【調査】回帰分析におけるバイアス推定量の選択*

この論文の目的は，線型回帰モデル $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}$ における係数 $\boldsymbol{\beta}$ の推定問題を，決定理論的視点から眺め，これまで支配的であり続けてきた最小2乗推定量の性質を再吟味するとともに，バイアス推定量を展望し，その性質と適用可能性を考えることで，新しく推定量を提案することにある。ただし，誤差項 $\boldsymbol{u}$ には，主として平均 $\mathbf{0}$，分散行列 $\sigma^2\boldsymbol{I}_n$ の正規分布を仮定し，$\boldsymbol{\beta}$ については先験的情報がないと仮定する。

最小2乗推定量が基本的にその正当性を主張する性質は，いうまでもなく不偏性である。しかし，不偏性は，推定量の標本分布の平均値が「真」の係数 $\boldsymbol{\beta}$ と一致することであって，それ自体推定量が $\boldsymbol{\beta}$ の近くに落る可能性を保証するものでない。とくに，繰返し実験が不可能な経済分析においては，推定量の標本分布は，いわばアプリオリな期待分布というような意味しかもたない場合が多く，そこにおいては，その分布の平均値が $\boldsymbol{\beta}$ であることを要求するより，最初から推定量と $\boldsymbol{\beta}$ との何らかの意味での平均距離(リスク)を小さくすることで，推定量が $\boldsymbol{\beta}$ の近くに落る可能性(期待)を高めた方がよいと考える。すなわち，不偏推定量のクラスに推定量を制限し，その中で分散行列(ちらばり)を最小にするという2段構えをとるより，最初からすべての推定量の中で，1つの指定した平均距離函数をなるべく小さくした方が適切であると考える。経済分析では，他の最適性(optimalities)をもつかもしれないバイアス推定量の選択の可能性をすべて排除してしまうことができるほど，不偏性は重要な性質だとは思われない。不偏推定量のクラスは推定量全体の中では相対的に小さく，とくに $\boldsymbol{u}$ の正規性のもとでは，十分統計量に基く不偏推定量は，最小2乗推定量ただ1つ(a. e.)である。

他方，推定量全体の中で，一つの平均距離函数(リスク函数)をなるべく小さくする推定量を選択するという立場に立った場合，問題となるのは，(1)距離函数(損失函数)の指定，及び(2) 1つのリスク函数が指定されても，一般にそれを一様に小さくする推定量が存在せず，推定量が一意に決まらないことである。(1)については，$\boldsymbol{u}$ の正規性のもとでは，問題の構造から自然な距離として，(a)推定量 $\hat{\boldsymbol{\beta}}$ と $\boldsymbol{\beta}$ とのユークリッド的距離の2乗 $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$，あるいは(b) $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$，が考えられる。後者の距離函数(損失函数)の合理性については§1を参照されたい。この論文では，(b)が採用される。(2)の問題は，個別的な問題の性質に応じた推定量を，リスク函数(平均距離函数)の behavior との関連で選択すればよいと考える。決定理論的には，admissibility とミニマックス性 (minimaxity) が客観的な選択基準となる。推定量が admissible であるということは，他に一様にリスクが小さい推定量が存在しないことであり，ミニマックス推定量は，現在の問題ではリスクの意味で最小2乗推定量よりよい推定量か，またはそれと同等な推定量である。従って，ここではこの2つの視点から問題を眺めることになる。これは，形式的には Stein (1956)に始まる正規分布の平均値の推定問題の議論を，回帰分析の議論へ翻訳あるいは拡張することにほかならない。

内容を要約しておく。§1では決定理論的に扱うための framework，とくに損失函数を指定する。§2では，最小2乗推定量の依拠する線型性及び不偏性を吟味する。線型性については，不変性との対応から議論する。また，誤差項 $\boldsymbol{u}$ の正規性のもとでは，最小2乗推定量は最良不偏推定量(十分統計量に基く不偏推定量は最小2乗推定量ただ1つ)であるが，これに対して，最小2乗推定量は最良不変推定量であることを証明する。十分統計量に基く不変推定量は，最小2乗推定量以外にも存在するので，この命題の方が，最良不偏推定量というより強いものといえる。§3では，推定量の選択基準として，ミニマックス性，及び admissibility がとりあげられる。また関連して，ベイズ等の概念が用いられる。論文をなるべく self-contained にするため，これらの概念の定義を与えてある。§4では，ミニマックス推定量の候補として，Elliptotically Symmetric 推定量のクラスを導き，Efron-Morris (1976) に基いて，その部分クラスとしてミニマックス推定量のあるクラスが与えられる。こ

* この研究は，「日本経済研究奨励財団の助成による」研究の一部である。

のクラスに属するいくつかの推定量を吟味し，一つの新しい推定量を提案する。しかし，admissibility についてはこれまで十分研究されていないため，十分議論されていない。また，いわゆる James–Stein 推定量の平均，及び分散行列を求めておいた。§5では，$\boldsymbol{\beta}$ について仮説検定をしてモデルまたは変数選択し，その後に最小2乗推定量を用いる場合を考察する。1つの仮説が採択された場合，それが正しいとしてそのもとで最小2乗推定量が通常適用されるが，このような方式に基く推定量(予備検定推定量)は，結果的には不偏推定量でなく，ミニマックスでもなければ admissible でもないことが証明される。なお，alternative な推定量が与えられる。この議論は，Bock–Yancey–Judge (1973) および Sclove–Morris–Radlhakrishnan (1972) に負う。§7では，線型バイアス推定量を取りあげる。Cohen (1966) に基いて，線型推定量が推定量全体で admissible であるための必要十分条件を述べる。(証明は Appendix 3 にある。) 一定の $(\boldsymbol{\beta}, \sigma^2)$ の範囲では，最小2乗推定量よりもよくなる線型推定量を R–推定量とよび，そのクラスに属する推定量として，Ridge 推定量をとりあげ，それは admissible であることを示す。Appendix 1 では，多変量正規分布に関する補助定理が，2では，Efron–Morris (1976) の証明の概略と James–Stein 推定量のリスク函数について述べられている。

なお，リスク函数の代わりに距離とか平均距離函数という言葉を上で用いたが，ここでいう距離は直観的な意味で推定量と $\boldsymbol{\beta}$ との平均的な"近さ"を表わす尺度という意味で，数学的な意味でのそれでない。以下も同様である。

§1 問題の設定

1.1 モデル 線型回帰モデル

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u} \tag{1.1}$$

において 係数 $\boldsymbol{\beta}(k\times 1)$ の推定問題を考察する。ここで，$\boldsymbol{X}$ は $n\times k$, $\operatorname{rank}(\boldsymbol{X})=k$ の既知行列，誤差項 $\boldsymbol{u}$ は，平均 $\mathbf{0}$，分散行列 $E(\boldsymbol{uu}')=\sigma^2\boldsymbol{I}_n$ をもつ正規分布 $N(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$ に従う。

$$\boldsymbol{u}\sim \boldsymbol{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_n) \tag{1.2}$$

誤差項 $\boldsymbol{u}$ についての正規性の仮定は以下の議論で積極的に用いられるが，いくつかの結果についてはその仮定は弱められる。$\boldsymbol{X}$ は既知行列であるから観察可能な確率変数(ベクトル)は $\boldsymbol{y}$ のみであって，標本空間は $\boldsymbol{y}$ の空間である $\mathcal{X}\equiv\boldsymbol{R}^n$ (n 次元ユークリッド空間) である。$(\boldsymbol{\beta}, \sigma^2)$ の空間であるパラメータ空間 Θ としては，$\boldsymbol{\beta}$ について全く先験的情報がない場合を主として扱うため，断りのない限り $\Theta=\boldsymbol{R}^k\times\boldsymbol{R}^+$ $(\boldsymbol{R}^+=\{x\in\boldsymbol{R}|x>0\})$ とする。従って，行動空間 (action space) Δ は $\Delta=\boldsymbol{R}^k$ となる。

1.2 推定量のクラスと十分統計量 以下の議論では，損失函数 (loss function) として $\boldsymbol{a}\in\Delta$ に関する凸函数 (2次形式) を用いるから，$\boldsymbol{\beta}$ の推定量として一般性を失うことなく nonrandomized 推定量のみを考察する。(詳細は Ferguson (1967) 第3章参照。) 従って推定量全体をボレル可測函数

$$\hat{\boldsymbol{\beta}}:\mathcal{X}=\boldsymbol{R}^n\longrightarrow\Delta=\boldsymbol{R}^k$$

全体(それを以下 $\mathcal{T}$ で示す)と考えることができる。次に一般性を失うことなく問題を十分性 (sufficiency) によって簡単にする。(1.1), (1.2) から $\boldsymbol{y}\sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}_n)$ となり

$$\begin{cases}\boldsymbol{w}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{X}'\boldsymbol{y}\\ \boldsymbol{v}=(\boldsymbol{y}-\boldsymbol{Xb})'(\boldsymbol{y}-\boldsymbol{Xb})\end{cases} \tag{1.3}$$

は十分統計量である。ただし，$\boldsymbol{b}$ は最小2乗推定量

$$\boldsymbol{b}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{w} \tag{1.4}$$

であり，$(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}$ は $[(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}]^2=(\boldsymbol{X}'\boldsymbol{X})^{-1}$ となる対称行列とする。勿論，$(\boldsymbol{b}, \boldsymbol{v})$ 自体も十分統計量である。従って，$\boldsymbol{\beta}$ の推定量として $(\boldsymbol{w}, \boldsymbol{v})$ (あるいは同じことだが $(\boldsymbol{b}, \boldsymbol{v})$) のボレル可測函数

$$\boldsymbol{\phi}:\boldsymbol{R}^k\times\boldsymbol{R}^+\longrightarrow\boldsymbol{R}^k$$

全体を考えれば十分である。この十分統計量 $(\boldsymbol{w}, \boldsymbol{v})$ に基く推定量全体を以下 $\mathcal{D}$ で示す。このときモデルは

$$\begin{cases}\boldsymbol{w}\sim N(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\beta}, \sigma^2 I_k)\\ \boldsymbol{v}\sim\sigma^2\chi^2{}_{n-k},\ \boldsymbol{w}\ \text{と}\ \boldsymbol{v}\ \text{は独立}\end{cases} \tag{1.5}$$

となる。ここに $\chi_m{}^2$ は自由度 m のカイ2乗分布を示す。

1.3 損失函数 決定理論的に問題を取り扱うために，損失函数を指定しなければならない。本論文では

$$\begin{aligned}&L(\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v}), (\boldsymbol{\beta}, \sigma^2))\\ &\quad=[\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})-\boldsymbol{\beta}]'X'X[\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})-\boldsymbol{\beta}]/\sigma^2\end{aligned} \tag{1.6}$$

$(\boldsymbol{\phi}\in\mathcal{D})$ を採用する。損失函数として

$$L(\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v}), (\boldsymbol{\beta}, \sigma^2))=\|\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})-\boldsymbol{\beta}\|^2/\sigma^2 \tag{1.7}$$

(ただし $\|\boldsymbol{a}\|^2=\boldsymbol{a}'\boldsymbol{a}$) でなく (1.6) を用いる理由は，既存の諸結果の適用可能性という技術的理由の他に，$(\boldsymbol{b}, \boldsymbol{v})$ が十分統計量であり，$\boldsymbol{b}\sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$ であるから，$\operatorname{cov}(\boldsymbol{b})=[\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}$ で正規化したという点があげられる。勿論，(1.7) の損失函数は，推定量 $\boldsymbol{\phi}$ とパラメータ $\boldsymbol{\beta}$ の通常の意味での距離の2乗を σ^2 で正規化したという意味で直観的に受け入れやすいが，(1.6) の損失函数自体も一つの合理的な損失函数と考えられる。実際，

経済分析における回帰モデルの多くでは，定数項以外の回帰係数は，例えば消費性向のように，一単位の所得の増加当りの消費の増加の割合いといった物理的単位をもち，被説明変数である $\boldsymbol{y}$ と同じ物理的単位をもたない場合が多く，$\boldsymbol{\beta}$ は一般に $\boldsymbol{y}$ に対して位置パラメータ (location parameter) の性格がうすい。あるいは逆に，$\boldsymbol{b}\sim N(\boldsymbol{\beta},\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$ から $\boldsymbol{\beta}$ が位置パラメータの意味をもつのは統計量 $\boldsymbol{b}$ に対してであって，その場合 $\boldsymbol{b}$ の分散行列が $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ であるから，損失函数(1.6)はその意味で一つの自然な損失函数ということができる。勿論ここで(1.7)の形の損失函数を否定しているのではなく，個別的な問題に応じて損失函数を選択していくのが望ましいと考える。しかし損失函数(1.7)を採用した場合，問題の構造におけるある種の不変性が失われ，技術的に非常に取り扱いにくくなる。Sclove(1966)は(1.7)の場合も(1.6)と同様に議論できるであろうと述べているが，実際はそうでない。ただしSclove(1966)自身が扱っている $\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}_k$(直交多項回帰)の場合は，(1.6)が(1.7)になる。なお，(1.6)の代わりに(1.7)を用いても，例えば最小2乗推定量が inadmissible であるというような結果は変らないことが随所ふれられる。

1.4 リスク函数 (1.6)の損失函数を用いて，真の値が $\boldsymbol{\beta}$ であるとき推定量 $\boldsymbol{\phi}\in\mathcal{D}$ を用いるときのリスクは

(1.8) $$R(\boldsymbol{\phi},(\boldsymbol{\beta},\sigma^2))=EL(\boldsymbol{\phi}(\boldsymbol{w},\boldsymbol{v}),(\boldsymbol{\beta},\sigma^2))$$

で定義される。これはパラメータ $\boldsymbol{\beta}$ と推定量 $\boldsymbol{\phi}$ の1つの平均距離である。われわれの問題は，何らかの意味でリスク函数 $R(\boldsymbol{\phi},(\boldsymbol{\beta},\sigma^2))$ を小さくする推定量 $\boldsymbol{\phi}$ を $\mathcal{D}$ の中からみつけることである。恒等的に $\boldsymbol{0}$ をとる推定量 $\boldsymbol{\phi}_0(\boldsymbol{w},\boldsymbol{v})\equiv\boldsymbol{0}$ を用いると，$\boldsymbol{\beta}=\boldsymbol{0}$ のときのリスクはゼロ ($R(\boldsymbol{\phi}_0,(\boldsymbol{0},\sigma^2))=0$) であるから，すべての $(\boldsymbol{\beta},\sigma^2)\in\Theta$ に対してリスク函数 $R(\boldsymbol{\phi},(\boldsymbol{\beta},\sigma^2))$ を一様に小さくする推定量は存在しない。従って，リスク函数から推定量を選択するための推定量選択基準を設定しなければならない。これを§3で扱う。

1.5 最後にわれわれの問題を多変量正規分布の平均値の推定問題と対応をつけておく。そのため(1.6)の損失函数を

(1.9) $$L(\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v}),(\boldsymbol{\beta},\sigma^2))=\|(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v})-(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{\beta}\|^2/\sigma^2$$

と書き直し

(1.10) $$\boldsymbol{\eta}=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{\beta},\quad \hat{\boldsymbol{\eta}}(\boldsymbol{w},\boldsymbol{v})=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v})$$

とおくと，$\boldsymbol{\eta}$ を $\hat{\boldsymbol{\eta}}$ で推定していると考えることができる。この場合モデル(1.1)を

(1.11) $$\boldsymbol{y}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\eta}+\boldsymbol{u}$$

と書きかえれば，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の任意の一つの推定量であるから，$\boldsymbol{\eta}$ の推定量を $\hat{\boldsymbol{\eta}}=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\hat{\boldsymbol{\beta}}$ とおけば，モデル(1.11)において $\boldsymbol{\eta}$ の推定問題を，損失函数

(1.12) $$L_0(\hat{\boldsymbol{\eta}}(\boldsymbol{w},\boldsymbol{v}),(\boldsymbol{\eta},\sigma^2))=\|\hat{\boldsymbol{\eta}}(\boldsymbol{w},\boldsymbol{v})-\boldsymbol{\eta}\|^2/\sigma^2$$

で考えていることになる。逆にこの問題が与えられているとき，$\boldsymbol{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\eta}$，$\boldsymbol{\beta}$ の推定量として $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\hat{\boldsymbol{\eta}}$ とおけば，損失函数

$$L_0(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v}),\ ((\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\beta},\sigma^2))$$

は(1,6)に等しくなり，$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}$ における $\boldsymbol{\beta}$ の推定問題となる。すなわち，損失函数(1.6)のもとでの $\hat{\boldsymbol{\beta}}$ のリスク函数 $R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))$ は，損失函数(1.12)のもとでの $\boldsymbol{\eta}=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{\beta}$，$\hat{\boldsymbol{\eta}}=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\hat{\boldsymbol{\beta}}$ のリスク函数

(1.13) $$R_0(\hat{\boldsymbol{\eta}},(\boldsymbol{\eta},\sigma^2))=EL_0(\hat{\boldsymbol{\eta}}(\boldsymbol{w},\boldsymbol{v}),(\boldsymbol{\eta},\sigma^2))$$

に恒等的に等しく，ある与えられた推定量 $\hat{\boldsymbol{\eta}}$ とパラメータ $\boldsymbol{\eta}$ のリスク函数 $R_0(\hat{\boldsymbol{\eta}},(\boldsymbol{\eta},\sigma^2))$ に関する性質は，$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\hat{\boldsymbol{\eta}}$ で定義される推定量と $\boldsymbol{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\eta}$ のリスク函数 $R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))$ に関する性質にそのまま翻訳される。勿論 $(\boldsymbol{w},\boldsymbol{v})$ は $(\boldsymbol{\eta},\sigma^2)$ に対する十分統計量でもあり，$\boldsymbol{\eta}$ の推定量全体も $\mathcal{D}$ である。以上のことを要約すれば，$\boldsymbol{\beta}$ の推定問題は次の $\boldsymbol{\eta}$ の推定問題と同等である。$\boldsymbol{w}$ と $\boldsymbol{v}$ が(1.5)の分布に従っているとき，$\boldsymbol{\eta}$ の推定量 $\hat{\boldsymbol{\eta}}\in\mathcal{D}$ で，リスク函数(1.13)を何らかの意味で小さくするような $\hat{\boldsymbol{\eta}}$ をみつけることである。従って問題はStein(1956)に始まったともいえる多変量正規分布の平均値の推定問題に帰着する。この意味で以下の内容はその問題の展望と拡張とみなすことができる。しかし，推定したいのは $\boldsymbol{\eta}$ でなく $\boldsymbol{\beta}$ であって，例えば不変性 (invariance) 等を適用する場合には注意が必要となることを後にみる.

