

# 小地域推定におけるベンチマーク問題と制約付き経験ベイズ推定の誤差評価

久保川 達也\*

## 1 はじめに

小地域の平均の推定においては、小地域からのデータ数が少ないため、標本平均で推定したのでは推定誤差が大きくなってしまい適切な推定が行えない。そこで、線形混合モデルなどベイズ的構造を導入して経験最良線形不偏予測量 (EBLUP) もしくは経験ベイズ推定量 (EB) を導出し、周辺地域のデータを組み込んだ安定した推定値の方向へ標本平均を縮小することにより、小地域の平均に対する推定精度を高めることができる。一方、広範な地域の平均の推定を考えると、十分なデータが利用できることから広範な標本平均は精度良く推定していると考えられるが、この広範な地域の推定を、個々の小地域の経験ベイズ推定値の(加重)平均で推定してみると、これは必ずしも広範な標本平均とは一致しない。これは、経験ベイズ推定量を利用する際の1つの問題点として知られている。もう1つの問題は、経験ベイズ推定値は縮小し過ぎている点であり、実際、経験ベイズ推定値の標本分散は事後分布の分散より小さくなっていることが指摘されている。そこで、これらの問題を解決する推定量として、制約付き経験ベイズ推定量 (CEB) を導出し、その平均2乗誤差 (MSE) の2次漸近近似とその2次不偏推定量を求める。また、県別家計調査データや市町村別胃がん死亡率データの解析を通してここで導出された手法の適用の様子を調べる。

## 2 小地域推定におけるベンチマーク問題

### 2.1 経験ベイズ推定 (EB) の有用性

小地域推定において地域レベルのモデルとして用いられるのが Fay-Herriot モデルである。これは、個々のデータではなく、地域毎の集計データのみが利用可能な場合に用いられる簡単なモデルである。いま全体で  $K$  個の小地域 (市町村など) があり、各小地域の平均を  $\mu_1, \dots, \mu_K$  とし、集計データから作られた直接的な推定値を  $y_1, \dots, y_K$  とする。このとき、

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, \dots, K \quad (2.1)$$

なるモデルを考える。ここで、 $\varepsilon_1, \dots, \varepsilon_K$  は互いに独立に正規分布  $\varepsilon_i \sim \mathcal{N}(0, d_i)$  に従うものとする。集計データのみが利用可能なため分散  $d_i$  の値を求めることができないが、通常は過去何年間かのデータから推定しておいて、モデル (2.1) を考える上では  $d_i$  を既知の値として扱っている。

\*東京都文京区本郷 7-3-1 東京大学大学院経済学研究科 113-0033,

E-Mail: tatsuya@e.u-tokyo.ac.jp

全領域の特性値を推定する目的で標本調査が行われ十分なデータが得られて精度よく全平均の推定がなされる。同じデータを市町村などの小地域の平均の推定に再利用しようとする、地域によっては十分なデータが利用できないため推定誤差が極めて大きくなってしまふ。これを小地域推定問題という。上のモデルに即していうならば、 $d_i$  が大きいとき推定値  $y_i$  のバラツキが大きくなり推定精度が低くなってしまふ。小地域の推定精度を改善するために、周辺地域からのデータや共変量のデータを組み込むことが考えられる。Fay and Herriot (1979) は、(2.1) の  $\mu_i$  に次のような事前分布を想定したベイズモデルを提案した。

$$\mu_i \sim \mathcal{N}(\mathbf{x}_i^t \boldsymbol{\beta}, \lambda), \quad i = 1, \dots, K \quad (2.2)$$

ここで  $\mathbf{x}_i$  は共変量データを含んだ  $p$ -次元のベクトル、 $\boldsymbol{\beta}$  は  $p$ -次元の未知のベクトル、 $\lambda$  は未知の分散である。(2.1) と (2.2) を合わせたモデルを Fay-Herriot モデルという。このとき、 $y_i$  の周辺分布は  $\mathcal{N}(\mathbf{x}_i^t \boldsymbol{\beta}, \lambda + d_i)$ 、 $y_i$  を与えたときの  $\mu_i$  の事後分布は  $\mathcal{N}(\hat{\mu}_i^B(\lambda, \boldsymbol{\beta}), d_i \lambda / (\lambda + d_i))$  となる。ただし、事後分布の平均  $\hat{\mu}_i^B$  は  $\mu_i$  のベイズ推定量であり  $\hat{\mu}_i^B(\lambda, \boldsymbol{\beta}) = (\lambda y_i + d_i \mathbf{x}_i^t \boldsymbol{\beta}) / (\lambda + d_i)$  で与えられる。これは、 $y_i$  と  $\mathbf{x}_i^t \boldsymbol{\beta}$  を精度  $d_i^{-1}$ 、 $\lambda^{-1}$  の比で内分した自然な形をしている。 $\lambda$  と  $\boldsymbol{\beta}$  は未知なので周辺分布  $y_i \sim \mathcal{N}(\mathbf{x}_i^t \boldsymbol{\beta}, \lambda + d_i)$ 、 $i = 1, \dots, K$  から推定する必要がある。 $\lambda$  については最尤推定法・制限最尤推定法により求めることができるが、不偏推定に基づいた簡単な推定量を用いることもできる。 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_K)^t$ 、 $\mathbf{y} = (y_1, \dots, y_K)^t$ 、 $\mathbf{D} = \text{diag}(d_1, \dots, d_K)$  とおき、 $\mathbf{E}_0 = \mathbf{I}_K - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$  とすると、Prasad-Rao の不偏推定量に基づいた打ち切り推定量として

$$\hat{\lambda} = \{(K - p)^{-1}(\mathbf{y}^t \mathbf{E}_0 \mathbf{y} - \text{tr}[\mathbf{D} \mathbf{E}_0]), 0\}$$

を使うこともできる。 $\boldsymbol{\beta}$  は重み付き最小2乗推定量

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^K \mathbf{x}_i \mathbf{x}_i^t / (\hat{\lambda} + d_i) \right)^{-1} \sum_{i=1}^K \mathbf{x}_i y_i / (\hat{\lambda} + d_i)$$

で推定できる。これらをベイズ推定量  $\hat{\mu}_i^B(\lambda, \boldsymbol{\beta})$  に代入したものが経験ベイズ推定量 (EB)

$$\hat{\mu}_i^{EB} = \hat{\mu}_i^B(\hat{\lambda}, \hat{\boldsymbol{\beta}}) = \mathbf{x}_i^t \hat{\boldsymbol{\beta}} + \frac{\hat{\lambda}}{\hat{\lambda} + d_i} (y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})$$

である。 $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$  は全データに基づいた安定した推定値であり、 $d_i$  が大きいときには、 $\hat{\mu}_i^{EB}$  は  $y_i$  を  $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$  の方向へ大きく縮小することによって、 $y_i$  より高い精度で推定できることが期待される。実際、推定誤差を平均2乗誤差  $MSE(\lambda, \boldsymbol{\beta}, \hat{\mu}_i^{EB}) = E[(\hat{\mu}_i^{EB} - \mu_i)^2]$  で測り、 $K$  を大きくしたときの MSE の2次漸近不偏推定値を調べてみると、多くの場合  $y_i$  の MSE 推定値より小さくなっていることがわかる。 $d_i$  がすべて等しく  $\boldsymbol{\beta} = \mathbf{0}$  のときには、 $\hat{\mu}_i^{EB}$  は数理統計学の分野で知られる James-Stein 推定量にほぼ近い形を与えており、 $\mu_1, \dots, \mu_K$  の同時推定の枠組みでは  $K \geq 3$  のとき James-Stein 推定量の MSE は  $y_i$  の MSE より一様に小さくなることが理論的に示される。応用的側面からも、経験ベイズ推定は、小地域平均の推定問題や疫学における地域別疾病率の推定など、実際のデータ解析の場面で利用されている。

## 2.2 経験ベイズ推定の問題点

経験ベイズ推定量は推定精度を高める推定手法として高く評価される一方で、 $K$  個の地域全体の平均を経験ベイズ推定値で構成してみると、全体の標本平均の値と一致しないという問題があること

が指摘される。例えば、 $w_i = d_i^{-1} / \sum_{j=1}^K d_j^{-1}$  を重みとして加重平均をとったもの  $\bar{y}_w = \sum_{i=1}^K w_i y_i$  で全体の平均を推定するとき、一般に経験ベイズ推定値の加重平均をとったものは  $\bar{y}_w$  と一致しない。即ち、

$$\sum_{i=1}^K w_i \hat{\mu}_i^{EB} \neq \bar{y}_w$$

である。 $K$  が大きいときには、 $\bar{y}_w$  は高い精度で全体平均を推定しているので、個々の小地域推定の累積によって全体の推定値  $\bar{y}_w$  が得られるのが自然である。

もう1つの問題は、7.2節の図1と図2を眺めてみるとわかるように、経験ベイズ推定値は縮小し過ぎてしまう傾向にある。Louis (1984) により指摘されたように、ベイズ推定値の加重平均と加重分散を事後分布の加重平均と加重分散と比較すると次の性質が成り立つ。

$$(1) \sum_{i=1}^K w_i \hat{\mu}_i^B = \sum_{i=1}^K w_i E[\mu_i | \mathbf{y}]$$

$$(2) \sum_{i=1}^K w_i \{\hat{\mu}_i^B - \bar{\mu}_w^B\}^2 \leq \sum_{i=1}^K w_i E\{[\mu_i - \bar{\mu}_w]^2 | \mathbf{y}\}$$

ただし、 $\mathbf{y} = (y_1, \dots, y_K)^t$ ,  $\bar{\mu}_w^B = \sum_{i=1}^K w_i \hat{\mu}_i^B$ ,  $\bar{\mu}_w = \sum_{i=1}^K w_i \mu_i$  である。等式(1)ではベイズ推定値の加重平均が事後分布の加重平均に一致しているが、(2)では分散についてはベイズ推定値の方が小さくなることを意味している。実際、不等式(2)は

$$\sum_{i=1}^K w_i E\{[\mu_i - \bar{\mu}_w]^2 | \mathbf{y}\} = \sum_{i=1}^K w_i \{\hat{\mu}_i^B(\lambda, \beta) - \bar{\mu}_w^B(\lambda, \beta)\}^2 + \sum_{i=1}^K w_i \text{Var}(\mu_i - \bar{\mu}_w | \mathbf{y}) \quad (2.3)$$

より従う。この不等式は、ベイズ推定値が縮小し過ぎていることを意味している。またこのことから、経験ベイズ推定値についても縮小し過ぎてしまうという問題点を引き継いでいることになる。

### 2.3 ベンチマーク問題と制約付き経験ベイズ推定 (CEB)

経験ベイズ推定量に関する上述の2つの問題点を解決するために、まずベイズ推定量を修正して制約付きベイズ推定量  $\hat{\mu}_i^{CB} = \hat{\mu}_i^{CB}(\lambda, \beta)$  を導出する。ベイズ推定量を  $\hat{\mu}_i^B = \hat{\mu}_i^B(\lambda, \beta)$  と記述して、次のような制約条件を考える。

$$(B1) \sum_{i=1}^K w_i \hat{\mu}_i^{CB} = \sum_{i=1}^K w_i y_i$$

$$(B2) \sum_{i=1}^K w_i \{\hat{\mu}_i^{CB} - \bar{\mu}_w^{CB}\}^2 = \sum_{i=1}^K w_i \{\hat{\mu}_i^B - \bar{\mu}_w^B\}^2 + \sum_{i=1}^K w_i (1 - w_i) d_i \lambda / (d_i + \lambda)$$

ここで、 $\bar{\mu}_w^{CB} = \sum_{j=1}^K w_j \hat{\mu}_j^{CB}$ ,  $\bar{\mu}_w^B = \sum_{j=1}^K w_j \hat{\mu}_j^B$  である。

$$\Delta_m^\dagger = \sum_{i=1}^K \frac{w_i d_i}{d_i + \lambda} (y_i - \mathbf{x}_i^t \beta)$$

とおくと, (B1) は  $\sum_{i=1}^K w_i \hat{\mu}_i^{CB} = \sum_{i=1}^K w_i \hat{\mu}_i^B + \Delta_m^\dagger$  と表される。また (2.3) の右辺の第2項を  $\Delta_v^\dagger$  とおくと,  $\mu_i | y_i \sim \mathcal{N}(\hat{\mu}_i^B, d_i \lambda / (\lambda + d_i))$  より

$$\begin{aligned} \Delta_v^\dagger &= \sum_{i=1}^K w_i \text{Var}(\mu_i - \bar{\mu}_w | \mathbf{y}) = \sum_{i=1}^K w_i E[\{(\mu_i - \hat{\mu}_i^B) - \sum_{j=1}^K w_j (\mu_j - \hat{\mu}_j^B)\}^2 | \mathbf{y}_i] \\ &= \sum_{i=1}^K w_i (1 - w_i) E[(\mu_i - \hat{\mu}_i^B)^2 | y_i] = \sum_{i=1}^K w_i (1 - w_i) \frac{d_i \lambda}{\lambda + d_i} \end{aligned}$$

となるので, 制約条件 (B2) は, 推定量の加重分散が事後分布の加重分散に等しいこと, 即ち,

$$\sum_{i=1}^K w_i \{\hat{\mu}_i^{CB} - \sum_{j=1}^K w_j \hat{\mu}_j^{CB}\}^2 = \sum_{i=1}^K w_i E[\{\mu_i - \bar{\mu}_w\}^2 | \mathbf{y}]$$

を意味している。

制約条件 (B1), (B2) のもとで条件付き MSE  $E[(\hat{\mu}_i^{CB}(\mathbf{y}) - \mu_i)^2 | \mathbf{y}]$  を最小にする  $\hat{\mu}_i^{CB}(\mathbf{y})$  を, 条件付き最適化問題として解くと

$$\hat{\mu}_i^{CB} = \hat{\mu}_i^B + \{a_B - 1\} \{\hat{\mu}_i^B - \sum_{j=1}^K w_j \hat{\mu}_j^B\} + \Delta_m^\dagger(\mathbf{y})$$

と表される。ここで,  $a_B$  は

$$\{a_B\}^2 = 1 + \frac{\Delta_v^\dagger}{\sum_{i=1}^K w_i \{\hat{\mu}_i^B - \sum_{j=1}^K w_j \hat{\mu}_j^B\}^2}$$

である。制約付きベイズ推定量  $\hat{\mu}_i^{CB} = \hat{\mu}_i^{CB}(\lambda, \beta)$  は  $\lambda, \beta$  の関数だから, それらの推定量を代入したものを制約付き経験ベイズ推定量 (Constrained Empirical Bayes, CEB) といい,  $\hat{\mu}_i^{CEB}$  と書くことにする。

制約付き経験ベイズ推定量は分散を大きくすることによって縮小し過ぎた経験ベイズ推定量の改善を図っている。経験ベイズ推定量の目的が推定精度を高めることにあったので, 制限付き経験ベイズ推定量の方向性はこれに反するものである。そこで, 実際, どの程度 MSE を大きくしてしまっているのか評価することは意味がある。次の節では, かなり一般的な混合モデルの枠組みで制限付き経験ベイズ推定量の誤差評価について統一的な結果を与える。

### 3 一般的な混合モデルでの MSE に関する統一的な結果

ここでは, 一般的な混合モデルにおいて経験ベイズ推定量や制約付き経験ベイズ推定量の MSE に関して, その漸近的性質と2次漸近不偏推定について統一的な結果を紹介しよう。

#### 3.1 経験ベイズ推定 (EB) の MSE

$\mathbf{y} = (y_1, \dots, y_N)^t$  を観測可能な  $N$ -次元確率ベクトル,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^t$  を観測できない  $p$ -次元確率ベクトルとする。  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)^t$  を未知母数に関する  $q$ -次元ベクトルとする。確率変数

$\mathbf{y}$ ,  $\boldsymbol{\theta}$  は連続的でも離散的でもよいが、記法の便宜上、連続的なものとして記述する。 $(\boldsymbol{\theta}, \boldsymbol{\eta})$  を与えたときの  $\mathbf{y}$  の条件付き確率密度関数を  $f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta})$  で示し、 $\boldsymbol{\eta}$  を与えたときの  $\boldsymbol{\theta}$  の条件付き確率密度関数を  $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$  で示す。即ち、

$$\begin{aligned} \mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta} &\sim f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta}) \\ \boldsymbol{\theta}|\boldsymbol{\eta} &\sim \pi(\boldsymbol{\theta}|\boldsymbol{\eta}) \end{aligned} \quad (3.1)$$

と書かれ、これは一般的な混合モデルを表している。与えられた  $\boldsymbol{\eta}$  に対して  $\mathbf{y}$  の周辺密度関数と、 $(\mathbf{y}, \boldsymbol{\eta})$  を与えたときの  $\boldsymbol{\theta}$  の条件付き密度関数（事後分布）は、それぞれ

$$\begin{aligned} m_\pi(\mathbf{y}|\boldsymbol{\eta}) &= \int f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})d\boldsymbol{\theta} \\ \pi(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta}) &= f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})/m_\pi(\mathbf{y}|\boldsymbol{\eta}) \end{aligned} \quad (3.2)$$

と書かれる。ここでは、一般的なスカラー関数  $\xi(\boldsymbol{\theta}, \boldsymbol{\eta})$  を  $\hat{\xi}(\mathbf{y})$  で推定（もしくは予測）する問題を扱う。推定量は平均2乗誤差 (MSE)  $MSE(\boldsymbol{\eta}, \hat{\xi}) = E[(\hat{\xi}(\mathbf{y}) - \xi(\boldsymbol{\theta}, \boldsymbol{\eta}))^2]$  で評価されるものとする。 $\boldsymbol{\eta}$  が既知のとき、MSE を最小にする意味で  $\xi(\boldsymbol{\theta}, \boldsymbol{\eta})$  の最良な推定量は条件付き平均値

$$\hat{\xi}(\mathbf{y}, \boldsymbol{\eta}) = E[\xi(\boldsymbol{\theta}, \boldsymbol{\eta})|\mathbf{y}]$$

で与えられ、これはベイズの枠組みではベイズ推定量である。本論文では、 $\boldsymbol{\eta}$  は未知母数なので、周辺密度  $m_\pi(\mathbf{y}|\boldsymbol{\eta})$  から推定する必要があり、得られる推定量を  $\hat{\boldsymbol{\eta}}$  と書く。これを  $\hat{\xi}(\mathbf{y}, \boldsymbol{\eta})$  に代入したものを  $\hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  が経験ベイズ推定量 (EB) である。

次に経験ベイズ推定量  $\hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  の MSE に対して  $N$  に関する2次近似を求めてみよう。そのために、次を仮定する。

(A1)  $\boldsymbol{\eta}$  の次元  $q$  は有界である。推定量  $\hat{\boldsymbol{\eta}}$  は、 $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta} = O_p(N^{-1/2})$ ,  $E[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}] = O(N^{-1})$  を満たす。

(A2) ベイズ推定量  $\hat{\xi}(\mathbf{y}, \boldsymbol{\eta})$  は  $\boldsymbol{\eta}$  に関して連続微分可能であり、 $\partial \hat{\xi}(\mathbf{y}, \boldsymbol{\eta})/\partial \eta_i = O_p(1)$ ,  $i = 1, \dots, q$ , である。

仮定 (A1), (A2) のもとで  $\hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  の MSE の2次近似が得られる。

$$g_1(\boldsymbol{\eta}) = E[\text{Var}(\xi(\boldsymbol{\theta}, \boldsymbol{\eta})|\mathbf{y})] \quad (3.3)$$

$$g_2(\boldsymbol{\eta}) = E\left[\left\{(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial \hat{\xi}(\mathbf{y}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right\}^2\right] \quad (3.4)$$

**定理 3.1** (A1), (A2) と  $g_1(\boldsymbol{\eta}) = O(1)$  を仮定すると、EB  $\hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  の MSE は

$$MSE(\boldsymbol{\eta}, \hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = g_1(\boldsymbol{\eta}) + g_2(\boldsymbol{\eta}) + o(N^{-1}) \quad (3.5)$$

と近似される。ここで  $g_2(\boldsymbol{\eta}) = O(N^{-1})$  である。

### 3.2 EB の MSE の推定

定理 3.1 から, MSE に対する 2 次不偏推定量を与えよう。これには 2 つの方法があり, 1 つはテーラー級数展開によるもの, もう 1 つはパラメトリック・ブートストラップによるものである。次を仮定する。

(A3)  $g_1(\boldsymbol{\eta})$  は連続微分可能で  $g_1(\boldsymbol{\eta}) = O(1)$ ,  $\partial g_1(\boldsymbol{\eta})/\partial \eta_i = O_p(1)$ ,  $\partial^2 g_1(\boldsymbol{\eta})/\partial \eta_i \partial \eta_j = O_p(1)$ ,  $i, j = 1, \dots, q$ , を仮定する。

仮定 (A3) のもとで

$$g_1(\hat{\boldsymbol{\eta}}) = g_1(\boldsymbol{\eta}) + \left(\frac{\partial g_1(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^t (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \frac{1}{2} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial^2 g_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^t} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$$

に注意すると,  $E[g_1(\hat{\boldsymbol{\eta}})] = g_1(\boldsymbol{\eta}) + g_{11}(\boldsymbol{\eta}) + g_{12}(\boldsymbol{\eta}) + o(N^{-1})$  が成り立つ。ここで

$$g_{11}(\boldsymbol{\eta}) = E\left[\left(\frac{\partial g_1(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^t (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right] \quad (3.6)$$

$$g_{12}(\boldsymbol{\eta}) = \frac{1}{2} E\left[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \frac{\partial^2 g_1(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^t} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right] \quad (3.7)$$

である。

**定理 3.2** (A1), (A2), (A3) を仮定する。このとき  $MSE(\boldsymbol{\eta}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}}))$  の 2 次不偏推定量は

$$mse(\mathbf{y}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = g_1(\hat{\boldsymbol{\eta}}) - g_{11}(\hat{\boldsymbol{\eta}}) - g_{12}(\hat{\boldsymbol{\eta}}) + g_2(\hat{\boldsymbol{\eta}}) \quad (3.8)$$

で与えられる。即ち,  $E[mse(\mathbf{y}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}})) + o(N^{-1})$  が成り立つ。

$q$  が大きい複雑なモデルにおいては, モーメント  $E[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})]$ ,  $E[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}]$  や  $\hat{\boldsymbol{\xi}}(\mathbf{y}, \boldsymbol{\eta})$  及び  $g_1(\boldsymbol{\eta})$  の微分の計算が厄介になる。そこで定理 3.2 の代わりに次のようなパラメトリック・ブートストラップ標本を用いると便利である。

$$\begin{aligned} \mathbf{y}^* | (\boldsymbol{\theta}^*, \hat{\boldsymbol{\eta}}) &\sim f(\mathbf{y}^* | \boldsymbol{\theta}^*, \hat{\boldsymbol{\eta}}), \\ \boldsymbol{\theta}^* | \hat{\boldsymbol{\eta}} &\sim \pi(\boldsymbol{\theta}^* | \hat{\boldsymbol{\eta}}) \end{aligned} \quad (3.9)$$

$\hat{\boldsymbol{\eta}}^*$  は  $\mathbf{y}^*$  に基づいた推定量で  $\hat{\boldsymbol{\eta}}$  と同じようにして与えられるものとする。 $E[g_1(\hat{\boldsymbol{\eta}})] = g_1(\boldsymbol{\eta}) + g_{11}(\boldsymbol{\eta}) + g_{12}(\boldsymbol{\eta}) + o(N^{-1})$  であるから,  $\mathbf{y}^*$  に関する期待値を  $E^*[\cdot]$  で表すと,

$$E^*[g_1(\hat{\boldsymbol{\eta}}^*)] = g_1(\hat{\boldsymbol{\eta}}) + g_{11}(\hat{\boldsymbol{\eta}}) + g_{12}(\hat{\boldsymbol{\eta}}) + o_p(N^{-1})$$

が成り立つ。従って,  $E[E^*[g_1(\hat{\boldsymbol{\eta}}^*)]] = E[2g_1(\hat{\boldsymbol{\eta}})] - g_1(\boldsymbol{\eta}) + o(N^{-1})$  もしくは

$$g_1(\boldsymbol{\eta}) = E[2g_1(\hat{\boldsymbol{\eta}}) - E^*[g_1(\hat{\boldsymbol{\eta}}^*)]] + o(N^{-1})$$

となる。 $g_2(\boldsymbol{\eta}) = O(N^{-1})$  より,

$$g_2^*(\hat{\boldsymbol{\eta}}) = E^*[\{\hat{\boldsymbol{\xi}}(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \hat{\boldsymbol{\xi}}(\mathbf{y}^*, \hat{\boldsymbol{\eta}})\}^2]$$

とおくと,  $E[g_2^*(\hat{\boldsymbol{\eta}})] = E[g_2(\hat{\boldsymbol{\eta}}) + o_p(N^{-1})] = g_2(\boldsymbol{\eta}) + o(N^{-1})$  となるので次の定理が成り立つ。

**定理 3.3** (A1), (A2), (A3) を仮定すると,

$$mse^*(\mathbf{y}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = 2g_1(\hat{\boldsymbol{\eta}}) - E^*[g_1(\hat{\boldsymbol{\eta}}^*)] + g_2^*(\hat{\boldsymbol{\eta}}) \quad (3.10)$$

は,  $MSE(\boldsymbol{\eta}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}}))$  の 2 次不偏推定量, 即ち,  $E[mse^*(\mathbf{y}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \hat{\boldsymbol{\xi}}(\mathbf{y}, \hat{\boldsymbol{\eta}})) + o(N^{-1})$  が成り立つ。

## 4 ベンチマーク問題への拡張

### 4.1 ベンチマーク問題と制限付き経験ベイズ推定量 (CEB)

一般的な混合モデル (3.1) について、全体が  $K$  個の小地域に分割されており、 $i = 1, \dots, K$  に対して  $\xi_i = \xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$  が推定したい小地域のスカラー関数とする。各  $\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$  に対する最良な推定量 (ベイズ推定量) は  $\hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) = E[\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})|\mathbf{y}]$  である。 $w_i$  を非負の定数で  $\sum_{i=1}^K w_i = 1$  を満たすとする。ここで次を仮定する。

(A4)  $K/N \rightarrow \gamma$ ,  $0 < \gamma \leq 1$ ,  $\sum_{i=1}^K w_i^2 = O(N^{-1})$ ,  $\sum_{i=1}^K w_i \{\hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) - E[\hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})]\} = O_p(N^{-1/2})$ ,  $Var(\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})|\mathbf{y}) = O_p(1)$

典型的な応用例では、 $N = K = p$  であり、各小地域の標本サイズ  $n_i$  に対して  $w_i = n_i / \sum_{j=1}^K n_j$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^t$  に対して  $\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta}) = \xi_i(\theta_i, \boldsymbol{\eta})$ ,  $\mathbf{y} = (y_1, \dots, y_K)^t$  に対して  $\hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) = \hat{\xi}_i(y_i, \boldsymbol{\eta})$  であり、 $y_1, \dots, y_K$  は互いに独立に分布している。この場合、仮定 (A4) は満たされる。

この節では、 $\xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})$  の推定量  $\delta_i(\mathbf{y})$  と推定量  $\hat{\boldsymbol{\eta}}$  について次を仮定する。

$$(C1) \sum_{i=1}^K w_i \delta_i(\mathbf{y}) = \sum_{i=1}^K w_i \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) + \Delta_m(\mathbf{y}), \Delta_m(\mathbf{y}) = O_p(N^{-1/2})$$

$$(C2) \sum_{i=1}^K w_i \{\delta_i(\mathbf{y}) - \bar{\delta}_w(\mathbf{y})\}^2 = \sum_{i=1}^K w_i \{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\}^2 + \Delta_v(\mathbf{y}), \Delta_v(\mathbf{y}) = O_p(1)$$

ここで、 $\bar{\delta}_w(\mathbf{y}) = \sum_{i=1}^K w_i \delta_i(\mathbf{y})$  であり、(C1), (C2) はそれぞれ平均制約, 分散制約と呼ぶことにする。(C1) については、例えば  $\Delta_m(\mathbf{y}) = \sum_{i=1}^K w_i \{\hat{\xi}_i(\mathbf{y}, \mathbf{0}) - \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}})\}$  は、 $E[\hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})] - E[\hat{\xi}_i(\mathbf{y}, \mathbf{0})] = O(N^{-1/2})$  と (A4) を仮定すると、 $\Delta_m(\mathbf{y}) = O_p(N^{-1/2})$  を満たす。(C2) については

$$\Delta_v(\mathbf{y}) = \sum_{i=1}^K w_i Var(\xi_i(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}})|\mathbf{y}) \quad (4.1)$$

