多重共線性と主成分回帰

井　上　勝　雄

§0

単一構造方程式の回帰分析を行なう際，説明変数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．これは最小二乗推定量の分散が多重共線性のために非常に大きくなるからである．このとき，最小二乗推定にかえて，より小さい平均二乗誤差をもつバイアス推定を採用することが可能である[1].

本稿は，説明変数間に多重共線性が存在するもとでの最小二乗推定と，これにかえて採用し得るバイアス推定とをシミュレーション実験によって比較することを主たる目的としている．§1及び§2で最小二乗推定と主成分回帰といえる上述のバイアス推定量の定式化を行う．§3は主成分回帰にある種の解釈を与えるため，主成分分析を考察する．§4はシミュレーション結果を報告することによって，最小二乗推定と主成分回帰の比較を行う．

§1

1.1　最初にわれわれがとり扱う単一回帰モデルの定式化をしておこう．

p 個の説明変数で被説変数を説明するモデルで，標本数は T 個あるとして，

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

で表現できる標準的なモデルである．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であり，u は $T\times 1$ の攪乱項ベクトルである．確率変数ベクトル u は，その期待値は0で相互に無

1)　参考文献〔2〕,〔5〕, 及び拙稿〔11〕参照.

相関，分散は一定σ^2であるとし，以下の分析では正規性の仮定は必らずしも要らないが，とりあえず正規分布するものと想定する．

われわれは本稿を通じて，説明変数間の多重共線性の問題を考察している．その際多重共線性の程度を測定するのに説明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について，基準化していると想定すると，積和行列$X'X$は原データの相関係数行列となる．

1.2　さて，以下の分析のために，考察の対象になっている（1.1）を，直交する説明変数への回帰分析に変換することを考える．

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）　$P'P=PP'=I_p$

を満たす直交行列とでき，$p\times p$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．いま，（1.2），（1.3）より

（1.4）　$P'X'XP=\Lambda$

が導ける．

上で定義したPを用いて，回帰モデル（1.1）を以下のように変形できる．

（1.5）　$y=XP\cdot P'\beta+u=Z\alpha+u$

　　　　ここで　$Z=XP,\ \ \alpha=P'\beta$

（1.5）で定義されたZについて

（1.4）　$Z'Z=P'X'XP=\Lambda$

であるからZの各列は互いに直交する．つまり，回帰モデル（1.1）を「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．

回帰係数については，

（1.6）　$\beta=P\alpha$

の関係が成り立つ．

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

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

より，（1.4）を用いて

$$\hat{\alpha}=\Lambda^{-1}Z'y \tag{1.7}$$

が得られる．

さて，（1.5），（1.7）より

$$\hat{\alpha}=\alpha+\Lambda^{-1}Z'u$$

が導かれこれから，直ちに，

$$\text{(1.8)}\quad E(\hat{\alpha})=\alpha$$

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

が明らかである．推定量 $\hat{\alpha}$ は不偏推定量であるが，その分散は $X'X$ の固有値の大きさに反比例し，したがって，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いものとなる．この場合、推定値は非常に精度が落ちることになる．

さて，モデル（1.1）で直接 β の最小二乗推定量を導出し，変形を行なうと次の様になる．

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

ゆえに

$$\hat{\beta}=P\hat{\alpha} \tag{1.9}$$

となり，最小二乗推定量 $\hat{\alpha}$，$\hat{\beta}$ の間には，母パラメタ α，β の関係（1.6）と同様の関係が成立する．さらに，（1.8），（1.9）より

$$\text{(1.10)}\quad E(\hat{\beta})=P\alpha=\beta$$

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

が導ける．いま，係数ベクトル β の第 i 要素の最小二乗推定量 $\hat{\beta}_i$ の精度をみるため，（1.10）より $\hat{\beta}_i$ の分散について，

$$var(\hat{\beta}_i)=\sigma^2\Sigma\frac{P_{ij}^2}{\lambda j}$$

が導ける．$\hat{\beta}_i$ の分散は $X'X$ の固有値と固有ベクトルの各素に依存する．分散を推定量の精度の尺度とする場合，0に近い固有値が存在するならば相対的に推定の精度は落ちることになる．いま $\lambda_k \doteqdot 0$ とすると，それに対応する固有ベクトルは正規性を保っているから $\Sigma P_{ik}^2 = 1$ である．かりにその固有ベクトルの大部分の要素が小さい値をとるとしても，特定の i に対して $\dfrac{P_{ik}^2}{\lambda_k}$ の値が非常に大きな値をとることになり，このとき $\hat{\beta}_i$ の精度は悪くなる．

§2

2.1　多重共線性によって，モデル（1.5）の α の最小二乗推定値 $\hat{\alpha}$ が不安定であったり，先述のように推定量の精度が悪くなるのは一つに，推定量の分散が大きくなるからである．このことを以下の考察のために明示的に定式化しておこう。

いま，$X'X$ の固有値について

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

とし，説明変数間の多重共線性によって，

（2.2）　$\lambda_p \doteqdot 0$

と想定する．換言すれば，固有値の対角行列 Λ の要素について（2.1），（2.2）を前提とする．このとき，§1の考察から明らかなように，α_p の最小二乗推定量 $\hat{\alpha}_p$ の分散は $\dfrac{\sigma^2}{\lambda_p}$ となり，（2.2）より，この分散は非常に大きい値をとることになる．また，モデル（1.1）の最小二乗推定量 $\hat{\beta}$ は，（1.9）より $\hat{\alpha}$ の一次結合であるから，個々の最小二乗推定量 $\hat{\beta}_i$ のいずれかは，相対的に大きい分散をもつことになる．