§2 最小2乗推定量

2.1 不偏性 (Unbiasedness) 本節では推定量選択基準を設定する準備として，支配的に用いられている最小2乗推定量(1.4)の性質を吟味する。推定量全体を $\mathcal{T}$ で表わし，十分統計量 $(\boldsymbol{w},\boldsymbol{v})$ (あるいは $(\boldsymbol{b},\boldsymbol{v})$) に基く推定量全体を $\mathcal{D}$ で表わすことを再記する。

周知の如く，$\boldsymbol{u}$ についての正規性(1.2)の仮定のもとでは(1.4)の最小2乗推定量 $\boldsymbol{b}$ は，不偏推定量のクラス

(2.1) $$\mathcal{U}=\{\hat{\boldsymbol{\beta}}\in\mathcal{T}\,|\,E\hat{\boldsymbol{\beta}}(\boldsymbol{y})=\boldsymbol{\beta}\}$$

の中で，分散行列

(2.2) $$\mathrm{cov}(\hat{\boldsymbol{\beta}})=E(\hat{\boldsymbol{\beta}}(\boldsymbol{y})-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}(\boldsymbol{y})-\boldsymbol{\beta})'$$

を非負値定符号の順序 (ordering) の意味で最小にする

唯一のものである。しかしこの事実は，(1)一般性を失わず推定量のクラスを十分統計量に基く推定量のクラス $\mathscr{D}$ に制限できること，及び(2) $(\boldsymbol{w}, \boldsymbol{v})$ は完備であるから $\mathscr{D}$ の中には不偏推定量は最小2乗推定量以外 (a.e の意味で)存在しないこと，を述べたものにほかならない。すなわち，

(2.3)　$$\mathscr{D} \cap \mathscr{U} = \{\boldsymbol{b}\} \quad \text{(a.e.)}.$$

また，上の結果は $(\boldsymbol{w}, \boldsymbol{v})$ が完備十分統計量であるという性質に依存し，正規性に直接には依存しない。(2.3) は $\mathscr{D}$ の中の不偏推定量は最小2乗推定量 $\boldsymbol{b}$ だけであることを示し，従ってそこではもはや分散行列 (2.2) を最小にすることは不必要になり，推定問題は trivial なものとなる。上の事実を，正規性の仮定のもとでは，最小2乗推定量 $\boldsymbol{b}$ は最良不偏推定量といわれるが，ここでの「最良」という形容詞は(本質的に)「唯一」という形容詞におきかえた方がわかりやすい。この意味で，不偏性の要求は最小2乗推定量の選択の要求にほかならず，非常に強い性質といえよう。勿論不偏性が，推定量の満たす必要欠くべからざる性質であることが示されれば，不偏性を課して最小2乗推定量を選択することに問題はない。しかし，推定量 $\hat{\boldsymbol{\beta}} \in \mathscr{D}$ の標本分布の平均値が真のパラメータ $\boldsymbol{\beta}$ と一致することを要求する不偏性は，推定量の良し悪しを示す性質でなく，1つの望ましい性質であることを認めたにしても，例えば何らかの意味で推定量 $\hat{\boldsymbol{\beta}}$ と真のパラメータ $\boldsymbol{\beta}$ との平均距離をなるべく小さくするといった性質に優先するものとは思われない。むしろ，繰り返し実験が不可能な経済分析においては，推定量の標本分布の平均値が真のパラメータ $\boldsymbol{\beta}$ と同じであることよりも，推定量の選択基準としては，真の値に平均的に近いといった性質の方が優先すると考える。勿論，そこでの平均距離として何を選ぶかという問題は別な問題であって，その距離函数の任意性，あるいは1つの距離函数(リスク函数)を採用しても推定量が一般に一意に決まらないことから不偏性を正当化できない。距離函数(リスク函数)の選択は，個別的な問題に応じて選択するのが適当であると考える。

2.2　線型性と不変性　$\boldsymbol{u}$ の正規性(1.2)の仮定がはずされても，最小2乗推定量 $\boldsymbol{b}$ は $E(\boldsymbol{u})=0, E(\boldsymbol{uu}')=\sigma^2\boldsymbol{I}_n$ である限り，線型推定量のクラス

(2.4)　$$\mathscr{L} = \{\boldsymbol{A} \mid \boldsymbol{A} \text{ は } k\times n \text{ の行列}\}$$

かつ不偏推定量のクラス $\mathscr{U}$ に属する推定量，すなわち

(2.5)　$$\mathscr{L} \cap \mathscr{U} = \{\boldsymbol{A} \in \mathscr{L} \mid \boldsymbol{AX} = \boldsymbol{I}_k\}$$

の中で，分散行列 (2.2) を非負値定符号の意味で最小にする。この結果は $\boldsymbol{u}$ の分布に依存せず，そこでは $(\boldsymbol{w}, \boldsymbol{v})$ は必ずしも十分統計量でない。線型推定量を考える理由は，ときにモデルの線型的構造から直観的に与えられる。しかしその直観的なものは，それを数学的に形式化した不変性原理と一致しない。推定量が $\boldsymbol{y}$ に関して線型であることは，標本空間で $\alpha_1\boldsymbol{y}_1+\alpha_2\boldsymbol{y}_2$ が観察されれば，$\boldsymbol{\beta}$ の推定値として $\alpha_1\hat{\boldsymbol{\beta}}(\boldsymbol{y}_1)+\alpha_2\hat{\boldsymbol{\beta}}(\boldsymbol{y}_2)$ を用いることを意味する。しかし§1で述べたように，実験計画等の場合と異って，経済分析では一般にパラメータは $\boldsymbol{y}$ と同じ物理的単位をもたず，$\boldsymbol{y}$ の変化が線型的に推定値に伝わるという性質は，最初からそれを要求するほど自然な構造といえないように思われる。他方，モデルが線型というのは，$\boldsymbol{\mu} \equiv E(\boldsymbol{y}) = \boldsymbol{X\beta}$ が $\boldsymbol{\beta}$ に関して線型であることであるが，これから $\boldsymbol{\beta}$ について解けば，$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mu}$ となり，$\boldsymbol{\mu}$ の推定量としては $\boldsymbol{y}$ が一つの自然な推定量であろうから，その意味で自然な $\boldsymbol{\beta}$ の推定量はただ1つしか存在しないことになる。次に不変性原理 (invariance principle) に基いた場合，線型推定量が構造的に自然な推定量であるかをみてみよう。(不変性については Ferguson (1967), Lehmann (1959) 参照。) 仮に損失函数を

(2.6)　$$L(\boldsymbol{a}, \boldsymbol{\beta}) = (\boldsymbol{a}-\boldsymbol{\beta})'\boldsymbol{Q}(\boldsymbol{a}-\boldsymbol{\beta})/\sigma^2$$

$\boldsymbol{Q}: k\times k$ 任意の既知の正値定符号行列

としよう。このとき変換群 (group of transformations) $G=\boldsymbol{R}^k\times\boldsymbol{R}^+$ をとり，$(\boldsymbol{g}, \alpha) \in G$ に対して標本空間の点 $\boldsymbol{y}$ を

(2.7)　$$\boldsymbol{y} \longrightarrow \alpha\boldsymbol{y}+\boldsymbol{Xg}$$

と変換した場合，対応してパラメータ空間の点 $(\boldsymbol{\beta}, \sigma^2) \in \Theta=\boldsymbol{R}^k\times\boldsymbol{R}^+$ 及び行動空間の点 $\boldsymbol{a} \in \Delta=\boldsymbol{R}^k$ を

(2.8)　$$\boldsymbol{\beta} \longrightarrow \alpha\boldsymbol{\beta}+\boldsymbol{g}, \quad \sigma^2 \longrightarrow \alpha^2\sigma^2$$

(2.9)　$$\boldsymbol{a} \longrightarrow \alpha\boldsymbol{a}+\boldsymbol{g}$$

と変換すれば，$\boldsymbol{\beta}$ の推定問題は不変である。ただし，ここでは $\boldsymbol{u}$ の分布が任意であるために，分布が (2.7) 及び (2.8) の変換に対して不変であることを仮定すれば通常の理論があてはまる。しかし，ここでは推定量の自然な構造にのみ興味があるために，分布の不変性の代わりに1次及び2次のモーメントの不変性でおきかえればよい。不変性原理によれば，$\boldsymbol{y}$ が観察されたとき $\boldsymbol{\beta}$ を (2.6) の損失函数に基いて $\boldsymbol{a}$ で推定することと，$\alpha\boldsymbol{y}+\boldsymbol{Xg}$ が観察されたとき $\alpha\boldsymbol{\beta}+\boldsymbol{g}$ を $\alpha\boldsymbol{a}+\boldsymbol{g}$ で推定することを同等とみなす。従ってこの場合，推定量 $\boldsymbol{\phi}$ の満たすべき自然な構造として

(2.10)　$$\boldsymbol{\phi}(\alpha\boldsymbol{y}+\boldsymbol{Xg}) = \alpha\boldsymbol{\phi}(\boldsymbol{y})+\boldsymbol{g}$$

を要求し，これを満たす推定量 $\boldsymbol{\phi}$ を不変推定量 (invariant estimator あるいは equivarient estimator) とよぶ。

以下不変推定量のクラスを $\mathscr{I}$ で示す。(2.10) では，$\boldsymbol{\phi}\in\mathscr{I}$ に対し標本空間に変換 (2.7) がほどこされたとき，対応して行動空間で変換 (2.9) を行えば両者は等しいことを示す。ここで注意したいのは，不変推定量のクラス $\mathscr{I}$ は正値定符号行列 $\boldsymbol{Q}$ のとり方に依存していない点である。このことは分散行列 (2.2) をリスク行列と考えてもよいことを示す。いま線型推定量 $\boldsymbol{A}\in\mathscr{L}$ が不変推定量であるとする。このとき (2.10) から

$$\boldsymbol{A}(\alpha\boldsymbol{y}+\boldsymbol{X}\boldsymbol{g})=\alpha\boldsymbol{A}\boldsymbol{y}+\boldsymbol{g}$$

が，すべての $\boldsymbol{y},\boldsymbol{g},\alpha$ に対して満たされなければならない。このことは明らかに $\boldsymbol{A}\boldsymbol{X}=\boldsymbol{I}_k$ と同等である。従って，線型不変推定量のクラスは，線型不偏推定量のクラス (2.5) に等しい：

$$\mathscr{L}\cap\mathscr{I}=\mathscr{L}\cap\mathscr{U} \tag{2.11}$$

しかし，例えば任意に固定した $k\times 1$ ベクトル $\boldsymbol{a}_0$ を用いて非線型推定量

$$\boldsymbol{\phi}_0(\boldsymbol{y})=\boldsymbol{b}+\|\boldsymbol{e}\|\boldsymbol{a}_0,\qquad \boldsymbol{e}=\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b} \tag{2.12}$$

を作ると，$\boldsymbol{\phi}_0\in\mathscr{I}$ であり，それゆえ $\mathscr{L}\cap\mathscr{I}\neq\mathscr{I}$ である。従って，

$$\mathscr{L}\cap\mathscr{U}\subset\mathscr{I}$$

となり，不変推定量のクラス $\mathscr{I}$ は，線型推定量のクラス $\mathscr{L}$ を含まないが，線型不偏推定量のクラス $\mathscr{L}\cap\mathscr{U}$ を含む。他方，(2.12) の推定量は不偏でないから，$\mathscr{I}$ は $\mathscr{U}$ に含まれない。

以上の議論では $\boldsymbol{u}$ の正規性を仮定しなかったが，それを仮定すると $(\boldsymbol{b},\boldsymbol{v})$ は十分統計量であり，$\boldsymbol{b}$ と $\boldsymbol{v}$ は独立に，

$$\boldsymbol{b}\sim N(\boldsymbol{\beta},\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}),\quad \boldsymbol{v}\sim\sigma^2\chi^2{}_{n-k} \tag{2.13}$$

であるから，これに変換群 $G=\boldsymbol{R}^k\times\boldsymbol{R}^+$

$$\begin{cases}\boldsymbol{b}\longrightarrow\alpha\boldsymbol{b}+\boldsymbol{g}, & \boldsymbol{v}\longrightarrow\alpha^2\boldsymbol{v}\\ \boldsymbol{\beta}\longrightarrow\alpha\boldsymbol{\beta}+\boldsymbol{g}, & \sigma^2\longrightarrow\alpha^2\sigma^2\end{cases} \tag{2.14}$$

を作用させても問題は不変である。この場合，不変推定量は

$$\boldsymbol{\phi}(\alpha\boldsymbol{b}+\boldsymbol{g},\alpha^2\boldsymbol{v})=\alpha\boldsymbol{\phi}(\boldsymbol{b})+\boldsymbol{g}$$

を満たすものであり，これから直ちに

$$\boldsymbol{\phi}(\boldsymbol{b},\boldsymbol{v})=\boldsymbol{b}+\sqrt{\boldsymbol{v}}\,\boldsymbol{c} \tag{2.15}$$

が導出される。ここで $\boldsymbol{c}$ は $k\times 1$ の固定したベクターである。変換 (2.14) は変換 (2.7)—(2.9) を十分統計量 $(\boldsymbol{b},\boldsymbol{v})$ で表わしたものに他ならないから，十分統計量 $(\boldsymbol{b},\boldsymbol{v})$ に基く不変推定量のクラス $\mathscr{D}\cap\mathscr{I}$ は，(2.15) と書き表わされる推定量のクラスとなる。すなわち，$\boldsymbol{u}$ についての正規性のもとでは

$$\mathscr{D}\cap\mathscr{U}=\{\boldsymbol{b}\}\subsetneqq\mathscr{D}\cap\mathscr{I}$$

となって，次の定理は不偏性の場合より意味が大きい。

[定理 1] 最小 2 乗推定量 $\boldsymbol{b}$ は最良不変推定量である。すなわち，$\boldsymbol{b}$ は $\mathscr{D}\cap\mathscr{I}$ の中で分散行列が最小となるものである。

証明は $\boldsymbol{\phi}\in\mathscr{D}\cap\mathscr{I}$ に対し

$$E[\boldsymbol{\phi}(\boldsymbol{b},\boldsymbol{v})-\boldsymbol{\beta}][\boldsymbol{\phi}(\boldsymbol{b},\boldsymbol{v})-\boldsymbol{\beta}]'$$
$$=\mathrm{cov}(\boldsymbol{b})+E(\boldsymbol{v})\boldsymbol{c}\boldsymbol{c}'\geqq\mathrm{cov}(\boldsymbol{b})$$

から明らかである。

最後に，$\boldsymbol{u}$ に正規性を仮定し，十分統計量 $(\boldsymbol{b},\boldsymbol{v})$ の term で問題を考えると，新しい変換群 $O^*(k)\times\boldsymbol{R}^+$，ただし

$$O^*(k)=\{\boldsymbol{A}:k\times k\,|\,|\boldsymbol{A}|\neq 0,\ \boldsymbol{A}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}'=\boldsymbol{X}'\boldsymbol{X}\}$$

が問題を不変にすることを指摘しておく。実際，$O^*(k)\times\boldsymbol{R}^+$ を，$(\boldsymbol{A},\alpha)\in O^*(k)\times\boldsymbol{R}^+$ に対し

$$\begin{cases}\boldsymbol{b}\longrightarrow\alpha\boldsymbol{A}\boldsymbol{b}, & \boldsymbol{v}\longrightarrow\alpha^2\boldsymbol{v}\\ \boldsymbol{\beta}\longrightarrow\alpha\boldsymbol{A}\boldsymbol{\beta}, & \sigma^2\longrightarrow\alpha^2\sigma^2\end{cases} \tag{2.16}$$

と作用させればよい。ここで $O^*(k)$ は $k\times k$ の直交行列群 $O(k)$ と同型で，従ってコンパクト群である。§3 では，$(\boldsymbol{b},\boldsymbol{v})$ の代わりに $(\boldsymbol{w},\boldsymbol{v})$ を用いるが，その場合 $O^*(k)\times\boldsymbol{R}^+$ は $O(k)\times\boldsymbol{R}^+$ となる。しかし，これらの変換群は，直接観察される $\boldsymbol{y}$ に対して作用するものでなく，十分統計量に要約して初めて群構造が生じたものである。その意味では自然な構造といいがたい。この点が正規分布の平均値の推定問題とニュアンスが異るところである。

§3 推定量の選択基準

3.1 推定量の選択は基本的にはリスク函数 (1.8) に基く。しかしすでにみたように，一様にリスク函数を最小にする推定量が存在しない。そのため一つの方法として，何らかの基準で $\mathscr{D}$ の一つの部分クラスをとり，そのクラスに推定量を制限し，その中でリスク函数 (1.8) をなるべく小さくするか，あるいはリスク函数の behavior が目的にかなうものを選ぶ。これは推定量 $\mathscr{D}$ の中でも，例えば恒等的に定数をとる推定量 $\boldsymbol{\phi}(\boldsymbol{w},\boldsymbol{v})\equiv\boldsymbol{c}$ のような初めから興味のない推定量をとり除くことになる。しかし，何が興味ある推定量かという点になると問題によって異るであろうし，決定理論的にはリスク函数の behavior から議論されるであろうが，ときにはそれ以外の基準が問題の性質から要請されるかもしれない。ここでは部分クラスをみつける基準として，便宜的に次の 4 つに分類しておく。

(1) 例えば，不偏性，一致性のように推定量が満たすことが望ましいと考えられる標本的性質をリスクと独立に課す。

(2) 線型性のように，リスクとは独立に推定量に何ら

かの構造を要求し，その構造をもつ推定量のみを考察する。

(3) 不変性のように，決定理論的な問題の構造と関連して自然な構造をもつ推定量を考える。

(4) 以下にみるミニマックス基準のように，リスク函数の behavior に何らかの性質を要求する。

われわれの基本的な立場として(4)の基準を採用するが，他の基準を全く捨てさるわけではない。

3.2 Admissibility と Minimaxity $\mathcal{T}$は推定量全体を示すことを再記する。

((定義1)) 2つの推定量 $\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 \in \mathcal{T}$ が与えられたとき，すべての $(\boldsymbol{\beta}, \sigma^2) \in \Theta$ に対し

$$R(\hat{\boldsymbol{\beta}}_2, (\boldsymbol{\beta}, \sigma^2)) \leq R(\hat{\boldsymbol{\beta}}_1, (\boldsymbol{\beta}, \sigma^2)) \tag{3.1}$$

が成立し，かつある $(\boldsymbol{\beta}_0, \sigma_0^2) \in \Theta$ に対して

$$R(\hat{\boldsymbol{\beta}}_2, (\boldsymbol{\beta}_0, \sigma_0^2)) < R(\hat{\boldsymbol{\beta}}_1, (\boldsymbol{\beta}_0, \sigma_0^2)) \tag{3.2}$$

が成立する場合，$\hat{\boldsymbol{\beta}}_2$ は $\hat{\boldsymbol{\beta}}_1$ よりよい推定量という。

((定義2)) 推定量 $\hat{\boldsymbol{\beta}}_1 \in \mathcal{T}$ に対し，$\hat{\boldsymbol{\beta}}_1$ よりよい推定量が存在するとき $\hat{\boldsymbol{\beta}}_1$ は inadmissible であるといい，inadmissible でない推定量を admissible 推定量という。

十分統計量 $(\boldsymbol{w}, \boldsymbol{v})$ に基く推定量のクラス $\mathcal{D}$ の中の admissible 推定量全体を $\mathcal{A}$ で示す。定義1,2より，$\hat{\boldsymbol{\beta}}_1$ が inadmissible であれば他にそれよりよい推定量の存在を許すから，一般性を失うことなく $\hat{\boldsymbol{\beta}}_1$ を選択の対象からはずすことができる。従って，リスクの視点から $\mathcal{D}$ の中で選択対象となる推定量は，admissible 推定量のクラス $\mathcal{A}$ である。しかし，$\mathcal{A}$ の中にも例えば，$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v}) \equiv \text{const.}$ となるような推定量も含まれており，admissibility は必要な性質であっても何ら optimality を示すものではない。

((定義3)) 推定量 $\boldsymbol{\phi}_0 \in \mathcal{T}$ が

$$\sup_{\Theta} R(\boldsymbol{\phi}_0, (\boldsymbol{\beta}, \sigma^2)) = \inf_{\boldsymbol{\phi}} \sup_{(\boldsymbol{\beta}, \sigma^2)} R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) < \infty \tag{3.3}$$

を満たすとき $\boldsymbol{\phi}_0$ をミニマックス推定量という。

以下 $\mathcal{D}$ の中のミニマックス推定量全体を $\mathcal{M}$ で示す。$\hat{\boldsymbol{\beta}} \in \mathcal{M}$ とし，(3.3)の右辺の値を l とすれば，すべての $(\boldsymbol{\beta}, \sigma^2) \in \Theta$ に対して

$$R(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \sigma^2)) \leq l \tag{3.4}$$

であり，ミニマックス推定量のクラス $\mathcal{M}$ の中にはリスクが l より大きくなる推定量は含まれない。例えば，恒等的に定数をとる推定量 $\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v}) \equiv \text{const.}$ や，線型バイアス推定量はミニマックス推定量となりえない。実際，線型推定量 $\boldsymbol{\phi}(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{y}$ に対してリスクは

$$R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) = \mathrm{tr}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{\beta}'(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{X})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{X})\boldsymbol{\beta}/\sigma^2 \tag{3.5}$$

となり，$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{I}$ でない限り $R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) \to \infty$ となりうる。$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{I}_k$ ならば

$$R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) = \mathrm{tr}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X} = \text{const.}$$

となるが，勿論これを最小にするのは最小2乗推定量 $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ の場合であって，そのリスクは

$$R(\boldsymbol{b}, (\boldsymbol{\beta}, \sigma^2)) = k$$

従って，(3.3)の右辺の値 l は k より大きくない。実際には $l=k$ であることを後にみる。これを先取りすれば，$\mathcal{D}$ の中のミニマックス推定量のクラス $\mathcal{M}$ は，最小2乗推定量を含むクラスとして

$$\mathcal{M} = \{\boldsymbol{\phi} \in \mathcal{D} | R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) \leq R(\boldsymbol{b}, (\boldsymbol{\beta}, \sigma^2)) = k, (\boldsymbol{\beta}, \sigma^2) \in \Theta\} \tag{3.6}$$