とおくと、 $\Delta_v(\mathbf{y}) = O_p(1)$  を満たす。このとき 制約は、

$$\sum_{i=1}^K w_i E[\{\delta_i(\mathbf{y}) - \bar{\delta}_w(\mathbf{y})\}^2|\mathbf{y}] = \sum_{i=1}^K w_i E[\{\xi_i(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}})\}^2|\mathbf{y}] \quad (4.2)$$

と表され、Ghosh (1992), Datta *et al.* (2011) により扱われた制約に対応する。

制約 (C1) と (C2) のもとでベンチマークを行うために、Ghosh (1992), Frey and Cressie (2003), Datta *et al.* (2011) は、制約条件 (C1), (C2) のもとで条件付き MSE  $E[\{\delta_i(\mathbf{y}) - \xi_i(\boldsymbol{\theta}, \boldsymbol{\eta})\}^2|\mathbf{y}]$  を最小化することを考えた。これを制約付きベイズ推定量 (CB) という。 $\hat{\boldsymbol{\eta}}$  を CB に代入して得られる推定量を、制約付き経験ベイズ推定量 (CEB) といい、

$$\delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}}) = \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) + \{a_B(\mathbf{y}) - 1\} \{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\} + \Delta_m(\mathbf{y}) \quad (4.3)$$

で与えられることがわかる。ここで

$$\{a_B(\mathbf{y})\}^2 = 1 + \frac{\Delta_v(\mathbf{y})}{\sum_{i=1}^K w_i \{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\}^2} \quad (4.4)$$

である。

## 4.2 CEB の MSE

さて, CEB の MSE を計算しよう。MSE は,

$$\begin{aligned}
MSE(\boldsymbol{\eta}, \delta_i^{CEB}) &= E[\{\widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) - \xi_i\}^2] \\
&\quad + E[\{(a_B(\mathbf{y}) - 1)(\widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) - \widetilde{\xi}_w(\mathbf{y}, \widehat{\boldsymbol{\eta}})) + \Delta_m(\mathbf{y})\}^2] \\
&\quad + 2E[\{\widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) - \xi_i\}\{(a_B(\mathbf{y}) - 1)(\widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) - \widetilde{\xi}_w(\mathbf{y}, \widehat{\boldsymbol{\eta}})) + \Delta_m(\mathbf{y})\}] \\
&= I_1(\boldsymbol{\eta}) + I_2(\boldsymbol{\eta}) + 2I_3(\boldsymbol{\eta})
\end{aligned} \tag{4.5}$$

と分解できる。\$I\_1(\boldsymbol{\eta}) = MSE(\boldsymbol{\eta}, \widehat{\xi}\_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}))\$ であるから, 定理 3.1 より, \$I\_1(\boldsymbol{\eta}) = g\_1(\boldsymbol{\eta}) + g\_2(\boldsymbol{\eta}) + o(N^{-1})\$ と近似できる。\$I\_2, I\_3\$ については, 平均制約と平均-分散制約の場合に分けて扱う。

[1] **平均制約.** 平均制約 (C1) のもとでは (4.3) において \$a\_B(\mathbf{y}) = 1\$ とおくことにより, CEB は

$$\delta_i^{CEB}(\mathbf{y}, \widehat{\boldsymbol{\eta}}) = \widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) + \Delta_m(\mathbf{y}) \tag{4.6}$$

と表され, \$I\_2(\boldsymbol{\eta}) = E[\{\Delta\_m(\mathbf{y})\}^2]\$, \$I\_3(\boldsymbol{\eta}) = E[\{\widehat{\xi}\_i(\mathbf{y}, \widehat{\boldsymbol{\eta}}) - \xi\_i(\mathbf{y}, \boldsymbol{\eta})\}\Delta\_m(\mathbf{y})]\$ となる。\$\Delta\_m(\mathbf{y}) = O\_p(N^{-1/2})\$ より, \$I\_2(\boldsymbol{\eta}) = O(N^{-1})\$ であり,

$$\bar{I}_3(\boldsymbol{\eta}) = E\left[\left(\frac{\partial \widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^t (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \Delta_m(\mathbf{y})\right] = O(N^{-1})$$

とおくと, \$I\_3(\boldsymbol{\eta}) = \bar{I}\_3(\boldsymbol{\eta}) + O(N^{-3/2})\$ となる。

**命題 4.1** 仮定 (A1)-(A4) のもとで, 平均制約 (C1) の CEB は (4.6) で与えられ, その MSE は次のように近似できる。

$$\begin{aligned}
MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\mathbf{y}, \widehat{\boldsymbol{\eta}})) &= MSE(\boldsymbol{\eta}, \widehat{\xi}_i(\mathbf{y}, \widehat{\boldsymbol{\eta}})) + E[\{\Delta_m(\mathbf{y})\}^2] \\
&\quad + 2E\left[\left(\frac{\partial \widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^t (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \Delta_m(\mathbf{y})\right] + o(N^{-1})
\end{aligned}$$

これより, CEB \$\delta\_i^{CEB}(\mathbf{y}, \widehat{\boldsymbol{\eta}})\$ と制約されていない EB \$\widehat{\xi}\_i(\mathbf{y}, \widehat{\boldsymbol{\eta}})\$ の MSE の違いは 2 次近似の項に現れることがわかる。

[2] **平均-分散制約.** この場合, CEB は (4.3) で与えられ, その MSE を近似するために次を仮定する。

$$(A5) \sum_{i=1}^K w_i \{(\widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) - \widetilde{\xi}_w(\mathbf{y}, \boldsymbol{\eta}))^2 - E[(\widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) - \widetilde{\xi}_w(\mathbf{y}, \boldsymbol{\eta}))^2]\} = O_p(N^{-1/2})$$

このとき \$I\_3(\boldsymbol{\eta}) = O(N^{-1/2})\$ であり

$$I_3(\boldsymbol{\eta}) = E\left[\left(\frac{\partial \widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^t (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) h(\mathbf{y}, \boldsymbol{\eta})\right] + O(N^{-1}) \tag{4.7}$$

と近似できる。また \$I\_2(\boldsymbol{\eta}) = O(1)\$ であり,

$$I_2(\boldsymbol{\eta}) = E\left[\{A_B(\mathbf{y}, \boldsymbol{\eta}) - 1\}^2 \left\{\widehat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) - \sum_{j=1}^K w_j E[\widehat{\xi}_j(\mathbf{y}, \boldsymbol{\eta})]\right\}^2\right] + O(N^{-1/2})$$

と近似できるので, 次の命題を得る。

**命題 4.2** (A1)-(A5) を仮定する。平均-分散制約 (C1), (C2) のもとで CEB は (4.3) で与えられ, MSE の近似値は次で与えられる。

$$\begin{aligned} & \lim_{N \rightarrow \infty} \{MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})) - MSE(\boldsymbol{\eta}, \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}))\} \\ &= \lim_{N \rightarrow \infty} E \left[ \{A_B(\mathbf{y}, \boldsymbol{\eta}) - 1\}^2 \left\{ \hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta}) - \sum_{j=1}^K w_j E[\hat{\xi}_j(\mathbf{y}, \boldsymbol{\eta})] \right\}^2 \right] \end{aligned}$$

この命題より, 平均-分散制約の CEB  $\delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  と非制約の EB  $\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}})$  の MSE の違いは 1 次近似において現れることになる。即ち, CEB と EB の MSE の違いは小さくないことを意味する。従って, CEB の MSE を見積もることが重要になってくる。

### 4.3 CEB の MSE の推定

最後に, 平均-分散制約 (C1), (C2) のもとで CEB の MSE に対する 2 次不偏推定量をパラメトリック・ブートストラップ法で与えよう。次を仮定する。

**(A6)**  $\Delta_v(\mathbf{y}) - E[\Delta_v(\mathbf{y})] = O_p(N^{-r-1/2})$ ,  $E[\{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})\} \hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})] = O(N^{-1})$ ,  $E[\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \hat{\xi}_i(\mathbf{y}, \boldsymbol{\eta})] = O(N^{-1})$

(4.5) における  $I_3(\boldsymbol{\eta})$  に対して, パラメトリック・ブートストラップによる 2 次不偏推定量は

$$I_3^*(\mathbf{y}) = E^*[\{\hat{\xi}_i(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \hat{\xi}_i(\mathbf{y}^*, \boldsymbol{\eta}^*)\} \{(a_B(\mathbf{y}^*) - 1)(\hat{\xi}_i(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \tilde{\xi}_w(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*)) + \Delta_m(\mathbf{y}^*)\}]$$

で与えられる。(4.5) における  $I_2(\boldsymbol{\eta})$  の正確な不偏推定量は

$$\hat{I}_2(\mathbf{y}) = \{(a_B(\mathbf{y}) - 1)(\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \tilde{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})) + \Delta_m(\mathbf{y})\}^2$$

となるので, 次の定理を得る。

**定理 4.1** (A1)-(A6) を仮定する。平均-分散制約 (C1), (C2) のもとで CEB の  $MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}}))$  に対する 2 次不偏推定量は, (3.10) の  $mse^*(\mathbf{y}, \hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}}))$  に対して

$$mse(\mathbf{y}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = mse^*(\mathbf{y}, \hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})) + \hat{I}_2(\mathbf{y}) + 2I_3^*(\mathbf{y}) \quad (4.8)$$

で与えられる。即ち,  $E[mse(\mathbf{y}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}}))] = MSE(\boldsymbol{\eta}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})) + o(N^{-1})$  が成り立つ。

## 5 2 次分散関数をもつ指数型分布族 (NEF-QVF) への応用

### 5.1 NEF-QVF における EB とその MSE

さて, これまでの節で得られた結果を 2 次分散関数をもつ指数型分布族 (NEF-QVF) へ適用しよう。この分布族は, 正規分布, ポアソン分布, 2 項分布などの基本的な分布を含んでおり, Ghosh and Maiti (2004) で得られた結果を用いて CEB の MSE の 2 次不偏推定量を導こう。

$y_1, \dots, y_K$  を互いに独立な確率変数で,  $\theta_i$  を与えたときの  $y_i$  の条件付き分布と  $\theta_i$  の周辺分布が次のような指数型分布族に属しているとする。

$$\begin{aligned} y_i|\theta_i &\sim f(y_i|\theta_i) = \exp[n_i(\theta_i y_i - \psi(\theta_i)) + c(y_i, n_i)] \\ \theta_i|\nu, m_i &\sim \pi(\theta_i|\nu, m_i) = \exp[\nu(m_i \theta_i - \psi(\theta_i))]C(\nu, m_i) \end{aligned} \quad (5.1)$$

ここで,  $n_i$  は既知のスカラー,  $\nu$  は未知のスカラーであり,  $\mathbf{y} = (y_1, \dots, y_K)^t$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^t$  とおく。  $f(y_i|\theta_i)$  は正則な 1 母数指数型分布族であり,  $\pi(\theta_i|\nu, m_i)$  は共役事前分布である。

$$\mu_i = E[y_i|\theta_i] = \psi'(\theta_i)$$

とおき,  $\psi''(\theta_i) = Q_i(\mu_i)$  と定義すると,

$$\text{Var}(y_i|\theta_i) = \frac{\psi''(\theta_i)}{n_i} = \frac{Q_i(\mu_i)}{n_i}$$

となる。ここで,  $Q_i(\cdot)$  は 2 次関数として  $Q_i(x) = v_{0,i} + v_{1,i}x + v_{2,i}x^2$  と表されるとし,  $v_{0,i}, v_{1,i}, v_{2,i}$  は既知で同時には 0 にならないとする。これを 2 次分散関数をもつ指数型分布族 (NEF-QVF) といい, Morris (1982, 83) により導入され議論された。2 項分布  $\text{Bin}(n_i, \mu_i/n_i)$ ,  $\mu_i = n_i p_i$ , は  $v_{0,i} = 0, v_{1,i} = n_i, v_{2,i} = -1$  に対応し, ポアソン分布  $\text{Po}(\mu_i)$ ,  $\mu_i = n_i \lambda_i$ , は  $v_{0,i} = v_{2,i} = 0, v_{1,i} = n_i$  に対応する。また既知の  $\sigma^2$  に対して正規分布  $\mathcal{N}(\mu_i, \sigma^2/n_i)$  は  $v_{0,i} = \sigma^2, v_{1,i} = v_{2,i} = 0$  に対応する。事前分布の平均と分散は

$$E[\mu_i|m_i, \nu] = m_i, \quad \text{Var}(\mu_i|m_i, \nu) = \frac{Q_i(m_i)}{\nu - v_{2,i}} \quad (5.2)$$

となる。この節では,  $m_i$  と  $\mathbf{x}_i^t \boldsymbol{\beta}$  の間に標準的な連結関数

$$m_i = \psi'(\mathbf{x}_i^t \boldsymbol{\beta}), \quad i = 1, \dots, K$$

を仮定する。前節の  $\boldsymbol{\eta}$  は  $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$  に対応する。

$(y_i, \theta_i)$  の同時分布は  $f(y_i|\theta_i)\pi(\theta_i|\nu, m_i) = \pi(\theta_i|y_i, \nu)f_\pi(y_i|\nu, m_i)$  と書き直される。ここで  $\pi(\theta_i|y_i, \nu)$  は  $y_i$  を与えたときの  $\theta_i$  の条件付き (事後) 分布,  $f_\pi(y_i|\nu, m_i)$  は  $y_i$  の周辺分布を表し,

$$\hat{\mu}_i = \hat{\mu}_i(y_i, \boldsymbol{\eta}) = E[\mu_i|y_i, \boldsymbol{\eta}] = \frac{n_i y_i + \nu m_i}{n_i + \nu}$$

に対して次で与えられる。

$$\begin{aligned} \pi(\theta_i|y_i, \nu, m_i) &= \exp[(n_i + \nu)(\hat{\mu}_i \theta_i - \psi(\theta_i))]C(n_i + \nu, \hat{\mu}_i) \\ f_\pi(y_i|\nu, m_i) &= \frac{C(\nu, m_i)}{C(n_i + \nu, \hat{\mu}_i)} \exp[c(y_i, n_i)] \end{aligned} \quad (5.3)$$

2 項分布やポアソン分布においては推定したい母数はそれぞれ  $p_i = \mu_i/n_i, \lambda_i = \mu_i/n_i$  であるので, 一般に既知の正の定数  $q_i$  に対して

$$\xi_i = q_i \mu_i \quad (5.4)$$

を推定する問題を考える。 $\nu, m_i$  が既知のときには  $\xi_i$  のベイズ推定量は

$$\hat{\xi}_i(y_i, \boldsymbol{\eta}) = q_i \hat{\mu}_i(y_i, \boldsymbol{\eta}) = q_i \frac{n_i y_i + \nu m_i}{n_i + \nu} \quad (5.5)$$

となる。Ghosh and Maiti (2004) で示されたように  $\phi_i = (1 + \nu/n_i)/(\nu - v_{2,i})$  に対して

$$\begin{aligned} E[y_i] &= E[\psi'(\theta_i)] = m_i \\ \text{Var}(y_i) &= \text{Var}(E[y_i|\theta_i]) + E[\text{Var}(y_i|\theta_i)] = \text{Var}(\mu_i) + E[Q_i(\mu_i)/n_i] = Q_i(m_i)\phi_i \\ \text{Cov}(y_i, \mu_i) &= E[\text{Cov}(y_i, \mu_i)|\theta_i] + \text{Cov}(E[y_i|\theta_i], \mu_i) = Q_i(m_i)/(\nu - v_{2,i}) \end{aligned}$$

である。これらを用いて Ghosh and Maiti (2004) はベイズ推定量  $\hat{\xi}_i$  が  $\xi_i$  の最良線形不偏推定量 (BLUP) であることを示した。

Godambe and Thompson (1989) の推定方程式に従って、Ghosh and Maiti (2004) は  $\beta, \nu$  の推定量を次の推定方程式の解として求めることを提案した。 $g_{1i} = y_i - m_i$ ,  $g_{2i} = (y_i - m_i)^2 - \phi_i Q_i(m_i)$  に対して  $\mathbf{g}_i = (g_{1i}, g_{2i})^t$  とおく。

$$\begin{aligned} \mathbf{D}_i^t &= Q_i(m_i) \begin{pmatrix} \mathbf{x}_i & Q_i'(m_i)\phi_i\mathbf{x}_i \\ \mathbf{0} & -(1 + v_{2,i}/n_i)(\nu - v_{2,i})^{-2} \end{pmatrix} \\ \Sigma_i &= \text{Cov}(\mathbf{g}_i) = \begin{pmatrix} \mu_{2i} & \mu_{3i} \\ \mu_{3i} & \mu_{4i} - \mu_{2i}^2 \end{pmatrix} \end{aligned}$$

とし、 $|\Sigma_i| = \mu_{4i}\mu_{2i} - \mu_{2i}^3 - \mu_{3i}^2$  とする。ここで  $\mu_{ri} = E[(y_i - m_i)^r]$ ,  $r = 1, 2, 3, 4$ , の正確な表現式は後ほど与えられる。このとき最適な推定方程式は  $\sum_{i=1}^K \mathbf{D}_i^t \Sigma_i^{-1} \mathbf{g}_i = \mathbf{0}$  で与えられ、

$$\begin{aligned} \sum_{i=1}^K \frac{1}{|\Sigma_i|} \left[ \{\mu_{4i} - \mu_{2i}^2 - \mu_{3i}\phi_i Q_i'(m_i)\} g_{1i} + \{\mu_{2i}\phi_i Q_i'(m_i) - \mu_{3i}\} g_{2i} \right] Q_i(m_i) \mathbf{x}_i &= \mathbf{0} \\ \sum_{i=1}^K \frac{1}{|\Sigma_i|} \{\mu_{2i} g_{2i} - \mu_{3i} g_{1i}\} Q_i(m_i) (1 + v_{2,i}/n_i) (\nu - v_{2,i})^{-2} &= 0 \end{aligned} \quad (5.6)$$

と表される。推定方程式の解を  $\hat{\boldsymbol{\eta}}^t = (\hat{\beta}^t, \hat{\nu})$  とし、 $\hat{m}_i = \psi'(\mathbf{x}_i^t \hat{\beta})$  とおいて (5.5) に代入すると、経験ベイズ推定量 (EB)

$$\hat{\xi}_i(y_i, \hat{\boldsymbol{\eta}}) = q_i \frac{n_i y_i + \hat{\nu} \hat{m}_i}{n_i + \hat{\nu}} \quad (5.7)$$

が得られる。Ghosh and Maiti (2004) で示されたように、 $\mathbf{s}_k(\boldsymbol{\eta}) = \sum_{i=1}^K \mathbf{D}_i^t \Sigma_i^{-1} \mathbf{g}_i$ ,  $\mathbf{U}_K = \text{Cov}(\mathbf{s}_K(\boldsymbol{\eta})) = \sum_{i=1}^K \mathbf{D}_i^t \Sigma_i^{-1} \mathbf{D}_i = O(K)$  に対して

$$\begin{aligned} \hat{\boldsymbol{\eta}} - \boldsymbol{\eta} &= \mathbf{U}_K^{-1} \mathbf{s}_k(\boldsymbol{\eta}) + o_p(K^{-1/2}) \\ E[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t] &= \mathbf{U}_K^{-1} + o(K^{-1}), \quad E[\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}] = O(K^{-1}) \end{aligned} \quad (5.8)$$

となるので、仮定 (A1) が満たされる。明らかに仮定 (A2) が成り立つ。 $\text{Var}(\mu_i|y_i) = Q_i(\mu_i)/(n_i + \nu - v_{2,i})$  より、

$$g_{1i}(\boldsymbol{\eta}) = \frac{E[Q_i(\mu_i)]}{n_i + \nu - v_{2,i}} = q_i^2 \frac{\nu}{(n_i + \nu)(\nu - v_{2,i})} Q_i(m_i)$$

となるので、仮定 (A3) が満たされる。従って、定理 3.1, 3.2, 3.3 が成り立つ。

最後に、Ghosh and Maiti (2004) に従って  $\mu_{ri}$  の正確な表現が次の命題で与えられる。

**命題 5.1** モーメント  $\mu_{ri} = E[(y_i - m_i)^r]$ ,  $r = 2, 3, 4$ , は,  $d_i = v_{2,i}/n_i$  に対して

$$\begin{aligned}\mu_{2i} &= \frac{Q_i(m_i)(\nu/n_i + 1)}{\nu - v_{2,i}}, \quad \mu_{3i} = \frac{Q_i(m_i)Q'_i(m_i)(\nu/n_i + 1)(\nu/n_i + 2)}{(\nu - v_{2,i})(\nu - 2v_{2,i})}, \\ \mu_{4i} &= (d_i + 1)(2d_i + 1)(3d_i + 1)E[(\mu_i - m_i)^4] + \frac{6}{n_i}Q'_i(m_i)(d_i + 1)(2d_i + 1)E[(\mu_i - m_i)^3] \\ &\quad + \frac{d_i + 1}{n_i^2} [7\{Q'_i(m_i)\}^2 + 2n_i(4d_i + 3)Q_i(m_i)]E[(\mu_i - m_i)^2] \\ &\quad + \frac{1}{n_i^3}Q_i(m_i)[n_i(2d_i + 3)Q_i(m_i) + \{Q'_i(m_i)\}^2]\end{aligned}$$

と書かれる。また *Morris (1983)*, *Ghosh and Maiti (2004)* に従って,  $E[(\mu_i - m_i)^2] = Q_i(m_i)/(\nu - v_{2,i})$ ,  $E[(\mu_i - m_i)^3] = 2Q_i(m_i)Q'_i(m_i)/\{(\nu - v_{2,i})(\nu - 2v_{2,i})\}$ ,

$$E[(\mu_i - m_i)^4] = \frac{3Q_i(m_i)[(\nu - 2v_{2,i})Q_i(m_i) + 2\{Q'_i(m_i)\}^2]}{(\nu - v_{2,i})(\nu - 2v_{2,i})(\nu - 3v_{2,i})}$$

と書ける。

## 5.2 CEB の MSE とその推定

さて, 4 節の結果を NEF-QVF に適用しよう。ここでは制約条件 (C1), (C2) については,  $\xi_i = q_i\mu_i$  の推定量  $\delta_i(\mathbf{y})$  について次のような典型的な制約を仮定する。

$$(C1') \sum_{i=1}^K w_i \delta_i(\mathbf{y}) = \sum_{i=1}^K w_i q_i y_i$$

$$(C2') \sum_{i=1}^K w_i \{\delta_i(\mathbf{y}) - \bar{\delta}_w(\mathbf{y})\}^2 = \sum_{i=1}^K w_i \{\hat{\xi}_i(y_i, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\}^2 + \Delta_v^\dagger(\mathbf{y}), \text{ ただし}$$

$$\Delta_v^\dagger(\mathbf{y}) = \sum_{i=1}^K w_i(1 - w_i)q_i^2 \frac{Q_i(\hat{\mu}_i(y_i, \hat{\boldsymbol{\eta}}))}{n_i + \hat{\nu} - v_{2,i}} \quad (5.9)$$

であり,  $\bar{\delta}_w(\mathbf{y}) = \sum_{i=1}^K w_i \delta_i(\mathbf{y})$ ,  $\bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}}) = \sum_{i=1}^K w_i \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) = \sum_{i=1}^K w_i q_i \hat{\mu}_i(\mathbf{y}, \hat{\boldsymbol{\eta}})$  である。

(C1) の  $\Delta_m(\mathbf{y})$  は

$$\Delta_m^\dagger(\mathbf{y}) = \sum_{i=1}^K \frac{w_i q_i}{n_i + \hat{\nu}} \hat{\nu}(y_i - \hat{m}_i) \quad (5.10)$$

と表され,  $O_p(K^{-1/2})$  であることが確かめられる。また  $\Delta_v^\dagger(\mathbf{y})$  は Fay-Herriot モデルの場合には (2.3) の右辺の第 2 項の推定量に対応する量である。実際,

$$\begin{aligned}\sum_{i=1}^K w_i \text{Var}(\xi_i - \bar{\xi}_w | \mathbf{y}) &= \sum_{i=1}^K w_i \left\{ q_i^2 (1 - 2w_i) \text{Var}(\mu_i | y_i) + \sum_{j=1}^K w_j^2 q_j^2 \text{Var}(\mu_j | y_j) \right\} \\ &= \sum_{i=1}^K (w_i q_i^2 - 2w_i^2 q_i^2 + w_i^2 q_i^2) \text{Var}(\mu_i | y_i) \\ &= \sum_{i=1}^K w_i (1 - w_i) q_i^2 Q_i(\mu) / (n_i + \nu - v_{2,i})\end{aligned}$$

であり, これは  $O_p(1)$  である。制約 (C1'), (C2') にもとで CEB は, (4.3), (4.4) より

$$\delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}}) = \hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) + \{a_B(\mathbf{y}) - 1\} \{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\} + \Delta_m^\dagger(\mathbf{y}) \quad (5.11)$$

と表される。ここで,  $a_B(\mathbf{y})$  は

$$\{a_B(\mathbf{y})\}^2 = 1 + \frac{\Delta_v^\dagger(\mathbf{y})}{\sum_{i=1}^K w_i \{\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})\}^2} \quad (5.12)$$

であり,  $\Delta_m^\dagger(\mathbf{y}), \Delta_v^\dagger(\mathbf{y})$  は (5.10), (5.9) で与えられている。

**定理 5.1**  $\max_i \{n_i\}, \max_i \{m_i\}, \max_i \{q_i\}$  が有界で,  $\sum_{i=1}^K w_i = 1, \sum_{i=1}^K w_i^2 = O(K^{-1})$  を仮定すると, 前節で設定された仮定 (A4), (A5), (A6) が満たされる。

この定理から, 推定量  $\delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  に関して命題 4.1, 4.2, 定理 4.1 のすべての結果が成り立つ。特に, CEB の MSE の推定は, 次のパラメトリック・ブートストラップ標本に基づいて与えられる。

$$\begin{aligned} y_i^* | \theta_i^* &\sim f(y_i^* | \theta_i^*) = \exp[n_i(\theta_i^* y_i^* - \psi(\theta_i^*)) + c(y_i^*, n_i)] \\ \theta_i^* | \hat{\nu}, \hat{m}_i &\sim \pi(\theta_i^* | \hat{\nu}, \hat{m}_i) = \exp[\hat{\nu}(\hat{m}_i \theta_i^* - \psi(\theta_i^*))] C(\hat{\nu}, \hat{m}_i) \end{aligned} \quad (5.13)$$

このとき, CEB  $\delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})$  の MSE の 2 次不偏推定量は

$$mse(\mathbf{y}, \delta_i^{CEB}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = mse^*(\mathbf{y}, \hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})) + \hat{I}_2(\mathbf{y}) + 2I_3^*(\mathbf{y}) \quad (5.14)$$

となる。ここで  $mse^*(\mathbf{y}, \hat{\xi}(\mathbf{y}, \hat{\boldsymbol{\eta}})) = 2g_1(\hat{\boldsymbol{\eta}}) - E^*[g_1(\hat{\boldsymbol{\eta}}^*)] + g_2^*(\hat{\boldsymbol{\eta}})$ ,  $g_2^*(\hat{\boldsymbol{\eta}}) = E^*[\{\hat{\xi}(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \hat{\xi}(\mathbf{y}^*, \hat{\boldsymbol{\eta}})\}^2]$ ,  $\hat{I}_2(\mathbf{y}) = \{(a_B(\mathbf{y}) - 1)(\hat{\xi}_i(\mathbf{y}, \hat{\boldsymbol{\eta}}) - \bar{\xi}_w(\mathbf{y}, \hat{\boldsymbol{\eta}})) + \Delta_m(\mathbf{y})\}^2$ ,

$$I_3^*(\mathbf{y}) = E^*[\{\hat{\xi}_i(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \hat{\xi}_i(\mathbf{y}^*, \hat{\boldsymbol{\eta}})\} \{(a_B(\mathbf{y}^*) - 1)(\hat{\xi}_i(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*) - \bar{\xi}_w(\mathbf{y}^*, \hat{\boldsymbol{\eta}}^*)) + \Delta_m(\mathbf{y}^*)\}]$$

