不均一分散線形回帰モデルにおける不偏推定量について

井　上　　　淳

1.　はじめに

階数が p で第 1 列が $\mathbf{1}_{n_i}\ (=(1,1,\ldots,1)^t)\ (n_i\times 1)$ であるような既知の計画行列 $X_i\ (n_i\times p)\ (i=1,\ldots,k)$，未知の回帰ベクトル $\beta\ (p\times 1)$，誤差ベクトル $\epsilon_i\ (n_i\times 1)\ (i=1,\ldots,k)$ によって観測ベクトル $y_i\ (n_i\times 1)\ (i=1,\ldots,k)$ が次のように表される線形回帰モデルを考え，このモデルの下で回帰ベクトル β を推定する問題を考える：

$$y_i=X_i\beta+\epsilon_i \quad (i=1,\ldots,k).$$

ここに，$\epsilon_i\ (i=1,\ldots,k)$ は互いに独立に平均 $\mathbf{0}$ の n_i 変量正規分布に従うものとし，ϵ_i の各成分の分散は未知の正数 σ_i^2 であり，各成分間の相関係数は成分の選び方によらずに一定の既知の値 $\rho_i\ (-(n_i-1)^{-1}<\rho_i<1)$ をとるものと仮定する．すなわち，$\epsilon=(\epsilon_1^t,\ldots,\epsilon_k^t)^t$,

$$\Omega_i=(1-\rho_i)I_{n_i}+\rho_i\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t \quad (i=1,\ldots,k) \tag{1.1}$$

とおくとき，ϵ_i，ϵ の分散共分散行列についてそれぞれ次が成り立つものと仮定する：

$$V(\epsilon_i)=\sigma_i^2\Omega_i \quad (i=1,\ldots,k),$$
$$V(\epsilon)=\mathrm{diag}(\sigma_1^2\Omega_1,\ldots,\sigma_k^2\Omega_k).$$

$\sigma_i^2\ (i=1,\ldots,k)$ が既知の場合は，β の一般化最小二乗推定量

$$\hat{\beta}_{gls}=\left(\sum_{i=1}^k\sigma_i^{-2}X_i^t\Omega_i^{-1}X_i\right)^{-1}\left(\sum_{i=1}^k\sigma_i^{-2}X_i^t\Omega_i^{-1}y_i\right)$$

が最良線形不偏推定量になっており，その分散 $V(\hat{\beta}_{gls})$ は次式で与えられる：

$$V(\hat{\beta}_{gls}) = \left(\sum_{i=1}^{k} \sigma_i^{-2} X_i^t \Omega_i^{-1} X_i\right)^{-1}. \tag{1.2}$$

現在の設定では σ_i^2 $(i=1,\ldots,k)$ は未知である．従って，一般化最小二乗推定量 $\hat{\beta}_{gls}$ を実際に用いることはできない．そこで，残差平方和を用いて σ_i^2 の不偏推定量

$$\hat{\sigma}_i^2 = \frac{y_i^t P_i y_i}{(1-\rho_i)(n_i-p)} \quad (P_i = I_{n_i} - X_i(X_i^t X_i)^{-1} X_i^t)$$

を構成し，この逆数 $\hat{\sigma}_i^{-2}$ を $\hat{\beta}_{gls}$ の中の σ_i^{-2} と置き換えることが考えられる．

本稿では $\hat{\sigma}_i^{-2}$ に偏りがあることを考慮し，この偏りを定数倍によって修正する余地を与えることにする．すなわち，正数 w_i を $\hat{\sigma}_i^{-2}$ に付加してできる推定量 $\hat{\beta}_w$ を考えることにする：

$$\hat{\beta}_w = \left(\sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} X_i^t \Omega_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} X_i^t \Omega_i^{-1} y_i\right).$$