いま，α の推定量 $\tilde{\alpha}$ を次の様に定義する．

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

上式で 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 α_p を0として，$\alpha_i (i=1, 2, \cdots p-1)$ は最小二乗推定を採用することである．換言すれば，

モデル(1.5)において y の Z への回帰を行なう場合，p 番目の説明変数 Z_p を欠落させて残りの $p-1$ 個の説明変数にだけ y を回帰させることに等しい．あるいは，p_p を $X'X$ の固有値 λ_p に対する固有ベクトル，つまり（1.2）で定義する行列 P の第 p 列とすると，

(2.4)　$Z_p'Z_p=p_p'X'Xp_p=\lambda_p p_p'p_p=\lambda_p$

が成立するが，上式において，$\lambda_p \fallingdotseq 0$ より Z_p を 0 ベクトルと近似することと同等である．さらに，(1.1) のモデルにおける β の推定量 $\tilde{\beta}$

(2.5)　$\tilde{\beta}=P\tilde{\alpha}$

と定義する．

2.2　この節で，先に定義した推定量 $\tilde{\alpha}$ の特性を考察しておこう．

推定量 $\tilde{\alpha}$ の定義（2.3）より

(2.6)　$E(\tilde{\alpha})=[I-J_p]\alpha=\alpha-J_p\alpha$

$bias(\tilde{\alpha})=-J_p\cdot\alpha$

が直ちに得られる．これより推定量 $\tilde{\alpha}_p$ のみに偏りがある．また分散については

(2.7)　$V(\tilde{\alpha})=\sigma^2[I-J_p]\Lambda^{-1}$

が容易に導ける．(2.6) および (2.7) の $\tilde{\alpha}$ についての性質は，その定義から簡単にわかることである．バイアス推定量 $\tilde{\alpha}_i$ は，$i=1,2,\cdots p-1$ については最小二乗推定量と同等であるからその期待値は母パラメタの値であり，分散についても最小二乗推定量の分散に等ししい．α_p については $\tilde{\alpha}_p=0$ と恒等的に定義するのであるから，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，これらの考察から，推定量 $\tilde{\alpha}_i$ の平均二乗誤差行列は，

(2.8)　$MSE(\tilde{\alpha})=\sigma^2[I-J_p]\Lambda^{-1}+J_p\alpha\alpha'J_p$

である．(2.8) を個々の推定量 $\tilde{\alpha}_i$ についての平均二乗誤差で表現すると

$$MSE(\tilde{\alpha}_i)=\begin{cases} V(\tilde{\alpha}_i)=\sigma^2\dfrac{1}{\lambda_i} & (i=1,2,\cdots,p-1) \\ bias(\tilde{\alpha}_p)=\alpha_p^2 & (i=p) \end{cases}$$

である．

次に，モデル（1.1）のβについてのバイアス推定量$\tilde{\beta}$について（2.5）を参考にして

（2.9）　$E(\tilde{\beta})=PE(\tilde{\alpha})=P[I-J_p]\alpha=P\alpha-PJ_p\alpha$

$$=\beta-PJ_pP'\beta=\beta-p_pp_p'\beta$$

$$bias(\tilde{\beta})=-PJ_pP'\beta=-p_pp_p'\beta$$

となり，推定量$\tilde{\beta}$のバイアスは特定の$\tilde{\beta}_i$だけでなく，一般的にはすべてのパラメタ推定量にバイアスが生じる．また，$\tilde{\beta}$の分散については，

（2.10）　$V(\tilde{\beta})=PV(\tilde{\alpha})P'=\sigma^2P[I-J_p]\Lambda^{-1}P'=\sigma^2[(X'X)^{-1}-\frac{1}{\lambda_p}P_pP_p']$

が得られる．

$\tilde{\alpha}$，及び$\tilde{\beta}$の全平均二乗誤差については，（2.8），（2.9），（2.10）より，

（2.11）　$TMSE(\tilde{\alpha})=TMSE(\tilde{\beta})=\sigma^2\sum^{p-1}\frac{1}{\lambda_i}+\alpha_p^2$

を導くことができる．

2.3　この節で最小二乗推定量$\hat{\alpha}$，本節のバイアス推定量$\tilde{\alpha}$の比較を考察する．この§の初めに述べたように，われわれは説明変数間に完全な多重共線がないにしてもかなりそれに近い状況を想定している．したがって（2.2）に示したように$X'X$の最小固有値が0に近い値をとる状況を前提している．したがって，さらにこの節では$\frac{\sigma^2}{\alpha_p^2}>\lambda_p\fallingdotseq 0$を前提にする．つまり，

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

を前提に考察をすすめる．

（2.12）から直ちに，（2.11）を参考に，

（2.13）　$TMSE(\tilde{\alpha})<TMSE(\hat{\alpha})$

が示せる．われわれは多重共線の状況を前提に本節では（2.2）を仮定しているが，このことから当然（2.13）の特性は容易に理解できることである．

§3

3.1　前節までは，モデル（1.1）をモデル（1.5）に変形し，つまり，変量yの経済的に意味のある説明変数Xへの回帰モデルを，Xの一次結合$Z=XP$へ

の回帰に変換し，議論をすすめてきた．最小二乗法による推定だけを考えるならその変換は必らずしも必要ではないが，§2で定義する推定法は，本来の説明変数 X の一次変換 Z への回帰を前提にして，導出されたものである．

p 個の変量 X の（1.2)，(1.3）を満たす P による一次結合 $Z=XP$ は，主成分分析における，変量 X の主成分と定義されるものである．この節で，主成分分析の二つの定義を考察する．その一つは Hotelling の定義であり[1]，もう一つは Anderson による定義[2]である．