であり,  $g_1(\hat{\boldsymbol{\eta}}) = q_i^2 \hat{\nu} [(n_i + \hat{\nu})(\hat{\nu} - v_{2,i})]^{-1} Q_i(\hat{m}_i)$ ,  $g_1(\hat{\boldsymbol{\eta}}^*) = q_i^2 \hat{\nu}^* [(n_i + \hat{\nu}^*)(\hat{\nu}^* - v_{2,i})]^{-1} Q_i(\hat{m}_i^*)$  である。

## 6 正規線形混合モデルにおける CEB とその誤差評価

### 6.1 Fay-Herriot モデルにおける CEB

2 節で紹介された Fay-Herriot モデル

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + v_i + \varepsilon_i, \quad i = 1, \dots, K \quad (6.1)$$

を考えてみる。ただし,  $v_i \sim \mathcal{N}(0, \lambda)$ ,  $\varepsilon_i \sim \mathcal{N}(0, d_i)$  に従う。これは (2.1), (2.2) のように分解できるので, 指数型分布族の枠組みで表現すると,  $m_i = \mathbf{x}_i' \boldsymbol{\beta}$  に対して

$$\begin{aligned} y_i | \mu_i &\sim (2\pi)^{-1/2} \exp \left[ d_i^{-1} \{ \mu_i y_i - \mu_i^2 / 2 \} - y_i^2 / (2d_i) \right] \\ \mu_i | \lambda, m_i &\sim \exp \left[ \lambda^{-1} \{ m_i \mu_i - \mu_i^2 / 2 \} - m_i^2 / (2\lambda) \right] \end{aligned} \quad (6.2)$$

となる。従って、 $\theta_i = \mu_i$ ,  $\psi(\mu_i) = \mu_i^2/2$ ,  $\nu = 1/\lambda$ ,  $n_i = 1/d_i$  とおくと、このモデルは (5.1) の枠組みに入り、 $Q_i(\cdot)$  は  $v_{0,i} = 1$ ,  $v_{1,i} = v_{2,i} = 0$  に対応する。

2 節で与えられているように、経験ベイズ推定量は

$$\hat{\mu}_i^{EB} = \mathbf{x}_i^t \hat{\boldsymbol{\beta}} + \frac{\hat{\lambda}}{\hat{\lambda} + d_i} (y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})$$

となる。平均制約，分散制約をそれぞれ，

$$(MC) \sum_{i=1}^K w_i \delta_i = \sum_{i=1}^K w_i \hat{\mu}_i^{EB} + \Delta_m^\dagger(\mathbf{y})$$

$$(VC) \sum_{i=1}^K w_i \{\delta_i - \bar{\delta}_w\}^2 = \sum_{i=1}^K w_i \{\hat{\mu}_i^{EB} - \bar{\mu}_w^{EB}\}^2 + \Delta_v^\dagger(\mathbf{y})$$

とする。ここで、 $\bar{\delta}_w = \sum_{j=1}^K w_j \delta_j$ ,  $\bar{\mu}_w^{EB} = \sum_{j=1}^K w_j \hat{\mu}_j^{EB}$  である。また

$$\Delta_m^\dagger(\mathbf{y}) = \sum_{i=1}^K \frac{w_i d_i}{d_i + \hat{\lambda}} (y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})$$

であり、 $\Delta_m(\mathbf{y}) = O_p(K^{-1/2})$  となる。 $\Delta_v^\dagger(\mathbf{y})$  については、ここではより一般的に

$$\Delta_v^\dagger(\mathbf{y}) = K^{-r} \sum_{i=1}^K w_i (1 - w_i) \frac{d_i \hat{\lambda}}{\hat{\lambda} + d_i} \quad (6.3)$$

を考えることができる。 $r$  については、 $r = 0, 0.5, 1$  を考えるが、Kubokawa (2012) で論じられているように、 $r = 0$  の場合が、推定量の加重分散が事後分布の加重分散の推定量に等しいことを意味しており、制約付き経験ベイズ推定量の MSE は 1 次のオーダーにおいて経験ベイズ推定量の MSE と異なってくる。これに対して、 $r = 0.5$ ,  $r = 1$  の場合には、両者の違いは MSE において 2 次のオーダーの項に現れる。

制約条件 (MC), (VC) のもとで制約付き経験ベイズ推定量 (CEB) は

$$\hat{\mu}_i^{CEB} = \hat{\mu}_i^{EB} + \{a_B(\mathbf{y}) - 1\} \left\{ \hat{\mu}_i^{EB} - \sum_{j=1}^K w_j \hat{\mu}_j^{EB} \right\} + \Delta_m^\dagger(\mathbf{y})$$

と表される。ここで、 $a_B(\mathbf{y})$  は

$$\{a_B(\mathbf{y})\}^2 = 1 + \frac{\Delta_v^\dagger(\mathbf{y})}{\sum_{i=1}^K w_i \left\{ \hat{\mu}_i^{EB} - \sum_{j=1}^K w_j \hat{\mu}_j^{EB} \right\}^2}$$

で与えられる。CEB の MSE の推定量は、(5.14) より得られる。

## 6.2 応用例

6.1 節の結果を家計調査のデータに当てはめてみよう。2011 年 11 月の家計調査のうち、‘教育’ についての支出について、47 都道府県の県庁所在都市における平均支出 (1,000 円単位) を  $y_i$ ,  $i = 1, \dots, 47$ , とする。分散  $d_i$  については、11 月の‘教育’支出についての過去 10 年間のデータから計算したものをを用いることにする。家計調査は毎月報告されているが、多くの都道府県につ

表 1: 経験ベイズ推定量 (EB) と 制約付き経験ベイズ推定量 (CEB) の値とそれらの MSE 推定値 (括弧内の値が MSE 推定値)

都県	$d_i$	$y_i$	$\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$	EB	CEB <sub>0</sub>	CEB <sub>0.5</sub>	CEB <sub>1</sub>
茨城	12.49	8.10	9.45	8.77 (5.34)	8.72 (5.90)	8.76 (5.45)	8.77 (5.38)
栃木	62.56	10.03	9.48	9.57 (7.05)	9.71 (7.76)	9.59 (7.18)	9.58 (7.09)
群馬	5.38	5.21	9.99	6.63 (3.48)	6.08 (4.20)	6.54 (3.57)	6.61 (3.51)
埼玉	9.01	12.33	14.30	13.14 (5.31)	14.12 (6.81)	13.30 (5.43)	13.17 (5.34)
千葉	91.77	30.71	12.17	14.43 (7.73)	15.71 (9.96)	14.63 (7.87)	14.46 (7.75)
東京	3.65	15.45	13.16	14.94 (2.79)	16.34 (5.01)	15.17 (2.90)	14.98 (2.81)
神奈川	27.48	23.25	12.54	15.93 (7.21)	17.56 (10.62)	16.19 (7.42)	15.97 (7.25)

いて標本サイズが 100 前後と小さく、しかも ‘教育’ 支出のデータのバラツキは非常に高くなっている。一方、全国消費実態調査のデータも 47 都道府県について利用可能である。これは、家計調査よりかなり大きい標本調査なため、平均支出の値はより信頼性が高いと考えられるが、調査自体は 5 年毎にしか実施されない。そこで、ここでは 2009 年の全国消費者実態調査の ‘教育’ 支出のデータを、 $X_i$  for  $i = 1, \dots, 47$ , としして用いることにする。 $\mathbf{x}_i^t = (1, X_i)$ ,  $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ ,  $K = 47$  としして、Fay-Herriot モデル

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta} + v_i + \varepsilon_i, \quad i = 1, \dots, K$$

を当てはめることにする。

母数推定を行うと、 $\hat{\lambda} = 12.752$ ,  $\hat{\beta}_1 = 2.209$ ,  $\beta_2 = 0.580$  となる。経験ベイズ推定値 EB と、 $r = 0, 0.5, 1$  に対応して 3 種類の制約付き経験ベイズ推定値 CEB<sub>0</sub>, CEB<sub>0.5</sub>, CEB<sub>1</sub> とそれらの MSE の推定値を計算することができる。東京周辺の都県についての値が表 1 に与えられている。この表からわかるように、 $d_i$  が大きいほど EB は  $y_i$  を  $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$  の方向へより大きく縮小している。CEB<sub>0</sub> は CEB<sub>0.5</sub>, CEB<sub>1</sub> よりバラツキが大きく、東京については CEB<sub>0</sub> の値が  $y_i$  と  $\mathbf{x}_i^t \hat{\boldsymbol{\beta}}$  の範囲を超えてしまっているが、EB と CEB<sub>0.5</sub>, CEB<sub>1</sub> の値はその範囲に入っている。また、東京の CEB<sub>0</sub> の MSE 推定値は EB の MSE 推定値の約 1.8 倍になっているので、この値には注意する必要があるようである。CEB<sub>0</sub> に比べて CEB<sub>0.5</sub>, CEB<sub>1</sub> の MSE 推定値は EB の MSE 推定値に近い値を与えている。

## 7 ポアソン-ガンマ混合モデルと死亡リスクの推定への応用

### 7.1 ポアソン-ガンマ混合モデル

混合モデル (5.1) の例として空間疫学における相対リスクの推定に用いられるポアソン-ガンマ混合モデルを扱う。いま全体が  $K$  個の地域から構成されているとする。 $i = 1, \dots, K$  に対して  $y_i$  は地域  $i$  における特定の疾患による死亡数を表し、 $n_i$  は年齢と性により調整された予想される期待死亡数を表す。 $y_1, \dots, y_K$  は互いに独立な確率変数で、 $y_i$  は平均  $\lambda_i n_i$  のポアソン分布  $\mathcal{P}o(\lambda_i n_i)$  に従い、 $\lambda_i$  は地域  $i$  における死亡リスクに対応する未知母数とする。 $\lambda_i$  の不偏推定量は

$$\hat{\lambda}_i^{SMR} = y_i/n_i \quad (7.1)$$

であり、標準化死亡比 (SMR) と呼ばれる。

$\hat{\lambda}_i^{SMR}$  の分散は  $Var(\hat{\lambda}_i^{SMR}) = \lambda_i/n_i$  であるから、 $n_i$  が小さいときには推定誤差が問題になることがわかる。そこで、ポアソン-ガンマ混合モデルを想定する。

$$\begin{aligned} y_i | \lambda_i &\sim \mathcal{P}o(n_i \lambda_i) \\ \lambda_i &\sim \mathcal{G}a(m_i \nu / n_i, 1/\nu) \end{aligned} \quad (7.2)$$

ここで  $\mathcal{G}a(\alpha, \beta)$  は平均  $\alpha\beta$ 、分散  $\alpha\beta^2$  のガンマ分布を表す。この分布を指数型分布族で表すと

$$\begin{aligned} y_i | \lambda_i &\sim \exp[n_i(y_i n_i^{-1} \log \lambda_i - \lambda_i) + (y_i \log n_i - \log y_i!)], \\ \lambda_i | \nu, m_i &\sim \exp\{[(m_i/n_i)\nu - 1] \log \lambda_i - \nu \lambda_i + (m_i/n_i)\nu \log \nu\} \{\Gamma(m_i \nu / n_i)\}^{-1} d\lambda_i \end{aligned} \quad (7.3)$$

となる。 $\theta_i = n_i^{-1} \log \lambda_i$  とおくと  $d\lambda_i = n_i \exp[n_i \theta_i] d\theta_i$  より、 $\psi(\theta_i) = \exp[n_i \theta_i] = \lambda_i$  に対して

$$\begin{aligned} y_i | \lambda_i &\sim \exp[n_i(y_i \theta_i - \psi(\theta_i)) + (y_i \log n_i - \log y_i!)] \\ \theta_i | \nu, m_i &\sim \exp[\nu(m_i \theta_i - \psi(\theta_i))] n_i^{-1} (\nu/n_i)^{m_i \nu / n_i} \{\Gamma(m_i \nu / n_i)\}^{-1} d\theta_i \end{aligned} \quad (7.4)$$

と書けるので、(5.1) の枠組みに入り、 $Q_i(m_i)$  については  $v_{0,i} = v_{2,i} = 0$ 、 $v_{1,i} = n_i$  となる。また  $\mu_i = \psi'(\theta_i) = n_i \lambda_i$  であり  $q_i = 1/n_i$  とおく。 $\mu_i$ 、 $\lambda_i$  のベイズ推定量は

$$\hat{\mu}_i^B(y_i, m_i, \nu) = \frac{n_i y_i + \nu m_i}{n_i + \nu}, \quad \hat{\lambda}_i^B(y_i, m_i, \nu) = \frac{y_i + \nu m_i / n_i}{n_i + \nu} \quad (7.5)$$

と書ける。

さて死亡リスクの推定に対してベンチマーク問題を考えよう。 $L = \sum_{i=1}^K n_i$ 、 $w_i = n_i / \sum_{j=1}^K n_j$ 、 $\mathbf{y} = (y_1, \dots, y_K)^t$  とおく。全地域での平均的死亡リスクは  $\bar{y}_w = L^{-1} \sum_{i=1}^K y_i$  であり、期待値は  $E[\bar{y}_w] = L^{-1} \sum_{i=1}^K E[y_i] = L^{-1} \sum_{i=1}^K \lambda_i n_i$  となる。そこで、(C1') に対応する平均制約として

$$(MC) \sum_{i=1}^K w_i \hat{\lambda}_i^C = L^{-1} \sum_{i=1}^K y_i$$

を考える。SMR  $\hat{\lambda}_i^{SMR} = y_i/n_i$  はこの制約を満たす。しかし、ベイズ推定量  $\hat{\lambda}_i^B(y_i, m_i, \nu)$  は満たしていない。(5.11)、(5.6) から  $\lambda_i$  の CEB は

$$\hat{\lambda}_i^{CEBm} = \hat{\lambda}_i^{EB} + \sum_{j=1}^K \frac{w_j}{n_j} \frac{\hat{\nu}(y_j - \hat{m}_j)}{n_j + \hat{\nu}} \quad (7.6)$$

となる。ここで  $\hat{\lambda}_i^{EB} = (y_i + \hat{\nu}\hat{m}_i/n_i)/(n_i + \hat{\nu})$  である。

分散制約 (C2') は,  $\bar{\lambda}_w^C = \sum_{j=1}^K w_j \hat{\lambda}_j^C$ ,  $\bar{\lambda}_w^{EB} = \sum_{j=1}^K w_j \hat{\lambda}_j^{EB}$  に対して次のように表される。

$$(VC) \sum_{i=1}^K w_i (\hat{\lambda}_i^C - \bar{\lambda}_w^C)^2 = \sum_{i=1}^K w_i (\hat{\lambda}_i^{EB} - \bar{\lambda}_w^{EB})^2 + \sum_{i=1}^K w_i (1 - w_i) \frac{\hat{\lambda}_i^{EB}}{n_i + \hat{\nu}}$$

(5.11) より, CEB は

$$\hat{\lambda}_i^{CEBv} = \hat{\lambda}_i^{EB} + \{a_B(\mathbf{y}) - 1\} \{\hat{\lambda}_i^{EB} - \bar{\lambda}_w^{EB}\} \quad (7.7)$$

と表される。ただし

$$\{a_B(\mathbf{y})\}^2 = 1 + \frac{\sum_{j=1}^K w_j (1 - w_j) \hat{\lambda}_j^{EB} / (n_j + \hat{\nu})}{\sum_{j=1}^K w_j (\hat{\lambda}_j^{EB} - \bar{\lambda}_w^{EB})^2}$$

である。平均-分散制約 (MC), (VC) が課されるときには, (5.11) より CEB は次で与えられる。

$$\hat{\lambda}_i^{CEBmv} = \hat{\lambda}_i^{EB} + \{a_B(\mathbf{y}) - 1\} \{\hat{\lambda}_i^{EB} - \bar{\lambda}_w^{EB}\} + \sum_{j=1}^K \frac{w_j \nu (y_j - \hat{m}_j)}{n_j (n_j + \hat{\nu})} \quad (7.8)$$

$m_i = \psi'(\mathbf{x}_i^t \boldsymbol{\beta})$ ,  $\psi(\theta_i) = \exp[n_i \theta_i] = \lambda_i$  であるから,  $m_i = n_i \exp[n_i \mathbf{x}_i \boldsymbol{\beta}]$  となる。未知母数  $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$  は推定方程式 (5.6) の解により推定される。このモデルでは,  $Q_i(m_i) = n_i m_i$ ,  $\phi_i = 1/\nu + 1/n_i$ ,  $v_{2,i} = 0$ ,  $g_{1i} = y_i - m_i$ ,  $g_{2i} = (y_i - m_i)^2 - m_i(1 + \tau_i)$ ,  $\tau_i = n_i/\nu$  であり,  $\mu_{\tau_i} = E[(y_i - m_i)^r]$ ,  $i = 2, 3, 4$ , は命題 5.1 より

$$\begin{aligned} \mu_{2i} &= m_i(1 + \tau_i), & \mu_{3i} &= m_i(1 + 3\tau_i + 2\tau_i^2) \\ \mu_{4i} &= m_i \{1 + 3m_i + 3(6m_i + 7)\tau_i + 3(m_i + 4)\tau_i^2 + 6\tau_i^3\} \end{aligned}$$

と書ける。故に,  $(\boldsymbol{\beta}, \nu)$  の推定量を求める推定方程式は

$$\begin{aligned} \sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|} \left[ \{\mu_{4i} - \mu_{2i}^2 - \mu_{3i}(1 + \tau_i)\} g_{1i} + \{\mu_{2i}(1 + \tau_i) - \mu_{3i}\} g_{2i} \right] n_i m_i \mathbf{x}_i &= \mathbf{0}, \\ \sum_{i=1}^K \frac{1}{|\boldsymbol{\Sigma}_i|} \{\mu_{2i} g_{2i} - \mu_{3i} g_{1i}\} n_i m_i &= 0 \end{aligned}$$

と書かれる。CEB  $\hat{\lambda}_i^{CEBm}$ ,  $\hat{\lambda}_i^{CEBv}$ ,  $\hat{\lambda}_i^{CEBmv}$  の MSE の 2 次不偏推定量は (5.14) により得られ,  $mse^*(\mathbf{y}, \hat{\lambda}_i)$  の中の  $g_1(\boldsymbol{\eta})$  は  $g_1(\boldsymbol{\eta}) = m_i / \{n_i(n_i + \nu)\}$  で与えられる。

## 7.2 死亡リスクの推定におけるベンチマークの例

7.1 節の結果を実際の死亡数データに応用し CEB がどのように作用するのか調べてみよう。

埼玉県の各市町村  $i$  に居住する女性の胃がんによる死亡数について 1995 年から 1999 年までの 5 年間の累計を  $y_i$ , 全体から期待される死亡数を  $n_i$  とし,  $y_i, n_i$  のデータが全市町村  $K = 92$  に對して得られていて,  $L = 3,953$  である。ポアソン-ガンマモデル (7.2) を適用して, 市町村別の

胃がんによる死亡リスク  $\lambda_i$  の推定を行う。回帰項  $\mathbf{x}_i^t \boldsymbol{\beta}$  については  $\mathbf{x}_i$  は 1 次元で  $\mathbf{x}_i = 1$  とし  $m_i = n_i \exp[n_i \beta_0]$ ,  $\boldsymbol{\beta} = \beta_0$  とする。

数値計算については Ox を用い、非線形方程式の解を求めるためのパッケージ SolveNLE を利用して推定方程式 (5.6) の解を求めると、 $\beta_0 = 1.53249 \times 10^{-4}$ ,  $\nu = 174.472$  となる。 $L = \sum_{i=1}^K n_i = 3,953$ ,  $w_i = n_i/L$  に対して、 $\sum_{i=1}^K w_i \hat{\lambda}_i^{SMR} = 1$ ,  $\sum_{i=1}^K w_i \hat{\lambda}_i^{EB} = 1.0086$  となり、それらの差は小さいことがわかる。一方、SMR の加重分散の値が  $\sum_{i=1}^K w_i (\hat{\lambda}_i^{SMR} - 1)^2 = 0.0302$  であるのに対して経験ベイズ推定量 EB の加重分散は 0.0014 となり、かなり小さくなっている。7.1 節の (VC) における右辺は

$$\sum_{i=1}^K w_i (\hat{\lambda}_i^{EB} - \bar{\lambda}^{EB})^2 + \sum_{i=1}^K w_i (1 - w_i) \frac{\hat{\lambda}_i^{EB}}{n_i + \hat{\nu}} = 0.0051$$

となるので、EB は縮小し過ぎていることがわかる。そこで、平均-分散制約のもとでの制約付き経験ベイズ推定量 CEB の挙動を調べてみる。ちなみに、 $a_B(\mathbf{y}) = 1.9242$  であり、CEB の加重分散の値は 0.0051 に一致している。SMR, EB, CEB の値をプロットした図が図 1, 2, 3 である。これらの図から、EB が SMR を縮小し過ぎていて、CEB はその縮小の程度を緩和していることがわかる。

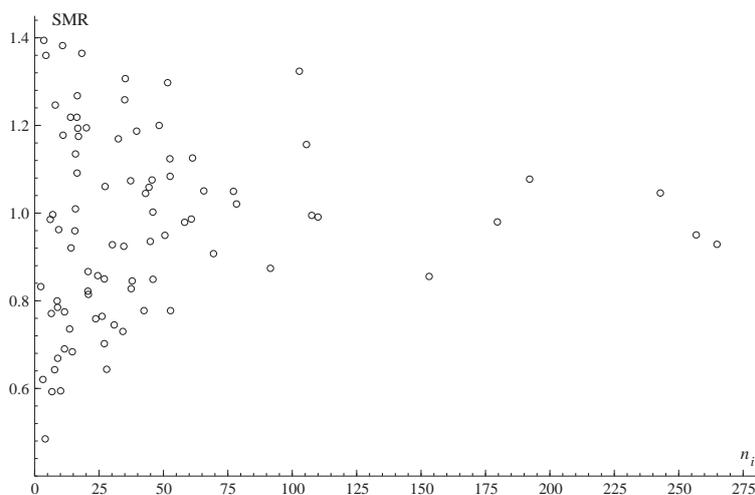


図 1: 埼玉県における胃がんによる女性死亡リスク SMR に関する市町村別プロット

埼玉県 92 市町村のうち 33 市町村を適当に選んで、SMR  $\hat{\lambda}_i^{SMR}$ , 経験ベイズ推定量 EB  $\hat{\lambda}_i^{EB}$  と平均-分散制約のもとでの 2 種類の制約付き経験ベイズ推定量  $CEB_0, CEB_{0.5}$  の値と、1,000 個のパラメトリック・ブートストラップ標本に基づいた MSE の推定値を与えたのが表 2 である。ここで、 $CEB_r$  は、(5.9) において  $\Delta_v^\dagger(\mathbf{y})$  を

$$\Delta_v^\dagger(\mathbf{y}) = \frac{1}{K^r} \sum_{i=1}^K w_i (1 - w_i) q_i^2 \frac{Q_i(\hat{\mu}_i(y_i, \hat{\boldsymbol{\eta}}))}{n_i + \hat{\nu} - v_{2,i}} \quad (7.9)$$

で置き換えたときの制約付き経験ベイズ推定量を表し、 $CEB_0, CEB_{0.5}$  は  $r = 0, r = 0.5$  に対応する。‘estimates’ と ‘estimated MSE’ の欄の数値は  $100 \times \hat{\lambda}_i, 100 \times mse(\hat{\lambda}_i)$  の値を報告

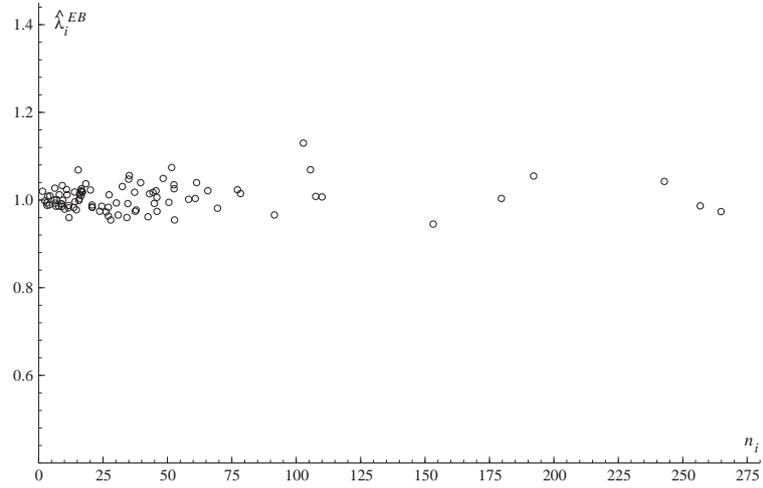


図 2: 埼玉県における胃がんによる女性死亡リスク EB に関する市町村別プロット

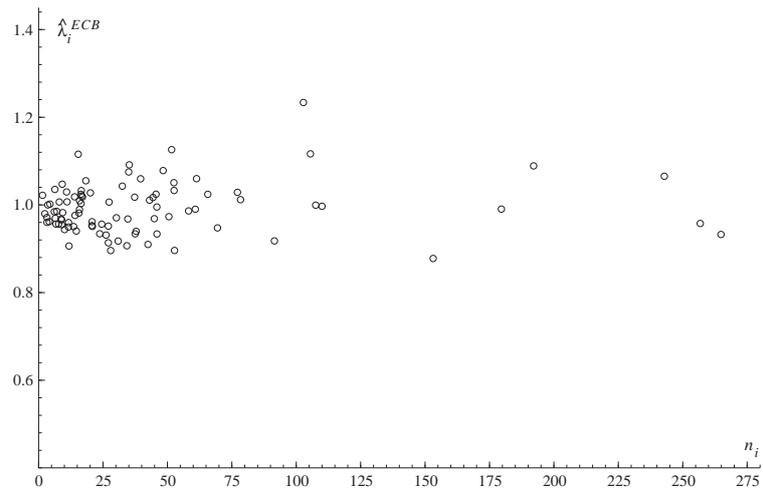


図 3: 埼玉県における胃がんによる女性死亡リスク CEB に関する市町村別プロット

表 2: 埼玉県 33 の市町村別 SMR, EB, CEB の値と MSE の推定値 ( $n_i$  は女性の胃がん死亡数についての期待数,  $CEB_0, CEB_{0.5}$  は  $r = 0, r = 0.5$  に対応した CEB の値, *estimates, estimated MSE* は  $100 \times \hat{\lambda}_i, 100 \times mse(\hat{\lambda}_i)$  の値を表している)

市町村	$n_i$	estimates				estimated MSE			
		SMR	EB	CEB <sub>0</sub>	CEB <sub>0.5</sub>	SMR	EB	CEB <sub>0</sub>	CEB <sub>0.5</sub>
Kawagoe	192.1	107	105	108	105	0.536	0.561	0.678	0.562
Kumagaya	102.7	132	112	123	113	0.989	0.732	1.804	0.738
Kawaguchi	242.8	104	104	106	103	0.427	0.497	0.549	0.499
Urawa	256.7	95	98	95	97	0.405	0.482	0.565	0.495
Oomiya	264.8	92	97	93	96	0.393	0.474	0.641	0.491
Gyouda	61.2	112	103	105	103	1.647	0.855	0.896	0.857
Tokorozawa	179.6	97	100	99	99	0.572	0.580	0.597	0.589
Honjyo	45.9	84	97	93	96	2.192	0.913	1.076	0.930
Higashimatsuyama	52.7	77	95	89	93	1.911	0.886	1.228	0.911
Kasukabe	105.5	115	106	111	106	0.963	0.725	0.949	0.725
Yono	48.3	120	104	107	104	2.084	0.903	0.987	0.904
Koshigaya	153.1	85	94	87	92	0.668	0.624	1.077	0.653
Hatogaya	35.2	130	105	109	105	2.856	0.958	1.082	0.958
Sakado	51.6	129	107	112	107	1.952	0.891	1.161	0.891
Satte	34.9	125	104	107	104	2.876	0.959	1.035	0.960
Yoshikawa	27.9	64	95	89	93	3.591	0.991	1.334	1.016
Ogose	10.8	138	102	102	101	9.229	1.080	1.083	1.084
Naguri	3.5	139	100	99	99	27.898	1.122	1.130	1.130
Namekawa	8.9	66	98	95	97	11.161	1.090	1.182	1.104
Ogawa	27.0	70	96	91	94	3.711	0.995	1.246	1.016
Kawashima	16.5	126	102	103	101	6.053	1.048	1.053	1.052
Yoshimi	15.3	182	106	111	106	6.517	1.055	1.275	1.055
Nagatoro	8.0	124	101	100	100	12.480	1.096	1.099	1.102
Okano	11.7	33	95	90	94	8.491	1.074	1.361	1.097
Ryoujin	3.2	30	98	96	97	30.852	1.125	1.201	1.137
Higashichichibu	4.4	135	100	100	100	22.678	1.117	1.123	1.124
Kamiizumi	1.5	328	101	102	101	65.719	1.135	1.136	1.140
Kamisato	18.3	136	103	105	103	5.473	1.039	1.071	1.041
Oosoto	6.3	174	102	103	102	15.864	1.106	1.113	1.110
Okabe	13.9	121	101	101	101	7.184	1.062	1.062	1.068
Kawamoto	9.1	163	103	104	102	10.913	1.089	1.109	1.092
Shiraoka	26.1	76	97	93	95	3.837	1.000	1.173	1.017
Shoubu	15.8	113	101	100	100	6.321	1.052	1.054	1.058

