回帰分析における検定仮説の識別可能性と OLSE と GLSE との間の選択*

刈屋 武昭

§1 序論と要約

線型回帰モデルにおける仮説検定問題は，回帰係数に関する線型仮説を検定する問題と，誤差項の分散行列の構造に関する仮説検定問題が代表的なものである。前者では F 検定(または t 検定)が用いられ，後者では時系列分析での1階マルコフ型の系列相関を検定する Durbin-Watson 検定が，OLSE(ordinary least squares estimator)を用いる目安として非常に頻繁に用いられている。しかしこれらの問題を考える場合，たてられた仮説がデータから一意的に識別されうるかという問題についてはこれまで議論されていない。これは，ある与えられた仮説に対して，データからみてこの仮説と同等に見える仮説(これを観察上同等な仮説と呼ぶ)が他に存在しないか，という問題である。これを仮説の識別可能性の問題と呼ぶことにする。この問題は，モデル(あるいは分布のクラス)の識別可能性の問題，特に計量経済学における同時方程式モデルの識別可能性の問題と区別すべきものである。いかなる統計的問題においても，このモデルの識別可能性は前もって保証されなければならない。従って仮説の識別可能性の問題を考察する場合，これを前提として議論する。1つの仮説の識別可能性は，事前的情報に基づいて選定されるモデルに依存することはいうまでもない。ある仮説が1つのモデルでは識別可能であるが，更に大きなモデルでは識別不可能であるということがしばしば起こる。他方，モデルは事前的情報に基づいてスペシファイされるが，その事前的情報は1つの確定的なモデルを選定するほど十分なものでない場合が多い。このような場合，いわゆる親モデルと呼ばれる大きなモデルを考えて，その親モデルの中で代替的なモデルを仮説検定したり，モデル判別を行ったりする。

この論文ではまず第2節で仮説の識別可能性を定義する。この定義は尤度原理(likelihood principle)に基づくものであり，2つの仮説(とくに帰無仮説)が恒等的に同じ最大尤度を与える場合，その2つの仮説は識別不可能とする。ここで仮説として考えるものは，与えられたパラメータ空間の中の(相対位相に関する)閉集合である。この識別性の定義は，通常の分布のクラスの識別性の定義の一般化とも考えられる。第3,4節では以上の定義に基づいて，線型回帰モデル

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

における仮説の識別可能性を考察する。ここで $\boldsymbol{X}$ は $n\times k$, $\mathrm{rank}(\boldsymbol{X})=k$ の既知行列，誤差項 $\boldsymbol{u}$ は平均 $\mathbf{0}$，分散行列 $E(\boldsymbol{uu}')=\boldsymbol{\Omega}$ をもつ正規分布に従うものとする。すなわち

(1.2) $\boldsymbol{y}\sim N(\boldsymbol{X\beta},\boldsymbol{\Omega})$, $\boldsymbol{\Omega}>0$

である。ただし，$\boldsymbol{\Omega}>0$ は $\boldsymbol{\Omega}$ が正値定符号行列であることを示す。第3節では $\boldsymbol{\Omega}=\sigma^2\boldsymbol{I}_n$(誤差項 $\boldsymbol{u}$ の要素 u_t は互いに独立で，分散一定)の仮定のもとで，一般線型仮説

(1.3) H_0: $\boldsymbol{R}_0\boldsymbol{\beta}=\boldsymbol{r}_0$, $\boldsymbol{R}_0$: $p\times k$, $\mathrm{rank}(\boldsymbol{R}_0)=p$

の識別可能性を示す。第4節では一般性を失うことなく $\boldsymbol{\Omega}=\sigma^2\boldsymbol{\Sigma}$, $|\boldsymbol{\Sigma}|=1$ として，$\boldsymbol{\Sigma}$ の構造に関する仮説は，モデルが観察上同等な仮説をすべて排除できるほど十分に小さくない限り，一般に識別不可能であることを示す。この $\boldsymbol{\Sigma}$ に関する仮説が一般に識別可能でないという性質は，その仮説を検定する場合 $\boldsymbol{\Sigma}$ の構造に関する情報は，残差

(1.4) $\boldsymbol{e}=\boldsymbol{Ny}$, $\boldsymbol{N}=\boldsymbol{I}-\boldsymbol{M}$, $\boldsymbol{M}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

という形でのみ利用可能であるという点に起因する。たとえば，帰無仮説 H_0: $\boldsymbol{\Sigma}=\boldsymbol{I}$ が対立仮説

* この研究は，「文部省特定研究経費」の助成を受けた。

$K:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\neq\boldsymbol{I}$ に対して識別可能でないということは，他に帰無仮説 $H_1:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1\neq\boldsymbol{I}$ が存在して，H_0 を K に対して検定する尤度比が H_1 を K に対して検定する尤度比に等しいことを意味するが，これは，$\boldsymbol{I}$ と $\boldsymbol{\Sigma}_K$ の相異が尤度に与える影響は残差(1.4)を通してしか測定されず，その結果 H_1 の介在を許すことになっている。従ってデータからみると H_0 と H_1 は区別できず，H_0 の採択は同時に H_1 の採択を意味することになる。このことは，H_0 のもとでは OLSE が適当であっても，H_1 のもとではそれは必ずしも有効でないかもしれない。この問題を扱うため，上の議論を OLSE と GLSE(generalized LSE, 一般化最小2乗推定量)の選択問題と結びつける。実際，$\boldsymbol{\Sigma}$ に関する1つの仮説に対してそれと観察上同等な仮説のクラスは，いわゆる Rao の分散行列の構造と関連している。Rao(1967)は，GLSE が OLSE と恒等的に等しいための必要十分条件は，$\boldsymbol{\Sigma}$ がクラス

(1.5)　$\mathcal{C}=\{\boldsymbol{\Phi}|\boldsymbol{\Phi}=\boldsymbol{X\Gamma X}'+\boldsymbol{Z\Delta Z}'+c\boldsymbol{I},\ \boldsymbol{X}'\boldsymbol{Z}=0,\ c>0,\ \boldsymbol{\Gamma},\boldsymbol{\Delta}$ は $\boldsymbol{\Phi}>0$ となる任意の対称行列$\}$