と表わされる。従って，ミニマックス推定量は最小2乗推定量に等しくない (a.e.) 限り，最小2乗推定量よりよい推定量である。また，minimaxity はリスク函数の behavior に関するひかえめな要求であって，パラメータ空間の構造に information がないという現在の仮定のもとでは，興味ある推定量はこのクラスに属していると考えられる。この意味で選択対象となる推定量のクラスは，$\mathcal{D}$ の中の admissible かつミニマックス推定量のクラス

$$\mathcal{A} \cap \mathcal{M}$$

といえよう。しかし $\mathcal{A} \cap \mathcal{M}$ が2つ以上の推定量を含む場合，その中から1つの推定量を選びだす基準を一般的に与えることは難しい。すなわち，$\hat{\boldsymbol{\beta}} \in \mathcal{A} \cap \mathcal{M}$ ならばそのリスク函数は k より大きくなく，また admissible であるから他にそれよりよい推定量が存在しない。このことは，リスク函数の $(\boldsymbol{\beta}, \sigma^2)$ に関する連続性から，他の任意の推定量 $\boldsymbol{\phi}_1$ に対して $\boldsymbol{\phi}_1$ のリスク函数が恒等的に $\hat{\boldsymbol{\beta}}$ のそれと等しくない限り，適当な点 $(\boldsymbol{\beta}_0, \sigma_0^2) \in \Theta$ の近傍 Δ^* をとれば

$$R(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \sigma^2)) < R(\boldsymbol{\phi}_1, (\boldsymbol{\beta}, \sigma^2)), \quad (\boldsymbol{\beta}, \sigma^2) \in \Delta^*$$

となることを意味する。この意味で，$\boldsymbol{\beta}$ に関する先験的情報がないという仮定のもとでは $\mathcal{A} \cap \mathcal{M}$ に属する推定量の間では，リスクの視点からは選択順序がつけられず，$\mathcal{A} \cap \mathcal{M}$ に属することが必要十分と考えられる。しかし，$\mathcal{A} \cap \mathcal{M}$ 自体を記述することは勿論，$\mathcal{A} \cap \mathcal{M}$ に属する推定量をみつけることも一般に容易でない。

以上の基準は，リスクの視点からのみの理想を述べたもので，分析の目的に応じた他の基準の選択をはばむものではない。更に，この論文では，$\mathcal{A} \cap \mathcal{M}$ に属することを理想としながらも，$\mathcal{A} \cap \mathcal{M}$ に属する適当な推定量をみつけることができないため，主として $\mathcal{M}$ に属する

推定量(最小2乗推定量よりよい推定量)に視点をあてている。

3.3 ベイズ推定量 $\mathscr{A}$ に属する推定量をみつける1つの方法として，ベイズの概念がある。$\boldsymbol{\Theta}=\boldsymbol{R}^k\times\boldsymbol{R}^+$ の空間をボレル可測空間として，$\boldsymbol{\Theta}$ の上に確率測度 π を考える。この π のことを事前確率分布とよび，事前確率分布全体を Π で示す。また，任意の $\boldsymbol{\beta}\in\mathscr{T}$ と任意の $\pi\in\Pi$ に対し

$$(3.7)\qquad r(\hat{\boldsymbol{\beta}},\pi)=\int R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))\,d\pi(\boldsymbol{\beta},\sigma^2)$$

で定義される函数 $r:\mathscr{T}\times\Pi\to\boldsymbol{R}^1$ を $\boldsymbol{\beta}$ の π に関するベイズリスクとよぶ。

((定義4)) 1つの $\pi\in\Pi$ に対して，$\hat{\boldsymbol{\beta}}\in\mathscr{T}$ が

$$(3.8)\qquad r(\hat{\boldsymbol{\beta}},\pi)=\inf_{\phi\in\mathscr{T}} r(\boldsymbol{\phi},\pi)<\infty$$

を満たすとき，$\hat{\boldsymbol{\beta}}$ を π に関するベイズ推定量という。また，$\hat{\boldsymbol{\beta}}$ が適当な $\pi\in\Pi$ をとれば(3.8)が成立するとき，$\hat{\boldsymbol{\beta}}$ をベイズ推定量という。

$\hat{\boldsymbol{\beta}}$ が π に関するベイズ推定量とする。このとき

$$\hat{\boldsymbol{\beta}}_0(\boldsymbol{w},\boldsymbol{v})=E[\hat{\boldsymbol{\beta}}/(\boldsymbol{w},\boldsymbol{v})]$$

と定義すると

$$R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))\leq R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))$$
$$r(\hat{\boldsymbol{\beta}}_0,\pi)=r(\hat{\boldsymbol{\beta}},\pi)$$

となるから，十分統計量 $(\boldsymbol{w},\boldsymbol{v})$ に基くベイズ推定量全体を考えればよく，それを $\mathscr{B}$ で示す。また，$(\boldsymbol{w},\boldsymbol{v})$ を与えたときの $(\boldsymbol{\beta},\sigma^2)$ の条件付分布である事後確率分布を，$\pi(\boldsymbol{\beta},\sigma^2|\boldsymbol{w},\boldsymbol{v})$ で示すと，π に関するベイズ推定量 $\hat{\boldsymbol{\beta}}$ は

$$(3.9)\qquad \hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v})=\frac{\int(\boldsymbol{\beta}/\sigma^2)\,d\pi(\boldsymbol{\beta},\sigma^2|\boldsymbol{w},\boldsymbol{v})}{\int(1/\sigma^2)\,d\pi(\boldsymbol{\beta},\sigma^2|\boldsymbol{w},\boldsymbol{v})}$$

で一意的に与えられる。

[補助定理1] $\pi\in\Pi$ に関してベイズ推定量が一意に定まれば，それは admissible である。

証明は Ferguson (1967) p. 60 参照。これは，$\mathscr{B}\subset\mathscr{A}$ を意味し，従って

$$(3.10)\qquad \mathscr{B}\cap\mathscr{M}\subset\mathscr{A}\cap\mathscr{M}$$

となる。それゆえ，$\mathscr{A}\cap\mathscr{M}$ に属する推定量をみつける1つの方法は，$\mathscr{B}\cap\mathscr{M}$ に属する推定量をみつけることである。

次に，推定量が $\mathscr{A}\cap\mathscr{M}$ に属するための必要条件を述べる。Π^* を Θ の上の σ-有限測度全体とする。$\pi\in\Pi^*$ に対し，形式的に(3.9)から計算した $\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v})$ が，$\|\hat{\boldsymbol{\beta}}(\boldsymbol{w},\boldsymbol{v})\|<\infty$ を満すとき，$\boldsymbol{\beta}$ を π に関する一般ベイズ(generalized Bayes)推定量とよぶ。一般ベイズ推定量全体を $\mathscr{B}^*$ で示す。Sacks (1963) は，平均2乗誤差のリスク函数のもとでは，ベイズ推定量の列の(適当な位相に関する)極限の推定量は，一般ベイズ推定量であることを示した。これは，$\mathscr{A}\subset\mathscr{B}^*$ を意味し，従って

$$(3.11)\qquad \mathscr{A}\cap\mathscr{B}\subset\mathscr{B}^*\cap\mathscr{M}$$

となる。すなわち，一般ベイズ推定量のクラス $\mathscr{B}^*$ に属さない推定量は admissible でなく，$\mathscr{A}\cap\mathscr{M}$ に属さない。

最後に，広義ベイズ (extendecl Bayes) の概念を定義しておく。ある推定量 $\boldsymbol{\phi}_0\in\mathscr{T}$ が広義ベイズであるというのは，任意の $\varepsilon>0$ に対し，適当に $\pi\in\Pi$ をとれば

$$r(\boldsymbol{\phi}_0,\pi)\leq\inf_{\phi} r(\boldsymbol{\phi},\pi)+\varepsilon<\infty$$

を満たす場合をいう。

[補助定理2] リスク函数が恒等的に定数となる推定量が広義ベイズならば，それはミニマックス推定量である。

証明は Ferguson (1967) p. 91 参照。これは次の節で用いられる。

§4 Elliptotically Symmetric 推定量

4.1 最小2乗推定量の性質 最小2乗推定量 $\boldsymbol{b}$ はミニマックス推定量，すなわち $\boldsymbol{b}\in\mathscr{M}$ であるが，$\boldsymbol{\beta}$ の次元 k が3以上のとき $\boldsymbol{b}$ は，inadmissible，すなわち $\boldsymbol{b}\notin\mathscr{A}$ であることを示す。$\boldsymbol{b}\in\mathscr{M}$ を示すためには，$R(\boldsymbol{b},(\boldsymbol{\beta},\sigma^2))=k$ であるから，補助定理2より広義ベイズ(extended Bayes)であることを示せばよい。いま(3.9)で，σ^2 の事前確率分布として $\sigma^2=1$ に確率1を与える分布をとり，$\boldsymbol{\beta}$ の事前確率分布として，$\boldsymbol{\beta}\sim N(\boldsymbol{0},a\boldsymbol{I})$ をとると，(3.9)よりベイズ推定量 $\boldsymbol{\phi}$ は

$$(4.1)\qquad \boldsymbol{\phi}(\boldsymbol{w},\boldsymbol{v})=\boldsymbol{A}\boldsymbol{b}=\left(\boldsymbol{X}'\boldsymbol{X}+\frac{1}{a}\boldsymbol{I}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$$

で与えられ，かつベイズリスクは

$$r(\boldsymbol{\phi},\pi)=\mathrm{tr}\,\boldsymbol{A}^2+\frac{1}{a}\mathrm{tr}\,(\boldsymbol{A}-\boldsymbol{I})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{A}-\boldsymbol{I})$$

$$\boldsymbol{A}=\left(\boldsymbol{X}'\boldsymbol{X}+\frac{1}{a}\boldsymbol{I}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}$$

となることから，任意の $\varepsilon>0$ に対し十分大きな $a>0$ をとることで

$$r(\boldsymbol{b},\pi)\leq r(\boldsymbol{\phi},\pi)+\varepsilon$$

となり，$\boldsymbol{b}$ は広義ベイズとなる。従って，すでに述べたように，ミニマックス推定量のクラス $\mathscr{M}$ は(3.6)で与えられる。

$k\geq3$ のとき $\boldsymbol{b}\notin\mathscr{A}$ を示すためには，$\boldsymbol{b}$ 以外にミニマックス推定量が存在することをいえばよい。Kudo (1955) 及び Kiefer (1957) は，一般の推定問題でミニマックス

推定量が存在すれば，適当な条件(これらは現在の問題で満たされる)のもとで不変な (invariant) ミニマックス推定量が存在することを証明した。そこで不変推定量を考察する。§2 で述べたように，十分統計量 $(\boldsymbol{b}, \boldsymbol{v})$ で問題を考えると，新しい変換群(2.16)は問題を不変にする。これは $(\boldsymbol{w}, \boldsymbol{v})$ の term では，変換群 $O(k)\times\boldsymbol{R}^+$ が

$$(4.2)\qquad \begin{cases} \boldsymbol{w}\longrightarrow\alpha\boldsymbol{p}\boldsymbol{w}, & \boldsymbol{v}\longrightarrow\alpha^2\boldsymbol{v} \\ \boldsymbol{\eta}\longrightarrow\alpha\boldsymbol{p}\boldsymbol{\eta}, & \sigma^2\longrightarrow\alpha^2\sigma^2 \end{cases}$$

と作用することと同等である。$((\boldsymbol{p}, \alpha)\in O(k)\times\boldsymbol{R}_0{}^+)$ このとき，$\boldsymbol{\eta}$ の不変推定量 $\boldsymbol{\psi}$ は，任意の $(\boldsymbol{w}, \boldsymbol{v})$ と任意の $(\boldsymbol{p}, \alpha)\in O(k)\times\boldsymbol{R}^+$ に対して

$$(4.3)\qquad \boldsymbol{\psi}(\alpha\boldsymbol{p}\boldsymbol{w}, \alpha^2\boldsymbol{v})=\alpha\boldsymbol{p}\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})$$

を満たすものであり，従って変換群(2.16)のもとでの $\boldsymbol{\beta}$ の不変推定量 $\boldsymbol{\phi}$ は，(4.3)を満たす $\boldsymbol{\psi}$ を用いて

$$(4.4)\qquad \boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})$$

と表わされる。

[補助定理 3]　変換群(2.16)のもとでの $\boldsymbol{\beta}$ の不変推定量 $\boldsymbol{\phi}$ は

$$(4.5)\qquad \boldsymbol{\phi}(\boldsymbol{b}, \boldsymbol{v})=g(\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}/\boldsymbol{v})\boldsymbol{b}$$

の形をもつ。ただし g は $\boldsymbol{R}^+\to\boldsymbol{R}$ のボレル可測函数である。

証明：$\|\boldsymbol{w}\|^2=\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}$ 及び(4.4)から，変換群(4.2)のもとでの $\boldsymbol{\eta}$ の不変推定量 $\boldsymbol{\psi}$ が

$$(4.6)\qquad \boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})=g(\|\boldsymbol{w}\|^2/\boldsymbol{v})\boldsymbol{w}$$

となることを示せばよい。(4.3)で $\alpha=1$, $\boldsymbol{w}=(c, 0, \cdots, 0)'\in\boldsymbol{R}^k\,(c\geq 0)$,

$$\boldsymbol{p}=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \varDelta & \\ 0 & & & \end{bmatrix}\qquad \varDelta\in O(k-1)$$

とおくと，$\boldsymbol{p}\boldsymbol{w}=\boldsymbol{w}$ 及び

$$(4.7)\qquad \boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})=\boldsymbol{p}\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})$$

となる。$\varDelta$ は任意だから，(4.6)は

$$\begin{aligned}\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})&=(\boldsymbol{\psi}_1(\boldsymbol{w}, \boldsymbol{v}), 0, \cdots, 0)'\in\boldsymbol{R}^k \\ &=(h(c, \boldsymbol{v}), 0, \cdots, 0)'\in\boldsymbol{R}^k\end{aligned}$$

を意味する。ただし h は $\boldsymbol{R}^+\times\boldsymbol{R}^+\to\boldsymbol{R}$ ボレル可測函数。次に任意の $\boldsymbol{w}\in\boldsymbol{R}^k$ に対して，$\alpha=1$，また $\boldsymbol{p}$ として最初の行が $\boldsymbol{w}'/\|\boldsymbol{w}\|$ である直交行列をとれば，$\boldsymbol{p}\boldsymbol{w}=(\|\boldsymbol{w}\|, 0, \cdots, 0)'\in\boldsymbol{R}^k$ となるから，(4.3)より

$$\begin{aligned}(4.8)\qquad \boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})&=\boldsymbol{p}'\boldsymbol{\psi}(\boldsymbol{p}\boldsymbol{w}, \boldsymbol{v}) \\ &=\boldsymbol{p}'(h(\|\boldsymbol{w}\|, \boldsymbol{v}), 0, \cdots, 0)'\in\boldsymbol{R}^k \\ &=h(\|\boldsymbol{w}\|, \boldsymbol{v})\boldsymbol{w}/\|\boldsymbol{w}\|\end{aligned}$$

となる。更にこれに(4.3)における α の不変性を要求すれば，(4.6)が結論される。

(4.15)の形の推定量を Elliptotically Symmetric 推定量とよび，このクラスを $\mathscr{E}$ で示す。明らかに $\boldsymbol{b}\in\mathscr{E}$ であるから，$\boldsymbol{b}$ 以外に $\mathscr{E}$ の中に $\mathscr{M}$ に属する推定量が存在するかどうかという問題が依然として残る。しかし Stein (1956) の議論に従えば，もし最小2乗推定量 $\boldsymbol{b}$ よりよい推定量が存在すれば，それは

$$(4.9)\qquad \boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=k(\|\boldsymbol{w}\|^2, \boldsymbol{v})\boldsymbol{b}$$

の形の推定量の中に存在する。実際，$\boldsymbol{\psi}(\boldsymbol{w}, \boldsymbol{v})$ を $\boldsymbol{w}$ よりよい $\boldsymbol{\eta}$ の推定量として

$$\hat{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{v})=\int\boldsymbol{p}'\boldsymbol{\psi}(\boldsymbol{p}\boldsymbol{w}, \boldsymbol{v})\,d\nu(\boldsymbol{p})$$

を定義すれば $\hat{\boldsymbol{\eta}}(\boldsymbol{Q}\boldsymbol{w}, \boldsymbol{v})=\boldsymbol{Q}\hat{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{v})$, $\boldsymbol{Q}\in O(k)$ を満たす。ただし ν は $O(k)$ の上の確率不変測度である。従って $\hat{\boldsymbol{\eta}}$ は(4.8)の形をもち

$$\begin{aligned}(4.9)'\qquad R_0(\hat{\boldsymbol{\eta}}, (\boldsymbol{\eta}, \sigma^2))&=E\|\hat{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{v})-\boldsymbol{\eta}\|^2/\sigma^2 \\ &\leq\int E\|\boldsymbol{\psi}(\boldsymbol{p}\boldsymbol{w}, \boldsymbol{v})-\boldsymbol{p}\boldsymbol{\eta}\|^2d\nu(\boldsymbol{p})/\sigma^2 \\ &\leq\int E\|\boldsymbol{w}-\boldsymbol{\eta}\|^2d\nu(\boldsymbol{p})/\sigma^2=k\end{aligned}$$

これは $\hat{\boldsymbol{\eta}}$ が $\boldsymbol{w}$ よりよい推定量であることを示す。スケール変換群 R^+ がコンパクトでないため，(4.9)の形を(4.5)の形に直接的には帰すことができない。次の定理は，$k\geq 3$ の場合，(4.9)でなく(4.5)の形の推定量で，$\boldsymbol{b}$ よりよいものが存在することを示している。

[定理 2]　(Stein (1956), James–Stein (1961))

(1)　$k=1, 2$ の場合

$$(4.10)\qquad \mathscr{A}\cap\mathscr{M}=\{\boldsymbol{b}\}\qquad \text{(a. e.)}$$

(2)　$k\geq 3$ の場合

$$(4.11)\qquad \boldsymbol{b}_1=[1-c\boldsymbol{v}/\|\boldsymbol{w}\|^2]\boldsymbol{b}$$

$$(4.12)\qquad \boldsymbol{b}_2=[1-c\boldsymbol{v}/\|\boldsymbol{w}\|^2]^+\boldsymbol{b}$$

はミニマックス推定量である。(i. e. $\boldsymbol{b}_1, \boldsymbol{b}_2\in\mathscr{M}$) また，$\boldsymbol{b}_1$ のリスク函数は

$$\begin{aligned}(4.13)\qquad &R(\boldsymbol{b}_1, (\boldsymbol{\beta}, \sigma^2)) \\ &=k-c(k-2)(n-k)E\left[\frac{1}{k-2+2\boldsymbol{J}}\right]\end{aligned}$$

である。ここで，$c=(k-2)/(n-k+2)$，また実数 a に対し $a^+=\max(a, 0)$ であり，$\boldsymbol{J}$ はパラメータ

$$(4.14)\qquad \lambda=\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}/2\sigma^2$$

をもつポアソン分布に従う確率変数である。

(3)　$\boldsymbol{b}_2$ は $\boldsymbol{b}_1$ よりよい推定量である。

定理2 (1)の証明は上記の論文に委ねる。(2)は定理3の特殊な場合である。(下の例1参照。) (3)の証明の概略は Appendix 2 で扱われている。また $\boldsymbol{b}_2$ のリスク函数については，下の例2及び Appendix 2 で扱われる。

なお，この節の終りで，(4. 13) より一般的に，$\boldsymbol{b}_1$ の平均，分散行列(及びリスク行列)が導かれる。$\boldsymbol{b}$ の inadmissibility は損失函数が，位置不変及び凸なものに対して一般の場合に証明されている。(Brown (1971) 参照。)

