多重共線性とバイアス推定

井　上　勝　雄

§0

われわれが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様々な困難にぶつかる．その典型的な事実は，単純最小二乗推定値が非常に不安定になるということであろう．このことは最小二乗推定量の分散が多重共線性が存在するとき非常に大きくなるという事実から推察できる．多重共線性の存在の一つの指標は説明変数間の積和行列が特異行列に近い，つまり積和行列の行列式の値が0に近いということである．換言すれば，積和行列の固有値の中に非常に0に近いものがある．このことと，最小二乗推定量の分散が大きくなるということと同値である．

さて，この最小二乗推定にかえて，より小さい分散をもつ推定量を求めることが，説明変数間の多重共線性が存在するときの課題となろう．しかし，最小二乗推定量は不偏推定量の中で最小分散をもつのであるから，バイアス推定量，つまり偏りのある推定量の中から出来るだけ分散の小さい推定量を求めなければならない．また，バイアス推定量を考察する際，その推定量の分散だけを効率基準にするのはよくない．分散とバイアスの程度の両者を考慮すべきで，そこから平均二乗誤差基準による比較をしなければならないことになる．

本稿の以下では説明変数間の多重共線性を前提に二つのバイアス推定量を考察する．§2 で考察するリッジ推定量も§3 のバイアス推定量も，平均二乗誤差基準で最小二乗推定量より精度がよいことを理論的に考察する．ただし，これらの推定量の精度は未知パラメタに依存するので最小二乗推定を採らないで，これらのバイアス推定を採用すべきという応用上の問題は本稿では扱わな

い．したがって，本稿の目的は，§ 2 で考察するリッジ推定と，§ 3 で考察するバイアス推定量との応用上の優劣を比較することではなく，両者が全く代替的な推定量であることを示すことである．

§ 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明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ベクトルであり，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 で相互に無相関，分散一定 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.

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

$$\text{ここで}\quad Z=XP,\ \alpha=P'\beta$$

(1.5) で定義された Z について

$$(1.4)\qquad Z'Z=P'X'XP=\Lambda$$

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[1]. 回帰係数については，

$$(1.6)\qquad \beta=P\alpha$$

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

$$(1.7)\qquad \hat{\alpha}=\Lambda^{-1}Z'y$$

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，

$$(1.8)\qquad E(\hat{\alpha})=\alpha$$

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

が明らかである．推定量 $\hat{\alpha}$ は不偏推定量であるが，その分数は $X'X$ 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

1)　「yの説明変数 X のすべての主成分 Z への回帰」ともいえる.

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

(1.9)　$\hat{\beta}=P\hat{\alpha}$

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 の間の関係(1.6)と同様の関係が成立する．さらに，(1.8)，(1.9)より

(1.10)　$E(\hat{\beta})=P\alpha=\beta$

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

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

$$var(\hat{\beta}_i)=\sigma^2\Sigma\frac{p_{ij}}{\lambda_j}$$

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要素に依存する．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 $\frac{p_{ik}^2}{\lambda_k}$ 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　$X'X$ のある固有値が0に近いとき，最小二乗推定量は分散が大きくなる．推定量の分散を推定精度の尺度とする限り，多重共線性が存在するときには，最小二乗推定の精度は悪いことになる．そこで，不偏性をもたないが，最小二乗推定よりも精度の良い推定量を求める方向が考えられる．Hoerl-Kennard [3] の提案するリッジ回帰は最小二乗推定量より精度のよい推定量が得られる．

リッジ推定量は不偏性を満たさない．つまり偏りのある推定量である．した

がって推定量の精度の尺度としては分散だけをみるのは不適当である．推定量の精度を平均二乗誤差基準で考察する必要がある．

一般に，p 個の母パラメタベクトル θ の推定量 $\hat{\theta}$ の平均二乗誤差行列は，

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

である．簡単な計算によって

$$MSE(\hat{\theta})=V(\hat{\theta})+bias(\hat{\theta})bias(\hat{\theta})'$$

が導ける．ここで，

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

である．

また，母パラメタ θ_i の推定量 θ_i について，

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

となり，推定量 $\hat{\theta}_i$ の平均二乗誤差は，推定量の分散と偏り（$bias$）の二乗和である．

さらに，推定量の全平均二乗誤差は，

$$TMSE(\hat{\theta})=trMSE(\hat{\theta})=trV(\hat{\theta})+bias(\hat{\theta})'bias(\hat{\theta})$$

