１



の主成分となす角は最小二乗推定法によるのである．このように第3主成分の情報を無視して得た平面と本来の説明変数の軸とがなす角の正接は，

$$\tilde{\beta} = P\tilde{\alpha} = \begin{pmatrix} 0.467 \\ 0.587 \\ -0.144 \end{pmatrix}$$

となる．

3.3　さて，以上の議論の中でわれわれが扱う説明変数，及び被説明変数は基準化していることを前提していた．しかしながら，計量経済分析で定式化される関係式は必ずしも基準化された変数のそれではない．本節の以下で変数が基準化されない定式化と，変数が基準化された定式化との関係を補足しよう．

いま，基準化されない変数間の関係式が，

$$(3.1)\quad g = \delta_o e + W\delta + u$$

であるとしよう．(3.1) において，T 次元ベクトル g は被説明変数ベクトル，δ_o は定数，e はすべての要素が1である T 次元ベクトル，$T\times p$ 次元行列 W は説明変数行列である．

T 個の標本による各変数の標本平均を m_g，m_i $(i=1,2,\cdots,p)$ とし，標本によって計算される標準偏差を s_g，s_i $(i=1,2,\cdots,p)$ とする．そして，すべての要素が被説明変数の平均 m_g である T 次元ベクトルを $\bar{g}$ とし，第 i 列のすべての要素が第 i 説明変数の標本平均 m_i である $T\times p$ 次元行列を $\overline{W}$ とする．このとき，定数項に関して，

$$(3.2)\quad \delta_o e = \bar{g} - \overline{W}\delta$$

とすると，つまり，

$$(3.3)\quad \delta_o = m_g - \Sigma_i m_i \delta_i$$

と定義すると，(3.1) は，

(3.4)　$g-\bar{g}=(W-\overline{W})\delta+u$

に変形できる．さらに，第 ii 要素を第 i 説明変数の標準偏差 s_i である対角行列を S とすると，前節での考察に用いてきた基準化された被説明変数 y，及び基準化された説明変数 X との間には，

(3.5)　$y=\dfrac{1}{s_g}(g-\bar{g}),\quad X=(W-\overline{W})S^{-1}$

の関係がある．上の（3.4)，(3.5）に注意して，(3.1）と§1の（1.1）とを対比して

(3.6)　$\delta=s_g\cdot S^{-1}\beta$

が得られる．さらに，(3.1）の最小二乗推定量 $\hat{\delta}$ と（1.1）の最小二乗推定量 $\hat{\beta}$ との間にも（3.6）と同様の関係

(3.7)　$\hat{\delta}=s_g\cdot S^{-1}\hat{\beta}$

が証明できる．

本節3.2の標本例では，最小二乗推定 $\hat{\beta}$，主成分回帰推定 $\tilde{\beta}$ に対して，説明変数と被説明変数の標本標準偏差による修正をするならば，

$$\hat{\delta}=\begin{pmatrix}0.100\\0.315\\-8.39\end{pmatrix},\quad \tilde{\delta}=\begin{pmatrix}0.062\\0.551\\-5.04\end{pmatrix}$$

が得られる．

§4

4.1　この節で最小二乗推定量と比較しながら，主成分回帰推定量の特性を考察しておこう．

推定量$\tilde{\alpha}$に関してその定義（1.14）より直ちに，

$$
\begin{aligned}
&E(\tilde{\alpha}) = [I - J_p]\alpha = \alpha - J_p\alpha \\
(4.1)\quad &bias(\tilde{\alpha}) = -J_p\alpha \\
&V(\tilde{\alpha}) = \sigma^2[I - J_p]\Lambda^{-1}
\end{aligned}
$$

が導ける[1]. 上の(4.1)に示した主成分回帰推定量$\tilde{\alpha}$の性質は容易に理解できる. 推定量$\tilde{\alpha}_i$（$i=1, 2, \cdots, p-1$）は最小二乗定量と同等であるから，その期待値は母パラメタに等しく，分散についても，最小二乗推定量のそれに等しい. 第pパラメタに関する推定量$\tilde{\alpha}_p$は恒等的にゼロとするのであるから，当然そのバイアスは母数そのものであり，分散は0である.