ここで，正数 w_i $(i=1,\ldots,k)$ の導入によって β の不偏推定量のクラスが生成されることに注意する．

$\Omega_i = I_{n_i}$ $(i=1,\ldots,k)$ のときは（すなわち，$\rho_i=0$ $(i=1,\ldots,k)$ のときは），不偏推定量 $\hat{\beta}_w$ のクラスの中で漸近的 $(k \to +\infty)$ に最適な推定量が存在すること，及びその推定量の漸近分散が次で与えられることが Inoue(2003) の Theorem 2.3 によって既に示されている：

$$\left(\sum_{i=1}^{k} \frac{n_i-p-4}{n_i-p-2} \sigma_i^{-2} X_i^t X_i\right)^{-1}. \tag{1.3}$$

β の不偏推定量の分散の下限(1.2)において $\Omega_i = I_{n_i}$ の場合を考え，それを(1.3)と比較すれば分かるように，k 個の標本のサイズ n_i $(i=1,\ldots,k)$ が小さいときは $\hat{\beta}_w$ の精度が低く，n_i $(i=1,\ldots,k)$ が大きいときは精度が高いというごく自然な現象を(1.3)は明らかにしている．

本稿では，この結果を(1.1)という形で与えられる Ω_i の場合に拡張する．これにより，誤差ベクトル ϵ_i $(i=1,\ldots,k)$ 内の無相関性の仮定をある程度緩め

ても Inoue(2003) の Theorem 2.3 と同様の結果が成立することが分かる．

2. 補助定理

この節では，補助定理を準備する．

補助定理 2.1　$\Omega_i\ (i=1,\ldots,k)$ について次が成り立つ．

(i)　$\Omega_i^{-1}=\dfrac{1}{1-\rho_i}\left\{I_{n_i}-\dfrac{\rho_i}{1+(n_i-1)\rho_i}\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t\right\}\quad (i=1,\ldots,k).$

(ii)　$\Omega_i^{1/2}=cI_{n_i}+d\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t\quad (i=1,\ldots,k)$

但し，$c^2=1-\rho_i$，$n_i d^2+2cd-\rho_i=0.$

証明の概略　直接的な計算により，直ちに結果を得る．　□

補助定理 2.2　$\mathrm{E}(\hat{\sigma}_i^{-2})=\sigma_i^{-2}\dfrac{n_i-p}{n_i-p-2}\quad (i=1,\ldots,k).$

証明の概略　$P_i\mathbf{1}_{n_i}=\mathbf{0}\ (i=1,\ldots,k)$ が成り立つことと，補助定理2.1(ii)の結果により次が成り立つ：

$$\begin{aligned}\Omega_i^{1/2}P_i\Omega_i^{1/2}&=(cI_{n_i}+d\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t)P_i(cI_{n_i}+d\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t)\\&=cP_i(cI_{n_i}+d\mathbf{1}_{n_i}\mathbf{1}_{n_i}^t)\\&=c^2P_i\\&=(1-\rho_i)P_i\quad (i=1,\ldots,k).\end{aligned}\tag{2.1}$$

次に，$f_i=\Omega_i^{-1/2}\epsilon_i\ (i=1,\ldots,k)$ とおく．f_i は平均 $\mathbf{0}$，分散共分散行列 $\sigma_i^2I_{n_i}$ の n_i 変量正規分布に従う．この $f_i\ (i=1,\ldots,k)$ と(2.1)を用いると，次の変形ができる：

$$y_i^tP_iy_i=(X_i\beta+\epsilon_i)^tP_i(X_i\beta+\epsilon_i)=\epsilon_i^tP_i\epsilon_i$$

$$=f_i^t(\Omega_i^{1/2}P_i\Omega_i^{1/2})f_i$$
$$=(1-\rho_i)\cdot f_i^tP_if_i \quad (i=1,\ldots,k). \tag{2.2}$$