に属することであることを示した。従って，帰無仮説 $H_0:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0\in\mathcal{C}$ を対立仮説 $K:\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\notin\mathcal{C}$ に対して検定する問題を考察する。この場合，実際には H_0 および K に対してそれぞれ観察上同等な仮説のクラス $\mathcal{D}_0,\mathcal{D}_K$ が存在するから，事前的情報がない限りこの問題は，H_0 と観察上同等な仮説のクラス $\mathcal{D}_0$ を K と観察上同等な仮説のクラス $\mathcal{D}_K$ に対して検定することになる。第5節ではこれらのクラスを具体的に記述し，H_0 と観察上同等な仮説は必ずしも $\mathcal{C}$ に属さないことを示す。従って，仮説 H_0 が採択されても，それが観察上同等な仮説のクラス $\mathcal{D}_0$ の採択を意味する限り，$\mathcal{D}_0\not\subset\mathcal{C}$ であるので必ずしも OLSE の適用ということにならない。他方，任意に与えられた $\mathcal{C}$ に属さない分散行列 $\boldsymbol{\Sigma}_K(\notin\mathcal{C})$ に対して，$\mathcal{C}$ に属する分散行列で，$\boldsymbol{\Sigma}_K$ と仮説として観察上同等なもの(それを $\hat{\boldsymbol{\Sigma}}_K$ とする)が存在することが示される。それゆえ，分散行列に何ら事前的情報がない場合は，$\boldsymbol{\Sigma}_K$ が仮に $\mathcal{C}$ の外にあっても，すなわちもし $\boldsymbol{\Sigma}_K$ が既知であれば OLSE の代りに GLSE を用いる方が適当な場合であっても，データからみると，$\boldsymbol{\Sigma}_K$ と同じ最大尤度を与える $\hat{\boldsymbol{\Sigma}}_K$ が $\mathcal{C}$ の中にあり，$\hat{\boldsymbol{\Sigma}}_K$ のもとでは OLSE と GLSE は全く等しいので($\boldsymbol{\Sigma}_K$ が未知であろうが既知であろうが)常に OLSE を用いてよいと議論できるかもしれない。しかし，これはある程度はっきりした $\boldsymbol{\Sigma}$ に関する事前的情報がない限り，$\boldsymbol{\Sigma}$ に関する仮説は検定不能であることを示すことに過ぎない。例えば，Durbin-Watson 検定では，誤差項の1階のマルコフ性を検定しようとするが，この1階のマルコフ性という $\boldsymbol{\Sigma}$ の構造に関する事前的情報が近似的にデータの生成過程をとらえていない限り，$\boldsymbol{\Sigma}=\boldsymbol{I}$ という仮説が採択されても，実際に $\boldsymbol{\Sigma}=\boldsymbol{I}$ と観察上同等な仮説のクラスに属するものが採択されたにすぎないかもしれない。この意味で $\boldsymbol{\Sigma}=\boldsymbol{I}$ という仮説は検定不能である。他方，$\boldsymbol{\Sigma}_K\notin\mathcal{C}$ であるとき OLSE を用いると，周知のようにその GLSE に対する相対的効率は低くなる。そこで，OLSE と GLSE の選択問題を仮説検定によらず，相対的効率の大きさから眺めることにする。相対的効率を OLSE $\boldsymbol{b}$ と GLSE $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K)$ の分散行列の行列式の比 $\eta\equiv|\mathrm{Var}(\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K))|/|\mathrm{Var}(\boldsymbol{b})|$ と定義し，η が1に近くなる場合を考察する。$\boldsymbol{\Sigma}_K$ がクラス $\mathcal{C}$ に属していれば $\eta=1$ となるから，$\boldsymbol{\Sigma}_K$ と $\mathcal{C}$ との距離を考え，その距離で OLSE を用いた場合の妥当性を測る1つの測度とする。更に，もう1つの方法として，$\boldsymbol{\Sigma}_K\in\mathcal{C}$ は $\boldsymbol{X}$ の列ベクトルの張る空間が $\boldsymbol{\Sigma}_K$ のある k 個の固有ベクトルの張る空間と同等であることから，k 個の固有ベクトルの選び方を与え，$\boldsymbol{X}$ の列ベクトル空間と選ばれた k 個の固有ベクトルの張る空間との近さを表わす測度を導入し，OLSE を用いる妥当性を測る基準とする。これら2つの測度は $\boldsymbol{X}$ と $\boldsymbol{\Sigma}_K$ のみに依存し，確率ベクトル $\boldsymbol{y}$ に依存していない。またこの測度を系列相関の場合に適用する。第6節では，上の議論を系列相関の問題に適用する。

§2　仮説の識別可能性

実際的な見地から，モデルが密度関数のクラスとして与えられている場合のみを扱う。いま

$$F(\Theta)=\{f(\boldsymbol{x}|\boldsymbol{\theta})|\boldsymbol{\theta}\in\Theta\}$$

で，$\boldsymbol{\theta}$ をパラメータとする R^n で定義された(ある σ 有限測度 μ に関する)密度関数のクラスとする。ただし Θ は R^k の空でない開集合とする。比較のために，クラス $F(\Theta)$(または簡単に Θ)の識別可能性の定義を与えておく。

定義1 $F(\Theta)$ の密度関数 $f(\boldsymbol{x}|\boldsymbol{\theta}_0)$，または単に $\boldsymbol{\theta}_0 \in \Theta$ が識別可能であるとは，ある $\boldsymbol{\theta}_1 \in \Theta$ に対して

(2.1) $f(\boldsymbol{x}|\boldsymbol{\theta}_0)=f(\boldsymbol{x}|\boldsymbol{\theta}_1)$ a. e. $\boldsymbol{x}(\mu)$

ならば，$\boldsymbol{\theta}_0=\boldsymbol{\theta}_1$ が成立する場合をいう。各 $\boldsymbol{\theta}\in\Theta$ が識別可能なとき，$F(\Theta)$ または Θ は識別可能であるという。

いかなる統計的問題においても，モデル(すなわち密度関数のクラス)の識別可能性が前もって保証されていなければならない。従って，以下では $F(\Theta)$ の識別可能性を仮定し，この仮定のもとで Θ に関する仮説の識別可能性を定義する。これを定義するにあたって最初に注意すべき点は，Θ に関する仮説とは何かという問題である。Θ の任意の部分集合は論理的には統計的仮説たりうるかもしれない。しかし，例えば $\boldsymbol{\theta}$ は有理点であるという仮説は，仮説としては意味がある場合があるであろうし，ある目的からチェックしたいという要請がでるかもしれない。しかしこのような仮説に対して，意味のある形で統計的な検定を行うことは不可能であろう。一般的に統計的検定問題は，ある帰無仮説 $H:\boldsymbol{\theta}\in\Theta_0$ をある対立仮説 $K:\boldsymbol{\theta}\in\Theta_K$ $(\Theta_0\cap\Theta_K=\phi)$ という形で与えられるが，その際帰無仮説を作る Θ の部分集合 Θ_0 は，多くの重要な仮説において Θ の(相対位相に関して)閉部分集合である。これに対して対立仮説を作る Θ の部分集合 Θ_K は，開または閉部分集合であることが多い。この点に注目して，Θ 自体と空集合を除く閉部分集合全体を ϑ で表し，集合族 ϑ に対して識別可能性を定義する。そしてこれを仮説の識別可能性の定義とする。従って主として帰無仮説の識別可能性を念頭においていることになるが，定義としては帰無仮説，対立仮説の区別がなく，Θ の閉部分集合に対して定義される。

定義2 仮説の識別可能性

$\Theta_0\in\vartheta$ が識別可能であることは，ある $\Theta_1\in\vartheta$ に対して

(2.2) $\sup\limits_{\theta\in\Theta_0} f(\boldsymbol{x}|\boldsymbol{\theta})=\sup\limits_{\theta\in\Theta_1} f(\boldsymbol{x}|\boldsymbol{\theta})$ a. e. $\boldsymbol{x}(\mu)$

ならば，$\Theta_0=\Theta_1$ が成立する場合をいう。識別可能でない場合，Θ_0 は識別不能であるという。

明らかにこの定義は定義1の一般化であって，その一般化が仮説検定の議論と関連づけられている。定義2の仮説検定論的な意味を考える。いま帰無仮説 $H:\boldsymbol{\theta}\in\Theta_0$ を簡単に Θ_0 で，対立仮説 $K:\boldsymbol{\theta}\in\Theta_K$ を Θ_K で表すことにする。帰無仮説 $\Theta_0\in\vartheta$ (従って Θ_0 は Θ の閉部分集合)を対立仮説 Θ_K に対して検定する問題(従って $\Theta_0\cap\Theta_K=\phi$ が自動的に仮定される)で，$\Theta_0\in\vartheta$ が識別不能であるとしよう。そのときには定義2から，ある $\Theta_1\in\vartheta$，$\Theta_1\neq\Theta_0$ で(2.2)を満たすものが存在する。すなわち，2つの異なる閉部分集合 Θ_0 と Θ_1 に対して，それぞれのもとでの最大尤度が常に等しい。ここで，Θ_1 は必ずしも $\Theta_1\cap\Theta_K=\phi$ を満たさないし，Θ_K が ϑ に属している場合 $\Theta_1=\Theta_K$ であるかもしれないことを注意しておく。この場合 Θ_0 を Θ_K に対して検定する尤度比(likelihood ratio)は，Θ_1 を Θ_K に対して検定する尤度比(ただし $\Theta_1\cap\Theta_K\neq\phi$ かもしれない)に等しい。すなわち

$$(2.3)\quad \frac{\sup\limits_{\Theta_0} f(\boldsymbol{x}|\boldsymbol{\theta})}{\sup\limits_{\Theta_K} f(\boldsymbol{x}|\boldsymbol{\theta})}=\frac{\sup\limits_{\Theta_1} f(\boldsymbol{x}|\boldsymbol{\theta})}{\sup\limits_{\Theta_K} f(\boldsymbol{x}|\boldsymbol{\theta})}$$