3.2　p 個の変量 $x_i(i=1,2,\cdots,p)$ はある試行の結果得られる観測値であるとする．そうして，各変量は平均が0で，分散が1となるよう規準化されているとする．通常，各変量 x_i は互いに相関している．そして，各変量が相互に相関しているのならば，これら変量 x_i のとる値を決定する，より基本的な独立変数 $z_j(j=1,2,\cdots)$ が存在すると考えられる．これらの変数 z_j を試行によって記述される複合体の成分という．成分 z は，観測される変量 x よりも本質的なものであるから，その個数は x の個数より少く，最大限 x の個数 p と等しいとする．

さて，変量 x_i は成分 z_j の関数であると考えられるから，その関数を線型関数と想定するならば，

(3.1)　$x'=z'B$

と表現できる．ここで，

$$x'=(x_1x_2\cdots x_p)$$

$$z'=(z_1z_2\cdots z_l)$$

$$B=\begin{pmatrix} b_{11} & b_{12}\cdots & b_{1p} \\ b_{21} & b_{22} & b_{2p} \\ \vdots & \vdots & \vdots \end{pmatrix}$$

である．

1)　文献〔4〕参照.
2)　文献〔1〕ch. 11参照.

また，xは規準化されているので

$$(3.2)\begin{cases} E(x)=0 \\ V(x)=E(xx')=R \end{cases}$$

であって R は x の相関係数行列である．次に成分 z は互いに独立で，変量 x と同様に規準化されていると，

$$(3.3)\begin{cases} E(z)=0 \\ V(z)=E(zz')=R \end{cases}$$

である．

観測値 x を成分 z で表現しようとすることを考えているのであるから，その重要度に従って，あるいはその存在が明確であるものから順に成分を抽出する方法が自然であると考えられる．主成分分析というのは，成分 z_1 が各変量 $x_i(i=1,2,\cdots p)$ の分散に寄与する部分を最大にするように決定し，これを第1主成分とする．各 x の分散のうち，z_1 が寄与した部分の残りを考え，それに対する寄与分を最大にする成分を第2主成分 z_2 とする．以下順に第3主成分以下を決定していく．

3.3　さて，第1主成分 z_1 が変量 x の分散に寄与する大きさを評価しよう．

(3.1) より，

$$x_i=\Sigma z_k b_{ki}$$

であり，成分 z については (3.3) を想定しているから

$$V(x_i)=\Sigma V(z_k)b^2_{ki}=\Sigma b^2_{ki}$$

である．したがって，第1主成分 z_1 の変量 x_i の分散への寄与分は b_{1i}^2 であり，z_1 の各 x の分散への寄与分全体は

$$\Sigma b_{1i}^2$$

である．いま，係数行列 B の第一行を $b_{1\cdot}'$ と表わすと，これは

(3.4)　$b_{1\cdot}' b_{1\cdot}$

である.

他方,(3.1)を(3.2)に代入して,(3.3)を用いると,

(3.5)　$R=V(x)=E(B'zz'B)=B'B$

となる. 上述の主成分分析の考え方を定式化するならば, 結局,(3.5)を制約条件として(3.4)を最大にするよう B を決定するという数学問題になる.

3.4　制約条件である(3.5)において, R も $B'B$ も対称行列であるから, 対称行列 M を未定乗数行列として, ラグランジェ関数

(3.6)　$S=b_{1\cdot}'b_{1\cdot}-t_r M(B'B-R)$

を最大化する問題の解が主成分分析の解である.

(3.6)を B, M で偏微分し, それを 0 とおくと

$$\frac{\partial S}{\partial B}=2\begin{bmatrix} b_{1\cdot}' \\ \cdots\cdots \\ 0 \end{bmatrix}-2BM=0$$

$$\frac{\partial S}{\partial \mathbf{M}}=B'B-R=0$$

が得られ,

(3.7)　$\begin{bmatrix} b_{1\cdot}' \\ 0 \end{bmatrix}=BM$

(3.8)　$B'B=R$

が導出できる.(3.7)の辺々前から B' をかけ,(3.8)を代入すると,

$$B\begin{bmatrix} b_{1\cdot}' \\ 0 \end{bmatrix}=B'BM=RM$$

上式辺々, 後から $b_{1\cdot}$ をかけると

$$B'\begin{bmatrix} b_{1\cdot}'b_{1\cdot} \\ 0 \end{bmatrix}=RMb_{1\cdot}$$

ここで,

(3.9)　$s_1=b_{1\cdot}'b_{1\cdot}$

とおくと,

(3.10)　$b_{1\cdot}s_1=RMb_{1\cdot}$

が得られる. さて,(3.7)の辺々第1行に注目すると

$$b_{1\cdot}'=b_{1\cdot}'M$$

であり，これより

$$b_{1\cdot}=M'b_{1\cdot}=Mb_{1\cdot}$$

が導け，これを（3.10）に代入すると

（3.11）　$b_{1\cdot}\cdot s_1=Rb_{1\cdot}$

が得られる．（3.11）より

$$(R-s_1I)b_{1\cdot}=0$$

であるから，s_1 は R の固有値で，$b_{1\cdot}$ はそれに対応する固有ベクトルでなければならない．他方，(3.9) からわかるように s_1 は最大にすべきものであるから，結局，s_1 は R の最大固有値で，$b_{1\cdot}$ は最大固有値に対応する固有ベクトルである．