で定義される．

さて，最小二乗推定量は不偏推定量であるから偏りは０である．つまり最小二乗推定量の平均二乗誤差は分散に等しい．しかし，多重共線性が存在するとき，最小二乗推定量は分散が大きくなることは§１で考察したとおりである．したがって，偏りのある推定量が有意味となるのは，偏りが０でなくとも，その推定量の分散が充分小さく，最小二乗推定の平均二乗誤差より小さい平均二乗誤差をもつときである．

2.2　まず，モデル（1.5）にリッジ推定を適用してみよう．（1.5）のリッジ推定は，$k>0$ に対して，

$$(2.1)\qquad [\Lambda+kI]\hat{\alpha}(k)=Z'y$$

を満たす $\hat{\alpha}(k)$ がその推定量である．つまり α の最小二乗推定量を得る正規方程式において $Z'Z=\Lambda$ の対角要素に $k>0$ を加えて得られる推定量がリッ

ジ推定量 $\hat{\alpha}(k)$ である．また，その定式化から明らかな様に，最小二乗推定量 $\hat{\alpha}$ は $\hat{\alpha}(0)$ であり，つまり $k=0$ としたときのリッジ推定となる．

さて，リッジ推定量 $\hat{\alpha}(k)$ と最小二乗推定量 $\hat{\alpha}$ の間には次の関係がある．

$$(2.2)\qquad \hat{\alpha}(k)=[\Lambda+kI]^{-1}Z'y=[\Lambda+kI]^{-1}\Lambda\hat{\alpha}$$
$$=[I+k\Lambda^{-1}]^{-1}\hat{\alpha}$$

上の (2.2) の関係式を利用して，リッジ推定量の期待値，及び偏りが次のように得られる．

$$(2.3)\qquad E(\hat{\alpha}(k))=diag\frac{\lambda_i}{\lambda_i+k}\alpha$$
$$bias(\hat{\alpha}(k))=diag\frac{-k}{\lambda_i+k}\alpha$$

(2.3) において，たとえば $diag\dfrac{\lambda_i}{\lambda_i+k}$ は，ii 要素が $\dfrac{\lambda_i}{\lambda_i+k}$ である対角行列を示している．

さらに，$\hat{\alpha}(k)$ の分散について，(1.8) および (2.2) より次のように得られる．

$$(2.4)\qquad V(\hat{\alpha}(k))=[I+k\Lambda^{-1}]^{-1}V(\hat{\alpha})[I+k\Lambda^{-1}]^{-1}$$
$$=\sigma^2\cdot diag\frac{\lambda_i}{(\lambda_i+k)^2}$$

平均二乗誤差については，(2.3) 及び (2.4) より，リッジ推定 $\hat{\alpha}(k)$ の平均二乗誤差行列は，

$$(2.5)\qquad MSE(\hat{\alpha}(k))=\sigma^2 diag\frac{\lambda_i}{(\lambda_i+k)^2}+diag\left(\frac{-k}{\lambda_i+k}\right)\alpha\alpha' diag\left(\frac{-k}{\lambda_i+k}\right)$$

である．

2.3　この節で全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リッジ推定量が必らず存在することを示しておく．

いま，リッジ推定量の全平均二乗誤差を $f(k)$ とおくと，(2.5) を参考に，

$$(2.6)\qquad f(k)=TMSE(\hat{\alpha}(k))=\sigma^2\Sigma\frac{\lambda_i}{(\lambda_i+k)^2}+\Sigma\frac{k^2}{(\lambda_i+k)^2}\alpha_i^2$$