さらに，モデル（1.1）の係数に関する主成分回帰推定（1.15）は，

$$
\begin{aligned}
&\mathrm{E}(\tilde{\beta}) = \beta - PJ_p\alpha = \beta - PJ_pP'\beta = \beta - p_pp_p'\beta \\
(4.2)\quad &bias(\tilde{\beta}) = -p_pp_p'\beta = -\alpha_pp_p \\
&V(\tilde{\beta}) = PV(\tilde{\alpha})P' = \sigma^2P[I - J_p]\Lambda^{-1}P'
\end{aligned}
$$

の性質を持つ. 推定量$\tilde{\beta}$は，推定量$\tilde{\alpha}$とは異なり，特定の$\tilde{\beta}_i$だけでなく，全べての推定係数にバイアスが生じる. 他方，全べての係数推定量の分散は最小二乗推定量の分散よりも小さくなる.

さて，説明変数間に多重共線性が存在するとき，つまり，説明変数間の積和行列の最小固有値が0に近いとき最小二乗推定が非常に大きい分散をもち，したがって，係数推定が不安定になることは§1において述べた. これに対して，主成分回帰は分散が小さく，安定的な推定が得られるが，他方，推定量にバイアスが生じる. そこで，不偏性を持たない推定量の精度基準として分散だけを挙げるのはよくない. 推定量の精度基準として平均二乗誤差基準を考えるべき

1）一般にp次元パラメタベクトルμの推定量$\hat{\mu}$について，

$$bias(\hat{\mu}) = E(\hat{\mu}) - \mu$$

である.

である．

4.2　一般に p 次元パラメタベクトル μ の推定量 $\hat{\mu}$ の平均二乗誤差行列は，

$$MSE(\hat{\mu}) = E(\hat{\mu}-\mu)(\hat{\mu}-\mu)'$$

で定義される．そして，簡単な計算によって，

$$\begin{aligned} MSE(\hat{\mu}) &= E(\hat{\mu}-E\hat{\mu}+E\hat{\mu}-\mu)(\hat{\mu}-E\hat{\mu}+E\hat{\mu}-\mu)' \\ &= E(\hat{\mu}-E\hat{\mu})(\hat{\mu}-E\hat{\mu})'+(E\hat{\mu}-\mu)(E\hat{\mu}-\mu)' \\ &= V(\hat{\mu})+bias(\hat{\mu})\,bias(\hat{\mu})' \end{aligned}$$

が導ける．最小二乗推定量は不偏推定量であるから，偏りは0であり平均二乗誤差行列は分散行列に等しい．したがって，(1.11) より，

$$(4.3)\quad MSE(\hat{\beta}) = V(\hat{\beta}) = \sigma^2 P\Lambda^{-1}P'$$

である．他方，主成分回帰推定量 $\tilde{\beta}$ の平均二乗誤差行列は (4.2) より，

$$(4.4)\quad MSE(\tilde{\beta}) = \sigma^2 P[I-J_p]\Lambda^{-1}P' + \alpha_p^2 p_p p_p'$$

となる．さらに，(4.3)，(4.4) より，

$$\begin{aligned} (4.5)\quad MSE(\hat{\beta})-MSE(\tilde{\beta}) &= \sigma^2 PJ_p\Lambda^{-1}P' - \alpha_p^2 p_p p_p' \\ &= \left(\frac{\sigma^2}{\lambda_p} - \alpha_p^2\right) p_p p_p' \end{aligned}$$

が得られる．

また，p 次元パラメタベクトル μ の第 i パラメタ μ_i に関する推定量 $\hat{\mu}_i$ の個別の平均二乗誤差は，

$$MSE(\hat{\mu}_i) = E(\hat{\mu}_i-\mu_i)^2$$

と定義される．そして，平均二乗誤差基準による推定量の良さは，推定量と母

パラメタとの距離の二乗を考え，その期待値が小さいことである．上の個別の平均二乗誤差は，先にみた推定量の平均二乗誤差行列の各対角要素を表わしているから，

$$MSE(\hat{\mu}_i)=V(\hat{\mu}_i)+bias(\hat{\mu}_i)^2$$

である．つまり，推定量の平均二乗誤差は推定量の分散とバイアスの二乗との和である．ここで，最小二乗推定量と主成分回帰推定量の平均二乗誤差を比較するため，(4.5) を参考に，次式を得る．

