

多様な関数を用いた経時変動の罰則付推定法

中京大学 国際教養学部 永井 勇

1. 導入

各個体において経時的に測定して得られるデータは経時測定データと呼ばれ、非常に多くの分野で収集・分析がされている。この経時測定データの分析では、平均や説明変数に対する経時的な変動（経時変動）を上手く捉えることが一つの目的である。本講演では、 i 番目の個体の j 時点 t_j での測定値を $y_i(t_j)$ と表し、 $t_1, \dots, t_p$ は全ての個体で共通とする。

経時変動全体を捉えるために広く使われる手法では、測定時点の多項式を用いて、平均的な経時変動と説明変数に対する経時変動を推定している。このとき、 q 個の未知の係数 $m_1, \dots, m_q$ を用いて $m_1 + m_2 t_j + \dots + m_q t_j^{q-1}$ とすることで、測定時点の $(q-1)$ 次多項式を用いた推定を考えることができる。ここで $\mathbf{m} = (m_1, \dots, m_q)'$ とすると、この項は $(1, t_j, \dots, t_j^{q-1})\mathbf{m}$ と書ける。この項を平均的な経時変動を表す項とする。一方で、各個体の説明変数に関する経時変動についても同様に測定時点の $(q-1)$ 次多項式での推定を考え

ると、未知の係数 ξ_{ij} を用いて $\left(\sum_{i=1}^q \xi_{1i} t_j^{i-1}, \dots, \sum_{i=1}^q \xi_{ki} t_j^{i-1} \right)$ となる。ここで (i, j) 成分が未知の ξ_{ij} からなる $k \times q$ 行列を Ξ とすると、この項は $(1, t_j, \dots, t_j^{q-1})\Xi'$ となり、これが i 番目の個体の k 個の説明変数 $\mathbf{a}_i$ (k 次元ベクトル) に関する経時変動を表現する項となる。この二つの項の表現より、 $\mathbf{X}$ の j 行目を $(1, t_j, \dots, t_j^{q-1})$ とし、 i 番目の個体の経時測定データ $\mathbf{y}_i = (y_i(t_1), \dots, y_i(t_p))'$ の経時変動全体を測定時点の $(q-1)$ 次多項式で捉えられるとすると、誤差ベクトル $\boldsymbol{\varepsilon}_i$ を用い、各個体に対し $\mathbf{y}'_i = \mathbf{X}\mathbf{m} + \mathbf{X}\Xi'\mathbf{a}_i + \boldsymbol{\varepsilon}_i$ とモデル化できる。

さらにここで、個体数を n 、全ての成分が 1 の n 次元ベクトルを $\mathbf{1}_n$ とし、 i 行目が $\mathbf{y}'_i$ からなる $n \times p$ 既知行列 $\mathbf{Y}$ 、 i 行目が $\mathbf{a}'_i$ からなる $n \times k$ 既知行列 $\mathbf{A}$ (ただし中心化したもの、つまり $\mathbf{A}'\mathbf{1}_n = \mathbf{0}_k$, $\mathbf{0}_k$ は k 次元ゼロベクトル)、 i 行目が $\boldsymbol{\varepsilon}'_i$ からなる $n \times p$ 誤差行列 $\boldsymbol{\varepsilon}$ を用いると、各個体に対するモデルは GMANOVA (一般化多変量分散分析) モデル (Pothoff & Roy, 1964) と同じ $\mathbf{Y} = \mathbf{1}_n \mathbf{m}' \mathbf{X}' + \mathbf{A} \Xi \mathbf{X}' + \boldsymbol{\varepsilon}$ という形にまとめることができる。このモデルにおいて q 次元未知ベクトル $\mathbf{m}$ と $k \times q$ 未知行列 Ξ を推定することで、 $\mathbf{Y}$ の経時変動全体を測定時点 $t_1, \dots, t_p$ の $(q-1)$ 次多項式で捉えることができる。さらに、 $\mathbf{X}$ の (i, j) 成分を t_j に関する柔軟な関数に置き換えることで、それらの関数の重み付和により経時変動全体を捉えることもできる。このとき、従来の推定法では経時変動全体ではなく経時測定データ $\mathbf{Y}$ へ過剰に適合してしまうが、この過剰適合を回避する推定法を Nagai (2011) で提案した。