1）$\hat{\alpha}(k)=diag\dfrac{\lambda_i}{\lambda_i+k}\hat{\alpha}$ であるから $k\to\infty$ に対して $\alpha(k)\to 0$ である．この観点からリッジ推定量は「縮小推定量（*Shrinkage Estimator*）」といえる。[1]，[5] 参照。

である．(2.6) において，$k=0$ とすると，

$$f(0)=\sigma^2\Sigma\frac{1}{\lambda_i}$$

となり，これは，最小二乗推定量の全平均二乗誤差である．さて，$k>0$ の範囲で，

$$f'(k)=2\Sigma\frac{\lambda_i(k\alpha_i^2-\sigma^2)}{(\lambda_i+k)^3}$$

が得られる．α_i^2の最大値を $\alpha^2_{\max}$ とし，また最小値を $\alpha^2_{\min}$ と表わすと，

$$0<k<\frac{\sigma^2}{\alpha^2_{max}} \quad \text{のとき} \quad f'(k)<0$$

$$\frac{\sigma^2}{\alpha^2_{min}}<k \quad \text{のとき} \quad f'(k)>0$$

である．さらに，実際に計算することより

$$\lim_{k\to+0} f'(k)=-2\sigma^2\Sigma\frac{1}{\lambda_i^2}<0$$

が得られ，$k>0$ の範囲で $f'(k)$ は連続であるから，

$$\frac{\sigma^2}{\alpha^2_{max}}<k<\frac{\sigma^2}{\alpha^2_{min}}$$

を満たすある k^* のとき，$f'(k^*)=0$ となり，このとき $f(k)$ は極小となる．つまり，

$0<k<k^*$ のとき　　$f(k)$ は減少関数

$k^*<k$ のとき　　$f(k)$ は増加関数

上述のことから結局，

$$f(k^*)<f(0)$$

が示せる．ゆえに，最小二乗推定の全平均二乗誤差よりも小さい全平均二乗誤差をもつリッジ推定量が得られる．

2.4　さて，Hoerl-Kennard の示したリッジ推定量は，モデル (1.1) に適用するものである．そうしてモデル (1.1)に適用したリッジ推定量 $\hat{\beta}(k)$ は，

$$(2.7) \qquad [X'X+kI]\hat{\beta}(k)=X'y$$

より導出できる．次に，上の $\hat{\beta}(k)$ と (2.1) より得られる $\hat{\alpha}(k)$ との関係をみ

みておく．(2.7) に前から P' を乗じて，変形すると

$$P'[X'X+kI]PP'\hat{\beta}(k)=P'X'y$$
$$\therefore\ [P'X'XP+kP'P]P'\hat{\beta}(k)=(XP)'y$$

これより (1.4)，(1.3) を参考して，

$$[\Lambda+kI]P'\hat{\beta}(k)=Z'y$$

である．ゆえに，$\hat{\alpha}(k)=P'\hat{\beta}(k)$ が得られ，

$$(2.8)\qquad \hat{\beta}(k)=P\hat{\alpha}(k)$$

が導ける．リッジ推定量 $\hat{\alpha}(k)$ と $\hat{\beta}(k)$ の間には母パラメタ α，β の関係 (1.6) と同様の関係がある．したがって，$\hat{\beta}(k)$ の全平均二乗誤差は $\hat{\alpha}(k)$ の全平均二乗誤差に等しいことが導ける．(1.6) および，(2.8) より

$$bias(\hat{\beta}(k))=bias(P\hat{\alpha}(k))=Pbias(\hat{\alpha}(k))$$

が導け，

$$bias(\hat{\beta}(k))'bias(\hat{\beta}(k)=bias(\hat{\alpha}(k))'P'Pbias(\hat{\alpha}(k))$$
$$=bias(\hat{\alpha}(k))'bias(\hat{\alpha}(k))$$

が得られる．また，

$$trV(\hat{\beta}(k))=trPV(\hat{\alpha}(k))P'=trV(\hat{\alpha}(k))P'P=trV(\hat{\alpha}(k))$$