次に，第2主成分 z_2 の各 x の分散への寄与分全体は，係数行列 B の第2行を $b_{2\cdot}'$ で表わすと，

$$b_{2\cdot}'b_{2\cdot}$$

となる．制約条件（3.5）のもとに上式を最大化する問題を考えれば前述の展開と同様に

$$(R-s_2I)b_{2\cdot}=0$$

$$s_2=b_{2\cdot}'b_{2\cdot}$$

が得られる．したがって先に求めた s_1 が R の最大固有値に対して，s_2 は2番目に大きい R の固有値であり，$b_{2\cdot}$ がその s_2 に対応する固有ベクトルである．

3.5　いま，x と同じ数だけの p 個の主成分を考えれば，結局，B の各行は，変量 x の相関係数行列 R の p 個の固有値に対する固有ベクトルである．そうして，それらの固有ベクトルに関して，各要素の2乗和は対応する固有値に等しくなければならない．

いま，対称行列 R を対角化する直交行列を P，R の固有値を対角要素にもつ正方行列を Λ とすると，

（3.11）　$RP=P\Lambda,\ PP'=P'P=I$

である．上述の係数 B は，

(3.12)　$B=\Lambda^{\frac{1}{2}}P'$

である．ゆえに，変量 x' は主成分 z によって

(3.13)　$x'=z'B=z'\Lambda^{\frac{1}{2}}P'$

で表現出来る．逆に，主成分は，変量 x' の一次結合で表現でき，

(3.14)　$z'=x'P\Lambda^{-\frac{1}{2}}$

である．

3.6　次に，上述の主成分の定義と異なる主成分分析を考察する．

p 個の変量 x_i の観測値がえられたとき，p 個の変量各々の変動を考察するより，それらから一つの変数を作り，その一つの変数に p 個の変量のもつ情報が充分組み込まれているなら，その一変数を考察する方が見易すいだろう．あるいは，実証研究で，変数の数を出来るだけ減らしたとしても，もとの変量のもつ情報の損失が少なければ，変数の数を減少させて考察することで充分であるといえる．したがって，p 個の変量を考察するより，それらを一つの代表的な変量に集約できるならば，それを総合特性値といってよい．

上述の考えを定式化すると，p 個の変量 $x'=(x_1x_2\cdots x_p)$ の一次結合を

(3.15)　$v_1=x'a$

と想定し，変量 v_1 の分散を最大にするよう a' を決める．ただし，標準化された一次結合（normalized linear combination）を考え，係数 a について

(3.16)　$a'a=1$

を制約条件とする．

(3.2) に注意して，v_1 の分散は

(3.17)　$V(v_1)=E(v_1^2)=E(a'xx'a)=a'Ra$

であるから，

(3.18)　$T=a'Ra-\lambda(a'a-1)$

を最大にするよう a を決定する数学的問題になる．ここで λ はラグランジェの未定乗数である．

Tを最大化するため

$$\frac{\partial T}{\partial a}=2Ra-2\lambda a=0$$

$$\frac{\partial T}{\partial \lambda}=a'a-1=0$$

が導かれ，結局

(3.19)　$Ra=\lambda a$

(3.20)　$a'a=1$

が得られる．(3.19)，(3.20) より

$$a'Ra=\lambda a'a=\lambda$$

となる．これは最大にすべき v_1 の分散である．さて，求めるべき係数 a は，(3.19) よりわかるように変量 x の相関係数行列 R の固有ベクトルでなければならない．しかも，固有値 λ が最大にすべきものであるから R の最大固有値に対応する固有ベクトルでなければならない．ここで，この様に決定された変量 v_1 を第１主成分という．

さて，変量 x の各分散の総和は，

$$t_r R=p$$

であり，上述のように第１主成分 v_1 の分散は R の最大固有値 λ であるから，v_1 の x の分散に対する寄与率は $\frac{\lambda}{p}$ と定義される．つまり，これは p 個の変量 x のもつ情報のうち，主成分 v_1 に組み込まれる割合である．第１主成分の寄与率 $\frac{\lambda}{p}$ が充分な大きさでなければ，主成分分析では第２主成分を求める．第２主成分 v_2 は第１主成分 v_1 と同様に標準化された x の一次結合であるが，第１主成分 v_1 と統計的に独立なものを選ぶ．第１主成分を求めたとほぼ同様の仕方で，第２主成分の係数は R の２番目の固有値に対する固有ベクトルであることが証明できる．このようにして，変量 x と同じ数だけの主成分を $v'=(v_1 v_2 \cdots v_p)$ とすると，

$$v'=x'A$$

とでき，係数行列 A は，結局，(3.11) の P に同等であることが明らかになる．つまり，主成分 v は x の一次結合であって，

(3.21)　$v'=x'P$

である[1]．

3.7　前節までに主成分分析の二つの定式化を考察してきた．3.2 でその考え方をみ，その定式化を 3.3 より 3.5 に示した Hotelling による主成分 z と，3.6 で考察した Anderson による主成分 v は，それぞれ (3.14)，(3.21) に示したように，

$$z=x'P\Lambda^{-\frac{1}{2}},\quad v=x'P$$

である[1]．主成分は試行結果の観測値 x の一次結合で表現でき，その係数は x の相関係数行列の固有ベクトルに対応しているという点では二つの定義は数学的に同等となる[2]．