この GMANOVA モデルを用いた場合、 $\mathbf{Y}$ の経時変動全体の推定に用いる関数は、平均的な経時変動の項 ($= \mathbf{1}_n \mathbf{m}' \mathbf{X}'$) も各説明変数に関する経時変動の項 ($= \mathbf{A} \Xi \mathbf{X}'$) も同じ関数系 (例えば、同じ次数の多項式) を用いた形に限られる。そこで本講演では、平均や説明変数により異なる関数系を用いた推定を考える。つまり例えば、平均的な経時変動は $(q_0 - 1)$ 次多項式で、いくつかの説明変数 $\mathbf{A}_1$ に関する経時変動は $(q_1 - 1)$ 次多項式で、残りの説明変数 $\mathbf{A}_2$ に関する経時変動は $(q_2 - 1)$ 次多項式で推定することを考える。このとき、上記の GMANOVA モデルと同様に考えると、次のような Extended GMANOVA モデル (Kollo & von Rosen (2005) Definition 4.1.3 を少し変形したモデル) による推定を考えることになる；

$$\mathbf{Y} = \mathbf{1}_n \boldsymbol{\mu}' \mathbf{X}'_0 + \mathbf{A}_1 \Xi_1 \mathbf{X}'_1 + \mathbf{A}_2 \Xi_2 \mathbf{X}'_2 + \boldsymbol{\varepsilon}, \quad (1)$$

ここで $\boldsymbol{\mu}$ は q_0 次元未知ベクトル、 $\mathbf{X}_i$ は j 行目が $(1, t_j, \dots, t_j^{q_i-1})$ などからなる $p \times q_i$ 既知行列 ($i = 0, 1, 2$; $\text{rank}(\mathbf{X}_i) = q_i \leq p$)、 $\mathbf{A}_i$ は $n \times k_i$ 既知行列 ($i = 1, 2$; $\text{rank}(\mathbf{A}_i) = k_i \leq n$)、た

だし $\mathbf{A}_i' \mathbf{1}_n = \mathbf{0}_{k_i}$), Ξ_i は $k_i \times q_i$ 未知行列 ($i = 1, 2$) であり, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ は $n \times p$ 誤差行列である. 本講演では, $E[\varepsilon_i] = \mathbf{0}_p$, $\text{Cov}[\varepsilon_i] = \boldsymbol{\Sigma}$ ($p \times p$ 未知正定値行列), $\varepsilon_i \perp \varepsilon_j$ ($j \neq i$), $n - k_1 - k_2 - 1 > 0$ とする. また, 従来の Kollo and von Rosen (2005) での定義や推定などでは $\mathbf{1}_n$, $\mathbf{A}_1$, $\mathbf{A}_2$ や $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$ に対応する部分などに階層構造を仮定しているが, 本講演ではこの構造を特に仮定しない. (1) 式のモデルにおいて $\boldsymbol{\mu}$, Ξ_1 , Ξ_2 を推定することで, それぞれに対して測定時点の $(q_i - 1)$ 次多項式などを用いて経時変動全体が推定できる. つまり例えば, 平均的な経時変動は 3 次多項式を, 性別に関する経時変動は 2 次多項式を, 他の変数に関する経時変動は 4 次多項式を用いた推定ができる.

本講演では (1) 式のモデルで経時変動全体を推定する際に, 平均的な経時変動の項 ($= \mathbf{1}_n \boldsymbol{\mu}' \mathbf{X}_0'$) や一部の説明変数に関する経時変動の項 ($= \mathbf{A}_1 \Xi_1 \mathbf{X}_1'$) にはそれぞれ測定時点の多項式を用いて, 他の説明変数に関する経時変動の項 ($= \mathbf{A}_2 \Xi_2 \mathbf{X}_2'$) を柔軟な関数を用いて推定することを考える. このときに起きる以下の問題の回避を本講演で考える;

- Ξ_i の推定量のために Ξ_j ($j \neq i$) の推定量が必要となる
(つまり, 罰則を付ける際に Ξ_2 の推定だけに着目した罰則を付けるのが難しい)
- 経時変動全体ではなく経時測定データ $\mathbf{Y}$ へ過剰に適合する

• 罰則を付けた場合の罰則パラメータの最適化のために初期値や反復計算が必要となる
本講演ではこれらの問題を回避するために, Ξ_1 の推定量を使わない形での Ξ_2 の推定量をまず構築する. その後, Nagai (2011) と同様に, 過剰適合の問題を回避するために罰則付推定量を構築する. そこで構築した推定量において, Mallows (1973) や Yanagihara and Satoh (2010) などと同様の手法で, 罰則パラメータなどの最適化のための C_p 型情報量規準を構築する. さらに, いくつかの仮定を追加した上で, この規準のバイアスを補正した情報量規準も構築する. 最後に, 構築した規準を最小にする罰則パラメータが陽に求まることを示す.