であるから，$\hat{\beta}(k)$ の全平均二乗誤差と $\hat{\alpha}(k)$ の全平均二乗誤差は等しい．

2.5　前節において，全平均二乗誤差基準に照らして，最小二乗推定よりも良いリッジ推定量が存在することが示された．全平均二乗誤差は，その定義からわかるように p 次元空間における推定量と母パラメタとの距離の平方が平均的にどれ程の大きさになるかを示している．したがって個々の限界係数である母パラメタ α_i と，推定量 $\hat{\alpha}_i$ や $\hat{\alpha}_i(k)$ との距離は考察の対象ではない．

この節で，個々のパラメタ α_i と，その最小二乗推定量 $\hat{\alpha}_i$，リッジ推定量 $\hat{\alpha}_i(k)$ をとりあげ，特に，α_i と $\hat{\alpha}_i(k)$ の平均二乗誤差の比較をする．

(2.5) よりリッジ推定量 $\alpha_i(k)$ の平均二乗誤差は，

$$(2.9)\qquad MSE(\hat{\alpha}_i(k))=\sigma^2\frac{\lambda_i}{(\lambda_i+k)^2}+\frac{k^2\alpha_i^2}{(\lambda_i+k)^2}$$

である．また，(1.8) より，最小二乗推定量 $\hat{\alpha}_i$ の平均二乗誤差は，その分散と等しく，

$$(2.10)\quad MSE(\hat{\alpha}_i)=\sigma^2\frac{1}{\lambda_i}$$

である．いま，最小二乗推定量とリッジ推定量の平均二乗誤差の差を計算すると，

$$MSE(\hat{\alpha}_i)-MSE(\hat{\alpha}_i(k))=\frac{g(k)}{\lambda_i(\lambda_i+k)^2}$$

$$g(k)=(\sigma^2-\lambda_i\alpha_i^2)k^2+2\lambda\sigma^2k$$

となる．したがって，

(i) $\lambda_i\leqq\dfrac{\sigma^2}{\alpha_i^2}$ ならば，$k>0$ のとき常に $g(k)>0$

(ii) $\dfrac{\sigma^2}{\alpha_i^2}<\lambda_i$ ならば，$0<k<\dfrac{2\lambda_i\sigma^2}{\lambda_i\alpha_i^2-\sigma^2}$ のとき $g(k)>0$

が導ける．$\lambda_i(\lambda_i+k)^2>0$ であるから(i)，(ii)のいずれの場合も，

$$MSE(\hat{\alpha}_i)>MSE(\hat{\alpha}_i(k))$$

となる k は存在することになる．

リッジ推定量 $\hat{\alpha}_i(k)$ の平均二乗誤差をもう少し詳細に考察してみよう．いま，

$$\varphi_i(k)=MSE(\hat{\alpha}_i(k))$$

とすると，

$$\varphi_i'(k)=\frac{2\lambda_i(\alpha_i^2k-\sigma^2)}{(\lambda_i+k)^3}$$

であるから $\varphi_i(k)$ の最小値は，

$$k=\frac{\sigma^2}{\alpha_i^2}$$

のとき得られる．さらに，

$$\lim_{k\to+0}\varphi_i(k)=\sigma^2\frac{1}{\lambda_i},\quad \lim_{k\to+\infty}\varphi_i(k)=\alpha_i^2$$

が計算できる．したがって，$\varphi_i(k)$ のグラフを描くと，図1，図2のようになる．

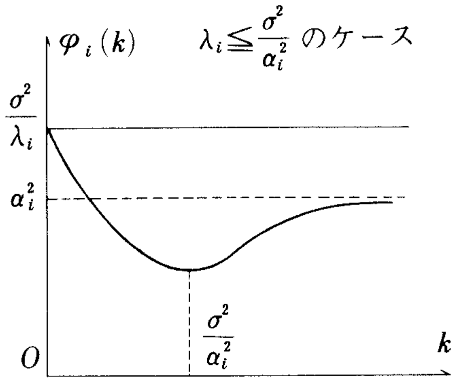

図 1

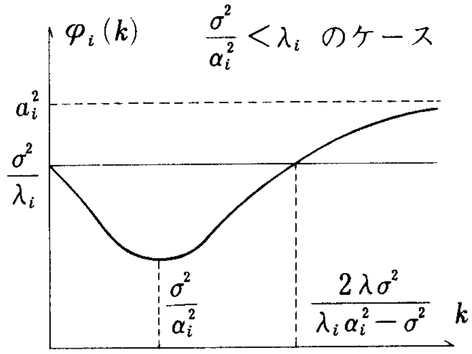

図 2

上述のように，$k=\frac{\sigma^2}{\alpha_i^2}$ が $\varphi_i(k)$ の最小値を与え，グラフからも明らかなように，少なくとも