している。ただし  $mse(\hat{\lambda}_i)$  は  $\hat{\lambda}_i$  の MSE の推定値を表す。SMR,  $y_i/n_i$ , の MSE の推定量は,  $E[(y_i/n_i - \lambda_i)^2] = n_i^{-2}E[(y_i - n_i\lambda_i)^2] = n_i^{-1}E[\lambda_i] = m_i/n_i^2 = n_i^{-1} \exp\{n_i\beta_0\}$  に注意すると, パラメトリック・ブートストラップ法により次で与えられる。

$$mse(\hat{\lambda}_i^{SMR}) = 2n_i^{-1} \exp\{n_i\hat{\beta}_0\} - n_i^{-1}E^*[\exp\{n_i\hat{\beta}_0^*\}]$$

小さい  $n_i$  に対しては SMR は大きな  $mse$  をもつものに対して EB, CEB<sub>0</sub>, CEB<sub>0.5</sub> の  $mse$  は小さく押さえられている。EB は SMR を大きく縮小しており, CEB<sub>0</sub>, CEB<sub>0.5</sub> は縮小の程度を緩和している。一方, 大きい  $n_i$  に対しては, EB, CEB<sub>0</sub>, CEB<sub>0.5</sub> は SMR をあまり縮小していない。大抵の市町村については, SMR の  $mse$  が大きいものに対して, EB, CEB<sub>0</sub>, CEB<sub>0.5</sub> の  $mse$  は小さい。川越, 川口, 浦和, 大宮, 所沢は,  $n_i$  が 179 以上の市であり, EB, CEB<sub>0</sub>, CEB<sub>0.5</sub> の  $mse$  の値が SMR の  $mse$  より若干大きくなっている。特に,  $n_i = 102.7$  の熊谷については, CEB<sub>0</sub> の  $mse$  は大きくなってしまいが, EB<sub>0.5</sub> の  $mse$  の値は EB の  $mse$  の値に近い。  $n_i = 105.5$  の春日部については, CEB<sub>0</sub>  $mse$  の値は SMR の  $mse$  の値に近い。CEB<sub>0.5</sub> と EB を比較すると, それらの推定値と  $mse$  の値は似通っていることがわかる。Louis (1984), Ghosh (1992) により与えられた CEB の動機付けを考慮に入れると, CEB<sub>0</sub> を使うのが望ましいと思われるが,  $mse$  の値を通して推定誤差がどの程度に抑えられているのか調べておく必要がある。特に, 熊谷のように比較的大きな  $mse$  が生ずる場合には, CEB<sub>0</sub> の値の妥当性に注意することが大切である。

## 8 2項-ベータ混合モデルとロジスティック回帰

最後に, ロジスティック回帰をもつ 2項-ベータ混合モデルについて簡単に説明しよう。  $i = 1, \dots, K$  に対して  $y_i$  は  $n_i$  の試行のうち, ある特定の事象が起こる回数とする。  $y_1, \dots, y_K$  は互いに独立な確率変数で,  $n_i y_i$  が 2項分布  $Bin(n_i, p_i)$ ,  $p_i$  がベータ分布に従うとする。具体的には,

$$\begin{aligned} y_i | p_i &\sim Bin(n_i, p_i) \\ p_i &\sim beta(m_i n_i^{-1} \nu, (1 - m_i n_i^{-1}) \nu) \end{aligned} \quad (8.1)$$

で与えられる。指数型分布族の枠組みで表現すると,

$$\begin{aligned} y_i | p_i &\sim \exp \left[ n_i \left\{ y_i n_i^{-1} \log \left( \frac{p_i}{1 - p_i} \right) + \log(1 - p_i) \right\} \right] \binom{n_i}{y_i} \\ p_i | \nu, m_i &\sim \exp \left[ m_i n_i^{-1} \nu \log \left( \frac{p_i}{1 - p_i} \right) + \nu \log(1 - p_i) \right] \frac{p_i^{-1} (1 - p_i)^{-1} \Gamma(\nu)}{\Gamma(m_i n_i^{-1} \nu) \Gamma((1 - m_i n_i^{-1}) \nu)} dp_i \end{aligned} \quad (8.2)$$

となる。  $\theta_i = n_i^{-1} \log(p_i/(1 - p_i))$ ,  $p_i^{-1} (1 - p_i)^{-1} dp_i = n_i d\theta_i$  とし,  $\psi(\theta_i) = \log(1 + \exp[n_i \theta_i]) = -\log(1 - p_i)$  とする。このとき, 混合モデル (8.2) は

$$\begin{aligned} y_i | p_i &\sim \exp \left[ n_i (y_i \theta_i - \psi(\theta_i)) \right] \binom{n_i}{y_i} \\ \theta_i | \nu, m_i &\sim \exp \left[ \nu (m_i \theta_i - \psi(\theta_i)) \right] \frac{\Gamma(\nu) n_i}{\Gamma(m_i n_i^{-1} \nu) \Gamma((1 - m_i n_i^{-1}) \nu)} d\theta_i \end{aligned} \quad (8.3)$$

と書き直されるので, (5.1) の枠組みに入り,  $Q_i(\cdot)$  は  $v_{0,i} = 0, v_{1,i} = n_i, v_{2,i} = -1$  に対応する。 $\mu_i = \psi'(\theta_i) = n_i p_i$  だから,  $\mu_i, p_i$  のベイズ推定量は

$$\hat{\mu}_i(y_i, m_i, \nu) = \frac{n_i y_i + \nu m_i}{n_i + \nu}, \quad \hat{p}_i(y_i, m_i, \nu) = \frac{y_i + \nu m_i / n_i}{n_i + \nu}$$

となる。 $m_i = \psi'(\mathbf{x}_i^t \boldsymbol{\beta})$ ,  $\psi(\theta_i) = \log(1 + \exp[n_i \theta_i]) = -\log(1 - p_i)$  より  $m_i = n_i \exp[n_i \mathbf{x}_i^t \boldsymbol{\beta}] / (1 + \exp[n_i \mathbf{x}_i^t \boldsymbol{\beta}])$  である。未知母数  $\boldsymbol{\eta}^t = (\boldsymbol{\beta}^t, \nu)$  は推定方程式 (5.6) により推定される。このモデルでは,  $Q_i(m_i) = n_i m_i - m_i^2$ ,  $\phi_i = (1 + \nu/n_i) / (\nu + 1)$ ,  $v_{2,i} = -1$ ,  $g_{1i} = y_i - m_i$ ,  $g_{2i} = (y_i - m_i)^2 - m_i(n_i - m_i)\phi_i$  であり,  $\mu_{ri} = E[(y_i - m_i)^r]$ ,  $i = 2, 3, 4$ , は命題 5.1 より

$$\begin{aligned} \mu_{2i} &= \frac{\nu/n_i + 1}{\nu + 1} m_i(n_i - m_i), \quad \mu_{3i} = \frac{(\nu/n_i + 1)(\nu/n_i + 2)}{(\nu + 1)(\nu + 2)} m_i(n_i - m_i)(n_i - 2m_i), \\ \mu_{4i} &= m_i^2(n_i - m_i)^2 \left[ 3 \frac{(d_i + 1)(2d_i + 1)(3d_i + 1)}{(\nu + 1)(\nu + 2)(\nu + 3)} \left\{ \nu + 2 + 2 \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} \right\} \right. \\ &\quad + \frac{12}{n_i} \frac{(d_i + 1)(2d_i + 1)}{(\nu + 1)(\nu + 2)} \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} + \frac{1}{n_i^2} \frac{d_i + 1}{\nu + 1} \left\{ 7 \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} + 2n_i(4d_i + 3) \right\} \\ &\quad \left. + \frac{2d_i + 3}{n_i^2} + \frac{1}{n_i^3} \frac{(n_i - 2m_i)^2}{m_i(n_i - m_i)} \right] \end{aligned}$$

となる。従って,  $(\boldsymbol{\beta}, \nu)$  の推定値は推定方程式 (5.6) の解として求めることができる。

以上の議論を用いて, 7.1 節で議論されたベンチマーク問題を扱うことができる。詳細は省略されるが, 平均制約 (MC), 分散制約 (VC) のもとでの CEB が与えられ, その MSE の 2 次不偏推定量が与えられる。

## 参考文献

- [1] Datta, G.S., Ghosh, M., Steorts, R., and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, **20**, 574-588.
- [2] Frey, J., and Cressie, N. (2003). Some results on constrained Bayes estimators. *Statist. Probab. Letters*, **65**, 389-399.
- [3] Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. American Statist. Assoc.*, **87**, 533-540.
- [4] Ghosh, M. and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, **91**, 95-112.
- [5] Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with Discussion). *J. Statist. Plan. Infer.*, **22**, 137-152.
- [6] Kubokawa, T. (2011). On measuring uncertainty of small area estimators with higher order accuracy. *J. Japan Statist. Soc.*, **41**, 93-119.
- [7] Kubokawa, T. (2012) Mixed effects prediction under benchmarking and applications to small area estimation. Discussion Paper Series, CIRJE-F-832.

- [8] Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Assoc.*, **79**, 393-398.
- [9] Morris, C. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65-80.
- [10] Morris, C. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.*, **11**, 515-529.



Which best describes the current and forecasted (for September 2012) conditions, excluding seasonal factors, out of three alternatives which best describes the current (from three months earlier) and forecasted (for September 2012)

Short-Term Economic Survey of Enterprises in Japan (Tankan)  
Sample Form of the June 2012 Survey

1. Independent Survey

(1) Business Conditions of Your Enterprise

1. Favorable	Actual	Forecast
2. Not so favorable		
3. Unfavorable		
	result	

2. Annual Projection

100% or more	90% to 99%	80% to 89%	70% to 79%	60% to 69%	50% to 59%	40% to 49%	30% to 39%	20% to 29%	10% to 19%	0% to 9%	Below 0%
100% or more	90% to 99%	80% to 89%	70% to 79%	60% to 69%	50% to 59%	40% to 49%	30% to 39%	20% to 29%	10% to 19%	0% to 9%	Below 0%

2012年9月29日

### 景況判断の理由は？

#### Multiple Answer

- 国内需要の動向
- 海外需要の動向
- 販売価格の動向
- 仕入れ価格の動向
- 仕入れ以外のコストの動向
- 為替レートの動向
- その他

#### Forced Choice

当てはまる ▼      当てはまらない ▼

- 国内需要の動向
- 海外需要の動向
- 販売価格の動向
- 仕入れ価格の動向
- 仕入れ価格以外の動向
- 為替レート動向

6

### 設備投資計画の有無

#### Branching

あり       なし

- 生産（販売）能力の拡大
- 製（商）品・サービスの質的向上
- 情報化への対応
- 省力化合理化
- 環境対策
- 海外投資
- 研究開発
- 新事業への進出
- 維持更新
- その他

#### Combined

- 生産（販売）能力の拡大
- 製（商）品・サービスの質的向上
- 情報化への対応
- 省力化合理化
- 環境対策
- 海外投資
- 研究開発
- 新事業への進出
- 維持更新
- その他
- 設備投資計画はない

7

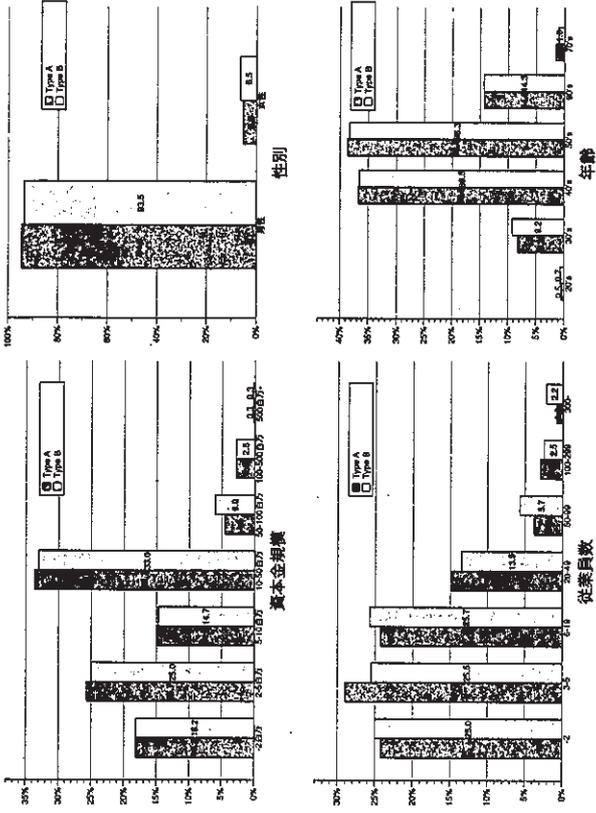
### 調査票設計にあたって

1. 季節的要因を除いた景況判断は可能か？
2. Multiple Answer 対 Forced Choice
3. Branching の影響

8

## Web実験調査の概要

- 調査票：2種類用意
- 調査対象：代表取締役（社長）
  - ネット調査会社の登録パネルから無作為に抽出
  - 各調査票に無作為に割当
- 調査期間：2012年2月17日 - 2月21日
- 標本サイズ：各調査票 600
  - 目標標本サイズが確保できるまでサイトにアクセス可



## 次の四半期の売上高は、直前と比べて？

季節変動要因を除いた実勢

- 大幅に増加する
- やや増加する
- 変わらない
- やや減少する
- 大幅に減少する

↑

↓

(季節変動要因を含めて)

- 大幅に増加する
- やや増加する
- 変わらない
- やや減少する
- 大幅に減少する

## 次の四半期の売上高は、直前と比べて？

季節変動要因を除いた実勢

- 大幅に増加する
- やや増加する
- 変わらない
- やや減少する
- 大幅に減少する

↑

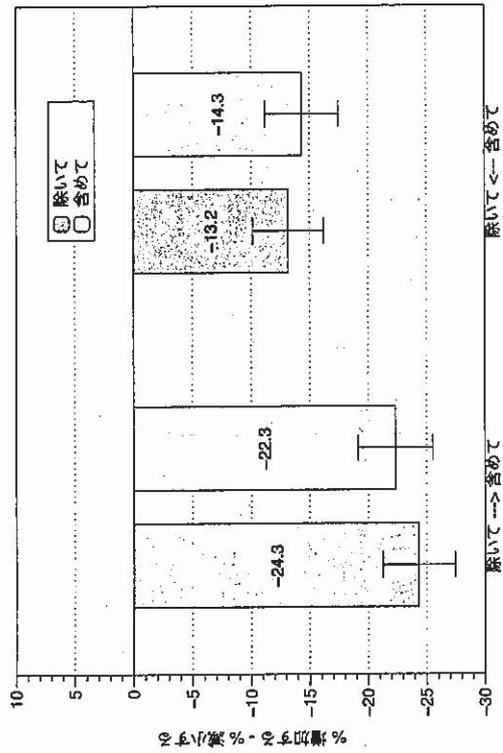
↓

(季節変動要因を含めて)

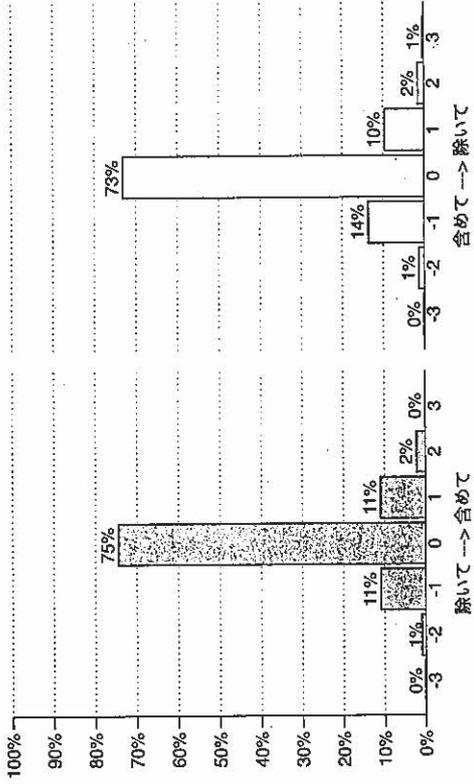
- ↑ 2.5%
- ↑ 17.0%
- ↑ 36.7%
- ↑ 31.8%
- ↑ 12.0%

19.5% - 43.8% = -24.3

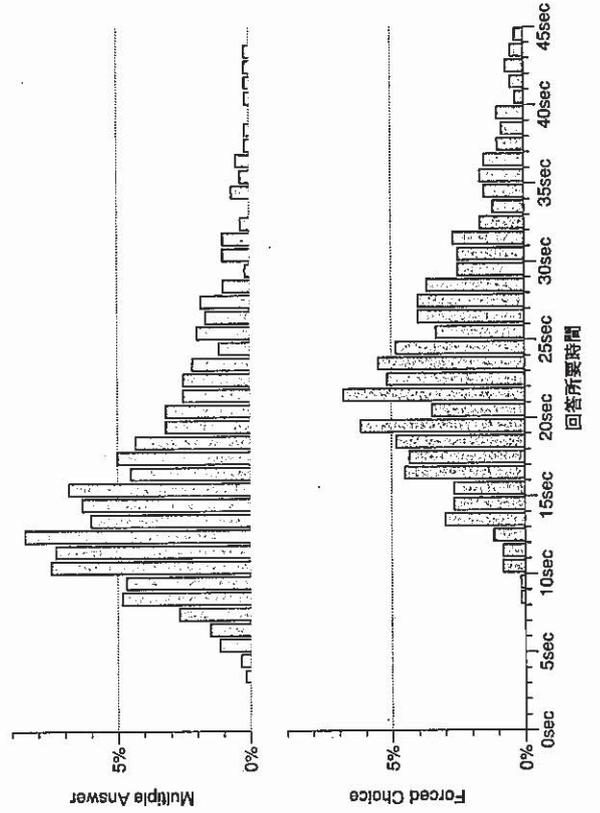
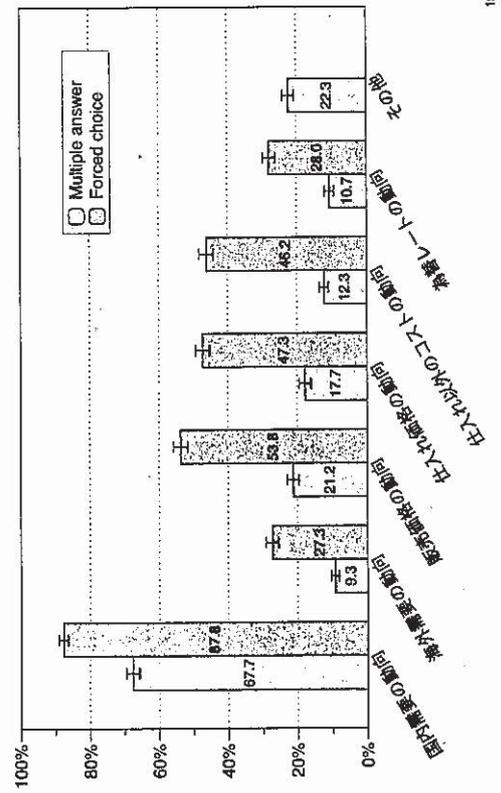
### 次の四半期の売上高



### 回答変化



### Multiple Answer 対 Forced Choice







# Lee-Carter 予測に関する時系列分析の理論的な長所と短所の解明と代替的予測法の提案

千木良 弘朗\*

山本 拓<sup>†</sup>

2012年9月13日

## 概要

死亡率の予測では、Lee and Carter (1992) が提案した予測法が広く使われている。しかし、その予測精度については、時系列分析の理論的な観点からの検証は厳密には行われていないようである。そこで、本稿では Lee-Carter の予測法の理論的な性質を明らかにし、それに基づいてどのような長所と短所があるのかを導く。そして、短所を補えるような予測法が構築できるか検討する。

## 1 はじめに

死亡率の将来の予測は、人口統計学の分野における大きな関心事である。死亡率の予測法は既に数多く提案されており、Keilman (2003) によるとそれらは demographic タイプと non-demographic タイプの2つに大別できる。前者は時系列予測等のデータに基づいた統計学的方法であり、後者は疫学、医学、ヘルスケア、環境といった知見に基づくものである。Keilman (2003) はどちらのタイプが優れているかは判断しかねるとしながらも、demographic タイプの方がよく使われていると述べている。

demographic タイプの1つに Lee and Carter (1992) が提案した方法がある。Lee-Carter の予測法はアメリカの死亡率を予測するために提案されたものだが、現在では他の国でも広く使われており、Deaton and Paxson (2004) によると死亡率予測の主流と呼べるものになっている。

しかし、Lee-Carter 予測の精度について、時系列分析の理論的な観点からの検証は厳密には行われていないようである。そこで、本稿では Lee-Carter 予測の理論的な性質を

---

\*東北大学大学院経済学研究科

<sup>†</sup>日本大学経済学部

明らかにする。そして、その性質から、予測の精度に関してどのような長所と短所があるのかを導く。そして、短所を補えるような予測法が構築できるか検討する。

本稿の構成は以下の通り。次の2節でDGPのモデルを導入する。3節でLee-Carter予測の理論的な性質を明らかにし、その長所と短所を整理する。続く4節では、それまでの議論に基づいてLee-Carter予測の改善策を検討する。5節でモンテカルロ実験を行い、Lee-Carter予測や改善策として提案した予測の精度を調べる。6節で実際のデータを使って死亡率を予測する。最後に、7節でまとめを行う。

## 2 Lee-Carter 予測の概観とモデルの設定

本節では、まずLee-Carter予測を概観する。そして、Lee-Carter予測が示す死亡率のデータの性質と実際のある死亡率のデータより、死亡率のモデルを設定する。

### 2.1 Lee-Carter 予測の概観

$t$ 年における年齢グループ  $a$  の死亡率を  $w_{at}$ ,  $a = 1, 2, \dots, m$ ,  $t = 1, 2, \dots, T$  とする。Lee-Carterの予測法では  $y_{at} = \log w_{at}$  と対数値に直して予測を行う。ここで、記法の簡単化のため、 $y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix}$  なる  $(m \times 1)$  のベクトルを導入する。Lee-Carter 予測を行うには、まず、平均  $\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$  からの偏差を  $\tilde{y}_t = y_t - \bar{y}$  として

$$\tilde{Y} = \begin{bmatrix} \tilde{y}'_1 \\ \vdots \\ \tilde{y}'_T \end{bmatrix}$$

なる行列を作る。そして、以下のような固有値分解を行う。

$$\tilde{Y}'\tilde{Y} = F\Lambda F', \quad F = \begin{bmatrix} f_1 & \dots & f_m \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_m \end{bmatrix} \quad (1)$$

ここで、 $\lambda_1 \geq \dots \geq \lambda_m$  が固有値であり、 $f_1, \dots, f_m$  がそれに対応する固有ベクトルである。後のために、 $\begin{bmatrix} f_2 & \dots & f_m \end{bmatrix} = F_{(m-1)}$  として  $F = \begin{bmatrix} f_1 & F_{(m-1)} \end{bmatrix}$  と書き直しておく。次

に、第1主成分  $f_1' \tilde{y}_t$  を計算し、 $f_1' \tilde{y}_t$  に対して単位根検定やモデルの特定化を行うこと無く<sup>1</sup>ARIMA(0,1,0)のドリフト付きランダムウォークに従うと仮定してドリフト項を

$$\frac{1}{T-1} \sum_{t=2}^{T-1} f_1' \Delta \tilde{y}_t$$

なる第1主成分の階差平均で推定する。ここで、後のために、これを

$$\frac{1}{T-1} \sum_{t=2}^{T-1} f_1' \Delta \tilde{y}_t = f_1' \frac{1}{T-1} \sum_{t=2}^{T-1} \Delta \tilde{y}_t = f_1' \frac{1}{T-1} \sum_{t=2}^{T-1} \Delta y_t = f_1' \overline{\Delta y} \quad (2)$$

と書き直しておく。そして、これより第1主成分の  $h$  期先予測を

$$f_1' \overline{\Delta y} h + f_1' \tilde{y}_T \quad (3)$$

として作成する。最後に、主成分を原データに戻すために主成分の予測量に  $f_1$  を掛け、引いておいた平均を足し戻して

$$\hat{y}_{T+h}^{LC} = f_1(f_1' \overline{\Delta y} h + f_1' \tilde{y}_T) + \bar{y} = f_1 f_1' \overline{\Delta y} h + f_1 f_1' \tilde{y}_T + \bar{y} = f_1 f_1' \overline{\Delta y} h + y_T^{LC} \quad (4)$$

とする。これが Lee-Carter 予測量である。

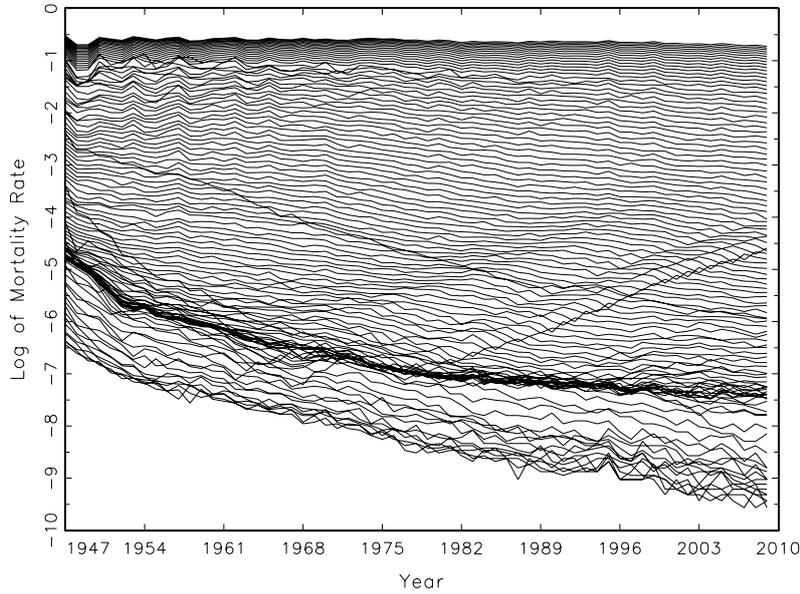
## 2.2 モデルの設定

Lee-Carter は  $y_t$  のモデルを仮定することなく  $y_t$  を予測しているので、理論上何を予測しているのかを調べるができない。そこで、ここでは  $y_t$  にモデルを仮定する。

Lee-Carter 予測は第1主成分をドリフト付きランダムウォークとしているが、このことは  $y_t$  自身も  $m$  変量のドリフト付きランダムウォークであることを示唆する。Lee-Carter の方法を多変量  $I(1)$  時系列過程  $y_t$  に主成分分析を適用したものであると考え、非定常時系列分析の文脈でいくつか参考になる先行研究が存在する。Stock and Watson (1988)、Harris (1997)、Snell (1999) 等がそれであり、Snell (1999) では共和分関係のある多変量  $I(1)$  時系列過程に主成分分析を適用すると、共和分ランクに応じて第1主成分を初めとするいくつかの主成分が  $I(1)$  になることが示されている。こうしたことから  $y_t$  のモデルとしては共和分過程が妥当であると考えられる。モデルを決めるには実際のデータを調