2. 提案する推定法

$\boldsymbol{\mu}$, Ξ_1 や Ξ_2 の従来の推定量 (最小二乗推定量) を求めると, $\mathbf{A}_i$ が中心化されているので,

$$\begin{cases} \hat{\boldsymbol{\mu}} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Y}' \mathbf{1}_n / n, \\ \hat{\Xi}_1 = (\mathbf{A}_1' \mathbf{A}_1)^{-1} \mathbf{A}_1' (\mathbf{Y} - \mathbf{A}_2 \hat{\Xi}_2 \mathbf{X}_2') \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1}, \\ \hat{\Xi}_2 = (\mathbf{A}_2' \mathbf{A}_2)^{-1} \mathbf{A}_2' (\mathbf{Y} - \mathbf{A}_1 \hat{\Xi}_1 \mathbf{X}_1') \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1}. \end{cases}$$

したがって, $\hat{\boldsymbol{\mu}}$ はそのまま求めることができるが, $\hat{\Xi}_1$ と $\hat{\Xi}_2$ はそれぞれ互いに影響し合っている. そのため, $\mathbf{X}_2$ に柔軟な関数を用いる $\mathbf{A}_2 \Xi_2 \mathbf{X}_2'$ の部分に対し, 過剰適合を回避しつつ経時変動全体を推定するため Ξ_2 の推定にだけ罰則を付ける推定方法を構築するのが難しい.

そこで本講演では, まず vec 作用素やクロネッカー積の性質 (例えば Lütkepohl (1996) 参照) を用いて, Ξ_1 の推定量を用いずに Ξ_2 の推定量を得る手法を提案する. その後, 提案する推定量において, 一般化リッジ回帰による推定法 (Hoerl & Kennard, 1970) などの手法を用いた罰則付推定量を構築する.

2.1. Ξ_2 に着目する推定法

まず, 上記の推定量 $\hat{\Xi}_2$ を変形し, $\hat{\Xi}_1$ を用いずに既知の $\mathbf{X}_1$, $\mathbf{A}_1$, $\mathbf{X}_2$, $\mathbf{A}_2$, $\mathbf{Y}$ で $\hat{\Xi}_2$ を得る形を求める. このとき, vec 作用素やクロネッカー積の性質 (例えば Lütkepohl (1996) 参照) を用いて変形を行う. 具体的には, 適当な行列 $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$ において, $\text{vec}(\mathbf{B} - \mathbf{C}) = \text{vec}(\mathbf{B}) - \text{vec}(\mathbf{C})$, $\text{vec}(\mathbf{BCD}) = (\mathbf{D}' \otimes \mathbf{B}) \text{vec}(\mathbf{C})$, $(\mathbf{BC} \otimes \mathbf{DE}) = (\mathbf{B} \otimes \mathbf{D})(\mathbf{C} \otimes \mathbf{E})$, $(\mathbf{B} \otimes \mathbf{C})' = \mathbf{B}' \otimes \mathbf{C}'$, $(\mathbf{B} \otimes \mathbf{C})^{-1} = \mathbf{B}^{-1} \otimes \mathbf{C}^{-1}$ という性質を用いる. その結果, $\mathbf{P}_M \stackrel{\text{def.}}{=} \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$, $\mathbf{Z} \stackrel{\text{def.}}{=} (\mathbf{I}_{np} - \mathbf{P}_{\mathbf{X}_1 \otimes \mathbf{A}_1})(\mathbf{X}_2 \otimes \mathbf{A}_2)$ を用いると, $\text{vec}(\hat{\Xi}_2) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\text{vec}(\mathbf{Y})$ となる. この変形によ

り, $\hat{\Xi}_2$ を並べ替えた $\text{vec}(\hat{\Xi}_2)$ が $\hat{\Xi}_1$ を用いずに得ることが出来る. またこの推定量の形は, 従来の重回帰モデルにおける未知回帰係数ベクトルの推定量と同じ形である. したがって, Hoerl and Kennard (1970) の一般化リッジ回帰による推定法を用いることで, $\text{vec}(\hat{\Xi}_2)$ を求める際に罰則を付けることができる.