次に，直交行列 Q_i，対角行列 $\Lambda_i=\mathrm{diag}(I_{n_i}-p,\ O_p)$ を用いて P_i を $Q_i^t\Lambda_iQ_i$ と分解し，$g_i=Q_if_i(i=1,\ldots,k)$ とおく．この g_i $(i=1,\ldots,k)$ と(2.2)を用いると，次の変形ができる：

$$\begin{aligned} y_i^tP_iy_i &= (1-\rho_i)\cdot f_i^tP_if_i=(1-\rho_i)\cdot f_i^tQ_i^t\Lambda_iQ_if_i \\ &= (1-\rho_i)\cdot g_i^t\Lambda_ig_i \quad (i=1,\ldots,k). \end{aligned} \tag{2.3}$$

ここで，g_i と f_i の分布は同一であることに注意する．このことと(2.3)により，結局次を得る：

$$\begin{aligned} E(\hat{\sigma}_i^{-2}) &= (1-\rho_i)(n_i-p)E\left(\frac{1}{y_i^tP_iy_i}\right) \\ &= (1-\rho_i)(n_i-p)E\left\{\frac{1}{(1-\rho_i)\cdot g_i^t\Lambda_ig_i}\right\} \\ &= (n_i-p)E\left(\frac{1}{g_i^t\Lambda_ig_i}\right) \\ &= \sigma_i^{-2}\frac{n_i-p}{n_i-p-2}. \end{aligned}$$

□

補助定理 2.3

$$E\left(\frac{\epsilon_i\epsilon_i^t}{\hat{\sigma}_i^4}\right)=\sigma_i^{-2}\frac{n_i-p}{n_i-p-2}\left[(1-\rho_i)P_i+\frac{n_i-p}{n_i-p-4}\{\Omega_i-(1-\rho_i)P_i\}\right] \quad (i=1,\ldots,k).$$

証明の概略　まず，$\hat{\sigma}_i^{-4}\epsilon_i\epsilon_i^t$ を次のように変形する：

$$\begin{aligned} \frac{\epsilon_i\epsilon_i^t}{\hat{\sigma}_i^4} &= \frac{\Omega_i^{1/2}f_if_i^t\Omega_i^{1/2}}{\{(n_i-p)^{-1}f_i^tP_if_i\}^2} \\ &= (n_i-p)^2\cdot\Omega_i^{1/2}\cdot\frac{Q_i^tQ_if_if_i^tQ_i^tQ_i}{(f_i^tQ_i^t\Lambda_iQ_if_i)^2}\cdot\Omega_i^{1/2} \\ &= (n_i-p)^2\cdot\Omega_i^{1/2}Q_i^t\cdot\frac{g_ig_i^t}{(g_i^t\Lambda_ig_i)^2}\cdot Q_i\Omega_i^{1/2}. \end{aligned} \tag{2.4}$$

次に，g_i を 2 つのベクトル $g_{i1}((n_i-p)\times 1)$，$g_{i2}(p\times 1)$ に分割し，$g_i^t=(g_{i1}^t,\ g_{i2}^t)^t$ と表すことにする．このとき，$g_i^t\Lambda_i g_i=\|g_{i1}\|^2$ と書くことができる．また，多変量正規分布の性質を用いて以下の式の成立を示すことができる：

$$
\begin{aligned}
&E\left(\frac{g_{i1}g_{i2}^t}{\|g_{i1}\|^4}\right)=O, \qquad E\left(\frac{g_{i2}g_{i1}^t}{\|g_{i1}\|^4}\right)=O,\\
&E\left(\frac{g_{i1}g_{i1}^t}{\|g_{i1}\|^4}\right)=\frac{1}{n_i-p}E\left(\frac{\|g_{i1}\|^2}{\|g_{i1}\|^4}\right)I_{n_i-p}=\frac{1}{n_i-p}E(\|g_{i1}\|^{-2})I_{n_i-p}\\
&\qquad\qquad\quad=\frac{\sigma_i^{-2}}{(n_i-p)(n_i-p-2)}I_{n_i-p},\\
&E\left(\frac{g_{i2}g_{i2}^t}{\|g_{i1}\|^4}\right)=\frac{1}{p}E\left(\frac{\|g_{i2}\|^2}{\|g_{i1}\|^4}\right)I_p=\frac{\sigma_i^{-2}}{(n_i-p-2)(n_i-p-4)}I_p \quad (i=1,\ldots,k).
\end{aligned}
$$