(4. 11) (4. 12) の $\boldsymbol{b}_1$ 及び $\boldsymbol{b}_2$，は通常 James–Stein 推定量とよばれる。この定理から，単純回帰式($k=1, 2$)の場合，最小2乗推定量 $\boldsymbol{b}$ は $\mathcal{A}\cap\mathcal{M}$ に属する唯一(a. e.)の推定量であるから，§3で述べた推定量選択基準によれば，これを採択する。他方，多重回帰式($k\geq 3$)の場合，最小2乗推定量は選択されない。また定理は，$k\geq 3$ のとき $\boldsymbol{b}_1$ は $\boldsymbol{b}$ よりよい推定量であり，$\boldsymbol{b}_2$ は $\boldsymbol{b}_1$ よりよい推定量であることを述べているが，$\boldsymbol{b}_2$ が admissible であることを述べるものではない。実際 $\boldsymbol{b}_2\notin\mathcal{A}$ であることが知られている。このことは，$\boldsymbol{b}_2$ のような形の推定量はベイズ推定量の列の極限となりえない，あるいは Sacks (1964) の定理を用いれば，それは一般ベイズ推定量となりえないであろうということから推察される。従って，$\boldsymbol{b}_1$ 及び $\boldsymbol{b}_2$ も $\mathcal{A}\cap\mathcal{M}$ に属さず，§3で述べた選択基準を適用すればそれらは選択からはずされることになる。しかし，$\mathcal{A}\cap\mathcal{M}$ に属する推定量をみつけることは一般に容易でなく，またみつけたにしても，必ずしもその推定量はリスク函数の behavior から言って望ましいものでないかもしれない。残念ながら，この論文では，$k\geq 3$ の場合に $\mathcal{A}\cap\mathcal{M}$ に属する推定量を例示していない。一方，$\mathcal{M}$ に属する推定量は，$\boldsymbol{b}$ を除けば $\boldsymbol{b}$ よりよい推定量であるから，形式的には $\boldsymbol{b}$ の代替物と考えてよいだろう。例えば，$\boldsymbol{b}_1$ についてみれば，リスク函数は $\lambda=0$ (i. e., $\boldsymbol{\beta}=0$) のとき最小値

$$R(\boldsymbol{b}_1, (0, \sigma^2))=k-c(n-k)\approx 2$$

をとる λ の単調増加な連続函数で，リスク函数の behavior からいえば望ましい性質を備えている。とくに，k が大きいとき，$\boldsymbol{\beta}=\boldsymbol{0}$ の近傍では $\boldsymbol{b}$ のリスクと $\boldsymbol{b}_1$ のリスクの差は，かなり大きなものになる。しかし，$\boldsymbol{b}_1$ の欠点は，もし $1<cv/\|\boldsymbol{w}\|^2$ ならば $\boldsymbol{b}$ の符号が変化し，無意味な推定値を生む可能性がある。他方，$\boldsymbol{b}_2$ は形式上これを避けた形になっている。実際，(4. 12) から $\boldsymbol{b}_2$ は

$$\boldsymbol{b}_2=I_{(c,\infty)}(\|\boldsymbol{w}\|^2/v)\,\boldsymbol{b}_1$$

と書ける。ただし，任意の集合 $A\subset\boldsymbol{R}$ に対し，$x\in A$ のとき $I_A(x)=1$, $x\notin A$ のとき $I_A(x)=0$ である。この式から，推定量 $\boldsymbol{b}_2$ は，$\|\boldsymbol{w}\|^2/v>c\equiv(k-2)/(n-k+2)$ のとき $\boldsymbol{\beta}$ を $\boldsymbol{b}_1$ で推定し，$\|\boldsymbol{w}\|^2/v\leq c$ のとき $\boldsymbol{\beta}$ を $\boldsymbol{0}$ で推定するいわゆる予備検定(仮説 $\boldsymbol{\beta}=\boldsymbol{0}$)に基く推定量であると考えることができる。この場合，有意水準は $P(\|\boldsymbol{w}\|^2/v>c|\boldsymbol{\beta}=\boldsymbol{0})$ であり，それは通常の場合より一般に大きく，仮説 $\boldsymbol{\beta}=\boldsymbol{0}$ が棄却された場合，最小2乗推定量 $\boldsymbol{b}$ でなく $\boldsymbol{b}_1$ が用いられるという点が通常の場合と異っている。以上から，$\boldsymbol{b}_2$ は単に定理2(3)の性質のみならず，上に述べた意味でも $\boldsymbol{b}_1$ よりよい $\boldsymbol{b}$ の代替物と考えられる。ただ，$\boldsymbol{b}_2$を用いる場合 $\|\boldsymbol{w}\|^2/v\leq c$ のとき，$\boldsymbol{\beta}$ の推定値として $\boldsymbol{0}$ を受け入れることができるかという問題が残る。なお，予備検定に基く推定量は次の§で扱われる。

最後に，$\boldsymbol{b}_1$ の平均，分散行列を導びく。$\boldsymbol{b}_1$ の平均は，Appendix 1 の補助定理2と5から

$$\begin{aligned}(4.15)\qquad E(\boldsymbol{b}_1)&=\boldsymbol{\beta}-c(n-k)(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}E(v\boldsymbol{w}/\|\boldsymbol{w}\|^2)\\&=\boldsymbol{\beta}-c(n-k)E[k+2\boldsymbol{J}]^{-1}\boldsymbol{\beta}\end{aligned}$$

となる。ただし $c\equiv(k-2)/(n-k+2)$。従ってバイアス行列は

$$\begin{aligned}(4.16)\qquad \mathrm{Bias}(\boldsymbol{b}_1)&=[\boldsymbol{\beta}-E(\boldsymbol{b}_1)][\boldsymbol{\beta}-E(\boldsymbol{b}_1)]'\\&=[c(n-k)E(k+2\boldsymbol{J})^{-1}]^2\boldsymbol{\beta}\boldsymbol{\beta}'\end{aligned}$$

となる。$\lambda=\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}/2\sigma^2$ であるから，もし $\boldsymbol{X}'\boldsymbol{X}/n$ が $n\to\infty$ のとき正値定符号行列に収束すれば，$\lambda\to\infty$ ($n\to\infty$) となり，$\mathrm{Bias}(\boldsymbol{b}_1)\to\boldsymbol{0}$ が証明される。$\boldsymbol{b}_1$ の分散行列は

$$\begin{aligned}(4.17)\qquad \mathrm{Var}(\boldsymbol{b}_1)&=\boldsymbol{\varDelta}-\mathrm{Bias}(\boldsymbol{b}_1)\\ \boldsymbol{\varDelta}&=E[\boldsymbol{b}_1-\boldsymbol{\beta}][\boldsymbol{b}_1-\boldsymbol{\beta}]'\end{aligned}$$

と表わされるから，$\boldsymbol{\varDelta}$ を計算する。$\hat{\boldsymbol{\eta}}_1=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{b}_1$ とおき，Appendix 1 補助定理 2, 4, 5 を用いると

$$\begin{aligned}(4.18)\qquad \boldsymbol{\varDelta}&=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}E[\hat{\boldsymbol{\eta}}_1-\boldsymbol{\eta}][\hat{\boldsymbol{\eta}}_1-\boldsymbol{\eta}]'(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\\
&=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}E[(\boldsymbol{w}-\boldsymbol{\eta})-cv\boldsymbol{w}/\|\boldsymbol{w}\|^2]\\
&\quad\times[(\boldsymbol{w}-\boldsymbol{\eta})-cv\boldsymbol{w}/\|\boldsymbol{w}\|^2]'(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\\
&=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}-2c(n-k)\\
&\quad\times\{E[\boldsymbol{\chi}^2(\lambda, k+2)]^{-1}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\\
&\quad+E[\boldsymbol{\chi}^2(\lambda, k+4)]^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\\
&\quad-E[\boldsymbol{\chi}^2(\lambda, k+2)]^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\}\\
&\quad+c^2(n-k)(n-k+2)\\
&\quad\times\{E[\boldsymbol{\chi}^2(\lambda, k+2)]^{-2}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\\
&\quad+E[\boldsymbol{\chi}^2(\lambda, k+4)]^{-2}\boldsymbol{\beta}\boldsymbol{\beta}'\}\\
&=\{1-2c(n-k)E[k+2\boldsymbol{J}]^{-1}\\
&\quad+c^2(n-k)(n-k+2)\\
&\quad\times E[k+2\boldsymbol{J})(k-2+2\boldsymbol{J})]^{-1}\}\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\\
&\quad+c(n-k)(k+2)\\
&\quad\times E[(k+2\boldsymbol{J})(k+2+2\boldsymbol{J})]^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\end{aligned}$$

となる。(4. 18) と (4. 16) から，$\boldsymbol{b}_1$ の分散行列が (4. 17) によって計算される。なお，(4. 18) とリスク函数の定義から

$$R(\boldsymbol{b}_1, (\boldsymbol{\beta}, \sigma^2)) = [\mathrm{tr}\, \boldsymbol{\Delta}(\boldsymbol{X}'\boldsymbol{X})]/\sigma^2$$

となる。これを Appendix 1 補助定理 5(12) を用いて評価すれば，(4. 13) ができる。$\boldsymbol{b}_2$ についても，例えば平均は，自由度 $(k+2, n-k)$，非心パラメータ λ をもつ非心 F 分布に従う確率変数 $\boldsymbol{z}$ を用いて，

$$E(\boldsymbol{b}_2) = \boldsymbol{\beta} P(\boldsymbol{z}>1) - \boldsymbol{\beta} E[I_{(1,\infty)}(\boldsymbol{z})/\boldsymbol{z}]$$

と表わされるが，分散行列は複雑すぎるのでここでは省略する。

4.2 Efron–Morris (1976) は，$k\geq 3$ の場合に比較的大きな $\mathcal{M}\cap\mathcal{E}$ の部分クラスを，興味ある議論によって導いている。まず (4. 5) の Elliptotically Symmetric 推定量を，一般性を失うことなく

(4. 19)　　$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v}) = [1-(k-2)\tau(\boldsymbol{F})/\boldsymbol{F}]\boldsymbol{b}$

と書き表わす。ここで τ は，$\boldsymbol{R}^+\rightarrow\boldsymbol{R}$ のボレル可測函数であり，また

$$\boldsymbol{F} = (n-k+2)\|\boldsymbol{w}\|^2/\boldsymbol{v}$$

である。$\boldsymbol{\phi}$ のリスク函数は (4. 14) における λ の函数であることが簡単に示されるし，また $\boldsymbol{\phi}$ は τ を与えれば決まる。このことを明示的に

$$\rho(\tau, \lambda) \equiv R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2))$$

と示し，与えられた τ に対し $\rho(\tau, \lambda)$ の推定問題を考える。τ が導函数 $\tau'(\boldsymbol{F})$ をもつ絶対連続函数で，$\rho(\tau, \lambda) < \infty$ であると仮定すると，$\boldsymbol{F}$ に基く $\rho(\tau, \lambda)$ の不偏推定量 $\boldsymbol{\rho}(\boldsymbol{F})$ が一意的に存在し，それは τ を用いて

(4. 20)
$$\boldsymbol{\rho}(\boldsymbol{F}) = k-(k-2) \times \{[(k-2)\tau(\boldsymbol{F})(2-\tau(\boldsymbol{F}))/\boldsymbol{F}] + 4\tau'(\boldsymbol{F})(1+c\tau(\boldsymbol{F}))\}$$

で与えられる。ただし (4. 20) の各項の期待値は存在するものとし，$c=(k-2)/(n-k+2)$ である。(Efron–Morris (1976) Theorem 1 参照。) 従って，(4. 20) の推定量がすべての $\boldsymbol{F}\geq 0$ に対して

(4. 21)　　$\boldsymbol{\rho}(\boldsymbol{F}) \leq k$

を満たせば

$$R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2)) = \rho(\tau, \lambda) = E\boldsymbol{\rho}(\boldsymbol{F}) \leq k$$

となり，$\boldsymbol{\phi}$ は Elliptotically Symmetric かつミニマックス推定量となる。実際 Efron–Morris は，(4. 21) が成立するための必要十分条件は，τ が絶対連続かつ以下の定理 3 の条件 (i) ～ (iii) を満たすことであることを証明している。従って，その条件を満たす τ をもつ推定量 $\boldsymbol{\phi}$ は $\mathcal{M}$ に属するが，もっと一般的に τ の絶対連続性を仮定しない次の定理が成立する。

[定理 3] $k\geq 3$ のとき，(4. 19) の推定量は，次の条件 (i) ～ (iii) を満たすときミニマックス推定量である。

(i)　　$0\leq\tau(\boldsymbol{F})\leq 2$

(ii)　$\tau(\boldsymbol{F})<2$ となるすべての $\boldsymbol{F}$ に対して

$$\gamma(\boldsymbol{F}) = \boldsymbol{F}^{\frac{1}{2}(k-2)}\tau(\boldsymbol{F})/(2-\tau(\boldsymbol{F}))^{1+2c}$$

は単調非減少函数である。

(iii)　もし $\tau(\boldsymbol{F}_1)=2$ となる $\boldsymbol{F}_1>0$ が存在すれば，すべての $\boldsymbol{F}\geq\boldsymbol{F}_1$ に対して $\tau(\boldsymbol{F})=2$ である。

Elliptotically Symmetric 推定量を (4. 19) の形に書いたとき，定理 3 の条件 (i) ～ (iii) を満たす推定量のクラスを $\mathcal{E}_0$ とすると，定理 3 は

$$\mathcal{E}_0 \subset \mathcal{M}\cap\mathcal{E}$$

であることを述べている。また上の定理から次の Baranchik (1970) の結果を得る。

[系] $k\geq 3$ のとき次の条件 (i) (ii) が満たされるならば，(4. 19) の推定量はミニマックス推定量である。

(i)　$0\leq\tau(\boldsymbol{F})\leq 2$

(ii)　$\tau(\boldsymbol{F})$ は単調非減少函数である。

$\mathcal{E}_0$ は $\mathcal{E}$ の中の比較的大きなミニマックス推定量のクラスであるが，$\mathcal{E}_0$ に属する推定量がいかなる条件のもとで admissible であるか，という問題については十分研究されていない。Efron–Morris (1976) によれば，次のような τ をもつ $\boldsymbol{\phi}\in\mathcal{E}_0$ は inadmissible である。

(a)　$\tau(\boldsymbol{F})$ は非増加で，$\tau(\boldsymbol{F})\geq -1/c$
この場合，James-Stein 推定量 $\boldsymbol{b}_1(\tau_1(\boldsymbol{F})=1)$ がよりよい推定量である。

(b)　$\tau'(\boldsymbol{F})\geq 0$ でかつ $\tau(\boldsymbol{F})\leq 1-\varepsilon$ $(\varepsilon>0)$
この場合，$\tau_2(\boldsymbol{F})=\tau(\boldsymbol{F})+\varepsilon$ をもつ推定量の方がよりよい。

(c)　$\tau(\boldsymbol{F})\geq 1+\varepsilon$，$\tau'(\boldsymbol{F})\leq (k-2)\varepsilon/4c\boldsymbol{F}$
この場合，$\tau_3(\boldsymbol{F})=\tau(\boldsymbol{F})-\varepsilon$ をもつ推定量がよりよい。

いくつかの例をみてみよう。

例 1　$\tau_1(\boldsymbol{F})=t\,(\mathrm{const})$,　$0\leq t\leq 2$

このとき上の系の条件が満たされ

(4. 22)　　$\boldsymbol{\phi}_1(\boldsymbol{w}, \boldsymbol{v}) = [1-(k-2)t/\boldsymbol{F}]\boldsymbol{b}$

は $\mathcal{M}$ に属する。(4. 20) から $\rho(\tau_1, \lambda)\equiv R(\boldsymbol{\phi}_1, (\boldsymbol{\beta}, \sigma^2))$ の不偏推定量は

(4. 23)　　$\boldsymbol{\rho}_1(\boldsymbol{F}) = k-(k-2)^2 t(2-t)/\boldsymbol{F}$

で与えられ，これを一様に小さくするのが James Stein 推定量 $\boldsymbol{b}_1$，すなわち $t=1$ の場合である。

例 2　$\tau_2(\boldsymbol{F})=\min(t, \boldsymbol{F}/(k-2))$,　$0\leq t\leq 2$

(4. 24)　　$\boldsymbol{\phi}_2(\boldsymbol{w}, \boldsymbol{v}) = [1-(k-2)t/\boldsymbol{F}]^+\boldsymbol{b}$

は $\mathcal{M}$ に属する。$t=1$ の場合が James–Stein 推定量 $\boldsymbol{b}_2$ である。定理 2(3) と同じように，同じ t の値に対しては，(4. 24) の推定量は (4. 22) の推定量よりよいことが示

される。従って，(4.22)の推定量は inadmissible である。また，$\boldsymbol{\phi}_2$ のリスク関数の不偏推定量は

(4.25) $$\boldsymbol{\rho}_2(\boldsymbol{F})=[k-(k-2)^2t(2-t)/\boldsymbol{F}]I_{[d,\infty)}(\boldsymbol{F})+[(n-k-2)\boldsymbol{F}/(n-k+2)-k]\times I_{(0,d)}(\boldsymbol{F})$$

$d=(k-2)t$ で与えられる。(Appendix 2 参照。) Efron-Morris (1973) (1976) は，t の選択として

$$t^*=\min(2,[1+1.34/(k-2)]/[1-2.66/(n-k+2)])$$

を提案している。これは $t(n-k)(k-2)/k(n-k+2)$ が，自由度 $(k, n-k)$ の F 分布のメディアンになる点である。(4.24)の推定量も inadmissible である。

例 3 $\tau(\boldsymbol{F})=tI_{((k-2)t,\infty)}(\boldsymbol{F})$,　$0\leq t\leq 2$

$$\boldsymbol{\phi}(\boldsymbol{w}.\boldsymbol{v})=[1-ctI_{(ct,\infty)}(\|\boldsymbol{w}\|^2/\boldsymbol{v})\boldsymbol{v}/\|\boldsymbol{w}\|^2]\boldsymbol{b}$$

も $\mathscr{M}$ に属する。この推定量は，$t\neq 1$ ならば，上の(b)(c)と同様にして inadmissible であることが証明される。$t=1$ のときの admissibility は明らかでないが，一般ベイズ推定量になりえないであろうから，inadmissible であることが推察される。

例 4 $\tau(\boldsymbol{F})=t\boldsymbol{F}/(\boldsymbol{F}+(k-2)t)$,　$0\leq t\leq 2$

(4.26) $$\boldsymbol{\phi}(\boldsymbol{w},\boldsymbol{v})=[\|\boldsymbol{w}\|^2/(\|\boldsymbol{w}\|^2+ct\boldsymbol{v})]\boldsymbol{b}$$

は $\mathscr{M}$ に属する。この推定量のリスク関数の不偏推定量は

(4.27) $$\boldsymbol{\rho}(\boldsymbol{F})=k-[(k-2)^2t/(\boldsymbol{F}+(k-2)t)^2]\times\{(2-t)\boldsymbol{F}+2(k-2)t+4t+4ct^2\boldsymbol{F}/(\boldsymbol{F}+(k-2)t)\}$$

となる。(4.26)の推定量は，$0\leq t<1$ ならば上の(b)が適用されて inadmissble となる。他方，$1\leq t\leq 2$ の適当な t に対しては，それは admissible であろうと推測する。

この節では，比較的広い $\mathscr{M}\cap\mathscr{E}$ の部分クラス $\mathscr{E}_0$ が与えられたにすぎず，$\mathscr{A}\cap\mathscr{M}$ に属する推定量という形では与えられていない。この節の結論として，例4の推定量で $t=1$ の場合，それを $\boldsymbol{b}$ の term で書けば

(4.28) $$\hat{\boldsymbol{\beta}}=[\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}/(\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}+c\boldsymbol{e}'\boldsymbol{e})]\boldsymbol{b}$$

$$\boldsymbol{e}=\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b},\quad c=(k-2)/(n-k+2)$$

となるが，これを $\boldsymbol{\beta}$ の推定量として提案する。この推定量のもつ意味をみるために，いま

(4.29) $$\boldsymbol{U}=\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}/d\boldsymbol{e}'\boldsymbol{e},\quad d=k/(n-k)$$

とおけば，$\boldsymbol{U}$ は仮説 $\boldsymbol{\beta}=\boldsymbol{0}$ を検定するいわゆる F 統計量であって，パラメータ $\lambda\equiv\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}/2\sigma^2$ をもつ自由度 $(k, n-k)$ の非心 F 分布をする。このとき(4.28)の $\hat{\boldsymbol{\beta}}$ は，$\boldsymbol{U}$ を用いて