<sup>1</sup>Girosi and King (2007) の p. 5 に “in practice the random walk with drift model for  $\gamma_t$  has been used almost exclusively.” とあるのでこれに従った。

図 1: 日本人男性の死亡率の対数（各年齢、0 歳～109 歳）



べる必要もあるので、Human Mortality Database<sup>2</sup>よりダウンロードした1947年～2009年における日本人男性の0歳～109歳の $y_t$ を図1にプロットした。各 $y_{at}$ は右下がりの線形トレンドを持ち、概ね似たような動きをしている。この図からも共和分関係があることが示唆されるため、本稿では共和分過程をモデルとして使うことにする。

具体的には、 $y_t$ の階差 $\Delta y_t$ が次のようなモデルに従うと仮定する。

$$\Delta y_t = C(L)(\varepsilon_t + \Phi D_t), \quad \varepsilon_t \sim i.i.d.(0, \Omega), \quad t = 1, \dots, T \quad (5)$$

ここで、 $C(L) = \sum_{i=0}^{\infty} C_i L^i$ であって $\delta$ をある正数として $C(z)$ は $|z| < 1 + \delta$ に対しconvrgentであるとし、 $D_t$ は確定項を表し、 $\Phi$ はそのパラメーター行列であり、 $\Omega$ は対称で正値定符号な分散共分散行列であるとする。(5)は次のように書き直せることが知られている。

$$\Delta y_t = C(\varepsilon_t + \Phi D_t) + C^*(L)(1 - L)(\varepsilon_t + \Phi D_t) \quad (6)$$

ここで、 $C = C(1)$ 、 $C(L)^* = \sum_{i=0}^{\infty} C_i^* L^i$ であり、 $C_i^* = -\sum_{j=i+1}^{\infty} C_j$ 、 $i = 0, 1, \dots$ である。 $\delta$ をある正数とすると、 $C^*(z)$ は $|z| < 1 + \delta$ に対しconvrgentになる。初期値を

<sup>2</sup>詳細は参考文献を参照のこと。

$y_0 = C^*(L)(\varepsilon_0 + \Phi D_0) = 0$  とすると、(5) は次のようにも書き直せる。

$$y_t = C \sum_{s=1}^t (\varepsilon_s + \Phi D_s) + C^*(L)(\varepsilon_t + \Phi D_t) \quad (7)$$

このモデルは各  $y_{at}$  が  $I(1)$  過程に従うことを意味し、いくつかの正則条件の下で  $(m \times r)$  の共和分行列  $\beta$  が存在することと  $\beta' C = 0$  となることが同値となる。確定項  $\Phi D_t$  については、

$$\Phi D_t = d_1 + d_2 t \quad (8)$$

とモデル化する。ここで、 $d_1$  と  $d_2$  は各々  $(m \times 1)$  のパラメーターベクトルである。階差表現の (6) を使うと、

$$\begin{aligned} \Delta y_t &= C(\varepsilon_t + d_1 + d_2 t) + C^*(L)(1 - L)(\varepsilon_t + d_1 + d_2 t) \\ &= C\varepsilon_t + C^*(L)(1 - L)\varepsilon_t + Cd_1 + Cd_2 t + C^*(1)d_2 \end{aligned}$$

のようになるが、ここで level に 2 次トレンドをがあると図 1 からして適切ではないので

$$Cd_2 = 0 \quad (9)$$

を仮定する。すると、

$$\mu = Cd_1 + C^*(1)d_2 = Cd_1 + d_2^* \quad (10)$$

として

$$\Delta y_t = \mu + C\varepsilon_t + C^*(L)(1 - L)\varepsilon_t = \mu + C(L)\varepsilon_t \quad (11)$$

となる。一方、level で表現した (7) の形では、(9) に注意すると

$$\begin{aligned} y_t &= C \sum_{s=1}^t (\varepsilon_s + d_1 + d_2 t) + C^*(L)(\varepsilon_t + d_1 + d_2 t) \\ &= Cd_1 t + C^*(1)d_1 + C^*(L)d_2 t + C \sum_{s=1}^t \varepsilon_s + C^*(L)\varepsilon_t \\ &= Cd_1 t + C^*(1)d_1 + C^*(1)d_2 t - \left( \sum_{i=1}^{\infty} i C_i^* \right) d_2 + C \sum_{s=1}^t \varepsilon_s + C^*(L)\varepsilon_t \\ &= \gamma + \mu t + C \sum_{s=1}^t \varepsilon_s + C^*(L)\varepsilon_t \end{aligned} \quad (12)$$

となる。ここで、 $\gamma = C^*(1)d_1 - (\sum_{i=1}^{\infty} iC_i^*)d_2$  である。この (11) もしくは (12) が、本稿で考えるモデルである。なお、こうした確定項のモデリングのインプリケーションは、(12) の左から  $\beta'$  を掛けると

$$\beta' y_t = \beta' \gamma + \beta' d_2^* t + \beta' C^*(L) \varepsilon_t \quad (13)$$

となって、共和分関係が平均定常ではなくトレンド定常になるということである。こうしたモデリングは Johansen (1995) の p. 81 にある  $H^*(r)$  モデルに沿ったものであり、Johansen (1995) が考えていた VEC (vector error correction) モデルの代わりに VMA (vector moving average) モデルを用いたと言える。VEC ではなく VMA を用いていることと共和分関係に定数項だけではなくトレンド項も入れたことでかなり一般的なモデルとなっていると考えられる。

### 2.3 予測精度の尺度

本稿は予測の精度に焦点を当てているので、ここで予測精度の尺度を導入する。予測の精度を測るには様々な尺度があるが、本稿では標準的に用いられている trace MSE を採用する。trace MSE は、ある予測法 (例えば予測法 i) により  $(m \times 1)$  の予測量  $\hat{y}_{T+h}^i$  を得た時に、その予測誤差  $e_{T+h}^i = y_{T+h} - \hat{y}_{T+h}^i$  の 2 乗和の期待値

$$\text{tr MSE}(\hat{y}_{T+h}^i) = E(e_{T+h}^{i'} e_{T+h}^i) = \text{tr} (E(e_{T+h}^i e_{T+h}^{i'}))$$

として定義される。複数の予測法 (例えば予測法 i と予測法 ii) の精度を比較するには、trace MSE を使って trace MSE ratio と呼ばれる比を

$$\frac{\text{tr MSE}(\hat{y}_{T+h}^i)}{\text{tr MSE}(\hat{y}_{T+h}^{ii})}$$

のように計算して、これが 1 より大きいかどうかで判断することが多い。本稿でもこの trace MSE ratio で予測の精度を比較する。

trace MSE ratio を用いた精度比較では、Chigira and Yamamoto (2012) が次の結果を示していることに注意されたい。

$$\lim_{h \rightarrow \infty} \frac{\text{tr MSE}(\hat{y}_{T+h}^{MISS})}{\text{tr MSE}(\hat{y}_{T+h}^{OPT})} = 1 \quad (14)$$

ここで、 $\hat{y}_{T+h}^{OPT}$  はモデルの全ての係数パラメーターを既知とした時に MSE を最小にする最適予測量、 $\hat{y}_{T+h}^{MISS}$  は  $\nu_{T+h}$  を  $\text{tr}(E(\nu_{T+h}\nu_{T+h}')) = O(1)^3$  を満たす任意の確率ベクトルとして  $\hat{y}_{T+h}^{MISS} = (T+h)\mu + \nu_{T+h}$  と書ける予測量である。(14) 式は、ドリフト  $\mu$  さえ正しく考慮すれば、長期予測ではどのような予測量も最適予測量と同等の精度を持つことを示す。この具体的な含意については、この後で折に触れて述べる。

### 3 Lee-Carter 予測の理論的な性質

本節では、モデル (11) もしくは (12) に Lee-Carter の予測法を適用した時に理論上どのような予測量が作られるのかを明らかにする。そして、その結果に基づいて Lee-Carter 予測の長所と短所を整理する。

Lee-Carter 予測量の理論的な性質を解明するには、(4) 式より、 $f_1$  と  $f_1 f_1' \overline{\Delta y}$  の挙動を明らかにする必要があることが解る。 $f_1$  については、次のことが言える。

**命題 1:**

$$f_1 \xrightarrow{p} \pm \frac{\mu}{\sqrt{\mu' \mu}} \quad (15)$$

**証明:** 付録を参照のこと。

このように、 $f_1$  は基本的にドリフト項  $\mu$  を一致推定量となっている。 $\sqrt{\mu' \mu}$  で割られているのは  $f_1$  が固有ベクトルなので  $f_1 f_1' = 1$  となるよう基準化されているためである。また、固有ベクトルは符号は定まらないので“±”が付いている。続いて  $f_1 f_1' \overline{\Delta y}$  については次のようになる。

**命題 2:**

$$f_1 f_1' \overline{\Delta y} \xrightarrow{p} \mu \quad (16)$$

**証明:** 付録を参照のこと。

このように、 $f_1 f_1' \overline{\Delta y}$  はドリフト項  $\mu$  の一致推定量になっている。

これらの結果より、Lee-Carter 予測は漸近的には

$$\hat{y}_{T+h}^{LC} \approx \mu h + y_T^{LC} \quad (17)$$

<sup>3</sup>ここでの“ $O(1)$ ”の意味は、 $T$  を fix した時に  $h$  について述べていることに注意されたい。

となる。ここで、“ $\approx$ ”は漸近的にという意味である。

### 3.1 Lee-Carter 予測の長所

(17) 式より、Lee-Carter 予測はドリフト項を正しく捉えていることが解る。さらに、Lee-Carter 予測は漸近的には (14) 式における  $\hat{y}_{T+h}^{MISS}$  の一種であることも示せる。よって、Lee-Carter 予測は長期予測では最適予測と同等に機能する。これが時系列分析の理論上の長所である。

他にも、推定するパラメーターの数が第 1 主成分のドリフトという 1 つだけと非常に少ないことも長所と言える。共和分過程の伝統的な予測法である VEC 予測は共和分システム全体を推定する必要があるのでパラメーター数が  $O(m^2)$  と多くなり、 $T$  が小さいとパラメーター推定の精度に問題が生じて予測精度が落ちる傾向があるのだが、Lee-Carter 予測はそうした問題は無いだろう。このことは、 $m$  に対して  $T$  が十分大きくない場合が多い死亡率のデータでは非常に重要である。さらに、Lee and Carter (1992) が主張しているように、予測の信頼区間が計算しやすいといった長所や、人口学的な観点でのいくつかの長所もある。

### 3.2 Lee-Carter 予測の短所

Lee-Carter 予測は第 1 主成分しか予測に使わないため、Giroso and King (2007) 等が指摘しているように、第 2~第  $m$  主成分の情報をロスするという問題がある。ただ、この情報のロスがどのような形で予測精度に悪影響を及ぼすかは明らかになっていないと思われる。本小節では、この情報のロスが

- 短期予測が大きく外れる
- 非効率的なドリフト推定

という 2 つの短所として表れることを明らかにする。さらに、Lee-Carter 予測ではモデルにドリフト項があるにもかかわらず、detrend を行わないで主成分分析をする。このことが、予測に共和分制約が掛からないという 3 つ目の短所に結びつく。

1 つ目の短期予測の問題を説明する。今、最も単純な予測として、 $y_t$  の各系列をドリフト付きの単系列 ARIMA(0,1,0) モデルで予測することを考える。この単系列 ARIMA 予

測は、ドリフトの推定量を  $\overline{\Delta y}$  として

$$\hat{y}_{T+h}^{ARIMA} = \overline{\Delta y}h + y_T \quad (18)$$

$$\approx \mu h + y_T \quad (19)$$

となる。(19)と(17)を比べると、 $\mu h$ は同じだが残りの項が異なる。(19)における $y_T$ という項はARIMAモデルの最適予測の理論から導かれたものであり正当性があるが、(17)の $y_T^{LC}$ なる項は、

$$\begin{aligned} y_T^{LC} &= f_1 f_1' \tilde{y}_T + \bar{y} \\ &= f_1 f_1' (y_T - \bar{y}) + \bar{y} + F_{(m-1)} F_{(m-1)}' y_T - F_{(m-1)} F_{(m-1)}' y_T \\ &= y_T + F_{(m-1)} F_{(m-1)}' (\bar{y} - y_T) \end{aligned} \quad (20)$$

のように $y_T$ と異なるため正統性は無い。このため、 $h$ が小さい短期予測では $y_T$ と $y_T^{LC}$ の違いが大きく影響してLee-Carter予測は単系列ARIMA予測より精度が悪くなる。ただ、 $h$ が大きい長期予測では、 $y_T$ と $y_T^{LC}$ の違いは $\mu h$ にdominateされるため精度は悪化しない<sup>4</sup>。こうした問題は、実はLee and Carter (1992)のpp. 665 - 666でも指摘されているのだが、彼らはそう大きな問題ではないとしていた。しかし、5節のモンテカルロ実験や6節の実証例で示すように、この問題は予測精度をかなり悪化させる。加えて(20)にある $(\bar{y} - y_T)$ は $O_p(T)$ であるので $T$ が大きくなるほど $y_T$ と $y_T^{LC}$ の差が大きくなってLee-Carter予測の精度は悪化することが解る。通常、予測量は $T$ が大きいほどパラメータ推定が正確にでき精度が上がるが、Lee-Carter予測はそうではない。このことも、5節のモンテカルロ実験で示される。なお、もしLee-Carter予測が第1主成分だけではなく全ての主成分を使い、各主成分をドリフト付きのARIMA(0,1,0)モデルで予測していたら、 $FF' = I_m$ に注意すると、

$$\begin{aligned} \hat{y}_{T+h}^{LC} &= F \frac{1}{T-1} \sum_{t=2}^{T-1} F' \Delta \tilde{y}_t h + FF' \tilde{y}_T + \bar{y} \\ &= FF' \frac{1}{T-1} \sum_{t=2}^{T-1} \Delta \tilde{y}_t h + FF' \tilde{y}_T + \bar{y} \\ &= \overline{\Delta y} h + y_T \end{aligned}$$

<sup>4</sup>(14)式で示したように、長期予測ではドリフト以外の情報は価値を持たないため、 $y_T$ と $y_T^{LC}$ の違いということも予測精度に影響しない。

のように Lee-Carter 予測は単系列 ARIMA 予測 (18) と完全に一致する。よって、乱暴な言い方をすれば、Lee-Carter 予測とは最も単純なドリフト付きの単系列 ARIMA(0,1,0) 予測の情報量を落とすだけのものである。

2つ目の短所である非効率的なドリフト推定については、(16)にあるように Lee-Carter 予測は  $f_1 f_1' \overline{\Delta y}$  なる項でドリフト  $\mu$  を一致推定はしているのだが、その分散が大きくなるということである。具体的には、以下のようなになる。

**命題 3:**

$$\sqrt{T}(f_1 f_1' \overline{\Delta y} - \mu) \xrightarrow{L} N\left(0, \Gamma + \frac{1}{5}\Gamma^+\right) \quad (21)$$

ここで、 $\Gamma = C\Omega C'$ 、 $\Gamma^+ = \bar{F}'_{(m-1)} \bar{F}'_{(m-1)} \Gamma \bar{F}_{(m-1)} \bar{F}_{(m-1)}$  である。

**証明:** 付録を参照のこと。

(21) 式が非効率性を表していることは、階差平均  $\overline{\Delta y}$  で  $\mu$  を推定した時の極限分布である

$$\sqrt{T}(\overline{\Delta y} - \mu) \xrightarrow{L} N(0, \Gamma)$$

と比べるとよく解る。 $f_1 f_1' \overline{\Delta y}$  の極限分散は階差平均  $\overline{\Delta y}$  のそれに  $\frac{1}{5}\Gamma^+$  なる半正値定符号行列を加えたものである。非効率性がどの程度であるかを調べるため、増大する分散の合計である  $\frac{1}{5}\text{tr}(\Gamma^+)$  を評価する。 $\bar{F}'_{(m-1)} \bar{F}_{(m-1)} = I_{m-1}$  であることに注意すると

$$\text{tr}(\bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Gamma \bar{F}_{(m-1)} \bar{F}'_{(m-1)}) = \text{tr}(\bar{F}'_{(m-1)} \Gamma \bar{F}_{(m-1)})$$

となるので、ここに Poincaré の不等式を適用すると

$$\frac{1}{5} \sum_{j=1}^{m-1} g_j \leq \frac{1}{5} \text{tr}(\Gamma^+) \leq \frac{1}{5} \sum_{j=2}^m g_j$$

を得る<sup>5</sup>。ここで、 $g_1 \geq \dots \geq g_m$  は  $\Gamma$  の固有値である。この不等式の持つ意味を理解するため例として  $m = 3$ 、 $r = 2$  の場合を考える。この時は  $g_2 = g_3 = 0$  となるので、階差平均  $\overline{\Delta y}$  の極限分散の合計  $\text{tr}(\Gamma) = g_1$  に対し増加分は最大で  $\frac{g_1}{5}$ 、最小で 0 であり、増加率は最大で  $100 \times \frac{g_1/5}{g_1} = 20\%$ 、最小で 0% である。5 節のモンテカルロ実験で  $m = 3$ 、 $r = 2$  のケースを行うが、いくつかのパラメーターを試してみたが増加率はせいぜい 1~2% 程度

<sup>5</sup>Magnus and Neudecker (1999) の p. 211 参照。

であった。このように、この非効率性の問題は、そう大きな問題にはならないこともある。ただ、 $\Gamma$  のランクが  $m - r$  なので、 $\frac{1}{5}\text{tr}(\Gamma^+)$  は  $m - r$  が増えると増える傾向にある。モンテカルロ実験では  $m = 30$ 、 $r = 26$  のケースも扱うが、こちらはいくつかのパラメータを試してみたところ増加率は 4~7%程度であった。しかし、図 1 を見る限りでは  $r$  はかなり高い ( $m - r$  は低い) と思われるので、この死亡率のデータに限って言えば非効率性の程度はそう高くないと思われる。なお、もし第 1 主成分だけではなく全ての主成分を使い、各主成分の階差平均でドリフト項を推定すれば

$$FF'\overline{\Delta y} = \overline{\Delta y}$$

となるので非効率性の問題は生じない。

3 つ目の共和分制約の問題について述べるには、共和分制約がどのようなもので予測の精度にどう関係するかを示さなければならない。これは 4.1 節で示すので、Lee-Carter 予測と共和分制約についてはその後の 4.3 節で議論する。

## 4 Lee-Carter 予測の改善

前節で示したように、Lee-Carter 予測にはいくつかの欠点がある。本節では、それらを補う予測法を提案する。

### 4.1 MTV 予測

Lee-Carter 予測の欠点を克服するため、MTV (multivariate time series variance component) 予測というものを使うことを考える。MTV 予測は元々は Kariya (1987) が定常時系列過程に対して考案したものだが、Chigira and Yamamoto (2009) が共和分過程に適合するよう拡張している。Chigira and Yamamoto (2009) の MTV 予測は主成分を使うという点では Lee-Carter 予測と同じだが、データを detrend してから主成分分析を行う点と、第 1 主成分だけではなく全ての主成分を使うという点が異なる。このため、Lee-Carter 予測のように情報のロスが生じないし、共和分制約も掛かることになる。

以下、MTV 予測を説明する。モデル (11) もしくは (12) に MTV 予測を適用するには、

まず demean と detrend を行うために  $y_t$  を定数項とトレンド項に OLS で回帰して残差

$$\check{y}_t = y_t - \hat{\gamma} - t\hat{\mu}_{trend} \quad (22)$$

を求める。ここで、 $\hat{\gamma}$  は定数項の OLS 推定量、 $\hat{\mu}_{trend}$  はトレンド項の OLS 推定量である。そして、

$$\check{Y}_{(T \times m)} = \begin{bmatrix} \check{y}'_1 \\ \vdots \\ \check{y}'_T \end{bmatrix}$$

なる行列を作り、以下のような固有値分解を行う。

$$\check{Y}'\check{Y} = B\Pi B', \quad B = [b_1 \ \dots \ b_m], \quad \Pi = \begin{bmatrix} \pi_1 & 0 & \dots & 0 \\ 0 & \pi_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \pi_m \end{bmatrix} \quad (23)$$

ここで、 $\pi_1 \geq \dots \geq \pi_m$  が固有値であり、 $b_1, \dots, b_m$  がそれに対応する固有ベクトルである。 $B$  の漸近的な挙動については、 $[b_1 \ \dots \ b_{m-r}] = B_{(m-r)}$ 、 $[b_{m-r+1} \ \dots \ b_m] = B_{(r)}$  とすると、

$$\begin{cases} B_{(m-r)} \xrightarrow{p} \beta_{\perp} P_{\perp} \\ B_{(r)} \xrightarrow{p} \beta P \end{cases} \quad (24)$$

となることが Harris (1997)、Snell (1999)、Chigira (2008) で示されている。ここで、 $\beta_{\perp}$  は  $\beta'\beta_{\perp} = 0$  なる性質を持つ  $\beta$  の直交補空間を形成する  $(m \times (m-r))$  の full rank 行列であり、 $P_{\perp}$  と  $P$  は full rank な正方行列である。この結果は、

$$\begin{cases} B'_{(m-r)}\check{y}_t \approx I(1) \\ B'_{(r)}\check{y}_t \approx I(0) \end{cases} \quad (25)$$

となることを意味する。これより、第 1~第  $m-r$  主成分を ARIMA( $p,1,q$ ) モデルで  $\widehat{B'_{(m-r)}\check{y}_{T+h}}$  と、第  $m-r+1$ ~第  $m$  主成分を ARIMA( $p,0,q$ ) モデルで  $\widehat{B'_{(r)}\check{y}_{T+h}}$  と予測する。この予測は主成分についてのものなので、次に、固有ベクトルをこれらの予測に掛けて主成分を原データに戻す。これで demean と detrend をした原データの予測量が作れるので、最後に、定数項とドリフト項を足し戻す。こうした手順により、MTV 予測は

具体的には

$$\hat{y}_{T+h}^{MTV} = (T+h)\hat{\mu}_{trend} + \hat{\gamma} + B_{(m-r)}\widehat{B'_{(m-r)}}\dot{y}_{T+h} + B_{(r)}\widehat{B'_{(r)}}\dot{y}_{T+h} \quad (26)$$

$$\begin{aligned} &= (T+h)\hat{\mu}_{trend} + \hat{\gamma} + B_{(m-r)}\widehat{B'_{(m-r)}}\dot{y}_{T+h} - B_{(m-r)}B'_{(m-r)}\hat{\gamma} + B_{(r)}\widehat{B'_{(r)}}\dot{y}_{T+h} - B_{(r)}B'_{(r)}\hat{\gamma} \\ &= (T+h)\hat{\mu}_{trend} + B_{(m-r)}\widehat{B'_{(m-r)}}\dot{y}_{T+h} + B_{(r)}\widehat{B'_{(r)}}\dot{y}_{T+h} \\ &\approx (T+h)\mu + \beta_{\perp}P_{\perp}\widehat{B'_{(m-r)}}\dot{y}_{T+h} + \beta P\widehat{B'_{(r)}}\dot{y}_{T+h} \end{aligned} \quad (27)$$

なる形になる。ここで、 $\dot{y}_t = y_t - t\hat{\mu}_{trend}$  である。

このように MTV 予測は各主成分を1つずつ予測するので推定するパラメーターの数は  $O(m)$  である。これに対し、VEC 予測は共和分システム全体を推定するのでパラメーターの数は  $O(m^2)$  となり、 $T$  が十分大きくないデータに対しては VEC 予測より MTV 予測の方がパフォーマンスが良い。こうしたことは、5節のモンテカルロ実験や6節の実証例で具体的に示される。

MTV 予測が共和分制約を満たすことは次のように示せる。まず、(27) に  $\beta'$  を左から掛けると

$$\begin{aligned} \beta'\hat{y}_{T+h}^{MTV} &= \beta'(T+h)\mu + \beta'\beta_{\perp}P_{\perp}\widehat{B'_{(m-r)}}\dot{y}_{T+h} + \beta'\beta P\widehat{B'_{(r)}}\dot{y}_{T+h} \\ &= \beta'd_2^*(T+h) + 0 + \beta'\beta P\widehat{B'_{(r)}}\dot{y}_{T+h} \end{aligned} \quad (28)$$

となる。次に、(13) を使うと

$$\begin{aligned} B'_{(r)}\dot{y}_t &\approx P'\beta'(y_t - \mu t) \\ &= P'(\beta'\gamma + \beta'C^*(L)\varepsilon_t) \end{aligned}$$

のように  $B'_{(r)}\dot{y}_t$  はトレンド項の無い  $I(0)$  過程となるので、 $\widehat{B'_{(r)}}\dot{y}_{T+h}$  は長期予測では  $P'\beta'\gamma$  になることが解る。よって、(28) より、

$$\lim_{h \rightarrow \infty} (\beta'\hat{y}_{T+h}^{MTV} - \beta'\beta P P'\beta'\gamma - \beta'd_2^*(T+h)) = 0$$

を得る。ここで、 $P$  はある対称な  $(r \times r)$  行列の  $r$  個の固有ベクトルからなる行列であることが Snell (1999) によって示されているので  $PP' = I_r$  となり、一般性を失うことなく  $\beta'\beta = I_r$  と基準化できることに注意すると

$$\lim_{h \rightarrow \infty} (\beta'\hat{y}_{T+h}^{MTV} - \beta'\gamma - \beta'd_2^*(T+h)) = 0$$

のように共和分制約を満たすことが解る。こうした議論の詳細は Chigira and Yamamoto (2009) を参照されたい<sup>6</sup>。Chigira and Yamamoto (2009) では、共和分制約は長期予測の精度は改善しないが短期～中期予測の精度は改善することがモンテカルロ実験で示されている。

## 4.2 修正 MTV 予測

本小節では、全小節で導入した MTV 予測にある修正を加える。MTV 予測では、(26) にあるように、ドリフトの推定量として  $\hat{\mu}_{trend}$  を使う。しかし、 $\hat{\mu}_{trend}$  は効率的な推定量ではないので、予測の精度に悪影響を与えられと考えられる。そこで、 $\hat{\mu}_{trend}$  の代わりに効率的な推定量である  $\overline{\Delta y}$  を使うことを考える。このためには、単純に (26) の  $\hat{\mu}_{trend}$  を  $\overline{\Delta y}$  で置き換えれば良いように見えるが、実はそうではない。MTV 予測は (22) 式のように demean と detrend をしているので、予測をする時に  $\hat{\mu}_{trend}$  を  $\overline{\Delta y}$  で置き換えるとある種の齟齬が生じる。具体的には、(22) 式で demean と detrend をした  $y_t$  を固有値分解して主成分にし、その主成分を原データに元に戻して  $\hat{\gamma}$  と  $\overline{\Delta y}$  を足し戻した際の  $y_T$  に注目すると、

$$BB'(y_T - \hat{\gamma} - T\hat{\mu}_{trend}) + \hat{\gamma} + T\overline{\Delta y} = y_T - T\hat{\mu}_{trend} + T\overline{\Delta y} \neq y_T$$

のように  $y_T$  に一致しない。この結果は、単純に (26) の  $\hat{\mu}_{trend}$  を  $\overline{\Delta y}$  で置き換えただけでは  $y_T$  が予測の出発点とならないことを意味し、予測の精度が悪化する原因となる。そこで、 $y_T$  が予測の出発点となるよう調整した

$$\hat{y}_{T+h}^{mMTV} = (T+h)\overline{\Delta y} + T(\hat{\mu}_{trend} - \overline{\Delta y}) + \hat{\gamma} + B_{(m-r)}\widehat{B'_{(m-r)}\check{y}_{T+h}} + B_{(r)}\widehat{B'_{(r)}\check{y}_{T+h}} \quad (29)$$

を修正 MTV 予測とする<sup>7</sup>。

修正 MTV 予測の作り方を整理しておく。

**Step.1** (22) 式のように demean と detrend をし、(23) のように固有値分解を行う。

**Step.2** 第 1～第  $m-r$  主成分  $B'_{(m-r)}\check{y}_t$  を ARIMA( $p,1,q$ ) モデルで  $\widehat{B'_{(m-r)}\check{y}_{T+h}}$  と、第

<sup>6</sup>Chigira and Yamamoto (2009) では共和分関係に確定項は入らないという定式化を使っているので  $\lim_{h \rightarrow \infty} \beta' \hat{y}_{T+h}^{MTV} = 0$  といった表現になっている。