これらにより，次が成り立つ：

$$
\begin{aligned}
E\left\{\frac{g_ig_i^t}{(g_i^t\Lambda_ig_i)^2}\right\}&=\mathrm{diag}\left\{\frac{\sigma_i^{-2}}{(n_i-p)(n_i-p-2)}I_{n_i-p},\ \frac{\sigma_i^{-2}}{(n_i-p-2)(n_i-p-4)}I_p\right\}\\
&=\frac{\sigma_i^{-2}}{(n_i-p)(n_i-p-2)}\Lambda_i+\frac{\sigma_i^{-2}}{(n_i-p-2)(n_i-p-4)}(I_{n_i}-\Lambda_i)\\
&\qquad (i=1,\ldots,k). \qquad (2.5)
\end{aligned}
$$

(2.4)，(2.5)，(2.1)を合わせることにより，定理の結論を得る：

$$
\begin{aligned}
&E\left(\frac{\epsilon_i\epsilon_i^t}{\hat{\sigma}_i^4}\right)\\
&=(n_i-p)^2\Omega_i^{1/2}Q_i^t\left\{\frac{\sigma_i^{-2}\Lambda_i}{(n_i-p)(n_i-p-2)}+\frac{\sigma_i^{-2}(I_{n_i}-\Lambda_i)}{(n_i-p-2)(n_i-p-4)}\right\}Q_i\Omega_i^{1/2}\\
&=(n_i-p)^2\Omega_i^{1/2}\left\{\frac{\sigma_i^{-2}P_i}{(n_i-p)(n_i-p-2)}+\frac{\sigma_i^{-2}(I_{n_i}-P_i)}{(n_i-p-2)(n_i-p-4)}\right\}\Omega_i^{1/2}\\
&=(n_i-p)^2\left[\frac{\sigma_i^{-2}(1-\rho_i)P_i}{(n_i-p)(n_i-p-2)}+\frac{\sigma_i^{-2}\{\Omega_i-(1-\rho_i)P_i\}}{(n_i-p-2)(n_i-p-4)}\right]\\
&=\sigma_i^{-2}\frac{n_i-p}{n_i-p-2}\left[(1-\rho_i)P_i+\frac{n_i-p}{n_i-p-4}\{\Omega_i-(1-\rho_i)P_i\}\right].
\end{aligned}
$$

□

補助定理 2.4

$$V=\left(\sum_{i=1}^{k} w_i\hat{\sigma}_i^{-2}X_i^t\Omega_i^{-1}\epsilon_i\right)=\sum_{i=1}^{k}\frac{w_i^2(n_i-p)^2}{(n_i-p-2)(n_i-p-4)}\sigma_i^{-2}X_i^t\Omega_i^{-1}X_i.$$

証明の概略　補助定理2.1(i)の結果により，次が成り立つことが分かる：

$$\Omega_i^{-1}P_i\Omega_i^{-1}=\frac{1}{(1-\rho_i)^2}P_i \quad (i=1,\ldots,k). \tag{2.6}$$

また，多変量正規分布の性質から，次が成り立つ：

$$E=\left(\sum_{i=1}^{k} w_i\hat{\sigma}_i^{-2}X_i^t\Omega_i^{-1}\epsilon_i\right)=\mathbf{0}. \tag{2.7}$$