(4.30) $$\hat{\boldsymbol{\beta}}=g(\boldsymbol{U})\boldsymbol{b},\quad g(\boldsymbol{U})=d\boldsymbol{U}/(d\boldsymbol{U}+c)$$

と書ける。g は $\boldsymbol{U}$ の単調増加函数であるから $g(\boldsymbol{U})$ 自体，$\boldsymbol{U}$ と同等な検定統計量であり，推定量 $\hat{\boldsymbol{\beta}}$ は，仮説 $\boldsymbol{\beta}=\boldsymbol{0}$ にウェイト(確率) $1-g(\boldsymbol{U})$ を与え，$\boldsymbol{\beta}(\neq 0)$ の推定量 $\boldsymbol{b}$ にウェイト(確率) $g(\boldsymbol{U})$ を与える推定量と考えることができる。他方，$\boldsymbol{U}$ は λ に関して単調尤度比 (monotone likelihood ratio) の性質をもつから，λ が大きければ大きい程($\boldsymbol{\beta}$ が仮説 $\boldsymbol{\beta}=\boldsymbol{0}$ から離れる程)，$\boldsymbol{U}$，従って $g(\boldsymbol{U})$ は確率的に大きな値 (stochastically larger) をとる。これから，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ を，常に $\boldsymbol{b}$ より絶対値において小さく推定するが，$\boldsymbol{\lambda}$ が大きくなるにつれて次第に $\boldsymbol{b}$ に近づくことがわかる。なお，例4で述べたように，$\hat{\boldsymbol{\beta}}$ の admissibility の問題は残されている。$\mathscr{E}_0$ に属する推定量が admissible であるためには，(4.9)の形をもつ推定量のクラスで admissible であればよいことが，(4.9)′と同様に示されるが，$\mathscr{E}$ の中での admissibility は推定量全体での admissibility を必ずしも意味しない。この点が σ^2 を既知とする場合と異る。

4.3 σ^2 が既知な場合に知られているいくつかの結果を要約しておく。これらの結果は，σ^2 が未知の場合にも同様な結果が成立するのではないかと推測させる意味をもつ。σ^2 が既知な場合，十分統計量は $\boldsymbol{w}$ だけとなり，$\boldsymbol{\beta}$ の推定量として $\boldsymbol{w}$ のみの函数を考える。この場合，Elliptotically Symmetric 推定量は

(4.31) $$\boldsymbol{\psi}(\boldsymbol{w})=h(\|\boldsymbol{w}\|^2)\boldsymbol{b}$$

である。(4.9)(4.9)′と同様にして，(1)もし $\boldsymbol{b}$ よりよい推定量が存在すれば，(4.31)の形をもつ推定量の中に存在する，(2)(4.31)の形をもつ推定量のクラスの中で admissible であれば，それは推定量全体の中で admissible であることが証明される。従って(4.31)の形をもつ admissible かつミニマックス推定量が1つの興味の対象となる。これに対して次の結果が成立する。

[定理 4] Strawderman (1970) (1971)

σ^2 を既知とする。(1) $k=1,2$ のとき，$\boldsymbol{b}$ は唯一の admissible かつミニマックス推定量である。(2) $k=3,4$ のとき，(4.26)の形をもつベイズかつミニマックス推定量は存在しない。(3) $k\geq 5$ のとき

$$\hat{\boldsymbol{\beta}}(\boldsymbol{w})=\left\{1-\frac{\sigma^2(k+2-2a)}{\|\boldsymbol{w}\|^2}+\frac{2\sigma^2\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{w}\|^2\right)}{\|\boldsymbol{w}\|^2\int_0^1 z^{\frac{1}{2}-a}\exp\left(-\frac{z}{2\sigma^2}\|\boldsymbol{w}\|^2\right)dz}\right\}\boldsymbol{b}$$

はベイズかつミニマックス推定量である。ただし，$\frac{1}{2}\leq\sigma\leq 1\,(k=5)$，$0\leq a\leq 1\,(k\geq 6)$

σ^2 が未知な場合に(2)(3)の結果を拡張するのは容易

でないように思われる。

§5　予備検定推定量

5.1　Testimator　変数選択，あるいはモデル選定問題を統計的に扱う場合，仮説検定を行い，その結果に基いたモデルにおいて未知の係数を推定する。その場合，採択されたモデルが正しいとして，その回帰係数の推定量として最小2乗推定量が選ばれ，それが最良線型不偏推定量といういい方がしばしばなされる。しかし，それは仮説検定によって得られた回帰式が「真」のモデルであるという条件付であって，他のモデルが正しい場合には，その最小2乗推定量は最良線型不偏推定量になりえない。更に問題になるのは，一体何を推定しているかということであり，それと関連して，このような仮説検定に基いたモデルにおける推定量の不偏性の概念である。上の場合，選ばれた回帰式が正しいものとしてその回帰係数に対する不偏性であって，1つのモデルが選択される以前には，パラメータ空間に2つの可能性があり(1つは他の1つの部分空間であるが)，不偏性の概念がはっきりしていない。以下この問題を取り扱う。回帰式 $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}$ において $\boldsymbol{u}$ の正規性 $N(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)$ を仮定し，一般線型仮説検定問題

(5.1)　$H_0:\boldsymbol{R\beta}=\boldsymbol{r}\quad \text{v.s.}\quad H_1:\boldsymbol{R\beta}\neq\boldsymbol{r}$

を考える。ここで $\boldsymbol{R}$ はランク q をもつ $q\times k$ の既知行列，$\boldsymbol{r}$ は $q\times 1$ の既知ベクターである。仮説 H_1 が正しい場合

(5.2)　$\boldsymbol{\delta}=\boldsymbol{R\beta}-\boldsymbol{r}$

は仮説のスペシフィケーションエラーを示すと考えられる。仮説 H_0 が正しい場合，制約付最小2乗推定量

(5.3)　$\hat{\boldsymbol{\beta}}=\boldsymbol{b}-\boldsymbol{A}(\boldsymbol{Rb}-\boldsymbol{r})$

(5.4)　$\boldsymbol{A}=(\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}$

が通常選択される。勿論この推定量は，仮説 H_0 のもとで最良不偏推定量である。また $\hat{\boldsymbol{\beta}}$ の標本分布は

(5.5)　$N(\boldsymbol{\beta}-\boldsymbol{A\delta},\ \sigma^2[(\boldsymbol{X'X})^{-1}-\boldsymbol{AR}(\boldsymbol{X'X})^{-1}])$

となり，線型仮説が正しくない場合 (i. e., $\boldsymbol{\delta}\neq\boldsymbol{0}$)，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ のバイアス推定量である。他方，対立仮説 H_1 のもとでは，最小2乗推定量 $\boldsymbol{b}$ が選択され，その標本分布は，どちらの仮説のもとでも

(5.6)　$N(\boldsymbol{\beta},\sigma^2(\boldsymbol{X'X})^{-1})$

である。従って $\boldsymbol{b}$ は H_0 のもとでも不偏推定量である。しかし (5.5) と (5.6) から直ちにわかるように，仮説の正しさに関係なく $\hat{\boldsymbol{\beta}}$ の分散行列は $\boldsymbol{b}$ の分散行列よりも，非負値定符号の意味で小さい。この視点に基いた変数選択の議論を 5.3 で扱う。

$\boldsymbol{\beta}$ の推定量として $\hat{\boldsymbol{\beta}}$ を用いるか $\boldsymbol{b}$ を用いるかは，検定統計量

(5.7)　$\boldsymbol{T}=(\boldsymbol{Rb}-\boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{Rb}-\boldsymbol{r})(n-k)$
$\div(\boldsymbol{y}-\boldsymbol{Xb})'(\boldsymbol{y}-\boldsymbol{Xb})q$

を用いて F 検定を行う。$\boldsymbol{T}$ は自由度 $(q, n-k)$，非心パラメータ

(5.8)　$\lambda=\boldsymbol{\delta}'[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{\delta}/2\sigma^2$

の F 分布をするから，有意水準 α に対応して決まる仮説 H_0 のもとでの F 値より $\boldsymbol{T}$ が大きければ，仮説 H_0 は棄却される。この F 値を c で示せば，仮説検定を行う前には $\boldsymbol{\beta}$ の推定量として

(5.9)　$\tilde{\boldsymbol{\beta}}=I_{(0,c]}(\boldsymbol{T})\hat{\boldsymbol{\beta}}+I_{(c,\infty)}(\boldsymbol{T})\boldsymbol{b}$

を選択していることになる。すなわちデータから計算される $\boldsymbol{T}$ の値が $(0, c]$ にあれば $\hat{\boldsymbol{\beta}}$ が選択され，そうでなければ $\boldsymbol{b}$ が選択される。推定量 $\tilde{\boldsymbol{\beta}}$ は予備検定推定量 (preliminary test estimator あるいは単に "testimator") とよばれる。$\tilde{\boldsymbol{\beta}}$ の期待値及び分散行列は Bock–Yancey–Judge (1973) が次のように導いている。

(5.10)　$E(\tilde{\boldsymbol{\beta}})=\boldsymbol{\beta}-EI_{(0,c)}(\boldsymbol{T})(\boldsymbol{b}-\hat{\boldsymbol{\beta}})$
$=\boldsymbol{\beta}-p_1(\lambda)\boldsymbol{A\delta}$

(5.11)　$\mathrm{Var}(\tilde{\boldsymbol{\beta}})=\sigma^2(\boldsymbol{X'X})^{-1}-\sigma^2p_1(\lambda)\boldsymbol{AR}(\boldsymbol{X'X})^{-1}$
$+[2p_1(\lambda)-p_1{}^2(\lambda)-p_2(\lambda)]\boldsymbol{A\delta\delta'A'}$

ただし

(5.12)　$p_j(\lambda)=P\left(\dfrac{\boldsymbol{\chi}^2(\lambda, q+2j)}{\boldsymbol{\chi}^2(n-k)}<\dfrac{cq}{n-k}\right).$

ここで，$\boldsymbol{\chi}^2(\lambda, q+2j)$ は自由度 $q+2j$ 非心パラメータ λ をもつカイ2乗分布に従う確率変数，$\boldsymbol{\chi}^2(n-k)$ は $\boldsymbol{\chi}^2(\lambda, q+2j)$ と独立な自由度 $n-k$ をもつカイ2乗分布に従う確率変数である。なお，(5.10) (5.11) の証明は Appendix の補助定理 2.4 を用いて示される。(5.10) から $\tilde{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の不偏推定量でない。すなわち $\hat{\boldsymbol{\beta}}$ は H_0 のもとで $\boldsymbol{\beta}$ の最良不偏推定量という理由で選択され，$\boldsymbol{b}$ は H_1 のもとで最良不偏推定量という理由で選択されるわけだが，結果的に選ばれる推定量は $\tilde{\boldsymbol{\beta}}$ であって，それはもはや $\boldsymbol{\beta}$ の不偏推定量でなく，従って $\boldsymbol{\beta}$ の推定という目的からは，不偏性をそれぞれの仮説のもとで推定量に課す意味がなくなる。不偏性の見地からは，$\tilde{\boldsymbol{\beta}}$ は平均的に (5.10) の $E(\tilde{\boldsymbol{\beta}})$ を推定していると考えられるが，勿論 $E(\tilde{\boldsymbol{\beta}})$ の推定に興味があるわけではない。他方，$\tilde{\boldsymbol{\beta}}$ を $\boldsymbol{\beta}$ の推定量と考える立場は，パラメータ空間を $\boldsymbol{R}^k$ 全体(あるいは $\boldsymbol{R}^k$ の開集合)として $\boldsymbol{\beta}$ の推定問題を考えていることになり，$\boldsymbol{R\beta}$ が近似的に $\boldsymbol{r}$ に等しいかもしれないという先験的情報を無視して，$\boldsymbol{\beta}$ の推定量を $\mathcal{D}$(十分統計量に基く推定量全体)の中からみつけよう。従って $\tilde{\boldsymbol{\beta}}$ も $\mathcal{D}$

の1つにすぎないとする立場であると考える。すなわち，(5.9)の形の推定量として結果的に表わされるにしても，そしてそれが5.2でみるように$\mathscr{D}$全体の中ではoptimalな性質をもたないにしても，それは先験的な情報を尊重した結果である。その際，もし仮説H_0が採択されたならば，モデルとしては$\boldsymbol{R\beta}=\boldsymbol{r}$の条件付のもとで$\boldsymbol{\beta}$の推定問題を考えることは，基本的に正しい態度であろう。問題は，1つのモデルが選択されたという条件付で考えるべき問題を，単に確率的な視点から条件をはずし，データをとる以前のアプリオリな推定量選択問題に帰している点にあるように思われる。同じ説明変数行列$\boldsymbol{X}$のもとで何組もの$\boldsymbol{y}$の値が観察可能であったにしても，上に述べたF検定を各組のデータに行い，それぞれの結果に対応して$\hat{\boldsymbol{\beta}}$または$\boldsymbol{b}$を選択していく想定は不自然である。$\tilde{\boldsymbol{\beta}}$の標本分布はこのような想定にたって初めて意味をもつものと考える。ただ一回限りのデータに対しても，標本分布をアプリオリな期待分布として解釈すれば別である。この問題は，同じデータで検定と推定を同時に行うという上の予備検定推定量という考え方自体に起因する。本来予備検定は別なデータを用いて行い，その結果に基いて手中のデータを推定するものであろう。しかし，経済分析においては予備検定のデータの利用可能性が問題になるであろう。とくに過去のデータでモデルを選択した場合，そこで選ばれたモデルはいわゆる構造的変化によって，もはや手中のデータのgeneratorとして考えるにはほど遠いという場合もしばしば起るであろう。従って，現実的には同じデータで検定と推定を行わざるをえないことになるかもしれない。このような場合には，$\tilde{\boldsymbol{\beta}}$の標本分布はアプリオリな期待分布という考え方になり，それを認めれば以下にみるように，そこから結論される$\tilde{\boldsymbol{\beta}}$の性質によって$\tilde{\boldsymbol{\beta}}$は選択されなくなる。

5.2 Nonoptimality of testimator　ここでは，予備検定推定量$\tilde{\boldsymbol{\beta}}$を，$\boldsymbol{\beta}$の1つの推定量としてリスクの視点からその性質を調べる。$\tilde{\boldsymbol{\beta}}$のリスク函数は，バイアス行列を

$$\mathrm{Bias}(\tilde{\boldsymbol{\beta}})=[E(\tilde{\boldsymbol{\beta}})-\boldsymbol{\beta}][E(\tilde{\boldsymbol{\beta}})-\boldsymbol{\beta}]'$$

とすれば，(5.10)(5.11)を用いて

$$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))=\{\mathrm{tr}\ \mathrm{Var}(\tilde{\boldsymbol{\beta}})\boldsymbol{X'X}+\mathrm{tr}\ \mathrm{Bias}(\tilde{\boldsymbol{\beta}})\boldsymbol{X'X}\}/\sigma^2$$

として次のように計算される。

$$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))=k-p_1(\lambda)q+[2p_1(\lambda)-p_2(\lambda)]2\lambda \tag{5.13}$$

ここで，$p_j(\lambda)$は(5.12)に定義されている。

［定理5］　(1)　$\lambda>q/2$ならば

$$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma_2))>k$$

従って，$\tilde{\boldsymbol{\beta}}$は$\boldsymbol{\beta}$の推定量としてミニマックス推定量でない。

(2)　$\lambda<q/4$ならば

$$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))<k.$$

証明：(1)　(5.12)から$p_1(\lambda)>p_2(\lambda)$が成立する。従って，(5.13)の$p_2(\lambda)$を$p_1(\lambda)$でおきかえれば

$$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))>k+(2\lambda-q)p_1(\lambda)>k$$

が成立する。(2)　$\Delta_j(\lambda)=1-p_j(\lambda)$とおくと

$$\begin{aligned}R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))&=k-q+\Delta_1(\lambda)(q-4\lambda)+\Delta_2(\lambda)2\lambda+2\lambda\\&<k-q+\Delta_2(\lambda)(q-4\lambda)+\Delta_2(\lambda)2\lambda+2\lambda\\&=k-(q-2\lambda)(1-\Delta_2(\lambda))<k\end{aligned}$$

が，$\Delta_1(\lambda)<\Delta_2(\lambda)$より成立する。

この結果はSclove–Morris–Radhakrishnan (1972)の定理1.2を一般化したものである。上の定理は，$\tilde{\boldsymbol{\beta}}$が$\mathscr{M}$に属さないことを示しているが，更に$q\geq3$のとき$\tilde{\boldsymbol{\beta}}$はinadmissibleである。これを示すため，$\tilde{\boldsymbol{\beta}}$を

$$\tilde{\boldsymbol{\beta}}=I_{(0,c)}(\boldsymbol{T})\hat{\boldsymbol{\beta}}_1+I_{(c,\infty)}(\boldsymbol{T})\tilde{\boldsymbol{b}}+\boldsymbol{a}_0 \tag{5.14}$$

と書き直す。ただし

$$\begin{cases}\hat{\boldsymbol{\beta}}_1=\hat{\boldsymbol{\beta}}-\boldsymbol{a}_0=\boldsymbol{b}-\boldsymbol{ARb}\\ \tilde{\boldsymbol{b}}=\boldsymbol{b}-\boldsymbol{a}_0=(\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{Xa}_0)\\ \boldsymbol{a}_0=\boldsymbol{Ar}=(\boldsymbol{X'X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{r}\end{cases} \tag{5.15}$$

である。これは非同次の制約$\boldsymbol{R\beta}=\boldsymbol{r}$を，同次の制約$\boldsymbol{R\gamma}=\boldsymbol{0}$，$\boldsymbol{\gamma}=\boldsymbol{\beta}-\boldsymbol{a}_0$に直したものである。すなわち，$\tilde{\boldsymbol{\beta}}-\boldsymbol{a}_0$は$\boldsymbol{\gamma}$の予備検定推定量である。ここで

$$\begin{cases}\boldsymbol{R}^*=\boldsymbol{R}(\boldsymbol{X'X})^{-\frac{1}{2}},\quad \boldsymbol{M}=\boldsymbol{R}^{*\prime}[\boldsymbol{R}^*\boldsymbol{R}^{*\prime}]^{-1}\boldsymbol{R}^*\\ \tilde{\boldsymbol{w}}=\boldsymbol{Q}(\boldsymbol{X'X})^{-\frac{1}{2}}\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{Xa}_0),\ \text{ただし}\ \boldsymbol{Q}\ \text{は}\\ \boldsymbol{QMQ}'=\begin{array}{c}\begin{bmatrix}\boldsymbol{I}_q & 0\\ 0 & 0\end{bmatrix}\end{array}\begin{array}{l}q\\ k-q\end{array}\ \text{なる直交行列,}\end{cases} \tag{5.16}$$

$$\tilde{\boldsymbol{w}}=\begin{pmatrix}\boldsymbol{w}_1\\ \boldsymbol{w}_2\end{pmatrix}\begin{array}{l}q\\ k-q\end{array} \tag{5.17}$$

とおくと

$$\begin{cases}\tilde{\boldsymbol{\gamma}}\equiv\tilde{\boldsymbol{\beta}}-\boldsymbol{a}_0=(\boldsymbol{X'X})^{-\frac{1}{2}}\boldsymbol{Q}'\{I_{(0,c)}(\boldsymbol{T})\tilde{\boldsymbol{\delta}}+I_{(c,\infty)}(\boldsymbol{T})\tilde{\boldsymbol{w}}\}\\ \tilde{\boldsymbol{\delta}}=\tilde{\boldsymbol{w}}-\boldsymbol{QMQ}'\tilde{\boldsymbol{w}}=(0',\boldsymbol{w}_2')'\\ \boldsymbol{T}=(n-k)\tilde{\boldsymbol{w}}'\boldsymbol{QMQ}'\tilde{\boldsymbol{w}}/q\boldsymbol{v}=(n-k)\boldsymbol{w}_1'\boldsymbol{w}_1/q\boldsymbol{v}\end{cases} \tag{5.18}$$