さて，われわれの問題意識は，回帰モデル (1.1) の分析であって，その際説明変数 X の多重共線性が回帰係数の推定に及ぼす影響であった．さて，ここで§2において考察したバイアス推定量に一つの解釈を与えておきたい．主成分分析の定式化を応用すると，p 個の説明変数 X が互いに相関しているならば，これらの説明変量のとる値を決定するようなより基本的な独立変数が存在すると考えられる．それが説明変数 X の主成分であり，モデル (1.1) の回帰

---

1) Girshick〔3〕は主成分の定義を以下のようにしている．
主成分 w は観測値 x' の一次結合で，基準化されているとする．したがって

$$V(w)=V(a'x)=E(a'xx'a)=a'Ra=1$$

主成分 w と x' の共分散の 2 乗和が最大なるよう w は決定される．つまり，最大にするものは，w と x' の共分散

$$E(wx')=E(a'xx')=a'R$$

より，$a'R(a'R)'=a'R^2a$ である．
結局，

$$Z=a'R^2a-\mu(a'Ra-1)$$

を最大にする数学問題に帰着する．

2) Pearson〔10〕によれば，幾何学的に主成分を求めることができる．p 次元空間内の観測値 x を $q(\leqq p)$ 次元空間に射影し，その距離の 2 乗和を最大にするような q 次元空間を決定する．求められた q 次元空間の座標軸が主成分になる．

係数の最小二乗推定量はすべての主成分に回帰させることによって得られる．これに対して，§2のバイアス推定量は，寄与率の最も低い主成分を無視して，他の寄与率の大きい主成分にのみ回帰させて，係数の推定を行なうことである．

§2で考察したバイアス推定は，説明変数の主成分を明示的に考慮することから主成分回帰といえる．

§4

さて，この節では最小二乗推定量と主成分回帰の推定量との比較をシミュレーション実験によって比較する．

多重共線性が存在し，厳密にいえば（2.12）が成立するならば，主成分回帰による方が最小二乗推定法によるよりも平均二乗誤差が小さい推定量が得られる．つまり，（3.12）のもとでは主成分回帰による方がより精度のよい計測が得られる．しかし，（3.12）は，母パラメタである α_p や σ^2 の関係であって，われわれが具体的に計測作業を行なうときには（2.12）の関係が成立しているのかどうかということは不明のままである．

以下のシミュレーションで，説明変数間の相関係数行列の固有値の大きさによって，主成分回帰による推定値のバイアスがどの程度であるかを調べたい．さらに，具体的な計測作業の際に，主成分回帰を用いる場合の採用基準を求めたい．もちろん，シミュレーションによっては普遍的な結論は導けない．しかし具体的な計量経済学的実証作業を行なう際の一つの指針を与えることにはなるだろうと思われる．

4.2　われわれのシミュレーションは次の手順に従って行なわれた．

(i)　各説明変数 X にかかる回帰係数 β と，その期待値が0である攪乱項 u の分散 σ^2 によって，被説明変数 y と説明変数間の真の回帰式を規定する．この条件のもとに被説明変数を生成する．

(ii)　上述のことに先立って，説明変数 X の構造を，各変数の平均 $E(X)$，標準偏差 $std.(X)$，及び変数間の相関係数行列 R によって規定する．こ

の条件のもとに説明変数 X を乱数発生する．

(iii) 標本（y, X）によって最小二乗推定と主成分回帰推定値を導出する．標本数は50又は30とした．

以上の手順に従うシミュレーションを100回行なった．

4.3　まず，最初のシミュレーションの真の回帰式は，

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+u$$

34.0　0.06　0.5　−5.4

であり，各係数の値は上の通りである．攪乱項 u は，

$$u\sim N(0,\ 4.3^2)$$

を満たしている．説明変数 X_1，X_2，X_3，の期待値 $E(X)$，標準偏差 $std.(X)$ は

	X_1	X_2	X_3
$E(X)$	850.0	84.0	7.8
$std.(X)$	190.0	24.0	0.7

であり，説明変数間の相関係数行列 R は

$$R=\begin{pmatrix}1.0 & 0.95 & 0.6\\ & 1.0 & 0.4\\ & & 1.0\end{pmatrix}$$

と設定された．

標本数は50として，100回のシミュレーションの結果，説明変数の相関係数行列の固有値について，

	第1固有値	第2固有値	第3固有値
平均：	2.336	0.641	0.022478
最大：	2.646	0.966	0.036692
最小：	2.004	0.333	0.010432

であった．最小固有値は0.01～0.03であるから充分小さいと考えられる．表1.a.1に，最小二乗推定（OLS）と主成分回帰推定（LAT）による各係数推定

値の分類をしている．表の第6行は係数の真の値を含む区間である．たとえば，β_1 については OLS では100回のうち26回，LAT では45回が0.055～0.065の区間に係数推定値があったことを示している．したがって第1行から第5行までは過少推定の，又第7行から第11行までは過大推定の回数を表わしている．表からわかることは，OLS の推定係数はかなり変化する．他方 LAT の推定係数は β_1 については下への，β_2 については上へのバイアスがあるものの，安定的な推定が得られている．β_3 については LAT の推定値も変動するが，これは説明変数 X_3 の標準偏差が小さいことに，その原因がある．