2.2. 罰則付推定法

次に, 先の $\text{vec}(\hat{\Xi}_2) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\text{vec}(\mathbf{Y})$ の形の推定量に Hoerl and Kennard (1970) で提案された一般化リッジ回帰による推定法を用いることで, Nagai (2011) と同様に柔軟な関数を用いても, 経時測定データ $\mathbf{Y}$ への過剰適合をしない推定量を提案する.

d_i を $\mathbf{Z}'\mathbf{Z}$ の固有値とし, $\mathbf{Q}$ を $\mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} = \text{diag}(d_1, \dots, d_{q_2k_2})$ (対角に $d_1, \dots, d_{q_2k_2}$ を並べた行列) となるような直交行列とし, $\theta_i \geq 0$ を罰則パラメータとし $\boldsymbol{\Theta} = \text{diag}(\theta_1, \dots, \theta_{q_2k_2})$ とすると, 一般化リッジ回帰による推定法を用いた $\text{vec}(\Xi_2)$ の罰則付推定量は, 次の形となる;

$$\text{vec}(\hat{\Xi}_2(\boldsymbol{\theta})) = (\mathbf{Z}'\mathbf{Z} + \mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}')^{-1}\mathbf{Z}'\text{vec}(\mathbf{Y}).$$

ここで $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{q_2k_2})'$. また, この推定量を並び替えることで $\hat{\Xi}_2(\boldsymbol{\theta})$ が得られ, それを用いて $\hat{\Xi}_1$ を得ることができる. 次に本講演では $\boldsymbol{\theta}$ の最適化を考える.

3. 情報量規準に基づく罰則パラメータなどの最適化

提案した推定量 $\text{vec}(\hat{\Xi}_2(\boldsymbol{\theta}))$ とそれを用いた Ξ_1 の推定量 ($\hat{\Xi}_1(\boldsymbol{\theta})$) と $\hat{\boldsymbol{\mu}}$ を用いた予測値を $\hat{\mathbf{Y}}(\boldsymbol{\theta}) = \mathbf{1}_n\hat{\boldsymbol{\mu}}'\mathbf{X}'_0 + \mathbf{A}_1\hat{\Xi}_1(\boldsymbol{\theta})\mathbf{X}'_1 + \mathbf{A}_2\hat{\Xi}_2(\boldsymbol{\theta})\mathbf{X}'_2$ として, 最適な $\boldsymbol{\theta}$ を求めることを考える. そのために, 次で定義される予測平均二乗誤差 (PMSE) を用いる;

$$\text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})] \stackrel{\text{def.}}{=} E_U \left[E_Y \left[\text{tr} \left\{ \left(\mathbf{U} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right) \boldsymbol{\Sigma}^{-1} \left(\mathbf{U} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right)' \right\} \right] \right],$$

ここで, $\mathbf{U}$ は $\mathbf{Y}$ と独立に同一の分布に従う確率変数であり, $E_U[\cdot]$ や $E_Y[\cdot]$ はそれぞれの確率変数に基づく期待値である. これを最小にする $\boldsymbol{\theta}$ を求めればよいが, そのまま最小化しても未知の $\boldsymbol{\Sigma}$ が残ってしまう. そこで, Mallows (1973), Yanagihara, Nagai and Satoh (2009), Yanagihara and Satoh (2010) と同様に, この PMSE を推定する関数として C_p 型情報量規準を構築し, その最小化を考える.

3.1. 情報量規準の構築

まずこの $\text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})]$ を展開し, C_p 型情報量規準を構築する. $\text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})]$ を展開すると,

$$\text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})] = E_Y \left[\text{tr} \left\{ \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right) \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right)' \right\} \right] + \sum_{i=1}^{q_2k_2} \frac{2d_i}{d_i + \theta_i} + 2(q_0 + q_1k_1),$$

となるため, C_p 型情報量規準は次の形となる;

$$C_p(\boldsymbol{\theta}) \stackrel{\text{def.}}{=} \text{tr} \left\{ \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right) \mathbf{S}^{-1} \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right)' \right\} + \sum_{i=1}^{q_2k_2} \frac{2d_i}{d_i + \theta_i} + 2(q_0 + q_1k_1),$$

ここで $\mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n} - \mathbf{P}_{\mathbf{A}_1} - \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}_1})\mathbf{A}_2})\mathbf{Y} / (n - k_1 - k_2 - 1)$ ($\det(\mathbf{S}) \neq 0$ を仮定) であり, これは $\boldsymbol{\Sigma}$ の不偏推定量である.