が成立する。従ってこのような尤度比に基づく限り，与えられたデータ $\boldsymbol{x}$ からは Θ_0 と Θ_1 が識別できず，Θ_0 が Θ_K に対して採択されたとしても $\boldsymbol{\theta}\in\Theta_0$ か $\boldsymbol{\theta}\in\Theta_1$ か判定できないことになる。データから識別されるものは，Θ_0 と観察上同等な(すなわち(2.2)が成立する) ϑ の仮説のクラスである。形式的には ϑ の上に同値関係 $\sim$ を

$$\Theta_0\sim\Theta_1 \Longleftrightarrow (2.2)$$

で定義し，同値類の集合を $\tilde{\vartheta}=\vartheta/\sim$，$\Theta_0$ を含む同値類を $\tilde{\Theta}_0$ で表せば，帰無仮説 Θ_0 は $\tilde{\Theta}_0$ に属する任意の閉集合 Θ_1 と識別されず，データから識別できるものは $\tilde{\vartheta}$ の点である各同値類である。

仮説の識別可能性の定義は，モデルのスペシフィケーションに依存している。ある1つのクラス $F(\Theta)$ の中では帰無仮説 $\Theta_0\in\vartheta$ が識別可能であっても，$F(\Theta)$ を含む更に大きなクラス $F(\Theta^*)$ $(\Theta^*$

$\supset\Theta$)の中では識別不能となるかもしれない。勿論モデル $F(\Theta^*)$は定義1の意味で識別可能であることを前提する。

次に，仮説の識別可能性の定義は尤度原理に基づいていることに注意しよう。この原理に基づいた理由として，Fisher, Barnard 等の尤度に基づく統計的立場からの議論と関連ができ，更に Neyman-Pearson 的な立場からも多くの問題で尤度比検定は最適性を示していることによる。Neyman-Pearson の検定論からは，Θ_0 を Θ_K に対して検定する場合の尤度比検定は

(2.4)　$[\sup_{\Theta_0} f(\boldsymbol{x}|\boldsymbol{\theta})/\sup_{\Theta_K} f(\boldsymbol{x}|\boldsymbol{\theta})]<k$

のとき帰無仮説 Θ_0 を棄却することになる。ここで k は有意水準から決められるので，Θ_0 に対して(2.2)の代りに適当な $c>0$ に対して

(2.5)　$\sup_{\Theta_0} f(\boldsymbol{x}|\theta)=c\sup_{\Theta_1} f(\boldsymbol{x}|\boldsymbol{\theta})$　a. e. $\boldsymbol{x}(\mu)$

を満たす $\Theta_1\in\vartheta$ が存在すれば，(2.4)に基づく棄却域と尤度比

(2.6)　$[\sup_{\Theta_1} f(\boldsymbol{x}|\boldsymbol{\theta})/\sup_{\Theta_K} f(\boldsymbol{x}|\boldsymbol{\theta})]<k'$

に基づく棄却域とは同じものになる。ただし，$\Theta_1\cap\Theta_K\neq\phi$ かもしれない。従って，このような視点からは，定義2の(2.2)を(2.5)で置きかえて仮説の識別可能性を定義できることになる。

定義3　$\Theta_0\in\vartheta$ が Neyman-Pearson の意味で識別可能であるとは，ある $\Theta_1\in\vartheta$ に対して(2.5)が成立するとき，$\Theta_0=\Theta_1$ である場合をいう。

最後に，モデル選択の問題にふれておく。いま $\Theta_0\in\vartheta$, $\Theta_1\in\vartheta$ での最尤推定量をそれぞれ $\hat{\boldsymbol{\theta}}_0=\hat{\boldsymbol{\theta}}_0(\boldsymbol{x})$, $\hat{\boldsymbol{\theta}}_1=\hat{\boldsymbol{\theta}}_1(\boldsymbol{x})$とすれば，(2.2)は

$$f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_0(\boldsymbol{x}))=f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_1(\boldsymbol{x}))\quad \text{a. e. } \boldsymbol{x}(\mu)$$

であり，従って最大尤度に基づくモデル選択の議論においても識別可能性が問題となる。特に赤池情報量基準(AIC)では，Θ_0 に対して情報量を

(2.7)　$I(\Theta_0)=-2\log f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}_0(\boldsymbol{x}))+2p$

(p は Θ_0 の自由度)

で定義し，Θ_0 と Θ_K のモデル選択は，$I(\Theta_0)$ と $I(\Theta_K)$の大きさに基づいて行われる。この基準は尤度比検定で棄却点を有意水準に依存させないでパラメータの自由度だけから決めるモデル判別方式と考えることもできる。この AIC においても，Θ_0 と Θ_K がモデルとして唯一の代替的なペアでない限り，モデル Θ_0 が選択されたとしても，Θ_0 が識別可能でなくてかつ Θ_0 と Θ_1 のパラメータの自由度が同じである場合には，$\boldsymbol{\theta}\in\Theta_0$ なのか $\boldsymbol{\theta}\in\Theta_1$ なのか識別されない。また，パラメータの自由度が異なっていても，Θ_0 が Neyman-Pearson の意味で識別可能でないならば，やはり問題が起こることは明らかである。

§3　一般線型仮説の識別可能性

線型回帰モデル(1.2)で一般線型仮説(1.3)の識別可能性を考察する。ただし $\boldsymbol{\Omega}=\sigma^2\boldsymbol{I}_n$ とする。$f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2)$でこの $\boldsymbol{y}$ の密度関数を表せば，モデルは

$$F(\Theta)=\{f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2)|(\boldsymbol{\beta},\sigma^2)\in\Theta\}$$

$\Theta=R^k\times R^+$, $R^+=\{c>0\}$ とかける。Θ_0 は(1.3)を満たす$(\boldsymbol{\beta},\sigma^2)$であり，対立仮説を $K:\boldsymbol{R}_0\boldsymbol{\beta}\neq\boldsymbol{r}_0$ とすれば $\Theta_K=\Theta-\Theta_0$ となる。従って

(3.1)　$\boldsymbol{b}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$,

$\boldsymbol{A}_0=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}_0'[\boldsymbol{R}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}_0']^{-1}$

(3.2)　$\hat{\sigma}_0{}^2=\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b}-\boldsymbol{X}\boldsymbol{A}_0(\boldsymbol{R}_0\boldsymbol{b}-\boldsymbol{r}_0)\|^2/n$

とおけば，容易にわかるように

$$\sup_{\Theta_0} f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2)=[2\pi\hat{\sigma}_0{}^2e]^{-n/2}$$

と(2.2)の意味で恒等的に等しくさせる閉集合 Θ_0 で，$\Theta_1\cap\Theta_K=\phi$ となるものは Θ_0 以外に存在しない。更に，(2.2)を満たす閉集合 Θ_1 は Θ_0 以外存在しないことが証明される。従って H_0 は常に識別可能である。しかし，これは $\boldsymbol{\Omega}=\sigma^2\boldsymbol{I}_n$ というモデルスペシフィケーションのもとに成立していることを注意しておく。

§4　分散行列の構造に関する仮説の識別可能性

正規回帰モデル(1.2)のパラメータ$(\boldsymbol{\beta},\boldsymbol{\Omega})$の空間を $R^k\times\mathcal{S}(n)$とする。ただし $\mathcal{S}(n)$は $n\times n$ の正値定符号行列全体である。ここで

(4.1)　$\boldsymbol{\Omega}\longrightarrow(\sigma^2,\boldsymbol{\Sigma})$

$\sigma^2=|\boldsymbol{\Omega}|^{1/n}$,　$\boldsymbol{\Sigma}=(\sigma^2)^{-1}\boldsymbol{\Omega}$

と変換すれば，$|\boldsymbol{\Sigma}|=1$ で $\boldsymbol{\Omega}$ と$(\sigma^2,\boldsymbol{\Sigma})$とは1対1対応をし，(4.1)は同相写像である。従って，以下

では一般性を失うことなく $\boldsymbol{\Omega}=\sigma^2\boldsymbol{\Sigma}$, $|\boldsymbol{\Sigma}|=1$ とする。このときの $\boldsymbol{y}$ の密度関数を $f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma})$ と書き，$\bar{\mathcal{S}}(n)=\{\boldsymbol{\Sigma}\in\mathcal{S}(n)||\boldsymbol{\Sigma}|=1\}$，$\Theta=R^k\times R^+\times\bar{\mathcal{S}}(n)$ とおくと，モデル(1.2)は

$$F(\Theta)=\{f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma})|(\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma})\in\Theta\}$$