表1.a.1

	β_1	β_2	β_3	β_1		β_2		β_3	
				OLS	LAT	OLS	LAT	OLS	LAT
(1)	～0.015	～0.05	～−9.9	0	0	0	0	0	0
(2)	～0.025	～0.15	～−8.9	2	0	1	0	1	0
(3)	～0.035	～0.25	～−7.9	7	0	0	0	6	0
(4)	～0.045	～0.35	～−6.9	12	2	8	0	16	3
(5)	～0.055	～0.45	～−5.9	12	53	30	0	21	17
(6)	0.055～0.065	0.45～0.55	−5.9～−4.9	26	45	33	67	21	37
(7)	0.065～	0.55～	−4.9～	18	0	18	32	18	28
(8)	0.075～	0.65～	−3.9～	16	0	7	1	9	10
(9)	0.085～	0.75～	−2.9～	6	0	3	0	5	5
(10)	0.095～	0.85～	−1.9～	0	0	0	0	2	0
(11)	0.105～	0.95～	−0.9～	1	0	0	0	1	0

表1.a.2

	OLS			LAT		
	mean	max.	min.	mean	max.	min.
β_1	0.06128	0.11413	0.01936	0.05417	0.06173	0.04173
β_2	0.49339	0.78928	0.14761	0.54295	0.65335	0.47592
β_3	−5.54474	−0.0934	−9.2955	−5.06607	−2.5373	−7.6568
const.	34.61	60.94	4.63	32.75	54.26	11.92
R^2	0.9595	0.9817	0.9107	0.9584	0.9807	0.9105
σ	4.265	5.43	3.29	4.323	5.50	3.31
D.W.	1.983	2.707	1.333	1.988	2.685	1.359

表1.a.2は，限界係数の他，定数項推定値（const.），決定係数（R^2），攪乱項の標準偏差（σ），ダービンワトソン比（D.W.）の100回シミュレーションの平均及びそれぞれの最大値，最小値を挙げている．

表1．a．2によれば，当然のことであるが，主成分回帰による場合OLS推定より決定係数 R^2 は微小であるが落ちる．定数項推定や攪乱項の標準偏差 σ，ダービンワトソン比D.W.は大きな違いはないように見られる．先の表1．a．1でもわかるように，限界係数についてはLAT推定はバイアスがあるものの係数推定値の変動幅からみてOLS推定に比してかなり安定した推定が得られ

表1．b．1

	β_1	β_2	β_3	β_1		β_2		β_3	
				OLS	LAT	OLS	LAT	OLS	LAT
(1)	～0.015	～0.05	～−9.9	3	0	0	0	1	1
(2)	～0.025	～0.15	～−8.9	2	0	0	0	3	0
(3)	～0.035	～0.25	～−7.9	5	0	5	0	7	3
(4)	～0.045	～0.35	～−6.9	11	3	10	0	18	9
(5)	～0.055	～0.45	～−5.9	13	52	28	0	22	19
(6)	0.055～0.065	0.45～0.55	−5.9～−4.9	18	45	26	58	17	29
(7)	0.065～	0.55～	−4.9～	25	0	16	42	15	19
(8)	0.075～	0.65～	−3.9～	8	0	11	0	10	12
(9)	0.085～	0.75～	−2.9～	10	0	2	0	1	5
(10)	0.095～	0.85～	−1.9～	2	0	1	0	2	2
(11)	0.105～	0.95～	−0.9～	3	0	1	0	4	1

表1．b．2

	OLS			LAT		
	mean	max.	min.	mean	max.	min.
β_1	0.06184	0.11078	0.00558	0.05437	0.0615	0.0380
β_2	0.49335	0.9574	0.1723	0.54591	0.6319	0.46936
β_3	−5.6864	0.7638	−11.362	−5.2135	− 0.779	−10.123
const.	35.31	70.11	−6.84	33.56	73.89	0.12
R^2	0.9579	0.9831	0.9001	0.9564	0.9826	0.8990
σ	4.24	6.13	3.23	4.32	6.14	3.24
D.W.	1.99	2.71	1.24	1.98	2.68	1.24

る．

以上の結果は，標本数が50の場合である．計量経済学における計測作業では，標本数がもっと小さい場合も多くある．したがって標本数が30の場合のシミュレーションを次に行なった．標本数が30になること以外，真の回帰式，説明変数の構造を同一にして，やはり100回のシミュレーションを行なった．その結果，表1．b．1及び表1．b．2が得られた．表1．b．1及び表1．b．2より，最小二乗推定は標本数が減少することによって，推定値の変動幅がさらに大きくなり，不安定な推定となることがわかる．主成分回帰では，標本数の減少によって推定値の分布は大きく変化しないし，バイアスの方向，程度もほとんど変化しないことが認められる．