<sup>7</sup>MTV 予測 (26) は、こうした調整をしなくても  $y_T$  が予測の出発点となっている。

$m - r + 1 \sim$  第  $m$  主成分  $B'_{(r)} \tilde{y}_t$  を ARIMA( $p, 0, q$ ) モデルで  $\widehat{B'_{(r)} \tilde{y}_{T+h}}$  と予測する。こ  
こまでは通常の MTV 予測と同じである。

**Step.3** Step.2 で得た予測量に固有ベクトルを掛けて原データに戻し、ドリフト推定量  $\overline{\Delta y}$ 、定数項推定量  $\hat{\gamma}$ 、調整項  $T(\hat{\mu}_{trend} - \overline{\Delta y})$  を加えた (29) 式を計算する。

この修正 MTV 予測は、予測に使うドリフト推定量が効率的な  $\overline{\Delta y}$  であるので MTV 予測より精度が高いと期待される。

### 4.3 Lee-Carter 全主成分予測

石井 (2008) 等では、Lee-Carter 予測を改善するために第 1 主成分だけでなく残りの主成分も使うことが試みられている。本小節では、こうしたことが予測精度の改善に役立つか検討する。

今、Lee-Carter 予測を行う際に第 1 主成分だけではなく全ての主成分を使うことを考え、こうした予測を Lee-Carter 全主成分予測と呼ぶことにする。この時には、全ての主成分を予測する必要があるため、各主成分がどのような過程に従うかを明らかにしなければならない。データを detrend すれば、各主成分は (25) 式で示したように  $I(1)$  と  $I(0)$  に分かれるが、detrend しないと次の命題のようになる。

**命題 4:**

$$f'_1 \tilde{y}_t \approx \pm \sqrt{\mu' \mu} t \pm \frac{1}{\sqrt{\mu' \mu}} \left( u'_\perp \beta'_\perp C \sum_{s=1}^t \varepsilon_s + \mu' C^*(L) \varepsilon_t - \mu' \bar{y} \right) + O_p(1) \quad (30)$$

$$F'_{(m-1)} \tilde{y}_t \approx V'_\perp \beta'_\perp C \sum_{s=1}^t \varepsilon_s + \bar{F}'_{(m-1)} C^*(L) \varepsilon_t - \bar{F}'_{(m-1)} \bar{y} + O_p(\sqrt{T}) \quad (31)$$

ここで、 $u_\perp$  はある nonzero の  $((m-r) \times 1)$  ベクトル、 $V_\perp$  は full rank の  $((m-r) \times (m-1))$  行列である。

**証明:** 要求に応じて提供する。

第 1 主成分には  $I(1)$  コンポーネントと  $I(0)$  コンポーネントが混じっているが、 $I(1) + I(0) = I(1)$  なので  $I(1)$  であり、また  $\pm \sqrt{\mu' \mu}$  なるドリフト項を持つことが解る。(30) は、Lee and Carter (1992) を初め多くの論文が第 1 主成分をドリフト付きの ARIMA(0,1,0) モデルで予測したことをある程度正当化している。ただ、かなり複雑な stochastic コンポーネ

ントを持つので、ドリフト付きの ARIMA( $p,1,q$ ) モデルを用いた方が良さだろう。第 2~第  $m$  主成分も  $I(1)$  と  $I(0)$  が混じった  $I(1)$  過程だが、第 1 主成分のようにドリフト項は無い。(31) は、Lee and Carter (1992) が第 1 主成分だけを用いたことをある程度正当化する。なぜならば、(14) 式で示したように  $I(1)$  過程の長期予測にはドリフト項が決定的に重要な役割を果たすが、第 2~第  $m$  主成分にはドリフト項が無いので無視してもあまり問題にならないからである。とはいえ、短期~中期予測では、3.2 節で示したようにこれらの主成分を無視することは予測精度の悪化につながる。第 2~第  $m$  主成分はドリフトの無い ARIMA( $p,1,q$ ) モデルで予測するのが良さだろう。(30) と (31) からは、データを detrend しないで主成分を算出すると、全ての主成分が  $I(1)$  になることが解る。これは、MTV 予測での (25) とは対照的である。

以上の結果より、Lee-Carter 全主成分予測では第 1 主成分をドリフト付きの ARIMA( $p,1,q$ ) モデルで

$$f_1' \overline{\Delta y} h + \widehat{f_1' \tilde{y}}_{T+h} \quad (32)$$

と予測する。ここで、 $f_1' \overline{\Delta y}$  は (2) にあるように第 1 主成分のドリフトの推定量であり、 $\widehat{f_1' \tilde{y}}_{T+h}$  はドリフト項以外の AR( $p$ ) と MA( $q$ ) による部分である。(30) より、 $f_1' \overline{\Delta y}$  は  $\pm\sqrt{\mu' \mu}$  の一致推定量であることが解る。(3) と (32) を見比べると違いは  $f_1' \tilde{y}_T$  か  $\widehat{f_1' \tilde{y}}_{T+h}$  かだけであり、もし  $p = q = 0$  ならば  $\widehat{f_1' \tilde{y}}_{T+h} = f_1' \tilde{y}_T$  となって両者は一致する。第 2~第  $m$  主成分については、(31) で示したようにドリフトの無い  $I(1)$  過程に従うドリフトの無い ARIMA( $p,1,q$ ) モデルで  $\widehat{F'_{(m-1)} \tilde{y}}_{T+h}$  と予測することにする。これで主成分の予測ができたので、後は固有ベクトルを掛けて原データに戻し、最後に平均を足し戻せば Lee-Carter 全主成分予測が作れる。具体的には、

$$\begin{aligned} \hat{y}_{T+h}^{LCA} &= f_1(f_1' \overline{\Delta y}(T+h) + \widehat{f_1' \tilde{y}}_{T+h}) + F_{(m-1)} \widehat{F'_{(m-1)} \tilde{y}}_{T+h} + \bar{y} \\ &= f_1(f_1' \overline{\Delta y}(T+h) + \widehat{f_1' y}_{T+h}) - f_1 f_1' \bar{y} + F_{(m-1)} \widehat{F'_{(m-1)} y}_{T+h} - F_{(m-1)} F'_{(m-1)} \bar{y} + \bar{y} \\ &= f_1(f_1' \overline{\Delta y}(T+h) + \widehat{f_1' y}_{T+h}) + F_{(m-1)} \widehat{F'_{(m-1)} y}_{T+h} \end{aligned} \quad (33)$$

となる。

この Lee-Carter 全主成分予測は全ての主成分を使っているので情報のロスは無くなる。具体的には、 $\bar{y}$  で demean した  $y_t$  を固有値分解して主成分にし、その主成分を原データに

元に戻して  $\bar{y}$  を足し戻した際の  $y_T$  に注目すると、

$$BB'(y_T - \bar{y}) + \bar{y} = y_T$$

のように  $y_T$  に一致する。よって、Lee-Carter 予測とは違い  $y_T$  が予測の出発点になっており、これは1つの長所である。また、パラメーター数が  $O(m)$  であることも長所と言える。しかし、ドリフト  $\mu$  の推定には  $f_1 f_1' \overline{\Delta y}$  を使うため、(21) で示したように漸近的には非効率的になる。また、次の命題に示すように Lee-Carter 全主成分予測は共和分制約を満たさない。

**命題 5:**

$$\lim_{h \rightarrow \infty} (\beta' \hat{y}_{T+h}^{LCA} - \beta' \gamma - \beta' d_2^*(T+h)) \neq 0 \quad (34)$$

**証明:** 付録を参照のこと。

なお、(34) 式は通常の Lee-Carter 予測に共和分制約が掛からないことも示している。

このように、Lee-Carter 予測の欠点を改善するために第1主成分以外の主成分を使うことは情報のロスを補う効果はある。しかし、第1主成分以外の主成分はドリフトの推定には貢献しないためドリフトの推定は漸的に非効率のままであるし、detrend をせずに主成分を計算するので共和分制約が掛からないという問題点は残る。

## 5 モンテカルロ実験

本節では、これまで示してきた予測の精度をモンテカルロ実験で調べる。本稿は死亡率の予測を考えているので、実験の DGP にはデータが図1に近くなるようなものを用いる。

### 5.1 実験のデザイン

実験に使う DGP は、以下のような VEC モデルとする。

$$\Delta y_t = \alpha \beta' y_{t-1} + \mu + \varepsilon_t, \quad \varepsilon_t \sim NID(0, I_m) \quad (35)$$

ここで、 $\alpha$  は  $(m \times r)$  の full rank 行列であり、 $\mu$  は (9) によって (10) のように書けるとする。(35) は、VMA 表現では、

$$\begin{aligned} C(L) &= I_m L^0 + (A - I_m)L + (A^2 - A)L^2 + (A^3 - A^2)L^3 + \dots \\ C &= C(1) = \beta_{\perp}(\alpha'_{\perp}\beta_{\perp})^{-1}\alpha'_{\perp} \\ C^*(L) &= (I - C)L^0 + (A - C)L + (A^2 - C)L^2 + (A^3 - C)L^3 + \dots \end{aligned}$$

として (11) や (12) のように書けることが知られている<sup>8</sup>。ここで、 $A = I_m + \alpha\beta'$  であり、 $\alpha_{\perp}$  は  $\alpha'\alpha_{\perp} = 0$  なる性質を持つ  $\alpha$  の直交補空間を形成する  $(m \times (m - r))$  の full rank 行列である。

パラメーター設定は、

$$\begin{cases} \text{ケース 1: } m = 3, r = 2 \\ \text{ケース 2: } m = 30, r = 26 \end{cases}$$

の 2 ケースを考え<sup>9</sup>、両ケース共  $T = 50, 100, 200, 500$ 、 $h = 1, 2, \dots, 5, 10, 20, 30, 40, 50$  とする。実験の繰り返し回数は 5000 回とする。具体的なパラメーターの値は、 $m = 30$  のケースは省略するが、 $m = 3$  では

$$\alpha = \begin{bmatrix} -1.35 & -4.05 \\ 2.25 & -2.25 \\ 0 & -0.90 \end{bmatrix}, \beta = \begin{bmatrix} 0.2 & 0.10 \\ -0.25 & 0.10 \\ 0 & -0.15 \end{bmatrix}, d_1 = \begin{bmatrix} -0.30 \\ -0.10 \\ -0.20 \end{bmatrix}, d_2 = \alpha_{\perp} \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix}$$

とした。 $d_2$  については、(9) を満たすために  $\alpha_{\perp}$  の空間に入るように設定している。これらのパラメーター設定のポイントは、

$$\mu = Cd_1 + C^*(L)d_2 \doteq \begin{bmatrix} -0.39 \\ -0.22 \\ -0.23 \end{bmatrix}$$

のようにドリフト項が全て負になるようにしていることである<sup>10</sup>。図 1 から見て、正のドリフトは適切ではないので、 $m = 30$  でも、ドリフトが全て負になるように注意してパラメーターを設定している。

<sup>8</sup>Johansen (1995) 参照。

<sup>9</sup>ケース 2 で  $r = 26$  というのは半端な数字の感じもするが、6.1 節で日本人男性の死亡率のデータを実際に分析した結果に照らしてこの値にした。ケース 2 は実際のデータでの予測精度をシミュレートするというモチベーションに立っている。

<sup>10</sup> $\mu$  を切りのよい数字にするのは難しいので四捨五入した値を書いている。

比較する予測は、VEC 予測、MTV 予測、修正 MTV 予測、Lee-Carter 予測、Lee-Carter 全主成分予測、単系列 ARIMA 予測とする。ここで、本稿が焦点を当てている Lee-Carter 予測では、第 1 主成分を特定化を行わずドリフト付きの ARIMA(0,1,0) として予測するので、 $\hat{y}_{T+h}^{LC}$  には必ずドリフトの存在と  $I(1)$  であることが impose されることに注意が必要である。今回の実験では、他の予測がこれに対して不利にならないように、ドリフトと  $I(1)$  であることを特定化を行わず impose することにする。各予測の具体的な計算法は以下ようになる。

**VEC 予測** Johansen の最尤法でモデルを推定し、VEC form で予測をする。今回の実験では理想的な予測として VEC 予測を扱うため、VAR の次数、 $r$ 、確定項の形は全て既知とする。VEC 予測は  $\hat{y}_{T+h}^{VEC}$  と書き、最尤法で推定された  $\mu$  を  $\hat{\mu}_{ML}$  と書く。VEC 予測は共和分制約を満たすことが知られている。

**MTV 予測** ドリフトがあることは既知として (22) 式のように detrend し、4.1 節のように予測量を作る。MTV 予測では  $r$  は未知とする。このため、各主成分に Phillips-Perron 検定を有意水準 1%で行って和分の次数を決める。ただし、 $I(1)$  であることを impose するため第 1 主成分が  $I(1)$  であることは既知とする。また、全ての主成分の AR と MA の次数、確定項の形は SBIC で決める。この予測量の具体的な式は (26) で与えられる。

**修正 MTV 予測** データの detrend と主成分の予測を MTV 予測と同様に行う。その後、4.2 節で説明したように、ドリフト項を足し戻す時に MTV 予測で使う  $\hat{\mu}_{trend}$  の代わりに  $\overline{\Delta y}$  を使う。こうしたことを行う理由は、 $\hat{\mu}_{trend}$  より  $\overline{\Delta y}$  の方が効率的なため、予測精度の向上が期待されるからである。この予測量の具体的な式は (29) で与えられる。

**Lee-Carter 予測** 3 節で説明した通りの予測を行う。具体的には、(4) で与えられる。

**Lee-Carter 全主成分予測** 4.3 節で説明した予測を行う。ドリフトと  $I(1)$  であることを impose するため、第 1 主成分がドリフト付きの  $I(1)$  であることは既知とする。第 2 主成分以降は  $I(1)$  として予測するが確定項の形は SBIC で決める。また、全ての主成分の AR と MA の次数は SBIC で決める。

**単系列 ARIMA 予測**  $y_{at}$ ,  $a = 1, 2, \dots, m$  の各系列に、通常の 1 変量 ARIMA モデルを当てはめて予測する。全系列がドリフト付きの  $I(1)$  であることは既知とし、AR と MA の次数は SBIC で決める。単系列 ARIMA 予測は  $\hat{y}_{T+h}^{ARIMA}$  と書く。なお、単系列 ARIMA 予測で用いられる  $\mu$  の推定量は  $\overline{\Delta y}$  である。

なお、パラメーターの数は VEC 予測が  $O(m^2)$ 、Lee-Carter 予測が 1 であり、それ以外の予測は全て  $O(m)$  である。

予測の精度を比較するには 2.3 節で述べた trace MSE ratio を使うが、この実験では最も primitive な単系列 ARIMA 予測をベンチマークとし、単系列 ARIMA 予測に対して各予測がどの程度の精度を持つのかを調べる。具体的には、以下の 5 つの trace MSE ratio を計算する。

$$\begin{aligned} \text{ratio(VEC)} &= \frac{\text{tr MSE}(\hat{y}_{T+h}^{VEC})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})} \\ \text{ratio(MTV)} &= \frac{\text{tr MSE}(\hat{y}_{T+h}^{MTV})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})} \\ \text{ratio(mMTV)} &= \frac{\text{tr MSE}(\hat{y}_{T+h}^{mMTV})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})} \\ \text{ratio(LC)} &= \frac{\text{tr MSE}(\hat{y}_{T+h}^{LC})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})} \\ \text{ratio(LCA)} &= \frac{\text{tr MSE}(\hat{y}_{T+h}^{LCA})}{\text{tr MSE}(\hat{y}_{T+h}^{ARIMA})} \end{aligned}$$

これらの trace MSE ratio が 1 より大きければ、単系列 ARIMA 予測より精度が悪いことを意味し、1 より小さければ単系列 ARIMA 予測より精度が良いことを意味する。ここで、(14) 式の意味について述べる。(14) 式における  $\hat{y}_{T+h}^{OPT}$  は (35) の下ではパラメーターを既知とした VEC 予測になり、本稿で考慮する VEC 予測以外の予測量も全て  $\mu$  を正しく考慮していることに注意すると、 $T$  が十分大きく  $\mu$  が概ね一致推定できていれば、長期予測においては

$$\text{ratio(VEC)} = \text{ratio(MTV)} = \text{ratio(mMTV)} = \text{ratio(LC)} = \text{ratio(LCA)} = 1$$

となる。MTV 予測や修正 MTV 予測が共和分制約を満たすといった長所や、Lee-Carter 予測が  $y_T$  を出発点としないといった短所は、長期予測の精度には影響せず短期～中期予測に影響するのである。

## 5.2 ケース 1 の実験結果

trace MSE ratio は表 1 のようになった。この表から、全ての  $T$  と  $h$  で、

- Lee-Carter 予測が最も悪い
- VEC 予測が最も良い
- MTV 予測と修正 MTV 予測が比較的良い
- Lee-Carter 全主成分予測は MTV 予測や修正 MTV 予測より良くなることはほぼ無いが、Lee-Carter 予測よりは良い

といったことが解る。

以下、各予測の精度についてコメントする。

### VEC 予測

正しい定式化で予測をしているので stochastic コンポーネントの予測が良く、これが  $h = 1 \sim 10$  の予測の精度を上げていると考えられる。長期予測に重要な枠割を果たすドリフト項についても、表 2 にあるように MSE が小さいため、 $h = 10$  以降の精度も良い。VEC 予測の精度が良い理由は、他の予測には取り入れられていない共和分関係にトレンドがあるという情報を明示的に取り入れていることも 1 つの理由かもしれない。

### MTV 予測

データを detrend してから主成分分析を行うことで共和分の情報を取り入れているため stochastic コンポーネントの予測が上手く行き、 $h = 1 \sim 10$  の精度が良いと考えられる。表 3 では、各主成分に Phillips-Perron 検定を行った結果  $I(1)$  主成分の数を何個と判定したかを示しているが、MTV 予測では  $T = 100$  で正しい数（つまり 1 個）を 5000 回全てで検知しており、これも共和分制約が適切に掛かって stochastic コンポーネントの予測が良く機能したことを裏付けている。共和分制約が短期～中期予測の精度を改善するという結果は Chigira and Yamamoto (2009) と同じである。ただ、表 2 にあるように、MTV 予測で使う  $\hat{\mu}_{trend}$  は MSE が大きく、これが  $h = 10$  以降の精度をやや下げていると思わ

表 1: trace MSE ratio (ケース 1)

$T = 50$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.68	0.64	0.66	0.72	0.75	0.84	0.90	0.92	0.94	0.94
ratio(MTV)	0.89	0.89	0.91	0.93	0.94	0.98	1.01	1.04	1.05	1.06
ratio(mMTV)	0.89	0.88	0.90	0.92	0.93	0.96	0.98	0.98	0.99	0.99
ratio(LC)	1.42	1.18	1.11	1.08	1.06	1.03	1.01	1.01	1.00	1.00
ratio(LCA)	0.92	0.92	0.94	0.95	0.95	0.97	0.98	0.99	0.99	0.99

$T = 100$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.65	0.61	0.64	0.69	0.73	0.85	0.92	0.94	0.95	0.95
ratio(MTV)	0.84	0.84	0.86	0.89	0.90	0.96	0.99	1.01	1.03	1.04
ratio(mMTV)	0.84	0.84	0.86	0.88	0.90	0.94	0.97	0.98	0.98	0.98
ratio(LC)	1.88	1.43	1.29	1.22	1.18	1.09	1.04	1.02	1.02	1.01
ratio(LCA)	0.91	0.91	0.92	0.94	0.95	0.97	0.98	0.99	0.99	0.99

$T = 200$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.64	0.61	0.63	0.69	0.73	0.85	0.92	0.94	0.96	0.96
ratio(MTV)	0.82	0.83	0.84	0.87	0.89	0.95	0.99	1.00	1.02	1.02
ratio(mMTV)	0.82	0.83	0.84	0.87	0.89	0.94	0.97	0.98	0.98	0.99
ratio(LC)	2.77	1.97	1.66	1.51	1.42	1.22	1.11	1.07	1.05	1.04
ratio(LCA)	0.89	0.91	0.92	0.94	0.94	0.97	0.98	0.99	0.99	0.99

$T = 500$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.63	0.60	0.62	0.68	0.72	0.85	0.93	0.95	0.96	0.97
ratio(MTV)	0.80	0.81	0.82	0.86	0.88	0.94	0.98	0.99	1.00	1.01
ratio(mMTV)	0.80	0.81	0.82	0.86	0.88	0.94	0.97	0.98	0.98	0.99
ratio(LC)	5.62	3.57	2.78	2.40	2.12	1.59	1.30	1.20	1.15	1.12
ratio(LCA)	0.88	0.90	0.91	0.93	0.94	0.97	0.99	0.99	0.99	1.00

れる。なお、表3は、(今回の実験では第1主成分を必ず  $I(1)$  にしていることを割り引く必要はあるが) Chigira (2008) の共和分検定が適切に work していることも表している。

### 修正 MTV 予測

stochastic コンポーネントの予測は MTV 予測と同じなので、MTV 予測と同じ理由で  $h = 1 \sim 10$  での精度が良い。ドリフトについては、MTV 予測と違って  $\overline{\Delta y}$  を使うが、表2にあるように  $\overline{\Delta y}$  の MSE は小さいため、 $h = 10$  以降でも予測の performance は良い。

### Lee-Carter 予測

3.2 節で説明したように、 $h = 1 \sim 10$  の予測がかなり悪い。ただ、表2にあるように、Lee-Carter 予測が使うドリフト推定量  $f_1 f_1' \overline{\Delta y}$  の MSE は小さいため、Chigira and Yamamoto (2012) が示したように長期予測は上手く行き、 $h = 10$  以降では予測精度は徐々に改善される。しかし、単系列 ARIMA 予測より良くなることはなかった。また、Lee-Carter 予測以外の予測は全て  $T$  が大きくなると trace MSE ratio が下がったが、Lee-Carter 予測は逆に上がった。この理由は 3.2 節で説明したが、 $T$  が大きいほど精度が悪化するという性質は大きな欠点である。

### Lee-Carter 全主成分予測

4.3 節で説明したように共和分制約が掛からないため  $h = 1 \sim 10$  は良くないが、そう悪いわけでもない。これは、 $u f_1' \tilde{y}_{T+h}$  と  $V F_{(m-1)}' \tilde{y}_{T+h}$  の1部がたまたま0に近い値になって accidentally に partial な共和分制約が掛かっていることを意味するかもしれない。このあたりはさらなる分析が必要だろう。いずれにせよ、Lee-Carter 全主成分予測は stochastic コンポーネントの予測は共和分制約が完全には掛からないという意味で大幅に良くはならないが、全ての主成分を使うことで Lee-Carter 予測よりは精度が上がり、ドリフト推定量  $f_1 f_1' \overline{\Delta y}$  の MSE は小さいので、全体では可もなく不可もないといった精度になっていると思われる。

表 2: 各ドリフト推定量の MSE の合計 (ケース 1)

$T$	$\overline{\Delta y}$	$\hat{\mu}_{ML}$	$\hat{\mu}_{trend}$	$f_1 f_1' \overline{\Delta y}$
50	0.075	0.072	0.086	0.074
100	0.036	0.035	0.042	0.035
200	0.019	0.018	0.022	0.018
500	0.008	0.007	0.009	0.008

Note: ドリフトの推定量は  $(3 \times 1)$  のベクトルになるので、各要素の MSE を計算してそれを合計した数値を載せている。

表 3: MTV 予測で判定された  $I(1)$  主成分の数 (ケース 1)

$T \setminus \#\{I(1)\}$	3	2	1
50	0.06	21.82	78.12
100	0	0	100
200	0	0	100
500	0	0	100

Note: この表の見方は、例えば  $T = 50$  の時  $I(1)$  主成分の数が 3 と判定されたのは 0.06%、2 と判定されたのが 21.82%、1 と判定されたのが 78.12%ということである。第 1 主成分を必ず  $I(1)$  にするので、 $I(1)$  主成分の数が 0 と判定されることはない。

表 4:  $\overline{\Delta y}$  と  $f_1 f_1' \overline{\Delta y}$  の分散の比較 (ケース 1)

$\text{tr}(\Gamma)$	$T$	$\text{tr}(\hat{G})$	$\text{tr}(\Gamma + \frac{1}{5}\Gamma^+)$	$T$	$\text{tr}(\widehat{G + \frac{1}{5}G^+})$
3.52	50	3.74	3.58	50	3.68
	100	3.50		100	3.42
	200	3.61		200	3.58
	500	3.60		500	3.63

Note:  $\hat{G}$  は 5000 個の  $\overline{\Delta y}$  から計算した  $(3 \times 3)$  の標本共分散行列に  $T$  を掛けたもので、 $\widehat{G + \frac{1}{5}G^+}$  は 5000 個の  $f_1 f_1' \overline{\Delta y}$  から計算した  $(3 \times 3)$  の標本共分散行列に  $T$  を掛けたものである。

予測精度の議論は以上だが、ここで命題 3 で示した  $f_1 f_1' \overline{\Delta y}$  の非効率性について考察する。今回の DGP では  $\text{tr}(\Gamma) \doteq 3.52$ 、 $\text{tr}(\Gamma + \frac{1}{5}\Gamma^+) \doteq 3.58$  であり増加率は約 1.9% である。表 4 では、 $\text{tr}(\Gamma)$  と  $\text{tr}(\Gamma + \frac{1}{5}\Gamma^+)$  を実験で得た 5000 個の  $\overline{\Delta y}$  と  $f_1 f_1' \overline{\Delta y}$  から推定した結果を載せている。 $\text{tr}(\hat{G})$ 、 $\text{tr}(\widehat{G + \frac{1}{5}G^+})$  共に真の値への収束は遅いようで、 $T = 50 \sim 200$  では  $\text{tr}(\widehat{G + \frac{1}{5}G^+})$  の方が  $\text{tr}(\hat{G})$  より小さくなり、命題 3 とは逆に  $f_1 f_1' \overline{\Delta y}$  の方が効率的になってしまっている。今回のように、 $m - r$  の値が 1 と小さい時はこの問題はそう大きな問題ではないようである。

### 5.3 ケース 2 の実験結果

trace MSE ratio は表 5 のようになった。この表から、

- ほぼ全ての  $T$  と  $h$  で Lee-Carter 予測が最も悪い
- VEC 予測は  $T = 50, 100$  ではかなり悪いものの  $T = 200, 500$  では最も良い
- ほぼ全ての  $T$  と  $h$  で MTV 予測と修正 MTV 予測が比較的良い
- 全ての  $h$  で、Lee-Carter 全主成分予測は MTV 予測と修正 MTV 予測より概ね悪いが、Lee-Carter 予測よりは良い

といったことが解る。

以下、各予測の精度についてコメントする。

#### VEC 予測

表 5: trace MSE ratio (ケース 2)

$T = 50$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	1.73	1.93	1.98	2.02	1.99	1.89	1.83	1.87	1.91	1.98
ratio(MTV)	1.06	1.02	0.99	0.98	0.98	0.99	1.02	1.04	1.05	1.06
ratio(mMTV)	1.06	1.02	0.99	0.98	0.98	0.98	0.99	0.99	0.99	0.99
ratio(LC)	2.17	1.54	1.34	1.26	1.20	1.11	1.06	1.05	1.04	1.04
ratio(LCA)	1.04	1.00	0.98	0.97	0.97	0.97	0.98	0.99	1.00	1.00

$T = 100$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.86	0.97	1.02	1.06	1.09	1.13	1.10	1.10	1.09	1.08
ratio(MTV)	1.00	0.96	0.93	0.93	0.93	0.96	0.99	1.00	1.02	1.03
ratio(mMTV)	1.00	0.95	0.93	0.92	0.92	0.94	0.97	0.97	0.98	0.98
ratio(LC)	3.37	2.19	1.81	1.63	1.52	1.29	1.17	1.12	1.10	1.09
ratio(LCA)	0.99	0.96	0.95	0.94	0.94	0.96	0.98	0.99	0.99	1.00

$T = 200$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.67	0.77	0.82	0.85	0.87	0.94	0.97	0.98	0.99	0.99
ratio(MTV)	0.94	0.90	0.88	0.88	0.89	0.93	0.97	0.99	1.00	1.01
ratio(mMTV)	0.93	0.90	0.88	0.88	0.88	0.92	0.95	0.97	0.97	0.98
ratio(LC)	5.86	3.58	2.80	2.42	2.17	1.67	1.38	1.27	1.21	1.18
ratio(LCA)	0.94	0.92	0.90	0.90	0.91	0.94	0.97	0.98	0.99	0.99