と表わされる。$\boldsymbol{\Sigma}$ に関する仮説の識別可能性を調べるために，与えられた1つの $\boldsymbol{\Sigma}_0\in\bar{\mathcal{S}}(n)$ に対して

$$(4.2)\quad \sup_{\beta,\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_0)=\sup_{\beta,\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_1)$$

を満たす $\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ のクラスを求めよう。いま $\boldsymbol{\Sigma}$ のもとでの GLSE と残差平方和をそれぞれ

$$(4.3)\quad \hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma})=(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

$$(4.4)\quad s^2(\boldsymbol{\Sigma})=(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}))$$
$$=\boldsymbol{y}'(\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})\boldsymbol{y}$$

とおくと，(4.2)は

$$(4.5)\quad s^2(\boldsymbol{\Sigma}_0)=s^2(\boldsymbol{\Sigma}_1)$$

と同等であることが簡単に示される。

補助定理 1(Khatri(1966)) 任意の $\boldsymbol{P}\in\mathcal{S}(n)$ と，$\boldsymbol{C}:n\times k$, $\boldsymbol{D}:n\times(n-k)$ で $\boldsymbol{C}'\boldsymbol{D}=\boldsymbol{0}$, $\mathrm{rank}([\boldsymbol{C},\boldsymbol{D}])=n$ を満たす任意の行列 $\boldsymbol{C}$ と $\boldsymbol{D}$ に対して

$$\boldsymbol{P}^{-1}-\boldsymbol{P}^{-1}\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{P}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{P}^{-1}$$
$$=\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{P}\boldsymbol{D})^{-1}\boldsymbol{D}'$$

が成立する。

この補題を用いると(4.5)は

$$(4.6)\quad \boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}=\boldsymbol{Z}'\boldsymbol{\Sigma}_1\boldsymbol{Z}$$

と同等である。ここで $\boldsymbol{Z}:n\times(n-k)$ は $\boldsymbol{Z}'\boldsymbol{X}=0$ を満たす任意の行列であるが，以下では(1.4)の $\boldsymbol{N}$ を用いて

$$(4.7)\quad \boldsymbol{Z}\boldsymbol{Z}'=\boldsymbol{N},\ \boldsymbol{Z}'\boldsymbol{Z}=\boldsymbol{I}_{n-k}$$

となる $\boldsymbol{Z}$ を選択する。(4.6)を $\boldsymbol{\Sigma}_1$ について解くと

$$(4.8)\quad \boldsymbol{\Sigma}_1=\boldsymbol{N}\boldsymbol{\Sigma}_0\boldsymbol{N}+\boldsymbol{B}-\boldsymbol{N}\boldsymbol{B}\boldsymbol{N}$$
$$=\boldsymbol{N}\boldsymbol{\Sigma}_0\boldsymbol{N}+\boldsymbol{M}\boldsymbol{B}\boldsymbol{N}+\boldsymbol{N}\boldsymbol{B}\boldsymbol{M}+\boldsymbol{M}\boldsymbol{B}\boldsymbol{M}$$

となる。ただし，$\boldsymbol{B}$ は $\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ となる任意の対称行列である。(Rao and Mitra(1971) pp. 24-25 を見よ。)以下では，与えられた $\boldsymbol{\Sigma}_0$ に対して(4.8)の形に書けて，$\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ となる $\boldsymbol{\Sigma}_1$ のクラスを $\mathcal{D}(\boldsymbol{\Sigma}_0)$ と表す。(4.8)で $\boldsymbol{B}=\boldsymbol{\Sigma}_0$ とおけば $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_0$ となる。$\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_0$ と $\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ となる $\boldsymbol{B}$ の存在を示すために，

$$(4.9)\quad \boldsymbol{B}=\tilde{\boldsymbol{X}}\boldsymbol{\Gamma}\tilde{\boldsymbol{X}}',\ \tilde{\boldsymbol{X}}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1/2},\boldsymbol{\Gamma}\in\mathcal{S}(k)$$

とおき，$\boldsymbol{Q}=[\tilde{\boldsymbol{X}},\boldsymbol{Z}]$ を用いて(4.8)を

$$(4.10)\quad \boldsymbol{Q}'\boldsymbol{\Sigma}_1\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{\Gamma} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}\end{bmatrix}$$

と変形する。従って，$|\boldsymbol{\Sigma}_1|=|\boldsymbol{Q}'\boldsymbol{\Sigma}_1\boldsymbol{Q}|=|\boldsymbol{\Gamma}||\boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}|=1$, すなわち $|\boldsymbol{\Gamma}|=1/|\boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}|$ となる任意の $\boldsymbol{\Gamma}\in\mathcal{S}(k)$ をもつ(4.9)の $\boldsymbol{B}$ を選ぶと $\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ となる。このとき $\boldsymbol{\Gamma}\neq\boldsymbol{X}'\boldsymbol{\Sigma}_0\boldsymbol{X}$ である限り $\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_0$ であることが，$\boldsymbol{Q}'\boldsymbol{\Sigma}_0\boldsymbol{Q}$ と(4.10)とを比較すれば示される。更に(4.9)で $\boldsymbol{\Gamma}=(\boldsymbol{X}'\boldsymbol{X})^{1/2}\boldsymbol{\Psi}(\boldsymbol{X}'\boldsymbol{X})^{1/2}$ とおき，$|\boldsymbol{\Psi}|=1/|\boldsymbol{X}'\boldsymbol{X}||\boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}|$ を満たす $\boldsymbol{\Psi}\in\mathcal{S}(k)$ とすれば，(4.8)は

$$(4.11)\quad \boldsymbol{\Sigma}_1=\boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{X}'+\boldsymbol{N}\boldsymbol{\Sigma}_0\boldsymbol{N}$$

と表わされ，$\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ となる。以上を次の形に要約しておく。

定理 1. 任意の $\boldsymbol{\Sigma}_0\in\bar{\mathcal{S}}(n)$ に対して，すべての $\boldsymbol{y}$ に対して(4.2)を満たす $\boldsymbol{\Sigma}_1\in\bar{\mathcal{S}}(n)$ で，$\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_0$ となるものが存在する。このような $\boldsymbol{\Sigma}_1$ は(4.8)の形をし，特に(4.11)の形をもつ $\boldsymbol{\Sigma}_1$ で，$|\boldsymbol{\Psi}|=1/|\boldsymbol{X}'\boldsymbol{X}||\boldsymbol{Z}'\boldsymbol{\Sigma}_0\boldsymbol{Z}|$, $\boldsymbol{\Psi}\in\mathcal{S}(k)$ をもつものは(4.2)を満たす。従って，任意の $\boldsymbol{\Sigma}_0\in\bar{\mathcal{S}}(n)$ に対して仮説 $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$ は識別不能である。

この定理は，最大尤度(4.2)の視点から見る限り $\boldsymbol{\Sigma}_0$ と $\mathcal{D}(\boldsymbol{\Sigma}_0)$ の任意の要素とは観察上仮説として同等である。これは任意の対立仮説 $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K$ に対しても，これと観察上同等な仮説のクラス $\mathcal{D}(\boldsymbol{\Sigma}_K)$ が存在することを意味し，$\boldsymbol{\Sigma}_0$ を $\boldsymbol{\Sigma}_K$ に対して検定する問題は，$\mathcal{D}(\boldsymbol{\Sigma}_0)$ を $\mathcal{D}(\boldsymbol{\Sigma}_K)$ に対して検定する問題と区別できない。$\boldsymbol{\Sigma}_0$ の採択は実は $\mathcal{D}(\boldsymbol{\Sigma}_0)$ の採択であるから，$\boldsymbol{\Sigma}_0$ が採択されても直ちに $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$ とするわけにはいかない。例えば $\boldsymbol{\Sigma}=\boldsymbol{I}$ としよう。このとき $\boldsymbol{I}$ と観察上同等な仮説のクラス $\mathcal{D}(\boldsymbol{I})$ は，(4.8)から

$$(4.12)\quad \boldsymbol{\Sigma}_1=\boldsymbol{N}+\boldsymbol{B}-\boldsymbol{N}\boldsymbol{B}\boldsymbol{N},\ |\boldsymbol{\Sigma}_1|=1$$