この $C_p(\boldsymbol{\theta})$ は $\text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})]$ の $\boldsymbol{\Sigma}^{-1}$ を $\mathbf{S}^{-1}$ に置き換えて $E_Y[\cdot]$ を単に外した形なので, $E_Y[C_p(\boldsymbol{\theta})] \neq \text{PMSE}[\hat{\mathbf{Y}}(\boldsymbol{\theta})]$, つまりバイアスがある. そこで本講演では, さらに $n - k_1 - k_2 - p - 2 > 0$ と $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N_p(\mathbf{0}_p, \boldsymbol{\Sigma})$ の二つの仮定を加えてバイアス補正を行う. そこで, 適当

な大きさの定数行列 $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ において, $\mathbf{Y}'\mathbf{F}\mathbf{Y}$ と $\mathbf{Y}'\mathbf{G}\mathbf{Y}$, $\mathbf{Y}'\mathbf{F}\mathbf{Y}$ と $\mathbf{H}\mathbf{Y}$ が独立な確率変数からなる行列かを考える. このとき, Schott (2005) Theorem 10.24 を用い, $\mathbf{A}_1$ や $\mathbf{A}_2$ が中心化されている ($\mathbf{A}_i'\mathbf{1}_n = \mathbf{0}_{k_i}$) ことに注意すると, 適当な大きさのゼロ行列を $\mathbf{O}$ とすると, $\mathbf{F}\mathbf{G} = \mathbf{O}$, $\mathbf{F}\mathbf{H} = \mathbf{O}$ であればよいことが分かる. この結果を用いると, Nagai (2011) と同様に C_p 型情報量規準のバイアス補正が可能であることが分かる. したがって, $E_{\mathbf{Y}}[C_p(\boldsymbol{\theta})]$ のバイアスを補正した情報量規準を導出すると次の形となる;

$$MC_p(\boldsymbol{\theta}) \stackrel{\text{def.}}{=} C_M \text{tr} \left\{ \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right) \mathbf{S}^{-1} \left(\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}) \right)' \right\} + \sum_{i=1}^{q_2 k_2} \frac{2d_i}{d_i + \theta_i} + 2(q_0 + q_1 k_1) + p(p+1),$$

ここで $C_M = 1 - (p+1)/(n - k_1 - k_2 - 1)$.

本講演では最後に, $C_p(\boldsymbol{\theta})$ や $MC_p(\boldsymbol{\theta})$ を最小にする $\boldsymbol{\theta}$ が陽に求まることを示す.

3.2. 罰則パラメータなどの最適化

$C_p(\boldsymbol{\theta})$ や $MC_p(\boldsymbol{\theta})$ において $\boldsymbol{\theta}$ に関係する項に着目すると, $GC_p(\boldsymbol{\theta}|\alpha) \stackrel{\text{def.}}{=} \alpha \text{tr} \{ (\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta})) \mathbf{S}^{-1} (\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta}))' \} + \sum_{i=1}^{q_2 k_2} \frac{2d_i}{d_i + \theta_i}$ を用いて, $C_p(\boldsymbol{\theta}) = GC_p(\boldsymbol{\theta}|1) + 2(q_0 + q_1 k_1)$, $MC_p(\boldsymbol{\theta}) = GC_p(\boldsymbol{\theta}|C_M) + 2(q_0 + q_1 k_1) + p(p+1)$ と表される. よって, $\boldsymbol{\theta}$ の各成分が非負という条件下で $\hat{\boldsymbol{\theta}}^{[G]}(\alpha) \stackrel{\text{def.}}{=} \arg \min_{\boldsymbol{\theta}} GC_p(\boldsymbol{\theta}|\alpha)$ とすると, $\hat{\boldsymbol{\theta}}^{[G]} \stackrel{\text{def.}}{=} \arg \min_{\boldsymbol{\theta}} C_p(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}^{[G]}(1)$, $\hat{\boldsymbol{\theta}}^{[M]} \stackrel{\text{def.}}{=} \arg \min_{\boldsymbol{\theta}} MC_p(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}^{[G]}(C_M)$ である. ここで, $GC_p(\boldsymbol{\theta}|\alpha)$ をそのまま展開すると, $\hat{\boldsymbol{\theta}}^{[G]}(\alpha)$ の第 i 成分 $\hat{\theta}_i^{[G]}(\alpha)$ のために他の $\hat{\theta}_j^{[G]}(\alpha)$ ($j \neq i$) が全て必要となるため, 初期値と反復計算が必要となる.