$$0<k<\frac{\sigma^2}{\alpha_i^2}$$

では　$\varphi_i(k)<\sigma^2\frac{1}{\lambda_i}=MSE(\hat{\alpha}_i)$

である．したがって，

$$k=\frac{\sigma^2}{\alpha_{max}^2}$$

は，すべての i について

$$MSE(\hat{\alpha}_i(k))<MSE(\hat{\alpha}_i)$$

なるための十分条件である．

さて，これまでのリッジ推定はすべての i に対して同一の k の値を考えていたが，

$$k_i=\frac{\sigma^2}{\alpha_i^2}$$

と定義するならば，すべての i に対して，$\alpha_i(k_i)$ の平均二乗誤差が同時に最小となる．これが一般リッジ推定として提示されるもので，そのとき推定量を得る方程式は，(2.1) に替えて，

$$[\Lambda+K]\hat{\alpha}(K)=Z'y$$

$$K=diag\left(\frac{\sigma^2}{\alpha_i^2}\right)$$

となる.

また,この一般リッジ推定によって得られる個々の推定量の平均二乗誤差は,

$$(2.12)\quad MSE(\hat{\alpha}_i(K))=\varphi_i\left(\frac{\sigma^2}{\alpha_i^2}\right)=\frac{\sigma^2\alpha_i^2}{\lambda_i\alpha_i^2+\sigma^2}$$

である. さらに上式から推定量 $\hat{\alpha}(K)$の全平均二乗誤差は,

$$(2.13)\quad TMSE(\hat{\alpha}(K))=\Sigma\frac{\sigma^2\alpha_i^2}{\lambda_i\alpha_i^2+\sigma^2}$$

である.

§ 3

3.1 この節で, リッジ推定とは別のバイアス推定量を考察する.

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

$$(3.1)\quad \lambda_1\geqq\lambda_2\geqq\cdots\cdots\geqq\lambda_p$$

とし, 説明変数間の多重共線性によって,

$$(3.2)\quad \lambda_p\fallingdotseq 0$$

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 (3.1), (3.2) 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, (3.2) 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.

$$(3.3)\quad \tilde{\alpha}=[I-J_p]\hat{\alpha}$$

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 $\widetilde{\alpha}$ 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，

$$(3.4) \qquad Z_p'Z_p = p_p'X'Xp_p = \lambda_p p_p'p_p = \lambda_p$$

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[1]．さらに，(1.1)のモデルにおける β の推定量 $\widetilde{\beta}$ を

$$(3.5) \qquad \widetilde{\beta} = P\widetilde{\alpha}$$

と定義する．

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

推定量 $\widetilde{\alpha}$ の定義(3.3)より

$$(3.6) \qquad E(\widetilde{\alpha}) = [I-J_p]\alpha = \alpha - J_p\alpha$$

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

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

$$(3.7) \qquad V(\widetilde{\alpha}) = \sigma^2[I-J_p]\Lambda^{-1}$$

が容易に導ける．(3.6)および(3.7)の $\widetilde{\alpha}$ 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 $\widetilde{\alpha}_i$ 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い．α_p については $\widetilde{\alpha}_p=0$ 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 $\widetilde{\alpha}_i$ の平均二乗誤差行列は，

1)　Z の各列は説明変数 X の主成分である．したがって第 p 主成分への y の回帰は行なわない．つまり，X の全分散に対する寄与率の大きい主成分に対して y を回帰することである．