となる。更に

$$\begin{cases}\tilde{\boldsymbol{\tau}}=\boldsymbol{Q}(\boldsymbol{X'X})^{\frac{1}{2}}\tilde{\boldsymbol{\gamma}}=I_{(0,c)}(\boldsymbol{T})\tilde{\boldsymbol{\delta}}+I_{(c,\infty)}(\boldsymbol{T})\tilde{\boldsymbol{w}}\\ \boldsymbol{\tau}=\boldsymbol{Q}(\boldsymbol{X'X})^{\frac{1}{2}}\boldsymbol{\gamma}=\boldsymbol{Q}(\boldsymbol{X'X})^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{a}_0)\end{cases} \tag{5.19}$$

とおくと，$\tilde{\boldsymbol{\beta}}$のリスク函数は

$$\begin{aligned}\boldsymbol{R}(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))&=E[(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\boldsymbol{X'X})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})/\sigma^2]\\&=E[(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'(\boldsymbol{X'X})(\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma})/\sigma^2]\end{aligned} \tag{5.20}$$

$$=E[(\tilde{\boldsymbol{\tau}}-\boldsymbol{\tau})'(\tilde{\boldsymbol{\tau}}-\boldsymbol{\tau})/\sigma^2]$$

となる。これを

$$\tilde{\boldsymbol{w}}\sim N(\boldsymbol{\tau},\sigma^2\boldsymbol{I}_k)$$

を用いて評価したものが(5.13)である。(5.18)(5.19)で

(5.21)　$\tilde{\boldsymbol{\tau}}=\begin{pmatrix}I_{(c,\infty)}(\boldsymbol{T})\boldsymbol{w}_1\\ \boldsymbol{w}_2\end{pmatrix}\equiv\begin{pmatrix}\tilde{\boldsymbol{\tau}}_1\\ \tilde{\boldsymbol{\tau}}_2\end{pmatrix},\ \boldsymbol{\tau}\equiv\begin{pmatrix}\boldsymbol{\tau}_1\\ \boldsymbol{\tau}_2\end{pmatrix}$

とおくと，(5.20)のリスク函数は

(5.22)　$R(\tilde{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))=\varDelta_1(\tilde{\boldsymbol{\tau}}_1)+\varDelta_2(\tilde{\boldsymbol{\tau}}_2)$

(5.23)　$\varDelta_i(\tilde{\boldsymbol{\tau}}_i)=E[(\tilde{\boldsymbol{\tau}}_i-\tau_i)'(\tilde{\boldsymbol{\tau}}_i-\tau_i)/\sigma^2]\quad(i=1,2)$

と書き表わせる。(5.22)は，$\tilde{\boldsymbol{\beta}}$ のリスクは(5.23)の意味での $\tilde{\boldsymbol{\tau}}_1$ のリスクと $\tilde{\boldsymbol{\tau}}_2$ のリスクの和であることを示している。$\tilde{\boldsymbol{\tau}}_2$ を $\boldsymbol{\tau}_2$ の推定量と考えると，$\tilde{\boldsymbol{\tau}}_2\sim N(\boldsymbol{\tau}_2,\sigma^2\boldsymbol{I}_{k-q})$ であるから，$k-q\geq 3$ のとき $\tilde{\boldsymbol{\tau}}_2$ よりよい推定量が存在する。(§4参照。) 例えば，James-Stein 推定量

(5.24)　$\boldsymbol{\tau}_2{}^*=[1-c_2\boldsymbol{v}/\|\tilde{\boldsymbol{\tau}}_2\|^2]\tilde{\boldsymbol{\tau}}_2$

$c_2=(k-q-2)/(n-k+2)$ はその1つであり，(4.13)から，すべての $(\boldsymbol{\beta},\sigma^2)$ に対して

$$\varDelta_2(\boldsymbol{\tau}_2{}^*)=k-q-c_2(n-k)(k-q-2)E[k-q-2+2\boldsymbol{J}]^{-1}$$
$$<\varDelta_2(\tilde{\boldsymbol{\tau}}_2)=k-q$$

となる。ただし，$\boldsymbol{J}$ はパラメータ $\lambda_2=\boldsymbol{\tau}_2{}'\boldsymbol{\tau}_2/2\sigma^2$ をもつポアソン分布に従う確率変数である。他方，$\tilde{\boldsymbol{\tau}}_1$ に対しては次の補助定理が成立する。

[補助定理 3]　(Sclove–Morris–Radhakrishnan (1972))
$q\geq 3$ のとき

(5.25)　$\boldsymbol{\tau}_1{}^*=[1-c_1\boldsymbol{v}/\|\boldsymbol{w}_1\|^2]\tilde{\boldsymbol{\tau}}_1,\ c_1=(q-2)/(n-k+2)$

とおくと，すべての $(\boldsymbol{\beta},\sigma^2)$ に対して

$$\varDelta_1(\boldsymbol{\tau}_1{}^*)<\varDelta_1(\tilde{\boldsymbol{\tau}}_1)$$

が成立する。

従って，$q\geq 3$ または $(k-q)\geq 3$ のとき，$\tilde{\boldsymbol{\tau}}$ よりよい $\boldsymbol{\tau}$ の推定量，それゆえ $\tilde{\boldsymbol{\beta}}$ よりよい $\boldsymbol{\beta}$ の推定量が存在することになり，$\tilde{\boldsymbol{\beta}}$ は inadmissible となる。すなわち，いま $m_1=q,\ m_2=k-q$ とし

(5.26)　$h_i\equiv h_i(\boldsymbol{w}_i,\boldsymbol{v})=\begin{cases}1 & (m_i=1,2)\\ 1-c_i\boldsymbol{v}/\|\boldsymbol{w}_i\|^2 & (m_i\geq 3)\end{cases}$

(5.27)　$\boldsymbol{H}=\begin{bmatrix}h_1\boldsymbol{I}_q & 0\\ 0 & h_2\boldsymbol{I}_{k-q}\end{bmatrix}:k\times k$

とおくと，$q\geq 3$ または $k-q\geq 3$ のとき，すべての $(\boldsymbol{\beta},\sigma^2)$ に対して

(5.28)　$E\|\boldsymbol{\tau}^*-\boldsymbol{\tau}\|^2<E\|\tilde{\boldsymbol{\tau}}-\boldsymbol{\tau}\|^2,\quad(\boldsymbol{\tau}^*\equiv\boldsymbol{H}\tilde{\boldsymbol{\tau}})$

が成立する。$\boldsymbol{\tau}^*$ に対応する $\boldsymbol{\beta}$ の推定量は，$\tilde{\boldsymbol{\tau}}$ と $\tilde{\boldsymbol{\beta}}$ の関係(5.15)—(5.19)から

(5.29)　$\boldsymbol{\beta}^*=I_{(0,c)}(\boldsymbol{T})\boldsymbol{G}\tilde{\boldsymbol{\beta}}+I_{(c,\infty)}(\boldsymbol{T})\boldsymbol{G}\boldsymbol{b}$

となることがわかる。ここで $\boldsymbol{G}$ は

(5.30)　$\boldsymbol{G}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{Q}'\boldsymbol{H}\boldsymbol{Q}(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}$

と表わされる $(\boldsymbol{w},\boldsymbol{v})$ に依存する $k\times k$ の行列である。ただし $\boldsymbol{Q}$ は(5.16)で，$\boldsymbol{H}$ は(5.27)で与えられる。勿論 $\boldsymbol{H}=\boldsymbol{I}$ の場合 ($q\leq 2$ かつ $k-q\leq 2$) は $\boldsymbol{\beta}^*\equiv\tilde{\boldsymbol{\beta}}$ となる。(5.20)及び(5.28)と上の議論から次の定理が証明された。

[定理 6]　$q\geq 3$ または $k-q\geq 3$ のとき，(5.9)で定義される予備検定推定量 $\tilde{\boldsymbol{\beta}}$ は，$\boldsymbol{\beta}$ の推定量として inadmissible である。$\tilde{\boldsymbol{\beta}}$ よりよい一つの推定量は，(5.29)の $\boldsymbol{\beta}^*$ で与えられる。

定理5,6から，$q\geq 3$ または $k-q\geq 3$ のとき，予備検定推定量 $\tilde{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の推定量として admissible でなくまたミニマックス推定量でもない。従って，$\tilde{\boldsymbol{\beta}}$ は選択対象からはずされることになる。しかしこれは，$\tilde{\boldsymbol{\beta}}$ を $\boldsymbol{\beta}$ の推定量とみなす場合であって，すでに述べたように $\tilde{\boldsymbol{\beta}}$ を用いるのは，近似的に先験的情報 $\boldsymbol{R\beta}=\boldsymbol{r}$ が成立しているかもしれない，そしてそれが近似的に成立している場合それを積極的に利用していこうという立場である。他方，例えば変数選択のやり方としては，このような先験的情報を重視するのでなく，いわゆる親モデルですべてのあるいはいくつかの変数の組み合せを考え，データに最もフィットするものを選ぶという立場もあるわけで，このような場合，どの仮説も同じウェイトをもち，$\tilde{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ を推定していると考えることもできる。このような場合，1つの alternative な推定量として(5.29)の形のものを用いることができるであろう。なお，(5.27)の h_2 を $h_3=\max(0,h_2)$ でおきかえて(5.29)を作れば，更によい推定量が得られることは，定理2並びに上の議論から明らかである。

また一方，先験的情報 $\boldsymbol{R\beta}=\boldsymbol{r}$ を仮説検定することで1つのモデルを選択しても，選択後のモデルの回帰係数の推定量の選択は，最小2乗推定量である必要はない。勿論，F-検定を分散分析的に解釈可能であるが，得られたモデルの推定量としては，仮説 $\boldsymbol{R\beta}=\boldsymbol{r}$ が採択されたならば

$$\hat{\boldsymbol{\beta}}^*=[1-d_1\boldsymbol{v}/\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}]\hat{\boldsymbol{\beta}}$$

を用い，それが棄却されたならば

$$\boldsymbol{b}^*=[1-d_2\boldsymbol{v}/\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}]\boldsymbol{b}$$

を用いることができる。ただし，$q\geq 3$ のとき $d_1=(q-2)/(n-k+2)$，$q\leq 2$ のとき $d_1=0$，また $k-q\geq 3$ のとき $d_2=(k-q-2)/(n-k+2)$，$k-q\leq 2$ のとき $d_2=0$。この場合，予備検定推定量は

(5.31)　$\tilde{\boldsymbol{\beta}}^*=\boldsymbol{I}_{(0,c)}(\boldsymbol{T})\hat{\boldsymbol{\beta}}^*+\boldsymbol{I}_{(c,\infty)}(\boldsymbol{T})\boldsymbol{b}^*$

となる。ここで注意に値するのは，この推定量は(5.29)の推定量と異っている点である。(5.31)の推定量の性質についての吟味は，別の機会にゆずる。

上の議論は $q=1$ の場合にもそのまま適用されるから，t 検定に基く t 値の値による変数選択の問題も含まれていることを付加えておく。

5.3 5.1で述べたように，仮説の正しさに関係なく制約 $\boldsymbol{R\beta}=\boldsymbol{r}$ を課した推定量 $\hat{\boldsymbol{\beta}}$ の分散行列は，最小2乗推定量 $\boldsymbol{b}$ の分散行列より非負値定符号の意味で小さい。また，それらの分散行列は $\boldsymbol{\beta}$ に依存していない。他方，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ のバイアス推定量であって，$\boldsymbol{\delta}=\boldsymbol{R\beta}-\boldsymbol{r}$ に依存する。このtrade-off関係に注目して，Toro-Vizcarrondo及びWallace (1968)，Wallace (1972) は次の議論を展開している。まず，リスク行列において

(5.32)　$E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\leq E(\boldsymbol{b}-\boldsymbol{\beta})(\boldsymbol{b}-\boldsymbol{\beta})'$

が成立するような $(\boldsymbol{\beta},\sigma^2)$ の値に対しては，$\hat{\boldsymbol{\beta}}$ の方が $\boldsymbol{b}$ よりよいから，$\boldsymbol{R\beta}=\boldsymbol{r}$ が成立していなくても $\hat{\boldsymbol{\beta}}$ を用い，それ以外の場合には $\boldsymbol{b}$ を用いることを提案している。この方式を，変数選択あるいはモデル選択における MSE (mean squared error) 基準とよんでいる。この場合，未知の $(\boldsymbol{\beta},\sigma^2)$ が (5.32) を満たしているという仮説を検定することになる。(5.32) は

$$E\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|^2\leq E\|\boldsymbol{b}-\boldsymbol{\beta}\|^2$$

あるいは

(5.33)　$R(\hat{\boldsymbol{\beta}},(\boldsymbol{\beta},\sigma^2))\leq R(\boldsymbol{b},(\boldsymbol{\beta},\sigma^2))$

を意味する。このような考え方をする理由として，(i) 通常の有意水準 $\alpha=0.05$ あるいは $\alpha=0.01$ では，データをプールして用いる場合等では F 値（または t 値）が常に有意になりやすい，(ii) 帰無仮説 $\boldsymbol{R\beta}=\boldsymbol{r}$ は，データの数 n を限りなく大きくすれば結局棄却されてしまう，ことを挙げている。(ii) は F 検定（または t 検定）が一致性をもつことを述べている。

[補助定理4]　(5.32) が成立する $(\boldsymbol{\beta},\sigma^2)$ の集合は

$$\lambda=\boldsymbol{\delta}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{\delta}/2\sigma^2\leq\frac{1}{2}$$

を満たす集合である。(5.33) が成立する $(\boldsymbol{\beta},\sigma^2)$ の集合は，$\lambda\leq q/2$ を満たす集合である。

証明：(5.5) (5.6) から，(5.32) は

$$\begin{aligned}&E(\boldsymbol{b}-\boldsymbol{\beta})(\boldsymbol{b}-\boldsymbol{\beta})'-E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\\&=\sigma^2\boldsymbol{AR}(\boldsymbol{X}'\boldsymbol{X})^{-1}-\boldsymbol{A}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{A}'\\&=\sigma^2\boldsymbol{A}[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'-\boldsymbol{\delta}\boldsymbol{\delta}'/\sigma^2]\boldsymbol{A}'\end{aligned}$$

が非負値定符号であることを示す。これは，[　] の中が非負値定符号であることと同等であり，それから $\lambda\leq\frac{1}{2}$ が結論される。(5.33) はその結果を用いて容易に導びかれる。

従って，仮説 $\boldsymbol{R\beta}=\boldsymbol{r}$ は仮説 $\lambda\leq\frac{1}{2}$ でおきかえられ，仮説 $\lambda\leq\frac{1}{2}$ が採択されたときは $\hat{\boldsymbol{\beta}}$ を用い，それが棄却されたときには $\boldsymbol{b}$ を用いることになる。この仮説に対しても，(5.7) の統計量 $\boldsymbol{T}$ に基く F 検定は，一様最強力不変検定である。この場合，有意水準を $\alpha=0.05$ としても，$\lambda=\frac{1}{2}$ で有意点を決めるため，有意点は $\lambda=0$ の場合よりかなり大きくなり，上に述べた (i) は解消されることになる。しかし，$\lambda>\frac{1}{2}$ であっても λ が $\frac{1}{2}$ に近いときは，F 検定の検出力は α に近いから特にデータ数が小さいときは間違って $\lambda\leq\frac{1}{2}$ を採択する可能性が高い。それゆえ，ぎりぎりの $\lambda=\frac{1}{2}$ を検定するのでなく，それより小さな値，例えば $\lambda\leq\frac{1}{4}$ を検定する方がよいと考えられる。

例として，$\boldsymbol{y}=\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{X}_2\boldsymbol{\beta}_2+\boldsymbol{u}$ において，仮説 $\boldsymbol{\beta}_2=\boldsymbol{0}$ を検定する場合を考えてみよう。この場合，

$\hat{\boldsymbol{\beta}}'=(\hat{\boldsymbol{\beta}}_1',\boldsymbol{0}')$　$(\hat{\boldsymbol{\beta}}_1=(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y})$，$\boldsymbol{b}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

また $\lambda=\boldsymbol{\beta}_2'[\boldsymbol{X}_2'\boldsymbol{X}_2-\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_2]\boldsymbol{\beta}_2/2\sigma^2$ となる。仮に $\boldsymbol{X}'\boldsymbol{X}=I_k$ とすれば，$\lambda\leq\frac{1}{2}$ は $\boldsymbol{\beta}_2'\boldsymbol{\beta}_2\leq\sigma^2$ に等しく，これを満たす範囲の $\boldsymbol{\beta}_2$ に対してはそれを $\boldsymbol{0}$ で推定した方が $\boldsymbol{\beta}$ との平均距離が近いことになる。しかし，この場合どの変数を $\boldsymbol{0}$ とするかで問題が残り，仮説が先験的情報としてある程度はっきりしていないと，重要な変数も落す結果になる可能性がある。

結局，上の議論は F 検定の有意点を大きくすることに他ならず，上の推定量は予備検定推定量 (5.9) の形に書き表わされるから，$\boldsymbol{\beta}$ の推定量としては **5.2** の議論があてはまる。とくにここでの議論は，上の例のように $\hat{\boldsymbol{\beta}}'=(\hat{\boldsymbol{\beta}}_1',\boldsymbol{0}')$ を $\boldsymbol{\beta}$ の推定量としてみなしている性格が強く，その意味ではよりよい推定量である (5.29) の $\boldsymbol{\beta}^*$ にとって代わられる可能性もなくはない。すなわち，仮説が単に $\boldsymbol{\beta}$ の推定量としてリスクを小さくする目的でたてられるのならば，最小2乗推定量と比較するのでは十分でなく，他の competitors と比較されなければならない。

最後に，最初から $\boldsymbol{\beta}$ の推定問題を考えるが，先験的情報 $\boldsymbol{R\beta}=\boldsymbol{r}$ が近似的に成立しているかもしれないという状況についてふれておく。この場合，すでに述べたように1つの立場として，仮説検定により1つのモデル選択（決定）を行い，そのもとで推定するいわゆる予備検定推定量の立場がある。他の1つとして，§4で (4.28) 式に関連して述べたように，仮説 $\boldsymbol{R\beta}=\boldsymbol{r}$ に対する信頼性の尺度をデータに依存させ，それを検定統計量の単調函数として表わし，それでもって $\boldsymbol{R\beta}=\boldsymbol{r}$ のもとでの推定量と，そうでない場合の推定量とにウェイトをつける方法

が考えられる。具体的には，(5.7)の $\boldsymbol{T}$ は，仮説 $\boldsymbol{R\beta}=\boldsymbol{r}$ の成立の程度を表す尺度と考えられるから，$\boldsymbol{T}$ の単調非減少函数 h で，$h(0)=0, h(\infty)=1, 0\leq h(\boldsymbol{T})\leq 1$ なるものをとり，

$$\bar{\boldsymbol{\beta}}=(1-h(\boldsymbol{T}))\hat{\boldsymbol{\beta}}+h(\boldsymbol{T})\boldsymbol{b}$$

で $\boldsymbol{\beta}$ を推定する方法である。ただし，$\hat{\boldsymbol{\beta}}$ は(5.3)で与えられる。この考えは，これまでとられた例を筆者は知らないが，予備検定推定量の考えの拡張とみなすこともできる。ウェイト函数 h の選択と，推定量の性質については別の機会にゆずる。

§6　線型バイアス推定量

6.1　線型推定量の admissibility　すでにみたように，線型推定量は最小2乗推定量を除いてミニマックス推定量になりえない。一方，$k\geq 3$ のとき最小2乗推定量は inadmissible であった。このことは，誤差項 $\boldsymbol{u}$ の正規性の仮定のもとでは，最小2乗推定量は唯一の最良不偏推定量であるから(§2参照)，線型不偏推定量を含む他の不偏推定量は，ミニマックス推定量でなければ，また admissible でもないことを意味する。従って，リスクの視点に立つ限り，$k\geq 3$ のとき admissible 推定量のクラス $\mathscr{A}$ に属する不偏推定量は1つもないことになる。この事実は，現在のリスク函数(あるいは損失函数)の形及び $\boldsymbol{u}$ の正規性の仮定に特別強く依存するものでないことが知られている。この節では線型推定量の admissibility を考察するが，上の議論から1つの線型推定量が admissible であるとすれば，その推定量はバイアス推定量でなければならないことがわかる。

以下この節では $\boldsymbol{u}$ の正規性を仮定する。その場合，十分統計量 $(\boldsymbol{w}, \boldsymbol{v})$ に基く線型推定量は，$k\times k$ の正方行列 $\boldsymbol{G}$ を用いて

$$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=\boldsymbol{Gw}=\boldsymbol{G}(\boldsymbol{X'X})^{-\frac{1}{2}}\boldsymbol{X'y} \tag{6.1}$$