4.4　上述のシミュレーションは，説明変数の相関係数行列の固有値のうち，最小固有値に対する主成分のみを欠落させて回帰を行なうケースであった．次に行なったシミュレーションでは主成分回帰において2つの固有値を問題とする．ここでのシミュレーションで説明変数間の相関係数行列を

$$R=\begin{pmatrix} 1.0 & 0.95 & 0.90 \\ & 1.0 & 0.80 \\ & & 1.0 \end{pmatrix}$$

として，これ以外の，真の回帰式，各説明変数の平均，標準偏差はすべて4.3のシミュレーションの場合と同様である．また，標本数は30である．

100回のシミュレーションの結果，説明変数間の相関係数行列の固有値は，

	第1固有値	第2固有値	第3固有値
平均：	2.779	0.1954	0.0246
最大：	2.898	0.4114	0.0501
最小：	2.551	0.0839	0.0082

であった．最小固有値は0.05～0.008であり，第2番目の固有値は0.4～0.08で平均はおよそ0.2である．したがって最小固有値に対応する主成分のみを欠落させる主成分回帰（LAT 2）と最小固有値に対応する第1主成分及び2番目

の固有値に対応する第2主成分の二つを欠落させる主成分回帰（LAT 1）の二つを考察の対象とした．

本節のシミュレーションの結果は，表2．1及び表2．2に示している通りである．LAT 2よりLAT 1の方が係数推定値の変動幅は小さい．しかし，各係数の真値からのバイアスはかなり大きくなる．LAT 2については限界係数β_1は下方に偏り，β_2およびβ_3は上へ偏る．他方，LAT 1では，β_1，β_2の推定値が下へ偏りβ_3は上へのバイアスが大きい．表2．2の各統計値の平均を見てもそのことは明らかである．さらに顕著なことは，LAT 1による定数項推定

表2.1

	β_1			β_2			β_3		
	OLS	LAT2	LAT1	OLS	LAT2	LAT1	OLS	LAT2	LAT1
(1)	2	0	0	0	0	0	6	0	0
(2)	1	0	0	0	0	0	3	1	0
(3)	5	8	35	1	0	3	10	3	0
(4)	7	46	65	10	0	97	15	6	0
(5)	17	43	0	27	0	0	15	3	0
(6)	18	3	0	37	23	0	14	18	0
(7)	25	0	0	17	65	0	19	19	0
(8)	14	0	0	5	11	0	5	18	0
(9)	8	0	0	3	1	0	7	18	0
(10)	2	0	0	0	0	0	3	7	0
(11)	1	0	0	0	0	0	3	7	100

表2.2

	OLS			LAT2			LAT1		
	mean	max.	min.	mean	max.	min.	mean	max.	min.
β_1	0.06255	0.11129	0.00325	0.04430	0.06288	0.03032	0.03607	0.04323	0.03109
β_2	0.49207	0.84824	0.19309	0.58836	0.75259	0.48086	0.27663	0.29896	0.23617
β_3	−5.7851	1.8180	−11.363	−3.9451	1.2253	−9.5585	9.3238	12.327	6.6783
const.	35.61	63.18	−2.58	28.66	54.89	−1.54	−41.63	−15.74	−67.29
R^2	0.9556	0.9839	0.8719	0.9533	0.9817	0.8719	0.8771	0.9461	0.7569
σ	4.201	5.892	2.706	4.316	5.893	2.941	7.079	9.068	5.066
D.W.	2.04	2.81	1.35	2.05	2.85	1.29	2.04	2.73	1.11

値をみても，LAT 1 による推定による推定は真の回帰直線からの乖離が大きい．決定係数も，最小二乗推定（OLS）と第 1 主成分のみを欠落させる主成分回帰（LAT 2）とではその差は無視し得る程であるが，LAT 2 のそれはかなり落ちる．結論的にいえることは，説明変数が 3 変数の場合， 2 つの主成分を欠落させる主成分回帰は許容し難いバイアスが生じるだろうということである．

4.5　本節のシミュレーションは，最小二乗推定（OLS）を採用しないで主成分回帰（LAT）による推定を行なう際の一つの限界を与える．

説明変数間の相関行列を

$$R=\begin{pmatrix} 1.0 & 0.7 & 0.6 \\ & 1.0 & 0.5 \\ & & 1.0 \end{pmatrix}$$

とする．これ以外の標本構造は以前と同様である．前節までのシミュレーションと同様100回の実験の結果，相関係数行列の固有値の分布は，

	第 1 固有値	第 2 固有値	第 3 固有値
平均：	2.1796	0.5527	0.2676
最大：	2.5015	1.0126	0.4587
最小：	1.5876	0.3045	0.1091

であり，最小固有値は0.1～0.45で 0 に充分近いとはいえない．

さて，このシミュレーションの結果は，各係数の分布は表 3．1 の通りである．表 3．1 に見られるように，主成分回帰の方が最小二乗推定より推定値の変動幅は大きい．推定値のちらばりによって主成分回帰による推定値のバイアスは以前のシミュレーションより明確ではない．しかし，このシミュレーションの結果より明らかなことは，説明変数間の相関係数行列の最小固有値が0.2～0.3以上ある場合，問題とすべき多重共線関係はない．したがって，最小二乗推定を放棄する根拠はないといえるだろう．

4.6　本節のシミュレーションでは，説明変数の構造は4.3でのそれと同様であ

表3.1

	β_1		β_2		β_3	
	OLS	LAT	OLS	LAT	OLS	LAT
(1)	0	0	0	0	1	0
(2)	0	4	0	0	0	0
(3)	0	8	0	2	5	2
(4)	1	20	0	2	8	4
(5)	28	29	17	8	24	10
(6)	52	28	68	43	24	16
(7)	16	10	15	39	21	19
(8)	3	1	0	6	10	12
(9)	0	0	0	0	7	16
(10)	0	0	0	0	0	8
(11)	0	0	0	0	0	13