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

である．(3.8) 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

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

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 (3.5) を参考にして

(3.9)
$$\begin{aligned} E(\tilde{\beta}) &= PE(\tilde{\alpha})=P[I-J_p]\alpha=P\alpha-PJ_p\alpha \\ &= \beta-PJ_pP'\beta \\ bias(\tilde{\beta}) &= -PJ_pP'\beta \end{aligned}$$

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，

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

が得られる．

$\tilde{\alpha}$，及び $\tilde{\beta}$ の全平均二乗誤差については，(3.8)，(3.9)，(3.10) より容易に，

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

と導くことが出来る．

3.3　この節で最小二乗推定量 $\hat{\alpha}$，リッジ推定量 $\hat{\alpha}(k)$，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 (3.2) 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 $\dfrac{\sigma^2}{\alpha_p^2}>\lambda_p\fallingdotseq 0$ を前提にする．つまり，

(3.12) $\alpha_p^2<\dfrac{\sigma^2}{\lambda_p}$

を前提に考察をすすめる．

(3.12) から直ちに，(3.11) を参考に，

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

が示せる．われわれは多重共線性の状況を前提に本節では (3.2) を仮定しているが，このことから当然 (3.13) の特性は容易に理解できることである．

次にリッジ推定量 $\hat{\alpha}(k)$ と本節でのバイアス推定量 $\widetilde{\alpha}$ との比較を全平均二乗誤差基準で考察しよう．

リッジ推定量 $\hat{\alpha}(k)$ の全平均二乗誤差の k に対する軌跡を描こうとする際，その一義的な軌跡を得ることは困難であると思われる．(3.12) 以外の母パラメタ α_i，$X'X$ の固有値 λ_i，及び σ^2 との関係についての種々の場合に依存するからである．そこで，以下で二つの特定化を行なう．

まず，リッジ推定量 $\hat{\alpha}(k)$ の全平均二乗誤差の最小値として，2.5 で考察した一般リッジ推定の全平均二乗誤差が採用できるとしよう．この場合，(2.12) と (3.10) より，

(3.13)　$TMSE(\hat{\alpha}(K)) < TMSE(\tilde{\alpha})$

が導ける．(3.13) は以下のように導出できる．$\hat{\alpha}(K)$ の個別の平均二乗誤差より，

$$\frac{\sigma^2\alpha_i^2}{\lambda_i\alpha_i^2+\sigma^2} = \frac{\sigma^2}{\lambda_i + \dfrac{\sigma^2}{\alpha_i^2}} < \frac{\sigma^2}{\lambda_i} \qquad i=1, 2, \cdots, p-1$$

$$\frac{\sigma^2\alpha_p^2}{\lambda_p\alpha_p^2+\sigma^2} = \frac{\alpha_p^2}{1+\lambda_p\dfrac{\alpha_p^2}{\sigma^2}} < \alpha_p^2$$

と変形でき，上二式の辺々を加えると，左辺は一般リッジ推定の全平均二乗誤差であり，右辺はバイアス推定量 $\widetilde{\alpha}$ の全平均二乗誤差である．したがって (3.13) が得られる．

われわれは，説明変数 X の多重共線性のために $X'X$ の最小固有値 λ_p が 0 に近似できる場合を想定し，(3.12) を前提したが，さらに，λ_p 以外の $X'X$ の固有値は充分大きいとして，

$$(3.14) \quad \frac{\sigma^2}{\alpha_i^2} < \lambda_i \qquad i=1,2,\cdots,p-1$$

とできるケースを考える．このとき 2.5 の考察より，

$$\lim_{k\to\infty} MSE(\hat{\alpha}_i(k)) = \lim_{k\to\infty} \varphi_i(k) = \alpha_i^2 > \frac{\sigma^2}{\lambda_i}$$

であるから，

$$\lim_{k\to\infty} TMSE(\hat{\alpha}(k)) = \lim_{k\to\infty} \Sigma MSE(\hat{\alpha}_i(k)) = \overset{p}{\Sigma}\alpha_i^2 > \overset{p-1}{\Sigma}\frac{\sigma^2}{\lambda_i} + \alpha_p^2$$