と書ける。$\boldsymbol{H}=(\boldsymbol{X'X})^{\frac{1}{2}}\boldsymbol{G}$ とおけば，$\boldsymbol{G}$ と $\boldsymbol{H}$ は一対一に対応するから，(6.1)のかわりに

$$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=(\boldsymbol{X'X})^{-\frac{1}{2}}\boldsymbol{H}(\boldsymbol{X'X})^{-\frac{1}{2}}\boldsymbol{X'y} \tag{6.2}$$

と最初から考えてもよい。このとき次の定理が成立する。

[定理7]　(Cohen (1966))

線型推定量(6.2)が admissible であるための必要十分条件は，(6.2)の $k\times k$ の行列 $\boldsymbol{H}$ が次の条件(i)(ii)を満たすことである。

(i)　$\boldsymbol{H}$ は対称行列である。

(ii)　$\boldsymbol{H}$ の固有値を h_i とすれば，$0\leq h_i\leq 1$ で，かつ $h_i=1$ なる固有値は高々2つである。

この定理の証明は Appendix 3 にある。

この定理の admissibility は，推定量全体のそれであって，線型推定量のクラスに限るものでないことを注意しておく。例として，$\boldsymbol{H}$ として対角行列で，対角要素 h_i が $0<h_i<1$ を満たすものをとれば，(6.2)の推定量は admissible 推定量となる。とくに，$\boldsymbol{H}=h\boldsymbol{I}_k\,(0<h<1)$ をとれば

$$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=h(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}=h\boldsymbol{b}$$

となり，これも admissible 推定量である。実際的な例としては次の節で扱う Ridge 推定量がある。

6.2　Ridge 推定量　最初は $\boldsymbol{u}$ の正規性を仮定しない。線型推定量 $\boldsymbol{A}\in\mathscr{L}(=\{k\times n$ の行列全体$\}$，§2参照)に対し，リスク行列を

$$\begin{aligned}\bar{R}(\boldsymbol{A}, (\boldsymbol{\beta}, \sigma^2)) &= E(\boldsymbol{Ay}-\boldsymbol{\beta})(\boldsymbol{Ay}-\boldsymbol{\beta})'/\sigma^2\\ &= \boldsymbol{AA}'+(\boldsymbol{AX}-\boldsymbol{I})\boldsymbol{\beta\beta}'(\boldsymbol{AX}-\boldsymbol{I})'/\sigma^2\end{aligned} \tag{6.3}$$

とする。また，与えられた $\boldsymbol{A}\in\mathscr{L}$ に対し，最小2乗推定量 $B=(\boldsymbol{X'X})^{-1}\boldsymbol{X}'$ のリスク行列とを比べて

$$\bar{R}(\boldsymbol{A}, (\boldsymbol{\beta}, \sigma^2))\leq\bar{R}(\boldsymbol{\beta}, (\boldsymbol{\beta}, \sigma^2))=(\boldsymbol{X'X})^{-1} \tag{6.4}$$

を満たす $(\boldsymbol{\beta}, \sigma^2)$ の集合を $\boldsymbol{\Phi}(\boldsymbol{A})$ で表わす。ただし，(6.4)の不等式は非負値定符号の意味である。$\boldsymbol{A}$ が不偏推定量 $(\boldsymbol{AX}=\boldsymbol{I}_k)$ ならば，最小2乗推定量は，非負値定符号の順序の意味でリスク行列の最小値を与える唯一のものであるから，$\boldsymbol{A}\neq\boldsymbol{B}$ である限り $\boldsymbol{\Phi}(\boldsymbol{A})=\phi$ (空集合)である。(6.3)の第2式第2項は非負値定符号行列であるから，$\boldsymbol{\Phi}(\boldsymbol{A})\neq\phi$ となるための条件は，$\boldsymbol{A}$ が

$$\mathscr{R}=\{\boldsymbol{A}\in\mathscr{L}\,|\,(\boldsymbol{X'X})^{-1}\geq\boldsymbol{AA}'\} \tag{6.5}$$

に属することである。これは，線型推定量で分散行列が最小2乗推定量の分散行列より大きくないクラスである。$\mathscr{R}$ に属する推定量を R-推定量とよぶことにする。いま，$\boldsymbol{A}$ を R-推定量としよう。このとき，(i)すべての σ^2 に対して $(\boldsymbol{0}, \sigma^2)\in\boldsymbol{\Phi}(\boldsymbol{A})$，(ii)各 σ^2 に対して $\boldsymbol{\Phi}(\boldsymbol{A})$ は $\boldsymbol{\beta}$ に関する凸集合，(iii)各 $\boldsymbol{\beta}$ に対して $\boldsymbol{\Phi}(\boldsymbol{A})$ は σ^2 に関する凸集合である。従って，ある $(\boldsymbol{\beta}_0, \sigma_0^2)\in\boldsymbol{\Phi}(\boldsymbol{A})$ ならば，$(\boldsymbol{0}, \sigma_0^2)$ と $(\boldsymbol{\beta}_0, \sigma_0^2)$ を結ぶ線分は $\boldsymbol{\Phi}(\boldsymbol{A})$ に属する。$\boldsymbol{\Phi}(\boldsymbol{A})$ は(6.4)を満たす $(\boldsymbol{\beta}, \sigma^2)$ の集合だから，これらの性質は任意の正値定符号行列 $\boldsymbol{Q}$ を用いて定義されるリスク函数

$$E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{Q}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \tag{6.6}$$

に対しても保存される。

$\mathscr{R}$ に属する推定量として，Hoerl–Kennard (1970) による Ridge 推定量

$$\boldsymbol{K}=(\boldsymbol{X'X}+\boldsymbol{D})^{-1}\boldsymbol{X}' \tag{6.7}$$

を取りあげよう。ただし，$\boldsymbol{D}$ は対角要素 $d_i\geq 0$ をもつ対角行列である。$\boldsymbol{K}\in\mathscr{R}$ であることは直ちにわかるが，

更に $\boldsymbol{K}$ の形から最小2乗推定量に比べて，重共線性に対してより安定的な推定量であることがわかる。$(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{D})^{-1}(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{D})=\boldsymbol{I}$ より，$\Phi(\boldsymbol{K})$ は，(6.3) から

(6.8) $$\boldsymbol{D\beta\beta}'\boldsymbol{D}/\sigma^2\leq 2\boldsymbol{D}+\boldsymbol{D}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{D}$$

を満たす $(\boldsymbol{\beta},\sigma^2)$ の集合である。$d_i>0\ (i=1,\cdots,k)$ ならば，(6.8) は

(6.9) $$\boldsymbol{\beta\beta}'/\sigma^2\leq 2\boldsymbol{D}^{-1}+(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

となり，$d=\max d_i$ とおけば，とくに

(6.10) $$\|\boldsymbol{\beta}\|^2/2\sigma^2\leq 1/d$$

なる集合は，$\Phi(\boldsymbol{K})$ に含まれる。明らかに (6.10) の $(\boldsymbol{\beta},\sigma^2)$ の集合は，d を小さくとれば大きくなる。従って，d_i をある程度小さくとれば，実際的と考えられる範囲の $(\boldsymbol{\beta},\sigma^2)$ に対して Ridge 推定量 $\boldsymbol{Ky}$ は最小2乗推定量よりよく，また重共線性に対してより安定的な推定量となる。とくに，先験的情報として

(6.11) $$\|\boldsymbol{\beta}\|\leq r,\quad \sigma_0{}^2\leq\sigma^2$$

が利用可能な場合，d_i として

$$0<d_i<\sigma_0{}^2/r\qquad(i=1,\cdots,k)$$

をとれば，(6.11) のすべての $(\boldsymbol{\beta},\sigma^2)$ に対して Ridge 推定量は最小2乗推定量よりよい推定量となる。(6.11) の $\|\boldsymbol{\beta}\|\leq r$ という情報は，多くの経済分析で利用可能であると考える。というのは，問題とする回帰方程式が何を表わす式かで，先験的に十分大きな r を与えることができる場合が多い。他方，$\sigma^2\geq\sigma_0{}^2$ の情報は，必ずしも先験的に得られるものでないが，$\sigma_0{}^2$ として被説明変数の有効数字より小さな正の値をとれば十分であろう。たとえば，日本の戦後のマクロ消費函数の場合，被説明変数は明らかに千円未満は意味がないと先験的に考えられるから，$\sigma_0=100$ あるいは $\sigma_0=10$ と与えることができる。最小2乗推定量に比べて Ridge 推定量の記すべきもう1つの性質は，$\boldsymbol{u}$ が正規分布 $N(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)$ をする場合，われわれのリスク函数 ((6.6) で $\boldsymbol{Q}=\boldsymbol{X}'\boldsymbol{X}$ とおいたもの) に関して Ridge 推定量は admissible であるという性質である。これは，(6.7) の $\boldsymbol{K}$ は

$$\boldsymbol{K}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}[(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{D})^{-1}(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}]\times(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{X}'$$

と書けるから，$\boldsymbol{H}\equiv(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{D})^{-1}(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}$ が定理7の条件 (i) (ii) を満たすことから証明される。Hoerl–Kennard (1970) は，対角行列 $\boldsymbol{D}$ として

$$d_i=la_i,\quad a_i \text{ は } \boldsymbol{X}'\boldsymbol{X} \text{ の第 } (i,i) \text{ 要素}$$

をとり，$l(>0)$ としていろいろな l の値に対する推定量をプロットすることによって決めることを提案している。Ridge 推定量の他の性質については，上記の論文の他に，Goldstein–Smith (1974)，Theobald (1974) 等を参照されたい。しかし，基本的な性質は上に述べた範囲をでるものではない。Ridge 推定量を実際に応用した例としては Brown–Beattie (1975) がある。

6.3 他の R-推定量の例をみてみよう。第1の例として，同次制約 $\boldsymbol{R\beta}=\boldsymbol{0}$ のもとでの制約付最小2乗推定量 ((5.3) で $\boldsymbol{r}=\boldsymbol{0}$ としたもの)

(6.12) $$\boldsymbol{C}=\boldsymbol{G}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

(6.13) $$\boldsymbol{G}=\boldsymbol{I}-(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']\boldsymbol{R}$$

がある。この推定量の分散行列は，(5.5) の分散行列に等しく，明らかに $\boldsymbol{C}\in\mathcal{R}$ であることがわかる。また，この推定量に関する議論，とくに $\boldsymbol{\beta}$ の推定量とみなした場合の議論はすでに§5.3でなされた。ここでは，補助定理4から

$$\Phi(\boldsymbol{C})=\left\{(\boldsymbol{\beta},\sigma^2)\,|\,\boldsymbol{\beta}'\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R\beta}/2\sigma^2\leq\frac{1}{2}\right\}$$

を再記しておく。$\boldsymbol{u}$ の正規性の仮定のもとで，推定量 $\boldsymbol{Cy}$ は $\boldsymbol{\beta}$ の推定量として $k-q\leq 2$ であれば admissible であるが，$k-q\geq 3$ ならば admissible でない。これは§5.3の議論の中で示されているが，定理7からも

$$\boldsymbol{C}=(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}[(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{G}(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}](\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{X}'$$

と書いたときに，$\boldsymbol{H}\equiv(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{G}(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}$ の固有値のうち1に等しいものの数が，$k-q\geq 3$ のとき3以上となることから証明される。明らかに $\boldsymbol{C}$ のランクは k より小さいが，一般に線型推定量 $\boldsymbol{A}\in\mathcal{L}$ のランクが k より小さいと，$\boldsymbol{\beta}$ の推定量として $\boldsymbol{Ay}$ は退化した標本分布をもち，$\boldsymbol{\beta}$ が近似的に k より小さな次元の部分空間に位置していると考えられる場合を除いては，そのような推定量は選択されないであろう。その意味では $\mathcal{R}$ の中で $\boldsymbol{\beta}$ の推定量として興味があるのはランク k をもつものである。しかし，次の例は，最小2乗推定量が選択されるところを，重共線性の懸念から結果的にランクが落ちた R-推定量を選んでいる。

第2の例は，分布ラグをもつ回帰式の場合である。分布ラグモデルについては，例えば Johnston (1972) を参照されたい。第 i 時点の回帰式は

$$y_i=\sum_{j=0}^{k-1}\beta_j x_{i-j}+u_i\qquad(i=1,\cdots,n)$$

で与えられるとする。この場合，第 i 時点の被説明変数は，1つの説明変数の時間のずれをもった k 個の値によって説明されるのであるから，説明変数の作る行列 $\boldsymbol{X}$ は，重共線性の可能性が強い。従って，たとえ $\boldsymbol{X}$ のランクが k であっても，通常最小2乗推定量は用いら

れず，より安定した推定量が用いられる。その1つとして Almon (1965) による推定量を取りあげる。そこでは，係数 β_j は近似的に1つの多項式

$$f(z)=\alpha_0+\alpha_1 z+\cdots+\alpha_{l-1}z^{l-1} \qquad (l<k)$$

のある特定な z_j に対応する値として考える。すなわち，$\beta_j=f(z_j)$ が近似的に成立すると考える。通常 $z_j=j$ が用いられる。このとき

$$\beta=Z\alpha \tag{6.14}$$

となるから，β の推定問題を α の推定問題に帰すことになる。(6.14)は，結局のところ β を l 次元 $(l<k)$ の部分空間におし込めたことを示している。l を適当にとることで，重共線性の問題は解消されると考えられる。(6.14) から，回帰式は $y=XZ\alpha+u$ となり，α の最小2乗推定量は

$$\hat{\alpha}=(Z'X'XZ)^{-1}Z'X'y$$

で与えられ，これから β の推定量が

$$\hat{\beta}=Z\hat{\alpha}=Z(Z'X'XZ)^{-1}Z'X'y\equiv Jy \tag{6.15}$$

として与えられる。この推定量は $\mathcal{R}$ に属し (i.e., $J\in\mathcal{R}$)，また $l\geq 3$ ならば定理7から admissible でないことが簡単に示される。なお，この議論は，(6.15) の $\hat{\beta}$ を β の1つの推定量とみなした場合で，(6.14) の成立を仮定するものでない。(6.14) が成立する場合は，$\hat{\beta}$ は β の最良線型不偏推定量であって，$b=(X'X)^{-1}X'y$ よりよい推定量であることを注意しておく。

APPENDIX

1. 以下 $\chi^2(\lambda, m)$ で自由度 m，非心パラメータ λ のカイ2乗分布をもつ確率変数を示す。この密度函数は次の式で与えられる。

$$g(w)=\sum_{j=0}^{\infty}(e^{-\lambda}\lambda^j/j!)f_{m+2j}(w) \tag{1}$$

ただし f_{m+2j} は自由度 $m+2j$ のカイ2乗分布の密度函数を示す。

補助定理 1. 確率変数 u が正規分布 $N(\theta,1)$ に従うとき，次の式が成立する。

$$Eh(u^2)u^2=Eh(\chi^2(\lambda,3))+\theta^2 Eh(\chi^2(\lambda,5)) \tag{2}$$

$$Eh(u^2)u=\theta Eh(\chi^2(\lambda,3)) \tag{3}$$

ただし，$\lambda=\theta^2/2$，h はボレル可測函数である。

証明： (2) の証明は，(1) で

$$wf_{m+2j}(w)=(m+2j)f_{m+2+2j}(w)$$

を用いれば直ちに計算できる。(3) は

$$\text{左辺}=\int_{\infty}^{\infty}h(u^2)u\exp\left[-\frac{1}{2}(u-\theta)^2\right]\Big/\sqrt{2\pi}\,du$$

$$=\int_0^{\infty}h(u^2)u\exp(-\lambda)\exp\left(-\frac{1}{2}u^2\right)\times[e^{u\theta}-e^{-u\theta}]/\sqrt{2\pi}\,du$$

で［ ］をテイラー展開し

$$\sqrt{\pi}\,\Gamma(2j)=\Gamma(j)\Gamma\left(j+\frac{1}{2}\right)2^{2j-1}$$

を用いて計算できる。

補助定理 2. $k\times 1$ の確率ベクター u が正規分布 $N(\theta, I)$ に従うとき，任意の定数 c に対して次の式が成立する。

$$Eh(c\|u\|^2)u=\theta Eh(c\chi^2(\lambda,k+2)) \tag{4}$$

$$Eh(c\|u\|^2)u'\theta=2\lambda Eh(c\chi^2(\lambda,k+2)) \tag{5}$$

ただし，$\lambda=\theta'\theta/2$ である。

証明： $u=(u_1,\cdots,u_k)'$ とすると，(3) を用いて

$$Eh(c\|u\|^2)u_1=EE[h(c(u_1^2+u_2^2+\cdots+u_k^2))|u_2,\cdots,u_k]$$
$$=\theta_1 Eh(c(\chi^2(\theta_1^2/2,3)+u_2^2+\cdots+u_k^2)$$
$$=\theta_1 Eh(c\chi^2(\lambda,n+2))$$

となる。(5) は (4) から明らかである。

(4) で定数 c を入れた理由は，c として u と独立な確率変数 v をとって，v が与えられたという条件付期待値と考えることができるからで，v について期待値をとれば無条件での期待値が与えられることになる。以下も同様である。

補助定理 3. $k\times 1$ の確率ベクター u が正規分布 $N(\theta, I)$ に従うものとする。このとき A を $k\times k$ の非負値定符号行列，c を定数として次の式が成立する。

$$Eh(c\|u\|^2)u'Au=trAEh(c\chi^2(\lambda,k+2))+\theta'A\theta Eh(c\chi^2(\lambda,k+4)) \tag{6}$$

ただし，$\lambda=\theta'\theta/2$ である。

証明： $c=1$ としてよい。P を A を対角化する直交行列とし，$PAP'=D=\mathrm{diag}\{d_1,\cdots,d_k\}$，$v=Pu$ とする。このとき

$$\text{左辺}=E[h(\|v\|^2)v'Dv]$$
$$=\sum_{i=1}^{k}d_i E\{E[h(v_1^2+\cdots+v_k^2)v_i^2|v_j, j\neq i]\}$$
$$=\sum_{i=1}^{k}d_i\{Eh\chi^2((p_i\theta)^2/2,3)+\sum_{j\neq i}v_j^2)$$
$$+(p_i\theta)^2 Eh(\chi^2((p_i\theta)^2/2,5)+\sum_{j\neq i}v_j^2)\}$$
$$=\text{右辺}$$

となる。ただし，p_i は P の第 i 行である。

補助定理 4. $k\times 1$ の確率ベクター u が正規分布 $N(\theta, I)$ に従うとき，次の式が成立する。

$$E[h(c\|u\|^2)uu']=E[h(c\chi^2(\lambda,k+2))]I+E[h(c\chi^2(\lambda,k+4))]\theta\theta' \tag{7}$$

ただし，c は定数，$\lambda=\boldsymbol{\theta}'\boldsymbol{\theta}/2$ である。

証明: (6) で $\boldsymbol{A}$ の行列として，第 (i, i) 要素 a_{ii} として 1 をとり，それ以外の要素をすべて 0 とすれば

$$(8)\qquad Eh(\|\boldsymbol{u}\|^2)u_i^2=Eh(\boldsymbol{\chi}^2(\lambda, k+2)) +\theta_i^2 Eh(\boldsymbol{\chi}^2(\lambda, k+4))$$

また，$a_{11}=a_{12}=a_{21}=a_{22}=1$，他の要素をすべて 0 とおけば，(6) から

$$(9)\qquad Eh(\|\boldsymbol{u}\|^2)(u_1^2+2u_1u_2+u_2^2) =2Eh(\boldsymbol{\chi}^2(\lambda, k+2)) +(\theta_1^2+2\theta_1\theta_2+\theta_2^2)Eh(\boldsymbol{\chi}^2(\lambda, k+4))$$

となる。(8) と (9) から

$$Eh(\|\boldsymbol{u}\|^2)u_1u_2=\theta_1\theta_2 Eh\boldsymbol{\chi}^2(\lambda, k+4))$$