る．真の回帰係数パラメタも同様であるが，撹乱項の分散が大きいケースを考察する．経済学の実証分析の中には，ある構造方程式の体系的部分に関する一つの仮説が承認されていて，そのような定式化に従う計測を行なっても，撹乱的要因の変動が大きいと考えられるケースもある．ここでの撹乱項 u は，

$$u \sim N(0,\ 21.5^2)$$

であって，u の標準偏差が以前のシミュレーションの場合の５倍を設定した．

説明変数の生成構造は4.3の場合と同じであるから，その相関係数行列の固有値に関しては大差ない結果である．回帰係数推定値の分布は，表４．１に挙げる通りで，一見して，最小二乗推定値（OLS）の変動幅は極端に大きくなることがわかる．それに対して，主成分回帰（LAT）による推定値はOLSに比してかなり安定したものが得られている．100回のシミュレーションの各推定値，及びそれら以外の統計値の平均を表４．２にまとめている．

説明変数間の多重共線性が存在する場合であっても最小二乗推定量は不偏推定量であることは周知のことであるが，各限界係数推定値の100回の平均は真値とかなり異なる．多重共線性のためと撹乱項の分散が大きいという二つの理由でOLS推定量の分散が非常に大きくなり，100回のシミュレーションでは

表4.1

	β_1		β_2		β_3	
	OLS	LAT	OLS	LAT	OLS	LAT
(1)	24	0	34	0	35	17
(2)	6	1	5	0	3	3
(3)	3	6	4	1	5	5
(4)	4	14	6	7	1	5
(5)	3	30	7	18	2	9
(6)	5	31	5	26	3	3
(7)	4	13	5	34	10	8
(8)	6	5	5	11	4	7
(9)	1	0	5	3	4	6
(10)	3	0	5	0	2	7
(11)	41	0	19	0	31	30

表4.2

	OLS mean	LAT mean
β_1	0.08551	0.0544
β_2	0.31398	0.52597
β_3	−6.36	−4.01
const.	35.585	25.955
R^2	0.4913	0.4702
σ	21.089	21.538
D.W.	1.995	2.009

推定値のちらばりが大きくなり過ぎてOLS推定量の不偏性を実験では示せなかったということになる.

主成分回帰による推定値について，β_1, β_2 は4.3のシミュレーションとほぼ同様であり，攪乱項の分散が大きくなることによる影響は大きくない．β_3 のバイアスはかなり大きく，したがって定数項も過少推計されている．これは以前にも述べたように，主成分回帰による推定量のバイアスに加えて，説明変数 X_3 の分散が小さいことにその原因があると思われる.

4.7　以上に示したシミュレーションの結果，3個の説明変数間の多重共線性

が存在する場合，最小二乗推定と主成分回帰の比較を次のように要約できる．

(i) 標本数が小さいとき，それが大きいときに比較して，OLS推定値のちらばりは大きい．他方，LAT推定値は標本数の違いによる影響はなく，バイアスはあるものの比較的安定した推計結果が得られる．

(ii) 第1主成分，第2主成分の二つの主成分を欠落させるLAT推定値のバイアスは非常に大きくなる．

(iii) 説明変数間の相関係数行列の最小固有値が0.1以下の場合に主成分回帰を考慮すべきであって，最小固有値が0.2以上の場合は問題とすべき多重共線性はないと考えてよい．したがって，このような場合は最小二乗推定で充分である．

(iv) 攪乱項の分散が大きい場合，最小二乗推定は非常に不安定な推計になる．他方，主成分回帰による推定はかなり安定している．

(v) ある一つの説明変数の分散が小さいとき，不安定な推定となり，これは主成分回帰によっても回復できない．

参考文献

〔1〕 Anderson, T. W., *An Introduction to Mulitivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

〔2〕 Farebrother, R. W., "Principal Component Estimators nad Minimum Mean Square Error Criteria in Regression Analysis," *Review of Economics and Statistics*, Vol. 54, 1972, pp. 332–336.

〔3〕 Girshick, M. A., "Principal Components," *Journal of American Statistical Association*, Vol. 31, 1936, pp. 519–5281.

〔4〕 Hotelling, H., "Analysis of A Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. 24, 1933, pp. 417–441, pp. 498–520.

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

〔6〕 Lowerre, J. M., "On the Mean Square Error of Parameter Estimates for Some Biased Estimators," *Technometrics*, Vol. 16, 1974, pp. 461–464.

〔7〕 Marquardt, D. W., "Generalized Inverse, Ridge Re Regression, Biased Linear Estimation, and Nonlinear Estimation," *Technometrics*, Vol. 12, 1970, pp. 591–612.

〔8〕 Massy, W. F., "Principal Components Regressinon in Exploratory Statistical Research," *Journal of the American Statistical Association*, Vol. 56, 1965, pp. 234–256.

〔9〕 McCallum, B. T., "Artificial Orthogonalization in Regression Analysis," *Review of Economics and Statistics*, Vol. 52, 1970, pp. 110–113.

〔10〕 Pearson, K., "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine*, sixth series, Vol. 2, 1901, pp. 559–572.

〔11〕 井上勝雄,「多重共線性とバイアス推定」,『経済学論究』第35巻第4号, 1982年1月.