である．他方，(3.11) より

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

であるから，結局，

$$(3.15) \quad \lim_{k\to\infty} TMSE(\hat{\alpha}(k)) > TMSE(\tilde{\alpha})$$

が導出できる．

リッジ推定量の全平均二乗誤差の最小値が一般リッジ推定量の全平均二乗誤差に等しいということ及び (3.14) の二つの特定化を行う場合，(3.15) を参考にして，図 3 が描ける．図 3 より $\bar{k} < k < \bar{\bar{k}}$ を満たす k についてはリッジ推定量 $\hat{\alpha}(k)$ が本節のバイアス推定量 α より精度が良い．$0 < k < \bar{k}$，あるい

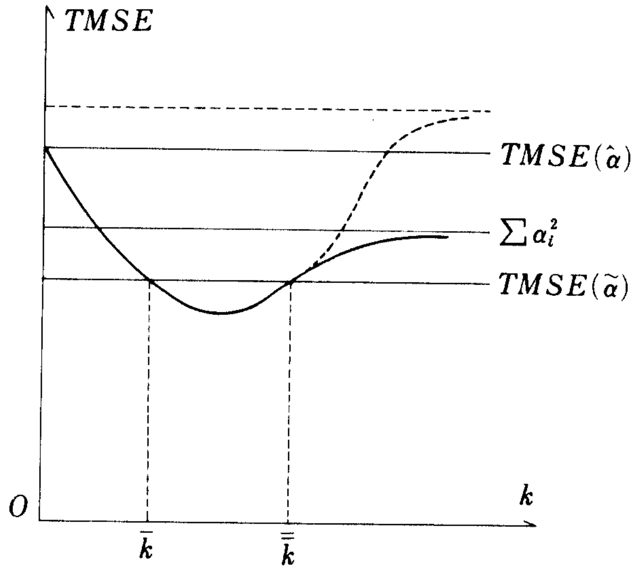

図 3　$TMSE(\hat{\alpha}(k))$ の軌跡

は$\overline{\overline{k}}<k$に対しては逆に，$\tilde{\alpha}$の方が$\hat{\alpha}(k)$より精度が良いことになる．

われわれは，図3を描くために，二つの特定化を行なった．以下でこの特定化を行なわない場合を二つのケースにわけて若干考察しておこう．

(i) リッジ推定量の全平均二乗誤差$TMSE(\hat{\alpha}(k))$の$k\to\infty$に対する漸近値が，$TMSE(\tilde{\alpha})$より小さい場合，$\overline{k}<k$を満たすkに関するリッジ推定量は，バイアス推定量$\tilde{\alpha}$より精度がよい．しかし，この場合でも$0<k<\overline{k}$を満たす$\overline{k}$に関するリッジ推定量は，αより精度が落ちる．

(ii) リッジ推定量$\hat{\alpha}(k)$のkに関する全平均二乗誤差の最小値が$\tilde{\alpha}$の全平均二乗誤差より大きい場合，常に，リッジ推定の精度が悪いことになる．

（関西学院大学経済学部助教授）

参考文献

[1] M. Goldstein and A. F. M. Smith, "Ridge-type Estimators for Regression Analysis", *Journal of the Royal Statistical Society*, ser. B, vol. 36, no. 2, 1974.

[2] W. J. Hemmerle, "An Explicit Solution for Generalized Ridge Regression," *Technometrics*, vol. 17, no. 3, 1975.

[3] A. E. Hoerl and R. W. Kennard, "Ridge Regression : Biased Estimation for Nonorthogonal Problem," *Technometrics* vol. 12, no. 1, 1970.

[4] ———, "Ridge Regression : Applications to Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, 1970.

[5] Hrichkesh D. Vinod, "A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares," *The Review of Economics and Statistics*, vol. 60, 1978.

[6] J. M. Lowerre, "On the Mean Square Error of Parametter Estimates for Some Biased Estimators," *Technometrics*, vol. 16, no. 3, 1974.

[7] S. D. Silvey, "Multicollinearity and Imprcise Estimation," *Journal of the Royal Statistical Society*, ser. B, vol. 32, no. 3, 1969.