そこで本講演では $GC_p(\boldsymbol{\theta}|\alpha)$ を工夫して展開する. ここで, $(\mathbf{J})_{ij}$ が行列 $\mathbf{J}$ の (i, j) 成分を表しているとし, $\mathbf{R} = \mathbf{P}_{\mathbf{X}_0 \otimes \mathbf{1}_n} + \mathbf{P}_{\mathbf{X}_1 \otimes \mathbf{A}_1}$ を用いて, 次の関数を定義する;

$$\begin{aligned} \varphi_i(\theta_i|\alpha) \stackrel{\text{def.}}{=} & -\frac{2\alpha}{d_i + \theta_i} (\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'(\mathbf{I}_{np} - \mathbf{R})(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii} \\ & + \frac{\alpha}{(d_i + \theta_i)^2} (\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ii} (\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii} \\ & + \frac{2\alpha}{d_i + \theta_i} \sum_{j>i} \frac{(\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ij} (\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ji}}{d_j + \theta_j} \\ & + \frac{2d_i}{d_i + \theta_i}. \end{aligned}$$

ただし, $\sum_{j>q_2 k_2} (\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ij} (\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ji} / (d_j + \theta_j) = 0$. これを用いると, $GC_p(\boldsymbol{\theta}|\alpha)$ を工夫して展開したものは次の形で得られる;

$$\begin{aligned} GC_p(\boldsymbol{\theta}|\alpha) = & \alpha \times \text{vec}(\mathbf{Y})'(\mathbf{I}_{np} - \mathbf{R})(\mathbf{S}^{-1} \otimes \mathbf{I}_n)(\mathbf{I}_{np} - \mathbf{R})\text{vec}(\mathbf{Y}) \\ & + \sum_{i=1}^{q_2 k_2} \varphi_i(\theta_i|\alpha). \end{aligned}$$

したがって, 各 i で $\hat{\theta}_i^{[G]}(\alpha) = \arg \min_{\theta_i \geq 0} \varphi_i(\theta_i|\alpha)$ となる. さらに $\varphi_i(\theta_i|\alpha)$ の形より, $\hat{\theta}_i^{[G]}(\alpha)$ ($i < q_2 k_2$) を得る際には $\hat{\theta}_j^{[G]}(\alpha)$ ($i < j$) だけで十分であり, $\varphi_{q_2 k_2}(\theta_{q_2 k_2}|\alpha)$ の第三行目の値が 0 となることに注意すると, $\hat{\theta}_{q_2 k_2}^{[G]}(\alpha)$ を得る際には他の $\hat{\theta}_i^{[G]}(\alpha)$ ($i \neq q_2 k_2$) が不要である.

また具体的に $\hat{\theta}_i^{[G]}(\alpha)$ ($i = 1, \dots, q_2 k_2$) を求めると、次の形となる;

$$\hat{\theta}_i^{[G]}(\alpha) = \begin{cases} 0 & \left(\text{if } 0 \leq \frac{d_i^3}{2} \dot{\varphi}_i(0|\alpha) \right) \\ \infty & \left(\text{if } \frac{d_i^3}{2} \dot{\varphi}_i(0|\alpha) + \alpha(\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ii}(\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii} \leq 0 \right) \\ \frac{-\frac{d_i^3}{2} \dot{\varphi}_i(0|\alpha) \times d_i}{\frac{d_i^3}{2} \dot{\varphi}_i(0|\alpha) + \alpha(\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ii}(\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii}} & (\text{上記以外}) \end{cases},$$