と書ける $\boldsymbol{\Sigma}_1$ のクラスである。そして対立仮説 K: $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\notin\mathcal{D}(\boldsymbol{I})$ に対して仮説 H: $\boldsymbol{\Sigma}=\boldsymbol{I}$ が尤度比検定に基づいて採択されたとしよう。このとき通常 $\boldsymbol{\Sigma}=\boldsymbol{I}$ のみが採択されたとして，(4.3)の GLSE の代りに(3.1)の OLSE $\boldsymbol{b}$ が用いられる。この場合，モデルとして可能な分散行列が $\boldsymbol{I}$ または $\boldsymbol{\Sigma}_K$ であるということがわかっていればこの方式は妥当であるが，そうでない場合は，採択されたもの

は $\boldsymbol{I}$ のみでなく $\mathcal{D}(\boldsymbol{I})$ であるから，必ずしも OLSE は適当でないかもしれない。このときもしすべての $\boldsymbol{\Sigma}_1\in\mathcal{D}(\boldsymbol{I})$ に対して恒等的に $\boldsymbol{b}=\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_1)$ (すなわち OLSE=GLSE) であれば，結果的に OLSE を用いたことになるが，後に見るように恒等的には $\boldsymbol{b}$ と等しくない $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_1)$ を生む $\boldsymbol{\Sigma}_1$ が $\mathcal{D}(\boldsymbol{I})$ の中に存在している。他方，$\boldsymbol{I}$ のもとで OLSE を用いた残差平方和(4.4)と，$\boldsymbol{\Sigma}_1\in\mathcal{D}(\boldsymbol{I}), \boldsymbol{\Sigma}_1\neq\boldsymbol{I}$ のもとで GLSE $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_1)$ を用いた残差平方和は，(4.5)に示されているように恒等的に等しい。この意味では OLSE を用いてもを用いても GLSE 同じであるが，$\boldsymbol{b}$ の分散行列と $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_1)$ の分散行列は必ずしも等しくない。従って，$\boldsymbol{I}$ に比べて $\boldsymbol{\Sigma}_1$ の方が与えられたデータを生成した $\boldsymbol{\Sigma}$ の構造として近い場合，t 統計量は t 分布をしなくなり係数の t 値はそのまま解釈できなくなる。この点については §5 で再論する。

具体例として系列相関の検定問題を考えよう。1階のマルコフ型モデルを含む多くの系列相関のモデルは，$\boldsymbol{u}\sim N(\boldsymbol{o}, \tau\boldsymbol{\Phi}(\rho))$

$$\text{(4.13)}\quad \boldsymbol{\Phi}(\rho)^{-1}=\boldsymbol{I}+\rho\boldsymbol{A},\quad \rho\in\Pi=\{\rho:\boldsymbol{\Phi}(\rho)^{-1}>0\}$$

と(近似的に)表わされる。$\tau\boldsymbol{\Phi}(\rho)=\sigma^2\boldsymbol{\Sigma}(\rho), \boldsymbol{\Sigma}(\rho)=|\boldsymbol{\Phi}(\rho)|^{-1/n}\boldsymbol{\Phi}(\rho)$ と書き直せば，$|\boldsymbol{\Sigma}(\rho)|=1$ となる。帰無仮説は H: $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(0)=\boldsymbol{I}$ であり，対立仮説は K: $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(\rho), \rho\neq 0$ としよう。この場合，モデルがデータの生成過程を十分よく反映しているとしても，$\boldsymbol{I}$ と観察上同等な仮説のクラス $\mathcal{D}(\boldsymbol{I})$ の中に適当な ρ_0 に対して $\boldsymbol{\Sigma}(\rho_0)$ が含まれるならば，$\boldsymbol{\Sigma}(0)=\boldsymbol{I}$ と $\boldsymbol{\Sigma}(\rho_0)$ が識別できないことになる。すなわち(4.6)から

$$\text{(4.14)}\quad \boldsymbol{Z}'\boldsymbol{\Sigma}(\rho_0)\boldsymbol{Z}=\boldsymbol{I}_{n-k}$$

となれば $\boldsymbol{I}$ と $\boldsymbol{\Sigma}(\rho_0)$ は識別不能となる。これは $\boldsymbol{X}$ と $\boldsymbol{A}$ に依存するが，一般に与えられたデータ $\boldsymbol{X}$ に対して(4.14)が正確に成立することはほとんどありえないかもしれない。しかし(4.14)が近似的に成立するような状況では，仮説 H を K に対して検定する問題で尤度比は $\rho=\rho_0$ の近くで1の近くの値をとることになる。従って Neyman-Pearson 流の検定方式を形式的にあてはめ H または K を採択しても，その結果は，表がでる確率が与えられた有意水準と同じ確率をもつ硬貨を投げて裏がでたら仮説 H を採択するという trivial な検定と差があまりないことになる。勿論，尤度比検定を近似したとみなしうる Durbin-Watson 検定の場合も状況は同じである。他方，データが(4.13)の構造から生成されたとみなしにくいような状況では，定理1にあるように $\boldsymbol{\Sigma}(\rho)$ と $\mathcal{D}(\boldsymbol{\Sigma}(\rho))$ の要素とは識別できない。

§5　OLSE と GLSE の選択

線型回帰モデル(1.1)で OLSE(3.1)と GLSE(4.3)の選択問題を考察する。$\boldsymbol{u}\sim N(\boldsymbol{o}, \sigma^2\boldsymbol{\Sigma})$ とし，$\mathrm{tr}\,\boldsymbol{\Sigma}=n$ または §4 のように $|\boldsymbol{\Sigma}|=1$ とする。GLSE は $\hat{\boldsymbol{\beta}}(\sigma^2\boldsymbol{\Sigma})=\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma})$ を満たすことに注意せよ。§1 で述べたように Rao(1967)は，OLSE $\boldsymbol{b}$ と GLSE $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma})$ とが恒等的に等しいための必要十分条件は，$\boldsymbol{u}$ の分散行列 $\sigma^2\boldsymbol{\Sigma}$ がクラス(1.5)に属することであることを証明した。(1.5)の $\boldsymbol{Z}$ として，一般性を失うことなく(4.7)を満たす $\boldsymbol{Z}$ をとる。Geisser(1970)は，(1.5)のクラスは

$$\text{(5.1)}\quad \mathcal{C}=\{\boldsymbol{\Phi}|\boldsymbol{\Phi}=\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{X}'+\boldsymbol{Z}\boldsymbol{\Delta}\boldsymbol{Z}',\ \boldsymbol{\Gamma}\in\mathcal{S}(k),\ \boldsymbol{\Delta}\in\mathcal{S}(n-k)\}$$

と同等であることを示した。以下この形で考える。OLSE を用いるか GLSE を用いるかという問題は，$\boldsymbol{\Sigma}$ が未知であってもとにかくそれがクラス C に入っていれば GLSE を用いても結果的に OLSE に等しくなる。従って最初から OLSE を用いることができることになる。$\boldsymbol{\Sigma}\in\mathcal{C}$ をみるための1つの方法として，仮説検定問題

$$H_0:\boldsymbol{\Sigma}\in\mathcal{C}\ \text{v.s.}\ K:\boldsymbol{\Sigma}\notin\mathcal{C}$$

が考えられる。しかし $\boldsymbol{\Sigma}$ が全く未知という状況では，データの数 n より $\boldsymbol{\Sigma}$ のパラメータの数 $[n(n+1)/2]-1$ の方が大きくなってこの問題は検定不能である。このため任意に $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\notin\mathcal{C}$ を与え，$\boldsymbol{\Sigma}_K$ と観察上同等な仮説のクラス $\mathcal{D}(\boldsymbol{\Sigma}_K)$ を考えてみよう。§4 の議論とあわせるために $|\boldsymbol{\Sigma}_K|=1$ としておく。(4.11)から，$\mathcal{D}(\boldsymbol{\Sigma}_K)$ の要素で，

$$\text{(5.2)}\quad \boldsymbol{\Sigma}_1=\boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{X}'+\boldsymbol{N}\boldsymbol{\Sigma}_K\boldsymbol{N}$$

なる形のものは明らかに $\boldsymbol{\Sigma}_1\in\mathcal{C}$ である。ここで $\boldsymbol{\Psi}\in\mathcal{S}(k)$ は $|\boldsymbol{\Psi}|=1/|\boldsymbol{X}'\boldsymbol{X}||\boldsymbol{Z}'\boldsymbol{\Sigma}_K\boldsymbol{Z}|$ を満たすものとする。従って任意の $\boldsymbol{\Sigma}_K\notin\mathcal{C}$ に対して，$\mathcal{C}$ の要