$T = 500$										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.58	0.67	0.71	0.74	0.76	0.84	0.91	0.94	0.95	0.96
ratio(MTV)	0.91	0.87	0.86	0.86	0.86	0.91	0.96	0.98	0.99	1.00
ratio(mMTV)	0.91	0.87	0.85	0.85	0.86	0.91	0.95	0.97	0.97	0.98
ratio(LC)	13.26	7.68	5.73	4.77	4.16	2.82	2.02	1.71	1.55	1.44
ratio(LCA)	0.91	0.87	0.86	0.86	0.86	0.91	0.95	0.97	0.98	0.98

VEC 予測はパラメーター数が  $m^2$  のオーダーで増えるので  $m = 30$  では十分な  $T$  が無いと大幅に精度が悪化する。しかし、 $T$  が大きければ正しい定式化で予測をしているので stochastic コンポーネントの予測が非常に良くなる。これが  $h = 1 \sim 10$  の予測の精度を上げていると考えられる。長期予測に重要な枠割を果たすドリフト項については表 6 にあるように MSE が大きくなってしまっているが、 $h = 1 \sim 10$  の予測が良いため  $h = 10$  以降も大幅に予測が悪くなることはないようである。

### MTV 予測

stochastic コンポーネントの予測が上手く行き、 $h = 1 \sim 10$  の精度が良いと考えられる。ただ、表 7 を見ると、 $m = 30$  では少なくとも  $T = 200$  は無いと正しい  $I(1)$  主成分の数 (つまり 4 個)、を検知できないことが解るので、共和分制約の掛かり方は  $m = 3$  の時より弱いと思われる。また、表 6 にあるように、 $\hat{\mu}_{trend}$  は MSE が大きいので、これが  $h = 10$  以降の精度を下げていると思われる。ただ、 $T$  が小さくても VEC 予測のように大幅に精度が悪化することはなかった。

### 修正 MTV 予測

stochastic コンポーネントの予測は MTV 予測と同じなので、 $T$  が大きければ共和分制約が掛かり  $h = 1 \sim 10$  での精度が良くなる。ドリフトについては、表 6 にあるように  $\overline{\Delta y}$  の MSE は小さいため、 $h = 10$  以降の予測の performance は良い。

### Lee-Carter 予測

ケース 1 と同様に、単系列 ARIMA 予測より精度が良くなることはなかった。また、 $T$  が大きいほど精度が悪化した。

### Lee-Carter 全主成分予測

ケース 1 と同様に全体では可もなく不可もないといった精度だが、 $T = 50$  では一部で MTV 予測や修正 MTV 予測を上回る精度になっている。これは、Lee-Carter 全主成分予測の短所である共和分制約が掛からないということが  $T = 50$  では MTV 予測や修正 MTV

表 6: 各ドリフト推定量の MSE の合計 (ケース 2)

$T$	$\overline{\Delta y}$	$\hat{\mu}_{ML}$	$\hat{\mu}_{trend}$	$f_1 f_1' \overline{\Delta y}$
50	0.654	3.529	0.751	0.671
100	0.317	2.098	0.367	0.331
200	0.159	0.576	0.188	0.167
500	0.066	0.076	0.079	0.070

Note: ドリフトの推定量は  $(30 \times 1)$  のベクトルになるので、各要素の MSE を計算してそれを合計した数値を載せている。

表 7: MTV 予測で判定された  $I(1)$  主成分の数 (ケース 2)

$T \setminus \#\{I(1)\}$	10	9	8	7	6	5	4	3	2
50	1.06	10.30	33.88	38.68	14.16	1.88	0.04	0	0
100	0	0	0.06	0.02	20.30	48.98	26.48	1.70	0
200	0	0	0	0	0	0	86.16	13.58	0.26
500	0	0	0	0	0	0	83.48	16.34	0.18

Note: この表の見方は 3 と同じである。値が 0 だったところは省略している。四捨五入のため合計が 100 にならない個所がある。

予測にもほぼ当てはまり、MTV 予測と修正 MTV 予測の優位性が崩れたためと思われる。

予測精度の議論の次に、命題 3 で示した  $f_1 f_1' \overline{\Delta y}$  の非効率性について考察する。今回の DGP では  $\text{tr}(\Gamma) \doteq 30.80$ 、 $\text{tr}(\Gamma + \frac{1}{5}\Gamma^+) \doteq 32.87$  であり増加率は約 6.7% である。表 8 では、 $\text{tr}(\Gamma)$  と  $\text{tr}(\Gamma + \frac{1}{5}\Gamma^+)$  を実験で得た 5000 個の  $\overline{\Delta y}$  と  $f_1 f_1' \overline{\Delta y}$  から推定した結果を載せている。 $\text{tr}(\hat{G})$ 、 $\text{tr}(\widehat{G + \frac{1}{5}G^+})$  共に真の値への収束は遅いようだが、全ての  $T$  で  $\text{tr}(\widehat{G + \frac{1}{5}G^+})$  の方が  $\text{tr}(\hat{G})$  より大きくなり、命題 3 で示した非効率性が観測されている。ただ、ratio(LCA) を見る限りでは、予測精度に悪影響を与えるほどの非効率性ではなかったようである。

## 5.4 実験結果のまとめ

本節の実験結果からは、全体的に修正 MTV 予測が最も精度が高く、次いで MTV 予測、3 番目に良いのが Lee-Carter 全主成分予測であることが示唆される。Lee-Carter 予

表 8:  $\overline{\Delta y}$  と  $f_1 f_1' \overline{\Delta y}$  の分散の比較 (ケース 2)

$\text{tr}(\Gamma)$	$T$	$\text{tr}(\hat{G})$	$\text{tr}(\Gamma + \frac{1}{5}\Gamma^+)$	$T$	$\text{tr}(\widehat{G + \frac{1}{5}G^+})$
30.80	50	32.57	32.87	50	33.43
	100	31.47		100	32.87
	200	31.18		200	32.84
	500	31.62		500	33.55

Note:  $\hat{G}$  は 5000 個の  $\overline{\Delta y}$  から計算した  $(30 \times 30)$  の標本共分散行列に  $T$  を掛けたもので、 $\widehat{G + \frac{1}{5}G^+}$  は 5000 個の  $f_1 f_1' \overline{\Delta y}$  から計算した  $(30 \times 30)$  の標本共分散行列に  $T$  を掛けたものである。

測は単系列 ARIMA 予測より精度が悪く、今回比較した予測法の中で最も精度が悪かった。VEC 予測は、正しい特定化を既知として計算したので全体的にはパフォーマンスが良かったが、 $m = 30$ 、 $T = 50$  のようにパラメーター数に対して  $T$  が小さいと精度が大幅に悪化した。

次節で示すように、実際の死亡率のデータでは  $m$  が大きく  $T$  が小さいので、VEC 予測のようにパラメーター数が多い予測法は適切ではないと思われる。MTV 予測、修正 MTV 予測、Lee-Carter 予測、Lee-Carter 全主成分予測は VEC 予測に比べてパラメーター数が少なく、本節の実験でも  $m$  が大きく  $T$  が小さい場合でも精度が大幅に悪化することはなかったので、実際の死亡率の予測に役立つだろう。

なお、(14) 式で示したように、 $I(1)$  過程の長期予測についてはドリフトを捉えている限りはどの予測も同程度の精度となる。今回の実験で  $h = 50$  では全ての trace MSE ratio が概ね 1 になっているのはこのため、どの予測法も精度に差があるのは短期～中期予測であることに注意が必要である。

## 6 実証例

本節では、これまでに提案した各予測法で実際の死亡率を予測し、外挿予測の精度を比較する。MTV 予測、修正 MTV 予測、Lee-Carter 予測、Lee-Carter 全主成分予測、単系列 ARIMA 予測の計算法は 5.1 節で説明した方法と同じである。VEC 予測については、VAR の次数は 1、確定項の形は 5.1 節でのモンテカルロ実験と同じ形とするが、共和分ランク  $r$  が未知であるので Chigira (2008) の共和分検定を使って求めている。

## 6.1 日本のデータ

Human Mortality Database よりダウンロードした日本人の死亡率データを使って予測を行い精度を調べる。データは、0 歳 ~109 歳の各年齢の死亡率と、0 歳 ~109 歳を 23 のグループにまとめた死亡率<sup>11</sup>という 2 種類を考える。この 2 種類のデータが男女別にあるのでデータは合計で 4 種類ある。全てのデータは 1947 年 ~2009 年の年次データで  $T = 63$  である。 $T$  が小さいので、1947 年 ~2004 年 ( $T = 58$ ) のデータでモデルの推定と予測を行い、2005 年 ~2009 年のデータで  $h = 5$  の短期予測について外挿予測の精度を計算する。

以下、データの種類毎に分析結果を述べる。

### 各年齢の死亡率 (男性)

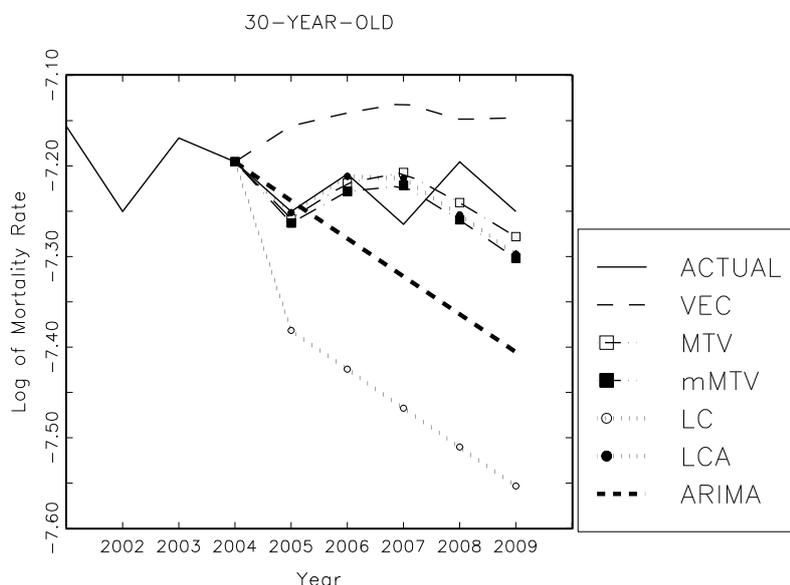
図 1 で示した、0 歳 ~109 歳の各年齢のデータで分析を行う。全ての年齢を 1 つの共和分モデルで表すと、 $m = 110$ 、 $T = 58$  となり、 $m$  が  $T$  より大きくなる。この時には VEC 予測は計算できないが、MTV 予測、修正 MTV 予測、Lee-Carter 予測、Lee-Carter 全主成分予測は計算できる。ただ、MTV 予測と修正 MTV 予測を計算する際には、(23) 式の  $\check{Y}\check{Y}$  が  $(110 \times 110)$  行列となるがランクが 58 なので第 1 ~ 第 58 主成分までで予測をする<sup>12</sup>。また、Lee-Carter 全主成分予測を計算する際には、(1) 式の  $\check{Y}\check{Y}$  が  $(110 \times 110)$  行列となるがランクが 58 なので第 1 ~ 第 58 主成分までで予測をする<sup>13</sup>。こうした  $m$  が非常に大きいモデルは、計算は可能ではあるがパラメータ推定の精度に問題があるかもしれないので、より  $m$  の小さいモデルの推定も行った。具体的には、0 歳 ~29 歳の  $m = 30$  のモデル、30 歳 ~59 歳の  $m = 30$  のモデル、60 歳 ~109 歳の  $m = 50$  の 3 つのモデルである。これらのモデルでは VEC 予測も可能であるが、60 歳 ~109 歳の  $m = 50$  のモデルでは  $m$  に比べて  $T$  が十分大きくないことが原因と思われる数値的な計算エラーが出たため VEC 予測は行っていない。trace MSE ratio は表 9 のようになった。この表からは、MTV 予測と修正 MTV 予測がほぼ同じ精度で良く work していることが解る。Lee-Carter 予測は 5 節でのモンテカルロ実験の結果と同じく精度がかなり悪かったが、Lee-Carter 全主成分予測

<sup>11</sup>0 歳は単独で 1 グループ、1 歳 ~4 歳で 1 グループ、以降は 5 歳 ~9 歳というように 5 歳毎で 1 グループとなる。最後のグループは 105 歳 ~109 歳であり、合計 23 のグループになる。

<sup>12</sup>実際には、第 57 主成分と第 58 主成分が  $10^{-7}$  程度の非常に小さい値になり計算が困難だったので第 56 主成分までを使った。

<sup>13</sup>実際には、第 58 主成分が  $10^{-7}$  程度の非常に小さい値になり計算が困難だったので第 57 主成分までを使った。

図 2: 30 歳日本人男性の死亡率の予測値と実績値



Note: 0 歳 ~109 歳の各年齢の死亡率の内、30 歳 ~59 歳を取り上げて  $m = 30$  のモデルを推定した時の、30 歳の予測値と実績値である。

はそう悪くはなかった。VEC 予測の精度は悪かったが、これは  $T$  が小さいためパラメータ推定の精度に問題があったためと思われる。予測値の具体的な値を示すため、上記の分析での典型的な例として、30 歳 ~59 歳のモデルにおける 30 歳の予測値と実績値を図 ref30jot にグラフ化した。この図からは、MTV 予測、修正 MTV 予測、Lee-Carter 全主成分予測が実績値に近い一方、VEC 予測、Lee-Carter 予測、単系列 ARIMA 予測は実績値から外れていることが解る。なお、 $I(1)$  主成分の数は表 9 に示したように 3、~5 個であり  $m$  に比べてかなり小さいので、共和分ランクがかなり高いことがうかがえる。

### グループ化された年齢の死亡率（男性）

0 歳 ~109 歳を 23 のグループにまとめたデータで分析を行う。データのプロットは図 3 のようになる。 $m = 23$ 、 $T = 58$  なので全ての予測量は問題無く計算でき、trace MSE ratio は表 10 のようになった。この結果は概ね表 9 と同じであり、MTV 予測と修正 MTV 予測が最も精度が良かった。また、 $I(1)$  主成分の数がかなり少ないので、共和分ランクがかなり高いと思われる。

表 9: trace MSE ratio (日本、男性、各年齢)

0 歳 ~109 歳 ( $m = 110$ 、 $\#\{I(1)\} = 5$ )

$h$	1	2	3	4	5
ratio(VEC)	.	.	.	.	.
ratio(MTV)	0.67	0.58	0.97	0.48	0.67
ratio(mMTV)	0.70	0.58	0.87	0.48	0.64
ratio(LC)	2.89	4.23	4.32	2.93	3.00
ratio(LCA)	1.40	1.62	1.06	0.86	0.93

0 歳 ~29 歳 ( $m = 30$ 、 $\#\{I(1)\} = 3$ )

$h$	1	2	3	4	5
ratio(VEC)	0.93	2.50	1.51	1.50	1.86
ratio(MTV)	0.62	0.58	0.64	0.58	0.68
ratio(mMTV)	0.67	0.72	0.77	0.82	0.92
ratio(LC)	2.00	2.57	2.63	2.01	1.93
ratio(LCA)	0.68	0.70	0.76	0.82	0.89

30 歳 ~59 歳 ( $m = 30$ 、 $\#\{I(1)\} = 3$ )

$h$	1	2	3	4	5
ratio(VEC)	1.73	2.22	1.51	1.33	1.06
ratio(MTV)	1.04	0.74	0.73	0.50	0.25
ratio(mMTV)	1.13	0.64	0.53	0.36	0.26
ratio(LC)	3.39	4.08	4.83	3.27	3.00
ratio(LCA)	1.07	0.80	1.06	0.67	0.32

60 歳 ~109 歳 ( $m = 50$ 、 $\#\{I(1)\} = 4$ )

$h$	1	2	3	4	5
ratio(VEC)	.	.	.	.	.
ratio(MTV)	0.73	0.64	0.63	0.53	0.61
ratio(mMTV)	0.78	0.64	0.62	0.64	0.60
ratio(LC)	1.09	1.98	2.52	1.56	2.15
ratio(LCA)	0.85	0.74	1.02	0.96	0.86

Note: 0 歳 ~109 歳の各年齢の死亡率で分析している。データで " $\#\{I(1)\}$ " は、 $m$  個の主成分の内、Phillips-Perron 検定で  $I(1)$  と判定された主成分の数である。

図 3: 日本人男性の死亡率の対数（グループ化された年齢）

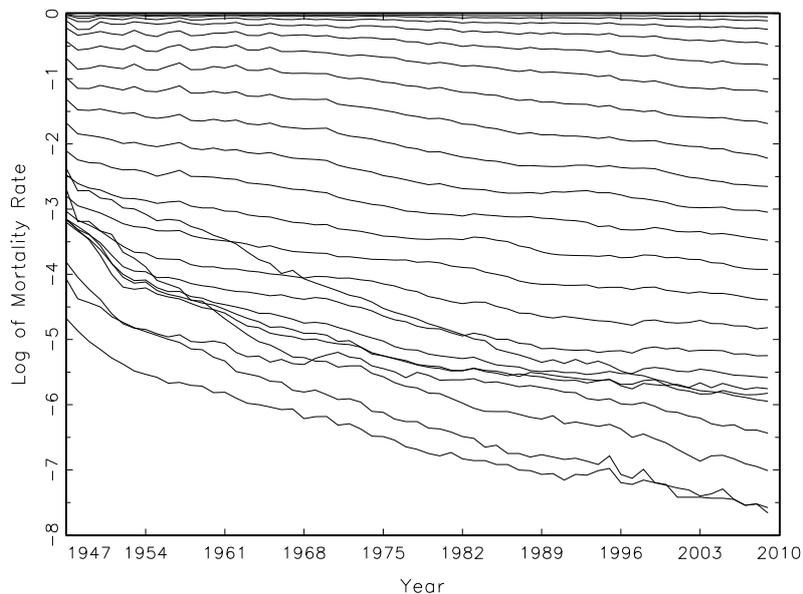


表 10: trace MSE ratio（日本、男性、グループ化された年齢）

$m = 23$ 、 $\#\{I(1)\} = 4$

$h$	1	2	3	4	5
ratio(VEC)	2.85	3.23	3.41	2.79	2.80
ratio(MTV)	0.56	0.40	0.84	0.33	0.52
ratio(mMTV)	0.62	0.45	0.67	0.32	0.47
ratio(LC)	4.94	6.68	8.09	4.16	3.91
ratio(LCA)	0.87	0.86	1.04	1.03	1.08

Note: 0歳を単独で1グループ、1歳～4歳を1グループ、以降は5歳～9歳というように5歳毎を1グループとし、最後のグループを105歳～109歳とした、合計23のグループ化された年齢の死亡率で分析している。“ $\#\{I(1)\}$ ”は、 $m$ 個の主成分の内、Phillips-Perron検定で $I(1)$ と判定された主成分の数である。

表 11: trace MSE ratio (日本、女性、各年齢)

30 歳 ~59 歳 ( $m = 30$ ,  $\#\{I(1)\} = 2$ )

$h$	1	2	3	4	5
ratio(VEC)	1.90	1.65	2.25	2.33	2.56
ratio(MTV)	0.42	0.37	0.19	0.15	0.10
ratio(mMTV)	0.43	0.37	0.19	0.18	0.13
ratio(LC)	1.65	1.13	1.20	1.28	1.19
ratio(LCA)	0.37	0.31	0.18	0.20	0.17

Note: 0 歳 ~109 歳の各年齢の死亡率の内、30 歳 ~59 歳を取り上げて分析している。“ $\#\{I(1)\}$ ” は、 $m$  個の主成分の内、Phillips-Perron 検定で  $I(1)$  と判定された主成分の数である。

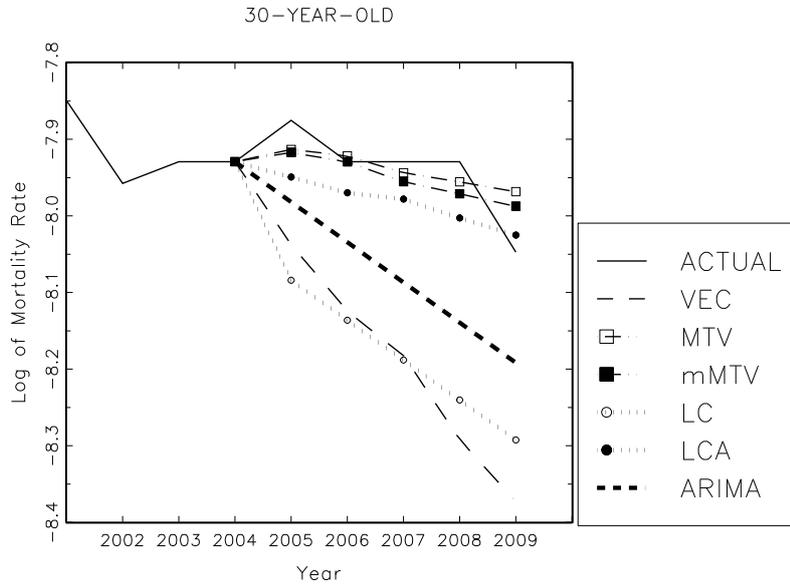
### 各年齢の死亡率 (女性)

男性のデータと同様に、0 歳 ~110 歳 ( $m = 110$ )、0 歳 ~29 歳 ( $m = 30$ )、30 歳 ~59 歳 ( $m = 30$ )、60 歳 ~109 歳 ( $m = 50$ ) の 4 つのモデルで予測を行った。紙幅の制限のため、30 歳 ~59 歳のモデルだけの trace MSE ratio を表 11 に示す。この表からは、男性の場合と同じく、MTV 予測、修正 MTV 予測、Lee-Carter 全主成分予測の精度が良く、VEC 予測と Lee-Carter 予測の精度が悪いことが解る。他の年齢のモデルにおける trace MSE ratio も概ね男性の場合と同じであった。男性の場合と同様に、30 歳 ~59 歳のモデルにおける 30 歳の予測値と実績値を図 4 に示す。この図からは、MTV 予測、修正 MTV 予測がかなり実績値に近く、Lee-Carter 全主成分予測も概ね実績値を追っていることが解る。一方、VEC 予測、Lee-Carter 予測、単系列 ARIMA 予測は実績値から外れている。なお、 $I(1)$  主成分の数は 2 個であり、男性の場合より少なかった。

### グループ化された年齢の死亡率 (女性)

trace MSE ratio は表 12 のようになった。この結果は、VEC 予測がやや良いことを除いては、概ね表 10 と同じであり、MTV 予測と修正 MTV 予測が最も精度が良かった。

図 4: 30 歳日本人女性の死亡率の予測値と実績値



Note: 0 歳 ~ 109 歳の各年齢の死亡率の内、30 歳 ~ 59 歳を取り上げて  $m = 30$  のモデルを推定した時の、30 歳の予測値と実績値である。

表 12: trace MSE ratio (日本、女性、グループ化された年齢)

$m = 23, \#\{I(1)\} = 4$

$h$	1	2	3	4	5
ratio(VEC)	1.43	0.37	1.43	0.84	0.68
ratio(MTV)	0.41	0.34	0.36	0.39	0.34
ratio(mMTV)	0.46	0.40	0.44	0.49	0.46
ratio(LC)	8.79	5.35	6.29	3.98	4.03
ratio(LCA)	0.64	0.43	0.40	0.61	0.79

Note: 0 歳を単独で 1 グループ、1 歳 ~ 4 歳を 1 グループ、以降は 5 歳 ~ 9 歳というように 5 歳毎を 1 グループとし、最後のグループを 105 歳 ~ 109 歳とした、合計 23 のグループ化された年齢の死亡率で分析している。“ $\#\{I(1)\}$ ” は、 $m$  個の主成分の内、Phillips-Perron 検定で  $I(1)$  と判定された主成分の数である。

## 6.2 スウェーデンのデータ

日本のデータでは  $T$  が小さく短期予測の評価しかできなかつたので、Human Mortality Database において  $T$  が最も大きいスウェーデンのデータを使い長期予測の評価を行う。日本の場合と同様、0 歳 ~109 歳の各年齢の死亡率と、0 歳 ~109 歳を 23 のグループにまとめた死亡率を男女別に扱う。全てのデータは 1751 年 ~2010 年の年次データで  $T = 260$  である。1751 年 ~1960 年 ( $T = 210$ ) のデータでモデルの推定と予測を行い、1961 年 ~2010 年のデータで  $h = 50$  の予測について外挿予測の精度を計算する。

以下、データの種類毎に分析結果を述べる。

### 各年齢の死亡率（男性）

日本の場合と同様に、0 歳 ~110 歳 ( $m = 110$ )、0 歳 ~29 歳 ( $m = 30$ )、30 歳 ~59 歳 ( $m = 30$ )、60 歳 ~109 歳 ( $m = 50$ ) の 4 つのモデルで予測を行った。紙幅の制限のため、30 歳 ~59 歳のモデルだけの trace MSE ratio を表 13 に示す。この表からは、日本の場合と同じく、概ね MTV 予測と修正 MTV 予測の精度が高いことが解るが、その優位性は日本の場合より低い。この理由としては、図 5 よりうかがえるように、1900 年頃を境にトレンドの傾きに構造変化が疑われること、第 1 次世界大戦で死亡率が大きく上がるという外れ値があることが考えられる<sup>14</sup>。また、Human Mortality Database で述べられているように、スウェーデンのデータでは 1751 年 ~1860 年のデータの質が悪いことも影響しているだろう。Lee-Carter 予測は短期予測では精度がかなり悪いが、長期予測では他の予測と概ね同じ程度の精度となっている。これは 5 節の実験結果や (14) 式と consistent である。日本と同様に 30 歳 ~59 歳のモデルにおける 30 歳の予測値と実績値を図 6 に示したが、短期 ~ 中期ではやはり Lee-Carter 予測が大きく外れる。なお、0 歳 ~110 歳と 0 歳 ~29 歳の trace MSE ratio も概ね表 13 と同じであったが、60 歳 ~109 歳においては MTV 予測と修正 MTV 予測のパフォーマンスが悪く、Lee-Carter 予測を下回ることもあった。

### グループ化された年齢の死亡率（男性）

0 歳 ~109 歳を 23 のグループにまとめたデータで分析を行う。紙幅の制限上 trace MSE

<sup>14</sup>本稿の予測法はいずれもトレンドに構造変化が無いことが前提となっている。また、今回の分析では第 1 次世界大戦時のデータに何も処置をせずそのまま使っている。