となり，(7) が証明される。

補助定理 5. $\boldsymbol{J}$ をパラメータ λ をもつポアソン分布に従う確率変数とする。このとき

$$(10)\qquad E[\boldsymbol{\chi}^2(\lambda, k+2)]^{-1}=E(k+2\boldsymbol{J})^{-1}$$

$$(11)\qquad E[\boldsymbol{\chi}^2(\lambda, k+2)]^{-2}=E[(k+2\boldsymbol{J})(k-2+2\boldsymbol{J})]^{-1}$$

$$(12)\qquad 1=kE[k+2\boldsymbol{J}]^{-1}+2\lambda E[k+2+2\boldsymbol{J}]^{-1}$$

が成立する。

証明: いずれも直接計算から証明される。

2. 定理 2, 3 について——Efron-Morris (1976)

補助定理 6. $h(x)$ は絶対連続とし，以下の各項の積分は存在して有限であるとする。

$$(1)\qquad \boldsymbol{X}\sim N(\theta, \sigma^2) \Longrightarrow E(\boldsymbol{X}-\theta)h(\boldsymbol{X})=Eh'(\boldsymbol{X})\cdot\sigma^2$$

$$(2)\qquad \boldsymbol{W}\sim\alpha\chi_n^2 \Longrightarrow E(\boldsymbol{W}-n\alpha)h(\boldsymbol{W}) =2\alpha E\boldsymbol{W}h'(\boldsymbol{W})$$

証明: (1) $f(x)$ を $\boldsymbol{X}$ の密度函数とし，$H(x)=h(x)f(x)$ とおく。$H(x)$ は絶対連続で $\int|H(x)|dx<\infty$ だから，部分積分

$$\int_{-\infty}^{\infty}h'(x)f(x)dx=h(x)f(x)\Big|_{-\infty}^{\infty} -\int_{-\infty}^{\infty}h(x)f'(x)dx$$

より (1) が結論される。(2) も同様である。

(4.19) で $B(\boldsymbol{F})=(k-2)\tau(\boldsymbol{F})/\boldsymbol{F}$ とおくと，$\boldsymbol{\phi}$ のリスク函数 $R(\boldsymbol{\phi}, (\boldsymbol{\beta}, \sigma^2))\equiv\rho(\lambda)$ は

$$(2.1)\qquad \rho(\lambda)=E\|\boldsymbol{w}-\boldsymbol{\eta}-B(\boldsymbol{F})\boldsymbol{w}\|^2/\sigma^2 =k-2E\sum_{i=1}^{k}B(\boldsymbol{F})w_i(w_i-\eta_i)/\sigma^2 +EB^2(\boldsymbol{F})\boldsymbol{w}'\boldsymbol{w}/\sigma^2$$

となる。補助定理 6 (1) で，$h(w_i)=w_iB(\boldsymbol{F})$ とおき $\partial\boldsymbol{F}/\partial w_i=2(n-k+2)w_i/\boldsymbol{v}$ を用いると

$$(2.2)\qquad \rho(\lambda)=k-2kEB(\boldsymbol{F})-4EB'(\boldsymbol{F})\boldsymbol{F} +EB^2(\boldsymbol{F})\boldsymbol{w}'\boldsymbol{w}/\sigma^2$$

となる。更に $h(\boldsymbol{v})=B^2(\|\boldsymbol{w}\|^2(n-k+2)/\boldsymbol{v})\|\boldsymbol{w}\|^2/\boldsymbol{v}$ に補助定理の (2) を適用して，(2.2) の最後の項を書き替えると

$$(2.3)\qquad EB^2(\boldsymbol{F})\|\boldsymbol{w}\|^2/\sigma^2=\frac{n-k-2}{n-k+2}E\boldsymbol{F}B^2(\boldsymbol{F}) -\frac{4}{n-k+2}E\boldsymbol{F}^2B(\boldsymbol{F})B'(\boldsymbol{F})$$

これを (2.2) に代入して，$B(\boldsymbol{F})$ をもどすと (4.20) が $\rho(\lambda)=E\boldsymbol{\rho}(\boldsymbol{F})$ としてでる。一意性は $\boldsymbol{F}$ の分布の完備性からいえる。定理 3 は，まず $\tau(\boldsymbol{F})$ が絶対連続として，$\rho(\lambda)$ の不偏推定量 $\boldsymbol{\rho}(\boldsymbol{F})\leq k$ が成立するための必要十分条件は，$\tau(\boldsymbol{F})$ が (i)～(iii) を満たすことであることを示し，絶対連続函数で $\gamma(\boldsymbol{F})$ を近似することにより証明される。

(2.3) を (2.2) に代入した場合，$B(\boldsymbol{F})$ によって $\rho(\lambda)$ の不偏推定量は

$$(2.4)\qquad \boldsymbol{\rho}(\boldsymbol{F})=k-2kB(\boldsymbol{F})-4\boldsymbol{F}B'(\boldsymbol{F})+\boldsymbol{F}B^2(\boldsymbol{F}) -4\boldsymbol{F}B(\boldsymbol{F})[B(\boldsymbol{F})+\boldsymbol{F}B'(\boldsymbol{F})]/(n-k+2)$$

とも書ける。これから，James–Stein 推定量 $\boldsymbol{b}_2$ のリスク函数の不偏推定量が，(4.25) (ただし $t=1$) となることを簡単にみよう。$\boldsymbol{b}_2$ を，$\boldsymbol{b}_2=(1-B(\boldsymbol{F}))\boldsymbol{b}$，ただし

$$B(\boldsymbol{F})=(k-2)/\boldsymbol{F}+[1-(k-2)/\boldsymbol{F}]I_{[0,k-2]}(\boldsymbol{F})$$

と書く。$B(\boldsymbol{F})$ は $\boldsymbol{F}=k-2$ で微分可能でないが，絶対連続であり，$\boldsymbol{F}=k-2$ となる確率は 0 だから，上の式が適用できて (4.25) がでる。

なお，定理 2 (3) は，$\boldsymbol{b}_1$ 及び $\boldsymbol{b}_2$ のリスクを損失函数の積分の形に書き，積分範囲を，$\|\boldsymbol{w}\|^2/\boldsymbol{v}>c$ と $\|\boldsymbol{w}\|^2/\boldsymbol{v}\leq c$ にわけて符号を考えると，(3) が証明される。

3. 定理 7 の証明——Cohen (1966) 参照

補助定理 7. $\boldsymbol{w}\sim N(\boldsymbol{\eta}, \sigma^2\boldsymbol{I}_k)$ のとき $\boldsymbol{\eta}$ の線型推定量 $\boldsymbol{Kw}$ が admissible であるための必要十分条件は任意の直交行列 $\boldsymbol{Q}(k\times k)$ に対し $\boldsymbol{\eta}$ の推定量 $\boldsymbol{Q}'\boldsymbol{KQw}$ が admissible であることである。

証明:
$$R_0(\boldsymbol{Q}'\boldsymbol{KQ}, (\boldsymbol{\eta}, \sigma^2))=E\|\boldsymbol{Q}'\boldsymbol{KQw}-\boldsymbol{\eta}\|^2/\sigma^2 =E\|\boldsymbol{KQw}-\boldsymbol{Q\eta}\|^2/\sigma^2=E\|\boldsymbol{Kw}-\boldsymbol{\eta}\|^2/\sigma^2 =R_0(\boldsymbol{K}, (\boldsymbol{\eta}, \sigma^2))$$

より明らかである。

$\boldsymbol{\beta}$ を $(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}\boldsymbol{Hw}$ で推定する問題は，$\boldsymbol{\eta}=(\boldsymbol{X}'\boldsymbol{X})^{\frac{1}{2}}\boldsymbol{\beta}$ を $\boldsymbol{Hw}$ で推定する問題と同等である。いま $\boldsymbol{H}$ を定理 7 の条件 (i) を満たす行列とする。$\boldsymbol{H}$ は対称行列だから，上の補助定理から一般性を失うことなく h_i を対角要素にもつ対角行列とする。このとき，もし $0\leq h_i<1$ $(i=1, \cdots, k)$ ならば，事前確率分布 $\pi\in\boldsymbol{\Pi}$ として

$$d\pi(\boldsymbol{\eta})=\left[(2\pi)^{-\frac{p}{2}}\prod_{i=1}^{k}\lambda_i^{\frac{1}{2}}e^{-\lambda_i\eta_i^2/2}d\eta_i\right]d\xi(\sigma^2)$$

(3. 1)　　$\xi(\sigma^2=1)=1, \lambda_i=(1-h_i)/h_i$

をとれば，$\boldsymbol{Hw}$ が π に関する一意に決まるベイズ推定量となり，$\boldsymbol{Hw}$ は admissible である。ただし，$h_i=0$ ならば，η_1 に対する π の周辺分布は $\eta_i=0$ に確率1を与えるものとする。もし h_i が1つでも負であれば，h_i を $-h_i$ でおきかえることでよりよい推定量が得られる。さらに，もし $h_i>1$ ならば，h_i を1でおきかえればよりよい推定量が得られる。また，もし3つ以上の $h_i=1$ で，他の h_i は $0\leq h_i<1$ ならば，$h_i=1$ となる h_i を $\left[1-(p-2)/(n-k+2)\sum_{i=1}^{p} w_i^2\right]$ でおきかえれば，Stein 推定量 $\boldsymbol{b}_1$ の議論と同じように，よりよい推定量が得られることになる。ただし，p は $h_i=1$ となる h_i の数である。さらに，$h_1=h_2=1$ かつ $0\leq h_i<1$ $(i=3, \cdots, k)$ としよう。このとき $\boldsymbol{Hw}$ が inadmissible と仮定して

$$\boldsymbol{\phi}(\boldsymbol{w}, \boldsymbol{v})=(\phi_1(\boldsymbol{w}, \boldsymbol{v}), \cdots, \phi_k(\boldsymbol{w}, \boldsymbol{v})) \in \mathcal{D}$$

が $\boldsymbol{Hw}$ よりよい推定量とする。

(3, 2)　　$R_0(\boldsymbol{\phi}, (\boldsymbol{\eta}, \sigma^2)) \leq R_0(\boldsymbol{H}, (\boldsymbol{\eta}, \sigma^2))$

がすべての $(\boldsymbol{\beta}, \sigma^2) \in \boldsymbol{\Theta}$ に対して成立し，ある $(\boldsymbol{\eta}_0, \sigma_0^2) \in \boldsymbol{\Theta}$ に対し不等号が strict に成立する。(3. 2) から

$$\begin{aligned} R_0(\boldsymbol{\phi}, \boldsymbol{\eta}, \sigma^2)) &= \sum_{i=1}^{k} E(\phi_i(\boldsymbol{w}, \boldsymbol{v})-\eta_i)^2/\sigma^2 \\ &\leq 2+\sum_{i=3}^{k}[h_i^2+\eta_i^2(h_i-1)^2/\sigma^2] \\ &= R_0(\boldsymbol{H}, (\boldsymbol{\eta}, \sigma^2)) \end{aligned}$$

この両辺に $\prod_{i=3}^{k}(\lambda_i/2\pi)^{\frac{1}{2}}\exp(-\lambda_i\eta_i^2/\sigma^2)$ をかけ，η_i 及び $d\xi(\sigma^2)$ に関して積分すると

$$\text{(3. 3)} \qquad \begin{aligned} &\sum_{i=1}^{2} EE(\phi_i(\boldsymbol{w}, \boldsymbol{v})-\eta_i)^2 \\ &\quad +\sum_{j=3}^{k} EE(\phi_j(\boldsymbol{w}, \boldsymbol{v})-\eta_j)^2 \\ &\leq 2+\sum_{i=3}^{k} h_i \end{aligned}$$

となる。ただし，ξ 及び λ_i は (3. 1) で与えられる。他方，$0\leq h_i<1$ $(i=3, \cdots, k)$ だから，すでにみたように，$\boldsymbol{H}^*\boldsymbol{w}^*$ (ただし，$\boldsymbol{H}^*=\mathrm{diag}\{h_3, \cdots, h_k\}$，$\boldsymbol{w}^*=(w_3, \cdots, w_k)'$) は，上の事前確率分布に対して一意に決まる $(\eta_3, \cdots, \eta_k)'$ のベイズ推定量である。従って

$$\text{(3. 4)} \qquad \sum_{i=3}^{k} h_i<\sum_{j=3}^{k} EE(\phi_j(\boldsymbol{w}, \boldsymbol{v})-\eta_j)^2$$

となり，(3. 3) から

$$\text{(3. 5)} \qquad \sum_{i=1}^{2} EE(\phi_i(\boldsymbol{w}, \boldsymbol{v})-\eta_i)^2 \leq 2$$

となる。左辺の期待値は，平均 $(\eta_1, \eta_2, 0, \cdots, 0)'$，分散行列 $\mathrm{diag}\{1, 1, (h_3+1)/h_3, \cdots, (h_k+1)/h_k\}$ をもつ正規分布，及び $\boldsymbol{v}\sim\chi^2_{n-k}$ に関する期待値であるから，$\phi_i(\boldsymbol{w}, \boldsymbol{v})$ $(i=1, 2)$ は (η_1, η_2) の十分統計量 (w_1, w_2) に基く推定量のみを考えればよく，その場合 Stein の結果である定理2 (ただし σ^2 が既知) より $\phi_1(\boldsymbol{w}, \boldsymbol{v})=w_1, \phi_2(\boldsymbol{w}, \boldsymbol{v})=w_2$ となり，(3. 5) で等号が成立する。従って，(3. 4) (3. 5) から (3. 3) に矛盾する結果がでる。

次に (i) が成立しない，すなわち $\boldsymbol{H}$ が対称でないとする。このとき

$$\boldsymbol{H}^*=\boldsymbol{I}-[(\boldsymbol{H}-\boldsymbol{I})'(\boldsymbol{H}-\boldsymbol{I})]^{\frac{1}{2}}$$

とおくと，$\boldsymbol{H}^*$ は対称で，$\boldsymbol{H}^*\boldsymbol{w}$ は $\boldsymbol{Hw}$ よりよい推定量となる。実際，これを示すには，

$$R_0(\boldsymbol{H}, (\boldsymbol{\eta}, \sigma^2))=\mathrm{tr}\boldsymbol{HH}'+\boldsymbol{\eta}'(\boldsymbol{H}-\boldsymbol{I})'(\boldsymbol{H}-\boldsymbol{I})\boldsymbol{\eta}/\sigma^2,$$

$$\text{(3. 6)} \qquad (\boldsymbol{H}^*-\boldsymbol{I})'(\boldsymbol{H}^*-\boldsymbol{I})=(\boldsymbol{H}-\boldsymbol{I})'(\boldsymbol{H}-\boldsymbol{I})$$

だから，$\mathrm{tr}\,\boldsymbol{H}^{*\prime}\boldsymbol{H}^*<\mathrm{tr}\,\boldsymbol{H}'\boldsymbol{H}$ を示せば十分である。これは (3. 6) より，$\mathrm{tr}(\boldsymbol{H}-\boldsymbol{I})>\mathrm{tr}(\boldsymbol{H}^*-\boldsymbol{I})=-\mathrm{tr}\{(\boldsymbol{H}-\boldsymbol{I})'(\boldsymbol{H}-\boldsymbol{I})\}^{\frac{1}{2}}$ を，あるいは $\mathrm{tr}\{(\boldsymbol{H}-\boldsymbol{I})'(\boldsymbol{H}-\boldsymbol{I})\}^{\frac{1}{2}}>\mathrm{tr}(\boldsymbol{H}-\boldsymbol{I})$ を示すことに等しい。これは，一般に任意の非対称行列 $\boldsymbol{A}$ に対して $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{A})^{\frac{1}{2}}>\mathrm{tr}\,\boldsymbol{A}$ が成立することからいえる。

【一橋大学経済研究所：刈屋武昭】

参　考　文　献

[1] Alam, K. (1973), "A family of admissible minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, 517–525.

[2] Almon, S. (1965), "The distributed lag between capital appropriation and expenditures," *Econometrica*, **30**, 178–196.

[3] Baranchik, A. J. (1970), "A family of minimax estimators of the mean of a multivariate normal distribution," *Ann. Math. Statist.*, **41**, 642–645.

[4] Baranchik, A. J. (1973), "Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables," *Ann. Statist.*, **1**, 312–321.

[5] Bock, M. E., Yancey, T. A. and Judge, G. G. (1973), "Statistical consequences of preliminary test estimators in regression," *JASA*, **68**, 109–116.

[6] Brown, L. (1966), "On the admissibility of invariant estimators of one or more location parameters," *Ann. Math. Statist.*, **37**, 1087–1136.

[7] Brown, L. (1971), "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *Ann. Math. Statist.*, **42**,

[8] Brown, W. G. and Beattie, B. R. (1975), "Improving estimator of economic parameters by use of redge regression with production function

applications," *Amer. Jour. Agr. Econ.*, 22–32.

[9] Cohen, A. (1966), "All admissible linear estimates of the mean vector of a normal distribution," *Ann. Math. Statist.*, 37, 458–463.

[10] Efron, B and Morris, C. (1973), "Stein's estimation rule and its competitors–An empirical Bayes approach," *JASA*, **68,** 117–130.

[11] Efron, B and Morris, C. (1976), "Families of minimax estimators of a multivariate normal distribution," *Ann. Statist.*, **4,** 11–21.

[12] Ferguson, T. S. (1967), *Mathematical Statistics: Decision theoretic approach,* Academic Press, New York.

[13] Goldstein, M. and Smith, A. F. M. (1974), "Ridge type estimators for regression analysis," *Jour. Royal Statist. Soc.*, **B, 36,** 284–291.

[14] Hoerl, A. E. and Kennard, R. W. (1970), "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, **12,** 55–67.

[15] Hoerl, A. E. and Kennard, R. W. (1970), "Ridge regression: Applications to nonorthogonal problems," *Technometrics*, **12,** 69–82.

[16] James, W. and Stein, C. (1961), "Estimation with quadratic loss," *Fourth Berk. Symp., Math. Prob.*, 1, 361–380. Univ. of California, Berkley.

[17] Johnston, J. (1972), *Econometric Methods*, 2nd edition, McGraw–Hill, New York.

[18] Kiefer, J. (1957), " Invariance, minimax sequential estimation and continuous time process," *Ann. Math. Statist.*, 28, 573–601.

[19] Kudo, H. (1955), "On minimax invariant estimates of the transformation parameter," *Natural Science Report*, **6,** 31–73. Ochanomizu University.

[20] Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, Wiley, New York.

[21] Morris, K. W. (1961), "Admissible and minimax estimates of parameters in truncated spaces," *Ann. Math. Statist.*, **32,** 136–142.

[22] Sacks, J. (1963), "Generalized Bayes solutions in estimation problems," *Ann. Math. Statist.*, **34,** 751–768.

[23] Sclove, S. L. (1968), "Improved estimators for linear coefficients in linear regression," *JASA*, **63,** 596–606.

[24] Sclove, S. L. and Morris, C. and Radhakrshnan, R. (1972), "Non–optimality of preliminary–test estimators for the mean of multivariate normal distribution," *Ann. Math. Statist.*, **43,** 1481–1490.

[25] Stein, C. (1956), "Inadmissibility of the usual estimates for the mean of a normal distribution," *Third Berk. Symp. Math. Statist. Prob.*, **1,** 197–206, Univ. of California, Berkley.

[26] Stein, C. (1960), "Multiple regression," *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, 424–433, Stanford University Press, Stanford.

[27] Stein, C. (1966), "An approach to the recovery of interblock information in balanced incomplete block designs," *Research Papers in Statistics: Festshrift for NJ. Neyman* (F. N. David ed.,), 351–366. Wiley, New York.

[28] Strawderman, W. (1970), "On the existence of proper Bayes minimax estimators of the mean of a multivariate normal distribution," *Sixth Berk. Symp. Math. Statist.*, **1,** 51–55.

[29] Strawderman, W. (1971), "Proper Bayes minimax estimators of the multivariate normal distribution mean, *Ann. Math. Statist.*, **42,** 385–388.

[30] Strawderman, W. and Cohen, A. (1971), "Admissibility of the mean vector of a multivariate normal distribution with quadratic loss," *Ann. Math. Statist.*, **42,** 270–296.

[31] Theobald, C. M. (1974), "Generalization of mean square error applied to ridge regression," *Jour. Royal Statist.*, **B, 36,** 103–106.

[32] Toro–Vizcarrondo, C. and Wallace, T. D. (1968), "A test of the mean squared error criterion for restrictions in linear regression," *JASA*, **63,** 558–572.

[33] Wallce, T. D. (1972), "Weaker criteria and tests for linear restrictions in regression," *Econometrica*, **40,** 689–709.