補助定理2.3，(2.6)，(2.7)の結果を合わせて次を得る：

$$\begin{aligned}
&V\left(\sum_{i=1}^{k} w_i\hat{\sigma}_i^{-2}X_i^t\Omega_i^{-1}\epsilon_i\right)\\
&=\sum_{i=1}^{k} w_i^2X_i^t\Omega_i^{-1}\cdot E\left(\frac{\epsilon_i\epsilon_i^t}{\hat{\sigma}_i^4}\right)\cdot\Omega_i^{-1}X_i\\
&=\sum_{i=1}^{k} w_i^2X_i^t\Omega_i^{-1}\cdot\sigma_i^{-2}\frac{n_i-p}{n_i-p-2}\left[(1-\rho_i)P_i+\frac{n_i-p}{n_i-p-4}\{\Omega_i-(1-\rho_i)P_i\}\right]\times\Omega_i^{-1}X_i\\
&=\sum_{i=1}^{k}\frac{w_i^2\sigma_i^{-2}(n_i-p)}{n_i-p-2}X_i^t\left\{\frac{P_i}{1-\rho_i}+\frac{n_i-p}{n_i-p-4}\left(\Omega_i^{-1}-\frac{P_i}{1-\rho_i}\right)\right\}X_i\\
&=\sum_{i=1}^{k}\frac{w_i^2(n_i-p)^2}{(n_i-p-2)(n_i-p-4)}\sigma_i^{-2}X_i^t\Omega_i^{-1}X_i.
\end{aligned}$$

□

3.　主要定理

この節では，前節で準備した補助定理の結果を利用して $\hat{\beta}_w$ の漸近分散を求める．また，$w_i\ (i=1,\ldots,k)$ を適当に選択することによって，$\hat{\beta}_w$ タイプの不偏推定量の中で漸近最適なものを見つけ出す．その結果，たとえば $w_1=w_2=\cdots=w_k$ という一見自然な選択は，一般に漸近最適な推定量を導かないことが分かる．

定理 3.1　適当な有界性の条件の下で，次の関係式が成り立つ：

$$\left(\sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} X_i^t \Omega_i^{-1} X_i\right)^{-1} = \left(\sum_{i=1}^{k} w_i \frac{n_i - p}{n_i - p - 2} \sigma^{-2} X_i^t \Omega_i^{-1} X_i\right)^{-1} + o_p(k^{-1}).$$

証明の概略

まず，補助定理 2.2 の結果と大数の法則を用いて次の関係式の成立を示すことができる：

$$\sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} X_i^t \Omega_i^{-1} X_i = \sum_{i=1}^{k} w_i \frac{n_i - p}{n_i - p - 2} \sigma^{-2} X_i^t \Omega_i^{-1} X_i + o_p(k).$$

この関係式を用いて逆行列間の漸近的関係を導くことができる．具体的な手順は Inoue(2003) の Lemma 2.1 の証明の流れと同様である．

□

定理 3.2　適当な有界性の条件の下で，次の関係式が成り立つ：

$$\eta^t \cdot V_w^{-1/2} \cdot B_w \left(\sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} X_i^t \Omega_i^{-1} \epsilon_i\right) \xrightarrow{d} \mathrm{N}(0,\ 1^2) \quad (k \to +\infty).$$

但し，$\eta(p \times 1)$ は任意の単位ベクトル，$V_w = B_w C_w B_w$,

$$B_w = \left(\sum_{i=1}^{k} w_i \frac{n_i - p}{n_i - p - 2} \sigma^{-2} X_i^t \Omega_i^{-1} X_i\right)^{-1},$$

$$C_w = \sum_{i=1}^{k} \frac{w_i^2 (n_i - p)^2}{(n_i - p - 2)(n_i - p - 4)} \sigma_i^{-2} X_i^t \Omega_i^{-1} X_i$$