図 5: スウェーデン人男性の死亡率の対数（各年齢、0 歳～109 歳）

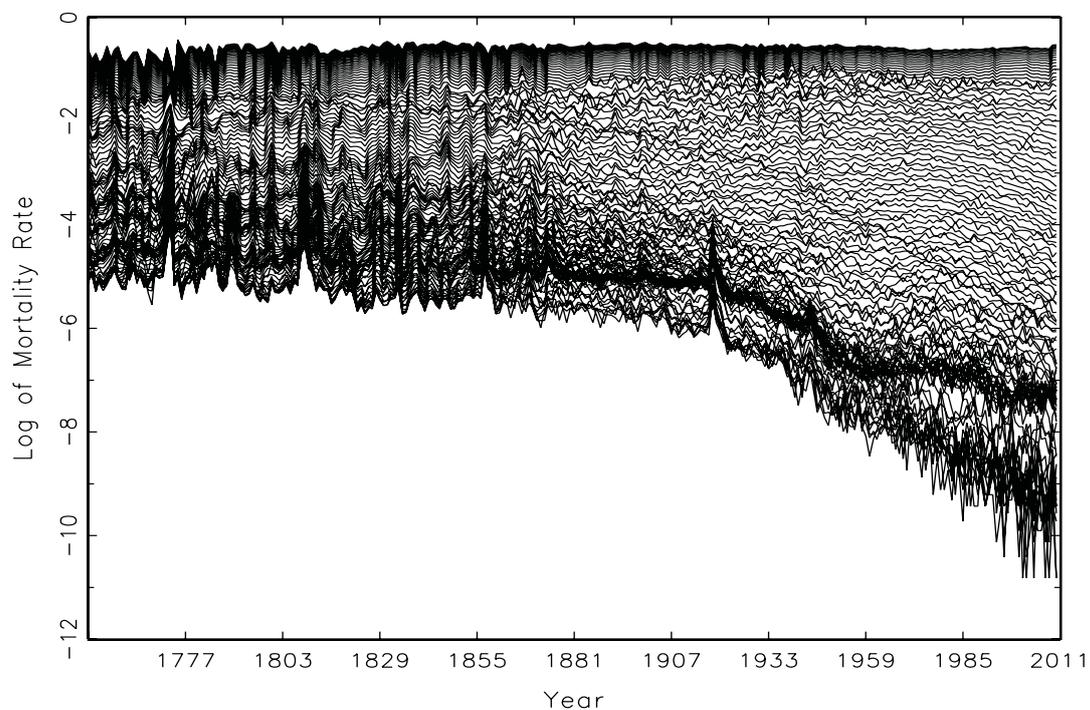


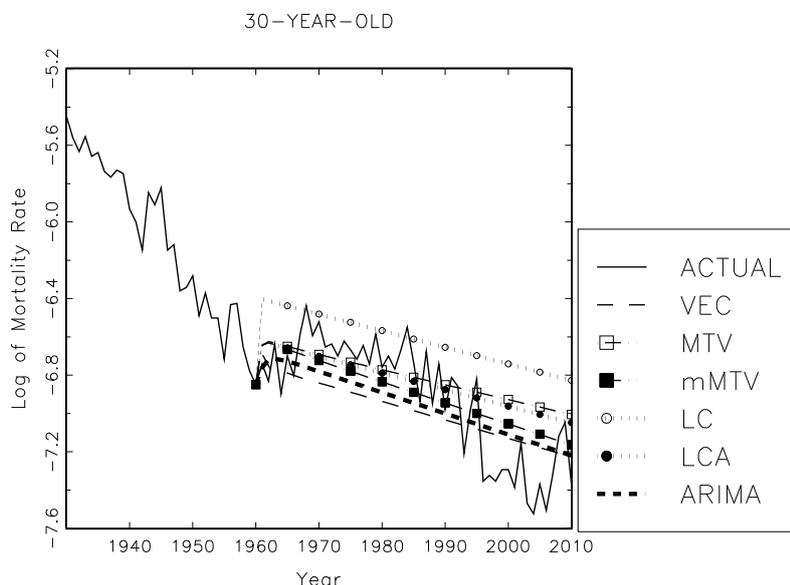
表 13: trace MSE ratio（スウェーデン、男性、各年齢）

30 歳～59 歳 ( $m = 30$ 、 $\#\{I(1)\} = 2$ )

$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	1.80	0.91	0.99	1.28	1.81	2.34	1.55	1.86	0.97	0.98
ratio(MTV)	0.80	1.04	0.97	1.12	0.92	1.13	1.09	0.82	0.96	0.97
ratio(mMTV)	0.80	1.02	0.95	1.09	0.96	1.23	1.18	0.93	0.86	0.91
ratio(LC)	2.80	3.34	2.65	3.54	2.26	2.81	1.99	2.13	0.90	0.80
ratio(LCA)	0.74	1.05	0.98	1.16	0.97	1.29	1.23	0.90	0.82	0.86

Note: 0 歳～109 歳の各年齢の死亡率の内、30 歳～59 歳を取り上げて分析している。“ $\#\{I(1)\}$ ” は、 $m$  個の主成分の内、Phillips-Perron 検定で  $I(1)$  と判定された主成分の数である。

図 6: 30 歳スウェーデン人男性の死亡率の予測値と実績値



Note: 0 歳 ~109 歳の各年齢の死亡率の内、30 歳 ~59 歳を取り上げて  $m = 30$  のモデルを推定した時の、30 歳の予測値と実績値である。

ratio の表は省略するが、全体的には表 13 と概ね同じだった。

### 各年齢の死亡率（女性）

0 歳 ~110 歳 ( $m = 110$ )、0 歳 ~29 歳 ( $m = 30$ )、30 歳 ~59 歳 ( $m = 30$ )、60 歳 ~109 歳 ( $m = 50$ ) の 4 つのモデルで予測を行った。紙幅の制限のため、30 歳 ~59 歳のモデルだけの trace MSE ratio を表 14 に示す。この表からは、VEC 予測、MTV 予測、修正 MTV 予測、Lee-Carter 全主成分予測の精度が良く、Lee-Carter 予測の精度が悪いことが解る。スウェーデンのデータは  $T$  が大きいため、VEC 予測が良く work する場合もあると思われる。日本と同様に 30 歳 ~59 歳のモデルにおける 30 歳の予測値と実績値を図 6 に示したが、短期 ~ 中期ではやはり Lee-Carter 予測が大きく外れる。なお、0 歳 ~110 歳と 0 歳 ~29 歳の trace MSE ratio も概ね表 14 と同じであったが、60 歳 ~109 歳においては MTV 予測と修正 MTV 予測より Lee-Carter 予測の方が精度が高かった。男性の場合と同様に、30 歳 ~59 歳のモデルにおける 30 歳の予測値と実績値を図 7 に示す。この図からは、Lee-Carter 予測が実績値から外れているが、それ以外の予測は概ね実績値を追っていることが解る。MTV 予測もやや実績値から外れているが、これはドリフトの推定に

表 14: trace MSE ratio (スウェーデン、女性、各年齢)

30 歳 ~59 歳 ( $m = 30$ 、 $\#\{I(1)\} = 2$ )										
$h$	1	2	3	4	5	10	20	30	40	50
ratio(VEC)	0.57	0.46	0.64	0.65	0.87	0.92	0.89	1.08	1.07	1.04
ratio(MTV)	0.84	0.71	0.78	0.72	0.92	0.88	0.80	1.17	1.33	1.19
ratio(mMTV)	0.82	0.69	0.75	0.68	0.90	0.82	0.77	0.84	0.93	0.91
ratio(LC)	1.99	1.35	1.64	1.57	1.65	1.38	0.99	0.86	1.16	0.78
ratio(LCA)	0.83	0.64	0.76	0.64	0.84	0.73	0.61	0.77	0.91	0.85

Note: 0 歳 ~109 歳の各年齢の死亡率の内、30 歳 ~59 歳を取り上げて分析している。“ $\#\{I(1)\}$ ” は、 $m$  個の主成分の内、Phillips-Perron 検定で  $I(1)$  と判定された主成分の数である。

トレンドへ回帰した OLS 推定量を使っているためかもしれない。

### グループ化された年齢の死亡率 (女性)

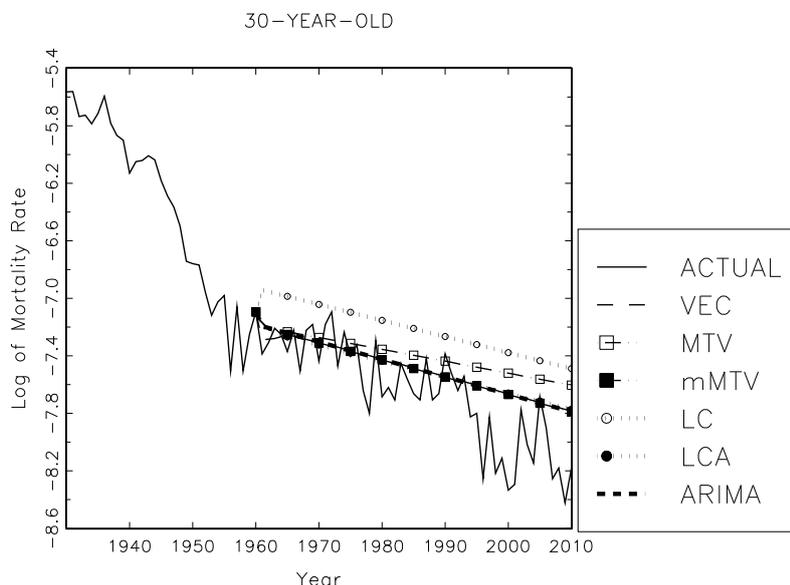
0 歳 ~109 歳を 23 のグループにまとめたデータで分析を行う。紙幅の制限上 trace MSE ratio の表は省略するが、全体的には表 14 と概ね同じだった。

## 6.3 実証例のまとめ

本節の実証結果では、全体的に MTV 予測と修正 MTV 予測の精度がかなり高く、次いで良いのが Lee-Carter 全主成分予測であった。VEC 予測は、 $T$  の大きいスウェーデンのデータで精度が高くなる場合もあったが、全体的には単系列 ARIMA 予測より精度が低かった。これは、VEC 予測では推定するパラメーターの数が多いにもかかわらず  $T$  が十分大きくなかったためと考えられる。Lee-Carter 予測は概ね単系列 ARIMA 予測より精度が悪く、今回比較した予測法の中で最も精度が悪かった。ただ、Lee-Carter 予測の精度が悪いのは概ね  $h = 30$  程度までで、 $h = 30$  以上では他の予測とほぼ同じ精度だった。こうした結果は 5 節の結果と consistent である。

日本のデータでは  $T$  が小さいにもかかわらず MTV 予測と修正 MTV 予測の精度がかなり高かった。これは日本のデータの標本期間が 1947 年 ~2009 年という安定した時期であったためと考えられ、 $T$  が小さくても構造が安定した時期のデータならば MTV 予測

図 7: 30 歳スウェーデン人女性の死亡率の予測値と実績値



Note: 0 歳 ~109 歳の各年齢の死亡率の内、30 歳 ~59 歳を取り上げて  $m = 30$  のモデルを推定した時の、30 歳の予測値と実績値である。

と修正 MTV 予測は有用であることを示唆している。一方、スウェーデンのデータでは  $T$  は大きいですが MTV 予測と修正 MTV 予測の他の予測に対する優位性は日本ほどではなかった。これは、1751 年 ~2010 年という長い標本期間の内にトレンドに構造変化が疑われることや 18 世紀 ~19 世紀の古いデータの質が悪いことが影響していると思われる。第 1 次世界大戦による死亡率の大幅な上昇をどう扱うかも難しい問題である。

## 7 まとめ

本稿では、死亡率の対数値にを共和分過程としてモデル化し、Lee-Carter 予測について時系列分析の理論上の性質を調べた。その結果、Lee-Carter 予測は  $I(1)$  過程の長期予測に大きな役割を果たすドリフト項を捉えるという長所を持つものの、情報のロスが大きく共和分制約が課せられずドリフトの推定も非効率的になるといった短所があることが解った。さらに、情報のロスは  $T$  と共に大きくなり、Lee-Carter 予測は  $T$  が大きいほど予測精度が悪化することが示された。そして、こうした短所を補うため、MTV 予測、修正 MTV 予測、Lee-Carter 全主成分予測を提案した。

モンテカルロ実験や日本とスウェーデンのデータを使った実証例では、全体的には修

正 MTV 予測が最も精度が高いことが確かめられた。MTV 予測も概ね良く機能するが、ドリフト推定量が非効率的なため精度が落ちる場合もあった。Lee-Carter 全主成分予測の精度は低くはなく、修正 MTV 予測と同等に機能する場合もあったが、全体としては MTV 予測や修正 MTV 予測に劣った。これは、ドリフト推定量の非効率性や共和分制約が掛からないといった問題があるためだと思われる。Lee-Carter 予測は本稿で比較の対象とした予測法の中では最も精度が悪く、特に短期～中期予測では大幅に精度が悪化した。ただ、長期予測では他の予測法と概ね同程度の精度だった。なお、本稿では共和分過程の伝統的な予測法である VEC 予測の精度も調べたが、 $T$  が  $m$  より十分大きい時には精度が高いものの、 $T$  が小さいと精度が極端に悪化するという欠点が明らかになった。死亡率のデータは  $m$  に比べて  $T$  が十分大きくないことが多いので、VEC 予測は薦められないだろう。

最後に、Lee-Carter 予測は短期～中期予測では理論的に単系列 ARIMA 予測にさえ劣ることを注意しておく。このことはモンテカルロ実験と実証例のほぼ全ての場合でも確かめられた。Lee-Carter 予測を改善するためには MTV 予測や修正 MTV 予測を用いるのが1つの理想的な選択だが、それが難しい場合には標準的な単系列 ARIMA 予測を用いるだけでも一定の改善になる。

## 参考文献

- 石井太 (2008): “近年のわが国の死亡動向に適合した将来生命表推計モデルの研究-年齢シフトモデルの開発-,” 人口問題研究, 64, 28-44.
- Christoffersen, P.F. and F.X. Diebold (1998): “Cointegration and Long-Horizon Forecasting,” *Journal of Business and Economic Statistics*, 16, 450-458.
- Chigira, H. (2008): “A Test of Cointegration Rank Based on Principal Component Analysis,” *Applied Economics Letters*, 15, 693-696.
- Chigira, H. and T. Yamamoto (2012): “The Effect of Estimating Parameters on Long-Term Forecasts for Cointegrated Systems,” *Journal of Forecasting*, 31, 344-360.
- Chigira, H. and T. Yamamoto (2009): “Forecasting in Large Cointegrated Processes,” *Journal of Forecasting*, 28, 631-650.

- Deaton, A. and C. Paxson (2004): “Mortality, Income, and Income Inequality Over Time in the Britain and the United States,” NBER Working Paper No. 8534.
- Giroi, F. and G. King (2007): “Understanding the Lee-Carter Mortality Forecasting Method,” Unpublished Manuscript.
- Harris, D. (1997): “Principal Components Analysis of Cointegrated Time Series,” *Econometric Theory*, 13, 529–557.
- Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at [www.mortality.org](http://www.mortality.org) or [www.humanmortality.de](http://www.humanmortality.de) (data downloaded on 22/08/2012).
- Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.
- Kariya, T. (1987): “MTV Model and Its Application to the Prediction of Stock Prices,” *Proceedings of the Second International Tampere Conference in Statistics*, 161–176.
- Keilman, N. (2003): “Types of Models for Projecting Mortality,” in Bengtsson, T. and N. Keilman eds. *Perspectives on Mortality Forecasting*, 19–27, Swedish National Social Insurance Board.
- Lee, R.D. and L.R. Carter (1992): “Modeling and Forecasting U.S. Mortality,” *Journal of the American Statistical Association*, 87, 659–671.
- Magnus, J.R. and H. Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley.
- Snell, A. (1999): “Testing for  $r$  versus  $r - 1$  Cointegrating Vectors,” *Journal of Econometrics*, 88, 151–191
- Stock, J.H. and M. Watson (1988): “Testing for Common Trends,” *Journal of the American Statistical Association*, 83, 1097–1107.

付録 命題の証明

命題1の証明

(12) 式を

$$y_t = \mu t + \xi_t + \varepsilon_t^*$$

と書き直す。ここで、 $\xi_t = C \sum_{s=1}^t \varepsilon_s$ 、 $\varepsilon_t^* = C^*(L)\varepsilon_t$  である。すると、 $\bar{\xi} = \frac{1}{T} \sum_{t=1}^T \xi_t$ 、 $\bar{\varepsilon}^* = \frac{1}{T} \sum_{t=1}^T \varepsilon_t^*$  として

$$\begin{aligned} \underset{(T \times m)}{\tilde{Y}} &= \begin{bmatrix} \mu' + \xi_1' + \varepsilon_1^{*'} - \mu' \frac{1}{T} \sum_{t=1}^T t - \bar{\xi}' - \bar{\varepsilon}^{*'} \\ \vdots \\ \mu'T + \xi_T' + \varepsilon_T^{*'} - \mu' \frac{1}{T} \sum_{t=1}^T t - \bar{\xi}' - \bar{\varepsilon}^{*'} \end{bmatrix} = \begin{bmatrix} \mu' + \xi_1' + \varepsilon_1^{*'} - \mu' \frac{T+1}{2} - \bar{\xi}' - \bar{\varepsilon}^{*'} \\ \vdots \\ \mu'T + \xi_T' + \varepsilon_T^{*'} - \mu' \frac{T+1}{2} - \bar{\xi}' - \bar{\varepsilon}^{*'} \end{bmatrix} \end{aligned} \quad (36)$$

と書ける。これより、

$$\tilde{Y}'\tilde{Y} = \mu\mu' \left( -\frac{1}{4}T(T+1)^2 + \frac{1}{6}T(T+1)(2T+1) \right) + o_p(T^3)$$

となるので、両辺を  $T^3$  で割って  $T \rightarrow \infty$  とすると

$$\tilde{Y}'\tilde{Y}/T^3 \xrightarrow{p} \frac{1}{12}\mu\mu' \equiv \dot{Y}$$

を得る。行列  $\dot{Y}$  は最大の固有値が  $\frac{1}{12}\mu'\mu$  でその他の固有値は0となるので、(1) 式の  $\lambda_1$  について

$$\lambda_1/T^3 \xrightarrow{p} \frac{1}{12}\mu'\mu \equiv \bar{\lambda}_1$$

となることが解る。  $\text{plim} f_1 = \bar{f}_1$  は

$$\dot{Y}\bar{f}_1 = \bar{\lambda}_1\bar{f}_1$$

を満たすベクトルなので、 $\bar{f}_1 = \pm \frac{\mu}{\sqrt{\mu'\mu}}$  となる。

命題2の証明

(15) 式と  $\overline{\Delta y} \xrightarrow{p} \mu$  より、スラツキーの定理を使って

$$f_1 f_1' \overline{\Delta y} \xrightarrow{p} \pm \frac{\mu}{\sqrt{\mu'\mu}} \left( \pm \frac{\mu}{\sqrt{\mu'\mu}} \right)' \mu = \mu$$

となる。

命題3の証明

この証明のために、次の補題を用いる。

**補題 1:**

$$f_1 - \bar{f}_1 \bar{f}'_1 f_1 = \bar{F}_{(m-1)} \left( \lambda_1 I_{m-1} - \bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{F}_{(m-1)} \right)^{-1} \left( \bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{f}_1 \right) \bar{f}'_1 f_1 \quad (37)$$

**証明:** この表現は Snell (1999) の p. 178 にある (P1.9) 式と本質的に同じである。Snell (1999) には (P1.9) 式の導出が詳述されているのでそれを参照されたい。

ここで、後のために、 $\lambda_1 I_{m-1} - \bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{F}_{(m-1)} = Z^*$ 、 $\bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{f}_1 = Z^\dagger$  と書いておく。(37) 式を使うと、さらに次の補題が導ける。

**補題 2:**

$$\sqrt{T}(f_1 f'_1 \Delta y - \mu) = \mu' f_1 \bar{F}_{(m-1)} (T^{-3} Z^*)^{-1} (T^{-5/2} Z^\dagger) \bar{f}'_1 f_1 + f_1 f'_1 \sqrt{T} \varepsilon^* \quad (38)$$

**証明:** (37) 式の両辺に  $\sqrt{T}$  を掛けると

$$\sqrt{T}(f_1 - \bar{f}_1 \bar{f}'_1 f_1) = \bar{F}_{(m-1)} (T^{-3} Z^*)^{-1} (T^{-5/2} Z^\dagger) \bar{f}'_1 f_1$$

となり、 $\lambda_1/T^3 \xrightarrow{p} \frac{1}{12} \mu' \mu$ 、 $\bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{F}_{(m-1)}/T^3 \xrightarrow{p} 0$ 、 $T^{-5/2} Z^\dagger$  は  $T \rightarrow \infty$  とするとある正規分布に収束する<sup>15</sup>ことから、この式の右辺は  $O_p(1)$  となることが解る。この式を

$$\begin{aligned} f_1 &= \bar{f}_1 \bar{f}'_1 f_1 + T^{-1/2} \bar{F}_{(m-1)} (T^{-3} Z^*)^{-1} (T^{-5/2} Z^\dagger) \bar{f}'_1 f_1 \\ &= \bar{f}_1 \bar{f}'_1 f_1 + Z^+ \end{aligned} \quad (39)$$

と書き直す。ここで、 $Z^+ = T^{-1/2} \bar{F}_{(m-1)} (T^{-3} Z^*)^{-1} (T^{-5/2} Z^\dagger) \bar{f}'_1 f_1 = O_p(T^{-1/2})$  である。

(39) 式より

$$\sqrt{T}(f_1 f'_1 \Delta y - \mu) = \sqrt{T} \left\{ \left( \frac{(\mu' f_1)^2}{(\mu' \mu)^2} \mu \mu' + \frac{(\mu' f_1)}{(\mu' \mu)} \mu Z^{+'} + \frac{(\mu' f_1)}{(\mu' \mu)} Z^+ \mu' + Z^+ Z^{+'} \right) (\mu + \varepsilon^*) - \mu \right\}$$

となり、 $Z^{+'} \mu = 0$  と  $Z^+ Z^{+'} = O_p(T^{-1})$  に注意して項を整理すると

$$\begin{aligned} \sqrt{T}(f_1 f'_1 \Delta y - \mu) &= \sqrt{T} \left\{ \left( \frac{(\mu' f_1)^2}{\mu' \mu} \mu + (\mu' f_1) Z^+ \right) + f_1 f'_1 \varepsilon^* - \mu \right\} \\ &= \sqrt{T} (\bar{f}'_1 f_1)^2 \mu + \sqrt{T} (\mu' f_1) Z^+ + f_1 f'_1 \sqrt{T} \varepsilon^* - \sqrt{T} \mu \end{aligned} \quad (40)$$

<sup>15</sup>このことはすぐ後に示す。

を得る。ここで、(39) 式より

$$f_1' f_1 = \frac{(\mu' f_1)^2}{(\mu' \mu)^2} \mu' \mu + \frac{2(\mu' f_1)}{(\mu' \mu)} \mu' Z^+ + Z^{+'} Z^+ = \frac{(\mu' f_1)^2}{(\mu' \mu)} + O_p(T^{-1}) = (\bar{f}_1' f_1)^2 + O_p(T^{-1})$$

となるが、 $f_1$  は固有ベクトルなので  $f_1' f_1 = 1$  であることに注意すると

$$(\bar{f}_1' f_1)^2 = 1 - O_p(T^{-1}) \quad (41)$$

と書ける。(41) を (40) に代入して項を整理すると

$$\sqrt{T}(f_1 f_1' \Delta y - \mu) = \sqrt{T}(\mu' f_1) Z^+ + f_1 f_1' \sqrt{T} \bar{\varepsilon}^*$$

となり、ここに  $Z^+ = T^{-1/2} \bar{F}_{(m-1)} (T^{-3} Z^*)^{-1} (T^{-5/2} Z^\dagger) \bar{f}_1' f_1$  を代入すれば (38) を得る。

(38) 式において  $T \rightarrow \infty$  とすれば求める結果が得られるが、(38) 式の内、 $T^{-5/2} Z^\dagger$  の収束先は obvious ではないので以下の補題で与えておく。

**補題 3:**

$$T^{-5/2} Z^\dagger \xrightarrow{L} \mu' \bar{f}_1 \bar{F}'_{(m-1)} \Theta \dot{S} \quad (42)$$

ここで、 $\Theta = C \Omega^{1/2}$ 、 $\Omega^{1/2}$  は  $\Omega$  の Cholesky 分解、 $\dot{S} = \int_0^1 k \left( W(k) - \int_0^1 W(l) dl \right) dk$ 、 $W(\cdot)$  は  $m$  次元標準ブラウン運動である。

**証明:** (36) 式を使うと

$$\tilde{Y} \bar{F}_{(m-1)} = \begin{bmatrix} \xi_1' \bar{F}_{(m-1)} + \varepsilon_1^* \bar{F}_{(m-1)} - \bar{\xi}' \bar{F}_{(m-1)} - \bar{\varepsilon}^* \bar{F}_{(m-1)} \\ \vdots \\ \xi_T' \bar{F}_{(m-1)} + \varepsilon_T^* \bar{F}_{(m-1)} - \bar{\xi}' \bar{F}_{(m-1)} - \bar{\varepsilon}^* \bar{F}_{(m-1)} \end{bmatrix}$$

$$\tilde{Y} \bar{f}_1 = \begin{bmatrix} \mu' \bar{f}_1 + \xi_1' \bar{f}_1 + \varepsilon_1^* \bar{f}_1 - \mu' \bar{f}_1 \frac{T+1}{2} - \bar{\xi}' \bar{f}_1 - \bar{\varepsilon}^* \bar{f}_1 \\ \vdots \\ \mu' \bar{f}_1 T + \xi_T' \bar{f}_1 + \varepsilon_T^* \bar{f}_1 - \mu' \bar{f}_1 \frac{T+1}{2} - \bar{\xi}' \bar{f}_1 - \bar{\varepsilon}^* \bar{f}_1 \end{bmatrix}$$

と書けるので、

$$\bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{f}_1 = \mu' \bar{f}_1 \bar{F}'_{(m-1)} \sum_{t=1}^T (\xi_t - \bar{\xi}) t + o_p(T^{5/2})$$

となる。よって、汎関数中心極限定理より

$$\bar{F}'_{(m-1)} \tilde{Y}' \tilde{Y} \bar{f}_1 / T^{5/2} \xrightarrow{L} \mu' \bar{f}_1 \bar{F}'_{(m-1)} \Theta \int_0^1 k \left( W(k) - \int_0^1 W(l) dl \right) dk$$

を得る。

(42) 式を使って (38) 式で  $T \rightarrow \infty$  とすると以下のようになる。

$$\begin{aligned} \sqrt{T}(f_1 f_1' \overline{\Delta y} - \mu) &\xrightarrow{L} \mu' f_1 \bar{F}_{(m-1)} \left( \frac{1}{12} \mu' \mu \right)^{-1} \left( \mu' \bar{f}_1 \bar{F}'_{(m-1)} \Theta \dot{S} \right) \bar{f}_1' \bar{f}_1 + \bar{f}_1 \bar{f}_1' \Theta \ddot{S} \\ &= 12 \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta \dot{S} + \bar{f}_1 \bar{f}_1' \Theta \ddot{S} \\ &\equiv \check{S} \end{aligned}$$

ここで、 $\dot{S} = W(1) \sim N(0, I_m)$  である。そして、簡単だが tedious な計算で  $\dot{S} \sim N(0, \frac{1}{120} I_m)$  であることと  $Cov(\dot{S}, \ddot{S}) = \frac{1}{12} I_m$  であることが求まるので、

$$\begin{aligned} Var(\check{S}) &= \begin{bmatrix} 12 \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta & \bar{f}_1 \bar{f}_1' \Theta \end{bmatrix} \begin{bmatrix} \frac{1}{120} I_m & \frac{1}{12} I_m \\ \frac{1}{12} I_m & I_m \end{bmatrix} \begin{bmatrix} 12 \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta \\ \bar{f}_1 \bar{f}_1' \Theta \end{bmatrix} \\ &= \frac{6}{5} \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta \Theta' \bar{F}_{(m-1)} \bar{F}'_{(m-1)} + \bar{f}_1 \bar{f}_1' \Theta \Theta' \bar{F}_{(m-1)} \bar{F}'_{(m-1)} + \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta \Theta' \bar{f}_1 \bar{f}_1' \\ &\quad + \bar{f}_1 \bar{f}_1' \Theta \Theta' \bar{f}_1 \bar{f}_1' \\ &= \frac{1}{5} \bar{F}_{(m-1)} \bar{F}'_{(m-1)} \Theta \Theta' \bar{F}_{(m-1)} \bar{F}'_{(m-1)} + \Theta \Theta' \end{aligned}$$

となり、求める結果を得る。

#### 命題5の証明

(33) に左から  $\beta'$  を掛け、一般性を失うことなく  $\beta' \beta = I_r$  と基準化できることに注意すると

$$\begin{aligned} \beta' \hat{y}_{T+h}^{LCA} &\approx \beta' \mu(T+h) + \beta' \beta u \widehat{f_1' y_{T+h}} + \beta' \beta V \widehat{F'_{(m-1)} y_{T+h}} \\ &= \beta' d_2^*(T+h) + u \widehat{f_1' y_{T+h}} + V \widehat{F'_{(m-1)} y_{T+h}} \end{aligned}$$

を得るが、 $\widehat{f_1' y_{T+h}}$  と  $\widehat{F'_{(m-1)} y_{T+h}}$  は ARIMA(p,1,q) での予測なので  $h \rightarrow \infty$  としても定数に収束することはない。よって、 $u \widehat{f_1' y_{T+h}} + V \widehat{F'_{(m-1)} y_{T+h}} \neq \beta' \gamma$  であり、求める結果を得る。