$$(4.6)\quad MSE(\hat{\beta}_i)-MSE(\tilde{\beta}_i)=\sigma^2\frac{p_{ip}^2}{\lambda_p}-\alpha_p^2 p_{ip}^2$$

$$=\left(\frac{\sigma^2}{\lambda_p}-\alpha_p^2\right)p_{ip}^2$$

次に，全平均二乗誤差は個別係数推定量の平均二乗誤差の和と定義される．つまり，推定量ベクトル $\hat{\mu}$ と母パラメタベクトル μ の距離の二乗を考え，その期待値である．つまり，

$$TMSE(\hat{\mu})=trMSE(\hat{\mu})=\Sigma MSE(\hat{\mu}_i)$$

である．全平均二乗誤差に関して，最小二乗推定量と主成分回帰推定量との差は，

$$(4.7)\quad TMSE(\hat{\beta})-TMSE(\tilde{\beta})=\left(\frac{\sigma^2}{\lambda_p}-\alpha_p^2\right)\Sigma p_{ip}^2$$

となる．

個別の平均二乗誤差基準においても，したがって全平均二乗誤差基準においても，主成分回帰推定量が最小二乗推定量よりも精度が良くなるのは，(4.6)，(4.7) より明らかなように，

$$\frac{\sigma^2}{\lambda_p}-\alpha_p^2>0$$

が満されたときである．上式を変形すると，

$$(4.8) \quad \lambda_p < \frac{\sigma^2}{\alpha_p^2}$$

とできる．説明変数間に多重共線性が存在し，したがって，説明変数間の積和行列の最小固有値 λ_p が 0 に近い場合，(4.8) の条件は決して厳しいものではない．つまり，平均二乗誤差基準に照らして，最小二乗推定より好ましい主成分回帰推定が採用可能となるのである.[1)]

4.3　上の議論は推定量の良し悪しを平均二乗誤差基準に求めた考察であった．以下では，最小二乗推定と主成分回帰推定のそれぞれの推定によって得られる計測結果の評価に係わる，決定係数の比較を考察しておこう．

決定係数は推定されたモデルがデータにどの程度適合しているかを示す尺度である．周知のことであるが，最小二乗推定の場合，説明変数によって説明されるモデルの体係的部分を，

$$\hat{y} = X\hat{\beta}$$

と定義する．そして，被説明変数の変動二乗和に対する $\hat{y}'\hat{y}$ の割合を最小二乗推定の決定係数 R_o^2 として，適合度の尺度としている．つまり，

$$R_o^2 = \frac{\hat{y}'\hat{y}}{y'y}$$

である．§1 で前提したように，被説明変数も基準化しているならば，

$$y'y = 1$$

であるから，(1.4)，(1.8)，(1.10) を参照することによって，

1）条件 (4.8) は未知パラメタを含むから，実際の分析の際に，この条件が成立するか否かは不明である．この条件の成立可否を問う統計的仮説検定方式については〔3，3 章〕を参照．

(4.9)　$R_o^2 = \hat{y}'\hat{y} = y'Z\Lambda^{-1}Z'y$

とできる．他方，主成分回帰推定による決定係数 R_p^2 については，まず，

$$\tilde{y} = X\tilde{\beta}$$

を主成分回帰推定による体系的部分と定義して，さらに，(1.4)，(1.8)，(1.14)，(1.15) を参照して，

(4.10)　$R_p^2 = \tilde{y}'\tilde{y} = y'Z\Lambda^{-1}[I - J_p]Z'y$

を主成分回帰推定の決定係数とできる．また，(4.9)，(4.10) より

(4.11)　$R_o^2 - R_p^2 = \hat{y}'\hat{y} - \tilde{y}'\tilde{y} = y'Z\Lambda^{-1}J_pZ'y$

$$= \frac{1}{\lambda_p}y'Z_p \cdot Z_p'y = y'\left(\frac{Z_p}{\sqrt{\lambda_p}}\right)\cdot\left(\frac{Z_p}{\sqrt{\lambda_p}}\right)'y$$

が得られる．ベクトル Z_p の絶対値の二乗は λ_p であるから，(4.11) 中の $\frac{Z_p}{\sqrt{\lambda_p}}$ は基準化された X の第 p 主成分である．そして，(4.11) に示した最小二乗推定と主成分回帰推定の決定係数の差は，結局，基準化された X の第 p 主成分と被説明変数との相関係数の二乗となることを意味している．