素 $\boldsymbol{\Sigma}_1$ で $\boldsymbol{\Sigma}_K$ と観察上同等，すなわち

$$(5.3)\quad \sup_{\beta,\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_1)/\sup_{\beta,\sigma^2} f(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Sigma}_K)=1$$

を満たすものが存在する。更に(4.4), (4.5)から

$$(5.4)\quad s^2(\boldsymbol{\Sigma}_1)=s^2(\boldsymbol{\Sigma}_K)$$

となる。尤度比(5.3)はデータから $\boldsymbol{\Sigma}_1$ と $\boldsymbol{\Sigma}_K$ が識別できないことを示し，(5.4)は $\boldsymbol{\Sigma}_1$ のもとでの残差平方和と $\boldsymbol{\Sigma}_K$ の残差平方和が等しいことを示している。従って，たとえ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\notin\mathcal{C}$ であっても，$\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1\in\mathcal{C}$ と装って常にOLSEを用いてもよいと議論しうるかもしれない。これは，$\boldsymbol{\Sigma}_1$ と $\boldsymbol{\Sigma}_K$ がデータから仮説として識別できず，更に(5.4)が成立するから，我々にとって便利な方を選択するという議論である。しかし，上で明確にされたことは，$\boldsymbol{\Sigma}_1$ と $\boldsymbol{\Sigma}_K$ とは仮説として観察上同等であるから検定不可能であるということであって，周知の如く $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_K\notin\mathcal{C}$ が「真」の分散行列であるとき $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1\in\mathcal{C}$ を装って OLSE を用いると，GLSE に比べて OLSE の相対的効率は小さくなる。この点を詳しくみる前に，正規化ルールについてふれておく。これまで正規化ルール $|\boldsymbol{\Sigma}_K|=1$ のもとで考えてきたが，勿論他の正規化ルール，例えば $\mathrm{tr}\,\boldsymbol{\Sigma}_K=n$, $\mathrm{tr}\,\boldsymbol{\Sigma}_K{}^2=n$ 等でおきかえてよい。この場合(5.3)の右辺はもはや1にならず $\boldsymbol{\Sigma}_K$ に依存した定数となる。従ってそこでは定義3の意味で，すなわち Neyman-Pearson の意味で $\boldsymbol{\Sigma}_1$ と $\boldsymbol{\Sigma}_K$ は識別不能となる。以下の議論でも一定の正規化ルールのもとで考えられている。

$\boldsymbol{\Sigma}$ に関するアプリオリな情報が十分でない場合，OLSE と GLSE の間の選択は仮説検定に基づくことができないことがわかった。以下では OLSE の GLSE に対する相対的効率を考察することで OLSE と GLSE の間の選択基準，選択の1つの尺度を作る。$\boldsymbol{\Sigma}_K\notin\mathcal{C}$ が「真」の分散行列であるとき，OLSE $\boldsymbol{b}$ の分散行列

$$(5.5)\quad \mathrm{Var}(\boldsymbol{b})=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

と，GLSE $\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K)$ の分散行列

$$(5.6)\quad \mathrm{Var}(\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K))=\sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}_K{}^{-1}\boldsymbol{X})^{-1}$$

に対し，その差は非負値定符号

$$(5.7)\quad \mathrm{Var}(\boldsymbol{b})-\mathrm{Var}(\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K))\geq 0$$

となる。(5.7)の左辺は補助定理1を用いると

$$\alpha^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{\Sigma}_K\boldsymbol{Z})^{-1}\times \boldsymbol{Z}'\boldsymbol{\Sigma}_K\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

となり，従って(5.7)で等号が成立するための必要十分条件は

$$\boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{Z}=\boldsymbol{0}$$

である。これは Rao(1967) が示しているように $\boldsymbol{\Sigma}_K\in\mathcal{C}$ と同等である。それゆえ $\boldsymbol{\Sigma}_K\notin\mathcal{C}$ のもとでは(5.7)は半正値定符号となる。相対的効率を

$$(5.8)\quad \eta=|\mathrm{Var}(\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K))|/|\mathrm{Var}(\boldsymbol{b})| =\{|\boldsymbol{X}'\boldsymbol{\Sigma}_K{}^{-1}\boldsymbol{X}||\boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{X}|/|\boldsymbol{X}'\boldsymbol{X}|^2\}^{-1}$$

の形で定義すると，ランク k をもつ任意の行列 $\boldsymbol{X}: n\times k$ に対して

$$(5.9)\quad 1\geq\eta\geq\prod_{i=1}^{l}[4\gamma_i\gamma_{n-i+1}/(\gamma_i+\gamma_{n-i+1})^2]$$

$l=\min(n-k,k)$ であることが証明されている。ただし，$\gamma_1\geq\gamma_2\geq\cdots\geq\gamma_n>0$ は $\boldsymbol{\Sigma}_K$ の固有値である (Bloomfield and Watson(1975), Knott(1975) を参照)。従って OLSE の GLSE に対する相対的効率 η の下限は $\boldsymbol{\Sigma}_K$ の固有値 γ_i のちらばりに依存し，$\boldsymbol{\Sigma}_K$ が単位行列 $\boldsymbol{I}$ に近くない限りその下限は1に近くなれない。しかし下限はすべての $\boldsymbol{X}$ に対して求められたものであるから，ある特定の $\boldsymbol{X}$ に対して効率 η は1に近くなりうる。実際，$\boldsymbol{X}$ の列ベクトル空間が $\boldsymbol{\Sigma}_K$ のある k 個の固有ベクトルに張られていれば，あるいは同じことだが $\boldsymbol{\Sigma}_K\in\mathcal{C}$ であれば，$\eta=1$ となる。それゆえたとえ $\boldsymbol{\Sigma}_K\notin\mathcal{C}$ であっても，$\boldsymbol{X}$ の列ベクトル空間が $\boldsymbol{\Sigma}_K$ のある k 個の固有ベクトルの張る空間に近ければ，相対的効率 η は1の近くにとどまることを示すことができる。そこで OLSE と GLSE の選択の判定尺度として次の2つの基準が考えられる。第1の基準は，$\boldsymbol{\Sigma}_K$ とクラス $\mathcal{C}$ との距離を考え，それが十分小さいとき OLSE を用いる方法である。第2のものは，$\boldsymbol{X}$ の列ベクトル空間と $\boldsymbol{\Sigma}_K$ の k 個の固有ベクトルの張る空間の近さを判定し，それが十分近いとき OLSE を用いる方法である。

第1の基準を考えよう。行列 $\boldsymbol{A}$ に対しユークリッドノルムを $\|\boldsymbol{A}\|=(\mathrm{tr}\,\boldsymbol{A}\boldsymbol{A}')^{1/2}$ とし，$\boldsymbol{\Sigma}_K$ と $\boldsymbol{\Phi}=\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{X}'+\boldsymbol{Z}\boldsymbol{\Delta}\boldsymbol{Z}'\in\mathcal{C}$ との距離

$$(5.10)\quad d(\boldsymbol{\Sigma}_K,\boldsymbol{\Phi})=\|\boldsymbol{\Sigma}_K-(\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{X}'+\boldsymbol{Z}\boldsymbol{\Delta}\boldsymbol{Z}')\|$$

を $\boldsymbol{\Gamma}\in\mathcal{S}(k)$, $\boldsymbol{\Delta}\in\mathcal{S}(n-k)$ に関して最小にする $\boldsymbol{\Phi}$ を求める。いま $\boldsymbol{\Sigma}_K$ と $\mathcal{C}$ の距離を

$$d(\boldsymbol{\Sigma}_K,\mathcal{C})=\inf_{\boldsymbol{\Phi}\in\mathcal{C}} d(\boldsymbol{\Sigma}_K,\boldsymbol{\Phi})$$

と定義する。

補助定理 2　距離(5.10)は

$$\hat{\boldsymbol{\Gamma}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},$$
$$\hat{\boldsymbol{\Delta}}=(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Sigma}_K\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$$

のとき，またそのときに限り最小となる。その最小値は

$$d(\boldsymbol{\Sigma}_K,\mathcal{C})=(\mathrm{tr}\boldsymbol{M}\boldsymbol{\Sigma}_K\boldsymbol{N}\boldsymbol{\Sigma}_K)^{1/2}>0$$