とする．

証明の概略

補助定理 2.4 の結果と中心極限定理を用いて示すことができる．具体的な手順は Inoue(2003) の Lemma 2.2 の証明の流れと同様である．

□

補助定理 2.2 の結果から，$\dfrac{n_i - p - 2}{n_i - p} \hat{\sigma}_i^{-2}$ が σ_i^2 の不偏推定量であることが分かる．この事実だけから判断すると

$$w_i \propto \frac{n_i - p - 2}{n_i - p} \quad (i = 1, \ldots, k)$$

を満たすように w_i $(i=1,\ldots,k)$ を選ぶのが最適なのではないかと思われる．しかし，このような選択が実際には最適ではないことが次の定理から分かる．

定理 3.3　適当な有界性の条件の下で，次の関係式：

$$V_w^{-1/2}(\hat{\beta}_w - \beta) \xrightarrow{d} \mathrm{N}(\mathbf{0},\ I_p) \quad (k \to +\infty)$$

が成り立つ．また，V_w には下限

$$V_* = \left(\sum_{i=1}^{k} \frac{n_i - p - 4}{n_i - p - 2} \sigma_i^{-2} X_i^t \Omega_i^{-1} X_i \right)^{-1}$$

が存在する．

V_w の下限 V_* が達成されるための必要十分条件は

$$w_i \propto \frac{n_i - p - 4}{n_i - p} \quad (i = 1, \ldots, k)$$

である．

証明の概略

漸近正規性は，定理 3.1，定理 3.2 の結果を用いて示すことができる．また，行列を適当に分解し，ベキ等行列の性質を用いれば，V_* が V_w の下限になっていることも容易に示される．

□

4.　結び

定理 3.3 の結果から分かるように，$\hat{\beta}_w$ タイプの不偏推定量の漸近分散の下限 V_* は一般化最小二乗推定量の分散(1.2)とは一致しない．先にも述べたように，n_i $(i=1,\ldots,k)$ が総じて小さいとき，V_* は(1.2)から乖離してしまう．このことから，残差平方和の逆数（の定数倍）を利用して β の推定量を構成する手法の限界が読み取れる．

不偏推定量 $\hat{\beta}_w$ のこのような漸近的な性質は，$\Omega_i=I_{n_i}$ なる強い条件下で Inoue (2003) が既に論じている．

本稿では，$\Omega_i=I_{n_i}$ という条件を(1.1)という緩い条件に置き換えた場合，それが $\hat{\beta}_w$ の挙動にどのような影響を及ぼすのかを調べた．そして，Ω_i $(i=1,\ldots,k)$ が既知であり，なおかつ条件(1.1)が満たされる場合，推定量の挙動は $\Omega_i=I_{n_i}$ なる条件下におけるそれと殆ど同じであることが観察できた．

なお，$\hat{\beta}_w$ の挙動を調べていく際，展開式の中に相関係数 ρ_i $(i=1,\ldots,k)$ が登場してくることはあったが，最終的な結果を見ると ρ_i $(i=1,\ldots,k)$ が Ω_i $(i=1,\ldots,k)$ 外に出てくることはなかった．定理 3.3 の結果を見れば分かるように，漸近分散を最小にする w_i $(i=1,\ldots,k)$ は n_i と p に依存するのみであり，相関係数 ρ_i には依存していない．

今回の結果は，既知の Ω_i $(i=1,\ldots,k)$ が(1.1)という形で与えられるという仮定の下で導かれたものである．しかし，たとえば ρ_i $(i=1,\ldots,k)$ が未知の場合，Ω_i $(i=1,\ldots,k)$ の構造が(1.1)とは異なっている場合にはどのような現象が観察できるのであろうか．その他にも，ϵ_i $(i=1,\ldots,k)$ に相関構造を想定することも考えられる．これらに関する議論については今後の課題としたい．

※本稿は平成14年度～平成15年度文部科学省科学研究費（若手研究Ｂ）（課題番号：14780166），及び平成15年度早稲田大学個人特定課題研究助成費（課題番号：2003A-010）による成果の一部である．

参考文献

Inoue, K. (2003). Iterative weighted least-squares estimates in a heteroscedastic linear regression model. *J. Statist. Plann. Inference* 110, 1-2, pp.133-146.