他方，説明変数間に多重共線性が存在するとき，§1 で考察したように，Z_p はゼロベクトルと近似でき，したがって，p 個の説明変数のもつ情報は第 p 主成分を除いても情報の損失はほとんど無いと考えてよい．換言すると，p 個の説明変数と被説明変数との相関の程度をみるとき，第1主成分から第 ($p-1$) 主成分までの ($p-1$) 個の主成分と被説明変数との相関係数の和で充分であると考えられるのである．というのは，第 p 主成分と被説明変数との相関は他の主成分の場合より低いと推測されるからである．したがって，最小二乗推定と主成分回帰推定の，二つの決定係数の差は微小であると考えてよい．結局，説明変数の間に多重共線性が存在するとき主成分回帰推定による推定を行って

も，最小二乗推定と比較して適合度の差はほとんど無いと考えられる.

4.4　前項までの考察から，多重共線性が存在するとき，主成分回帰推定は最小二乗推定より有効な推定方法と考えてよいが，その最大の難点は推定量のバイアスであろう．特に，説明変数が2個の場合，真の構造がどのようなものであれ特定方向のバイアスを持つ．このことを以下に示しておこう.

説明変数が2個の場合，その説明変数間の相関係数行列が如何なるものであれ，その固有ベクトルは,

$$p_1=\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \qquad p_2=\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

となることは容易に確かめられる．したがって，直ちに,

$$p_2 p_2'=\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

を導くことができ，これを（4.2）に代入して，主成分回帰推定量 $\tilde{\beta}$ の期待値

$$(4.12) \quad E(\tilde{\beta})=\beta-p_2 p_2'\beta=\begin{pmatrix} \frac{\beta_1+\beta_2}{2} \\ \frac{\beta_1+\beta_2}{2} \end{pmatrix}$$

が得られる．つまり，主成分回帰推定によって得られる二つの説明変数にかかる回帰係数は均等化する傾向を持つのである.

ただし，上の考察は基準化変数についての議論である．他方，（3.7）で明らかにしたように，基準化されない変数によって定義されるモデルの係数は基準化された変数モデルの推定係数に被説明変数の標準偏差を掛けて，更にそれぞれの説明変数の標準偏差で割る．したがって，主成分回帰推定による推定係数は,二つの説明変数の標準偏差の逆数に比例する傾向を持つといえるのである.

主成分回帰推定によってバイアスのある推定係数が得られるが，説明変数が3個以上の場合では説明変数間の相関係数行列に依存してそのバイアスの方向は一定ではない．しかし，説明変数が2個の場合，説明変数の相関係数行列の如何を問わず，特定方向のバイアスを持つのである．

参考文献

〔1〕 Anderson, T.W., *An Introduction to Multivariate Statistical Analysis* John Wiiey & Sons, 1958.

〔2〕 Farebrother, R.W., "Principal Component Estimators and Minimum Mean Square Error Criteria in Regression Analyss," *Review of Economics and Statistics*, vol. 54 (1972), pp. 32-36.

〔3〕 井上勝雄『計量経済学の理論と応用』有斐閣，1983年

〔4〕 ―――，「多重共線性と主成分回帰」，『経済学論究』36巻2号，1982年8月

〔5〕 ―――，「多重共線性とバイアス推定」，『経済学論究』35巻4号，1982年1月

〔6〕 Kendall, M.G., *A Course in Multivariate Analysis*, Charles Griffin and Co., 1957.

〔7〕 Lowerre, J.M., "On the Mean Square Error of Parameter Estimates for Some Biased Estimators," *Technometrics*, vol. 16 (1974), pp. 461-64.

〔8〕 Massy, W.F., "Principal Components Regression in Exploratory Statstical Research," *Journal of the American Statistcal Association*, vol. 56 (1965), pp. 234-56.

〔9〕 McCallum, B.T., "Artificial Orthogonalizaton in Regression Analysis," *Review of Economics and Statistics*, vol. 52 (1970), pp. 10-13.

〔10〕 Rao, C.R., *Linear Statistical Infrence and its Applpicatins*, second edition, John Wiley, 1973.

〔11〕 Silvey, S.D., "Multicollinearity and Imprecise Estimation," *Journal of the Royal Statstical Socety*, ser. B, vol. 32 (1969), pp. 539-52.