である。

証明　$\boldsymbol{X}'\boldsymbol{Z}=0$ から

$$d(\boldsymbol{\Sigma}_K,\boldsymbol{\Phi})^2=\|\boldsymbol{\Sigma}_K-\boldsymbol{X}\hat{\boldsymbol{\Gamma}}\boldsymbol{X}'-\boldsymbol{Z}\hat{\boldsymbol{\Delta}}\boldsymbol{Z}'\|^2$$
$$+\|\boldsymbol{X}(\boldsymbol{\Gamma}-\hat{\boldsymbol{\Gamma}})\boldsymbol{X}'\|^2+\|\boldsymbol{Z}(\boldsymbol{\Delta}-\hat{\boldsymbol{\Delta}})\boldsymbol{Z}'\|^2$$

であることが容易に示される。従って右辺の第2項と第3項がゼロになるとき，またそのときに限り(5.10)は最小となる。これから直ちに $d(\boldsymbol{\Sigma}_K,\mathcal{C})^2=\mathrm{tr}\,\boldsymbol{M}\boldsymbol{\Sigma}_K\boldsymbol{N}\boldsymbol{\Sigma}_K$ が計算される。更に

$$d(\boldsymbol{\Sigma}_K,\mathcal{C})=0\Leftrightarrow \boldsymbol{X}'\boldsymbol{\Sigma}_K\boldsymbol{Z}=0\Leftrightarrow \boldsymbol{\Sigma}_K\in\mathcal{C}$$

であるから，$\boldsymbol{\Sigma}_K\notin\mathcal{C}$ に対して $d(\boldsymbol{\Sigma}_K,\mathcal{C})>0$ である。

これから $\boldsymbol{\Sigma}_K$ に最も近い $\mathcal{C}$ の要素 $\hat{\boldsymbol{\Sigma}}_K$ は

$$\hat{\boldsymbol{\Sigma}}_K=\boldsymbol{M}\boldsymbol{\Sigma}_K\boldsymbol{M}+\boldsymbol{N}\boldsymbol{\Sigma}_K\boldsymbol{N}$$

と表わされる。距離 $d(\boldsymbol{\Sigma}_K,\mathcal{C})$ が十分小さいとき OLSE を用いても相対的効率がそれほど落ちないことになるが，どの程度小さければよいかという問題が残る。しかし，この距離を効率 η と関係づけることが難しいため，この問題を一般的に取扱うことは困難であろう。またそれは正規化ルールにも関係している。1つの実際的判断として，正規化ルールが $(\mathrm{tr}\,\boldsymbol{\Sigma}_K^2)^{1/2}=n$ の場合，$d(\boldsymbol{\Sigma}_K,\mathcal{C})\leq 0.1n$ であれば OLSE を用いるという判断を提案しておくが，あくまでもこれは便宜的なものである。ここでは Strand(1974) Theorem 5.1 の結果を修正して援用しておく。

定理 2　$d(\boldsymbol{\Sigma}_K,\mathcal{C})\leqq\varepsilon$ ならば

$$E[\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K)-\boldsymbol{b}]'\boldsymbol{X}'\boldsymbol{X}[\hat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}_K)-\boldsymbol{b}]/\sigma^2$$
$$\leqq\varepsilon^2/2+O(\varepsilon^3)$$

である。

第2の方法を考える。$\boldsymbol{\Sigma}_K$ の互いに直交する固有ベクトルを $\boldsymbol{\delta}_i(i=1,\cdots,n)$ とする。理論的には n 個の固有ベクトルから k 個を取り出し，その k 個の固有ベクトルの張る空間と $\boldsymbol{X}$ の列ベクトル空間の間に近さを測る尺度を導入し，その尺度の最小な値が十分小さければ OLSE を用いればよい。しかし n が k に比べて相対的に大きいと，計算量が莫大になるため次の方式を提案する。まず固有ベクトル $\boldsymbol{\delta}_i$ を $\boldsymbol{X}$ に回帰し，$\boldsymbol{X}$ の列ベクトル空間上への $\boldsymbol{\delta}_i$ 正射影による像 $\hat{\boldsymbol{\delta}}_i=\boldsymbol{M}\boldsymbol{\delta}_i$ と，残差 $\boldsymbol{e}_i=\boldsymbol{N}\boldsymbol{\delta}_i$ を求める。そのとき $\boldsymbol{X}$ の列ベクトル空間と $\boldsymbol{\delta}_i$ の角度を θ_i とすると

$$\tan\theta_i=\|\boldsymbol{e}_i\|/\|\hat{\boldsymbol{\delta}}_i\|$$

と表されるから

$$(5.11)\quad T_i=\|\hat{\boldsymbol{\delta}}_i\|^2/\|\boldsymbol{e}_i\|^2$$

を大きさの順に並べて，大きい方から k 個の添数に対応する固有ベクトル $\boldsymbol{\delta}_{i_j}(j=1,\cdots,k)$ を取りだす。一般性を失うことなくそれを $\boldsymbol{\delta}_i(i=1,\cdots,k)$ とし，$\boldsymbol{\Lambda}_1=[\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_k]:n\times k$ とおく。$\boldsymbol{X}$ の列ベクトル空間と $\boldsymbol{\Lambda}_1$ の列ベクトル空間の近さを測る尺度としては，(5.8)の η の定義との関係から

$$(5.12)\quad \tau_1=|\boldsymbol{X}'\boldsymbol{\Lambda}_1|^2/|\boldsymbol{X}'\boldsymbol{X}||\boldsymbol{\Lambda}_1'\boldsymbol{\Lambda}_1|$$
$$=|\boldsymbol{X}'\boldsymbol{\Lambda}_1|^2/|\boldsymbol{X}'\boldsymbol{X}|,\qquad (\|\boldsymbol{\delta}_i\|=1)$$

を採用する。勿論他の定義，例えば

$$\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Lambda}_1(\boldsymbol{\Lambda}_1'\boldsymbol{\Lambda}_1)^{-1}\boldsymbol{\Lambda}_1'\boldsymbol{X}$$

等も同じ意味をもつ尺度である。(5.12)の τ_1 は，$0<\tau_1\leq1$ で，$\tau_1=1$ のとき $\boldsymbol{X}$ の列ベクトル空間は $\boldsymbol{\Lambda}_1$ の列ベクトル空間，すなわち $\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_k$ の張る空間と一致するから，$\eta=1$ となる。τ_1 がどの程度大きければよいかという問題は，$\boldsymbol{\Sigma}_K$ の固有値の大きさとちらばりにも依存する。ここでは恣意的ではあるが，$\tau_1\geq0.9$ を1つの目安として与えておく。

以上の議論は $\boldsymbol{\Sigma}_K$ が既知でない場合，そのままの形では適用できない。他方，もし $\boldsymbol{\Sigma}_K$ が既知であれば GLSE の適用が可能になるので，上の議論は実際的な意味をもたないように思われるかもしれない。この点については次の節で具体例を示し，上の議論が実際的にも有効であることを示す。

最後に，1つの注意を述べておく。$\boldsymbol{\Sigma}_K\in\mathcal{C}$ であれば GLSE と OLSE は恒等的に等しく，従って最初から OLSE を用いた場合と同じ結果を生む。しかし，$\boldsymbol{\Sigma}_K=\boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{X}'+\boldsymbol{Z}\boldsymbol{\Delta}\boldsymbol{Z}'$ のもとでの OLSE $\boldsymbol{b}$

の分布は，$\boldsymbol{b}\sim N(\boldsymbol{\beta}, \sigma^2\boldsymbol{\Gamma})$ となり，$\boldsymbol{\Gamma}\doteqdot(\boldsymbol{X}'\boldsymbol{X})^{-1}$ でない限り，通常の各係数に付される t 値は意味をなさないことになる[1)]。

§6　系列相関の場合への応用

この節では(4.13)の形で与えられる系列相関をもつ分散行列の構造に対して第5節の結果を適用する。第1のアプローチとしては，(4.13)の $\boldsymbol{\Phi}(\rho)=(\boldsymbol{I}+\rho\boldsymbol{A})^{-1}$ とクラス $\mathcal{C}$ の距離が