ここで $\frac{d_i^3}{2} \dot{\varphi}_i(0|\alpha) = \alpha d_i (\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'(\mathbf{I}_{np} - \mathbf{R})(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii} - \alpha (\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ii}(\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ii} - \alpha d_i \sum_{j>i} \frac{(\mathbf{Q}'\mathbf{Z}'\text{vec}(\mathbf{Y})\text{vec}(\mathbf{Y})'\mathbf{Z}\mathbf{Q})_{ij}(\mathbf{Q}'\mathbf{Z}'(\mathbf{S}^{-1} \otimes \mathbf{I}_n)\mathbf{Z}\mathbf{Q})_{ji}}{d_j + \theta_j} - d_i^2$ であり、再掲の記号を列挙すると $\mathbf{Z} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}_1 \otimes \mathbf{A}_1})(\mathbf{X}_2 \otimes \mathbf{A}_2)$, d_i は $\mathbf{Z}'\mathbf{Z}$ の固有値, $\mathbf{Q}$ は $\mathbf{Q}'\mathbf{Z}'\mathbf{Z}\mathbf{Q} = \text{diag}(d_1, \dots, d_{q_2 k_2})$ となる直交行列, $\mathbf{S} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n} - \mathbf{P}_{\mathbf{A}_1} - \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}_1})\mathbf{A}_2})\mathbf{Y}/(n - k_1 - k_2 - 1)$, $\mathbf{R} = \mathbf{P}_{\mathbf{X}_0 \otimes \mathbf{1}_n} + \mathbf{P}_{\mathbf{X}_1 \otimes \mathbf{A}_1}$ である。

この結果から, $\varphi_i(\theta_i|\alpha)$ 定義よりまず $\hat{\theta}_{q_2 k_2}^{[G]}(\alpha)$ が陽に得られ, $\hat{\theta}_i^{[G]}(\alpha)$ ($i = 1, \dots, q_2 k_2 - 1$) は $\hat{\theta}_j^{[G]}(\alpha)$ ($i < j$) を使って順次陽に得られ, その結果から $\hat{\boldsymbol{\theta}}^{[G]}(\alpha)$ ($q_2 k_2$ 個のパラメータ) が全て陽に得られる. また α をそれぞれの情報量規準に対応する値に変更することで, $\hat{\boldsymbol{\theta}}^{[G]}(1)$ で $\hat{\boldsymbol{\theta}}^{[C]} (= C_p(\boldsymbol{\theta}))$ を最小にする非負の罰則パラメータ) が, $\hat{\boldsymbol{\theta}}^{[G]}(C_M)$ で $\hat{\boldsymbol{\theta}}^{[M]} (= MC_p(\boldsymbol{\theta}))$ を最小にする非負の罰則パラメータ) が得られる。

他のパラメータ q_0 (平均的な経時変動の推定に用いる多項式の次数), q_1 (一部の説明変数 $\mathbf{A}_1$ に関する経時変動の推定に用いる多項式の次数), q_2 (他の説明変数 $\mathbf{A}_2$ に関する経時変動の推定に用いる柔軟な関数の個数) は $C_p(\hat{\boldsymbol{\theta}}^{[C]})$ や $MC_p(\hat{\boldsymbol{\theta}}^{[M]})$ の最小化により選択できる。

数値実験を通じた比較などは講演で示す。

4. 参考文献

- [1] Hoerl, A. E. & Kennard R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67.
- [2] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Springer.
- [3] Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley & Sons.
- [4] Mallows, C. L. (1973). Some comments on C_p . *Technometrics*, **15**, 661–675.
- [5] Nagai, I. (2011). Modified C_p criterion for optimizing ridge and smooth parameters in the MGR estimator for the nonparametric GMANOVA model. *Open J. Stat.*, **1**, 1–14.
- [6] Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.
- [7] Schott, J. R. (2005). *Matrix Analysis for Statistics* (Second Edition). John Wiley & Sons.
- [8] Yanagihara, H., Nagai, I. & Satoh, K. (2009). A bias-corrected C_p criterion for optimizing ridge parameters in multivariate generalized ridge regression. *Jpn. J. Appl. Stat.*, **38**, 151–172 (in Japanese).
- [9] Yanagihara, H. & Satoh, K. (2010). An unbiased C_p criterion for multivariate ridge regression. *J. Multivariate Anal.*, **101**, 1226–1238.