(6.1)
$$d(\boldsymbol{\Phi}(\rho),\mathcal{C})=[\mathrm{tr}(\boldsymbol{I}+\rho\boldsymbol{A})^{-1}\boldsymbol{M}(\boldsymbol{I}+\rho\boldsymbol{A})^{-1}\boldsymbol{N}]^{1/2}\leq\varepsilon$$

を満たすとき，定理2の結果が成立する。(6.1)は，ρ が十分小さいときだけでなく，$\boldsymbol{X}$ の列ベクトル空間が $(\boldsymbol{I}+\rho\boldsymbol{A})$ の k 個の固有ベクトルによって張られる空間に近いときにも成立する。

次に第2のアプローチについて考えよう。$\boldsymbol{A}$ の固有値を $\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_n$ とする。$\boldsymbol{A}$ が既知であるから $\boldsymbol{\delta}_i$ は既知である。従って，$\boldsymbol{\delta}_i$ は $\boldsymbol{\Phi}(\rho)$ の既知の固有ベクトルである。与えられた X に対して $\hat{\boldsymbol{\delta}}_i=\boldsymbol{M}\boldsymbol{\delta}_i$, $\boldsymbol{e}_i=\boldsymbol{N}\boldsymbol{\delta}_i$ を計算し，$T_i=\|\hat{\boldsymbol{\delta}}_i\|^2/\|\boldsymbol{e}_i\|^2$ を大きい順に並べて大きい方から k 個の添数をもつ k 個の固有ベクトルを選び，$\boldsymbol{\Lambda}_1$ をつくる。一般性を失うことなく $\boldsymbol{\Lambda}_1=[\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_k]$ とする。$\boldsymbol{\Lambda}_1$ が求まれば，(5.11)から τ_1 が計算でき，それが0.9以上であればOLSEを用いてもそれほど相対的効率の損失は少ないと考える。この方法は $H:\rho=0$ を検定する場合に用いられる検定統計量

(6.2)　$d=\boldsymbol{y}'\boldsymbol{NANy}/\boldsymbol{y}'\boldsymbol{Ny}$

を補う意味をもつ。例えば，$\boldsymbol{NAN}$ が極めて0に近いとしよう。この場合，d に基づく検定によれば正の相関が極めて強いと判断されがちである。しかし，$\boldsymbol{A}$ の $\boldsymbol{\delta}_i$ に対応する固有値を λ_i, $\boldsymbol{\Lambda}_2=[\boldsymbol{\delta}_{k+1},\cdots,\boldsymbol{\delta}_n]$, $\boldsymbol{D}_1=\mathrm{diag}\{\lambda_1,\cdots,\lambda_k\}$, $\boldsymbol{D}_2=\mathrm{diag}\{\lambda_{k+1},\cdots,\lambda_n\}$, $\boldsymbol{N}=\boldsymbol{ZZ}'$ とすれば，

(6.3)　$\boldsymbol{Z}'\boldsymbol{AZ}=\boldsymbol{Z}'\boldsymbol{\Lambda}_1\boldsymbol{D}_1\boldsymbol{\Lambda}_1{}'\boldsymbol{Z}+\boldsymbol{Z}'\boldsymbol{\Lambda}_2\boldsymbol{D}_2\boldsymbol{\Lambda}_2{}'\boldsymbol{Z}$

と書けるから，$\boldsymbol{Z}'\boldsymbol{\Lambda}_1\doteqdot\boldsymbol{O}$, $\boldsymbol{D}_2\doteqdot\boldsymbol{O}$ の場合，$\boldsymbol{Z}'\boldsymbol{AZ}\doteqdot\boldsymbol{O}$, 従って $\boldsymbol{NAN}\doteqdot\boldsymbol{O}$ となる。それゆえ，$\boldsymbol{X}$ の列ベクトル空間が $\boldsymbol{\Lambda}_1$ の列ベクトルによって近似的に張られ，$\boldsymbol{\Lambda}_2$ の列ベクトル(固有ベクトル $\boldsymbol{\delta}_j$, $j\geq k+1$)に対応する固有値が O に近い場合，通常の検定統計量では正の系列相関があると判断されがちである。(6.3)の関係は，既知行列 $\boldsymbol{A}$ と $\boldsymbol{X}$ の関係であって，確率ベクトル $\boldsymbol{y}$ に依存していない。このことは，(6.2)の d 統計量が系列相関を必ずしも正しく判別できないことを示している。他方，(6.2)で $\boldsymbol{Z}'\boldsymbol{\Lambda}_1\doteqdot\boldsymbol{O}$ が成立することは，$\boldsymbol{X}$ の列ベクトル空間が $\boldsymbol{\Lambda}_1$ の列ベクトル空間で近似されることを意味し，それは(5.12)の τ_1 を用いて判断できる。また，そのとき $\boldsymbol{D}_2$，すなわち固有値 $\lambda_j(j\geq k+1)$ の大きさも既知である。ここで，λ_j は T_i を大きさの順に並べた場合の第 j 番の大きさに対応する固有値であることに注意されたい。更に，ある定数 c に対して $\boldsymbol{NAN}\doteqdot c\boldsymbol{N}$ となる場合を考えよう。この場合(6.2)の d は，すべての $\boldsymbol{y}$ に対して $d\doteqdot c$ となり，d に基づいて系列相関の検定をしても意味がないことになる。注意すべきことは，このような場合もコンピューターは機械的に $d\doteqdot c$ を計算するだけである。$\boldsymbol{NAN}\doteqdot c\boldsymbol{N}$ が成立するのは，$\boldsymbol{A}$ がある $n\times n$ の対称行列 $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ を用いて

$$\boldsymbol{A}\doteqdot c\boldsymbol{N}+\boldsymbol{MB}_1\boldsymbol{M}+\boldsymbol{NB}_2\boldsymbol{M}+\boldsymbol{MB}_2\boldsymbol{N}$$

と書ける場合である。これが成立するのは，(6.3)から $\boldsymbol{D}_1\doteqdot c\boldsymbol{I}_k$, $\boldsymbol{D}_2\doteqdot c\boldsymbol{I}_{n-k}$ か，$\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ の選び方から $\boldsymbol{Z}'\boldsymbol{\Lambda}_1\doteqdot\boldsymbol{O}$ で $\boldsymbol{D}_2\doteqdot c\boldsymbol{I}_{n-k}$ のいずれかの場合である。最初の場合は，全部の固有値 $\lambda_j(j=1,\cdots,n)$ のちらばりが小さい場合で，$\boldsymbol{\Phi}(\rho)$ の固有値は $\gamma_i\doteqdot1+\rho c(i=1,\cdots,n)$ とであるから，(5.8)から $\eta\doteqdot1$ となり，OLSEを用いても相対効率は落ちない。2番目の場合は，τ_1 が1に近く，$\lambda_j(j=k+1,\cdots,n)$ のちらばり

(6.4)
$$\tau_2=\sum_{j=k+1}^{n}(\lambda_j-\bar{\lambda})^2/(n-k),$$

が小さい場合である。ただし $\bar{\lambda}=\sum_{j=k+1}^{n}\lambda_j/(n-k)$ である。このような場合，d の値はあまり意味をもたず，OLSEを適用しても相対的効率はそれほど落ちない。

(一橋大学経済研究所)

1)　溝口敏行教授の指摘による。

参考文献

[1] Bloomfield, P. and Watson, G. S. (1975). "The inefficiency of least squares," *Biometrika*, **62,** 121–128.

[2] Geisser, S. (1970). "Bayesian analysis of growth curves," *Sankhyā*, **A32,** 53–64.

[3] Kadiyala, K. R. (1970). "Testing for the independence of regression disturbance," *Econometrica*, **38,** 97–117.

[4] Khatri, C. G. (1966). "A note on a MANOVA model applied to problems in growth curves," *Ann. Inst. Statist. Math.*, **18,** 75–86.

[5] Knott, M. (1975). "On the minimum efficiency of least squares," *Biometrika*, **62,** 129–132.

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

[7] Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Application.* Wiley, New York.

[8] Rao, C. R. (1967). "Least squares theory using an estimated dispersion matrix and its application to measurement of signals," *Fifth Berk. Symp. Math. Statist. Prob.*, **1,** 355–372. University of California Press.

[9] Strand, O. N. (1974). "Coefficient errors caused by using the wrong covariance matrix in the general linear model," *Ann. Statist.*, **2,** 935–949.