

日本のモデル生命表の開発と地域別生命表推計への 応用¹

慶應義塾大学経済学部 4年 13組 堀口侑²

¹ 本稿の執筆に際して、石井太教授（慶應義塾大学経済学部）から多くの有益な助言をいただいた。ここに記して感謝を申し上げる。

² 慶應義塾大学経済学部 4年 石井太研究会 1期

概要

本研究では、Wilmoth et.al(2012)および石井(2013)の2つの先行研究を踏襲しながら、「日本型死亡データベース(以下JMDという)」の都道府県別生命表を用いて、日本型モデル生命表を作成することを目的とする。オムランの「疫学的転換」は、歴史的な平均寿命伸長を、前近代の「伝染病と飢餓の時代」から産業革命と近代公衆衛生の発達によって「伝染病後退の時代」を経て「慢性退行性疾患及び人為的傷病の時代」へと3段階で以降するとした古典的理論であるが(Omran 1971)、1970年以降の死亡率改善はその大部分を高年齢死亡率改善に負うものであり(石井 2008)、Olshansky and Ault(1986)はこうした高年齢層での死亡率改善を、疫学的転換の第3段階に続く第4段階、即ち「慢性退行性疾患遅延の時代」と表現した。また、このような状況はわが国において最も顕著であり、それ故死亡率改善の高年齢へのシフトに適合した死亡率モデル開発の必要性が指摘されていた(石井 2008)。

本研究は、このようなわが国の死亡状況を表現可能な日本型モデル生命表を、JMDを用いて構築することを企図するものである。具体的には、はじめにWilmoth et.al(2012)による”flexible two-dimensional model”(以下flexible modelという)の推計式をJMDの都道府県別生命表に適用した。その際、推計式の係数は、Wilmoth et.al(2012)において、JMD作成の元となったHuman Mortality databaseに収録された719の生命表から推計されたもの(Wilmoth et.al 2012)を用いた。その結果、flexible modelによる推計結果は、特に65歳以上の高年齢層で推計誤差が大きくなることが確認された。そこで、この65歳以上の年齢階級におけるflexible modelの推計誤差を行列の形に収納し、この行列に特異値分解を施して第一特異値と左右特異ベクトルから推計誤差自体の理論値を得た。この理論値を高年齢層におけるflexible modelによる対数死亡率の理論値から減ずることによって高年齢層での死亡率推計の改善を達成した。さらに、本研究から得られたパラメータを用いて、市区町村別生命表から2つの自治体(埼玉県和光市・長野県天龍村)の2015年女性のデータにモデル生命表を当てはめたところ、その精度の高さが確認できた。

キーワード：高年齢死亡率の改善、JMD、都道府県別生命表、flexible model、日本型モデル生命表、SVD、市区町村別生命表

目次

1. 研究の背景と目的	4
1.1 長期的視座としての「疫学的転換」と現代の高齢死亡率改善の実際	4
1.2 死亡の年齢パターンの分類とリレーショナルモデル	4
2. 先行研究	5
2.1 リー・カーター・モデル	5
2.2 flexible model	5
2.3 わが国における死亡モデル研究—石井(2008)・石井(2013)	8
2.4 HMDとJMD.....	10
3. データと方法	11
3.1 修正モデル	11
3.2 修正モデルの作成.....	12
3.3 パラメータ r の効果.....	15
4. 結果と考察	15
4.1 修正モデルにおける推計の正確性.....	15
4.2 修正モデルの市区町村別生命表への応用とモデルの有効性.....	15
5. 議論.....	17
6. 結論.....	18
参考文献.....	19
図表.....	20

1. 研究の背景と目的

1.1 長期的視座としての「疫学的転換」と現代の高齢死亡率改善の実際

オムランの「疫学的転換」は、歴史的な平均寿命伸長を、前近代の「伝染病と飢餓の時代」から産業革命と近代公衆衛生の発達によって「伝染病後退の時代」を経て、「退行性疾患及び人為的傷病の時代」へと3段階で移行するとした理論である(Omran 1971)。実際20世紀前半までの主要国では乳幼児・若年死亡率の改善が見られたが、一方1970年以降の死亡率改善の大部分は高齢死亡率改善によっている(石井 2008)。

Olshansky and Ault (1986)は上述の高齢層での死亡率改善を、疫学的転換の第3段階に続く第4段階、即ち「退行性疾患遅延の時代」と捉えたが、このような状況はわが国において最も顕著であり、死亡率改善の高齢へのシフトに適合した死亡率モデルの開発が必要とされている(石井 2008)。本研究はこのようなわが国の死亡状況を表現可能な日本型モデル生命表を、「日本版死亡データベース」(以下JMDという)の都道府県データを用いて構築することを企図するものである。

1.2 死亡の年齢パターンの分類とリレーショナルモデル

一般に、人口動態事象の発生頻度は年齢に大きく依存する。死亡はその代表的な事象であるが、年齢別に死亡率を観察するためには多くの人口学的率を処理する必要があるため、分析が複雑化する。死亡率を含む生命表関数は、それゆえ、死亡の年齢パターンモデルから単純な規則・法則を見出して表現される。

死亡の年齢パターンモデルは(1)数学的関数によるモデル、(2)数表によるモデル、(3)リレーショナルモデルの3つに大別される。このうち(3)は、(1)の生命表関数を少数のパラメータで表現できるという利点および(2)の経験に基づく年齢パターンを表現できるという双方の長所を兼ね備えていることから近年様々な応用が行われており、最も代表的なものが現在、国などの将来推計で標準的に用いられているリー・カーター・モデル(Lee and Carter, 1992)である。

このリレーショナルモデルを用いて、モデル生命表を作成する研究が Wilmoth et al. (2012)によって行われている。この研究は、Human Mortality Database(以下HMDという)に収載されている生命表から、間接推計のための新たなモデル生命表(flexible model)を開発する研究であり、少ないパラメータで多くの生命表を表現できるモデルとなっている。

2. 先行研究

2.1 リー・カーター・モデル

リー・カーター・モデル(Lee and Carter, 1992)は現在、国などの将来推計で標準的に用いられている、最も代表的なリレーショナルモデルである。リー・カーター・モデルによる対数死亡率 $\ln[m(x, t)]$ は、以下の式(1)で推計される：

$$\ln[m(x, t)] = a_x + b_x k_t + \varepsilon_{x,t} \quad (1)$$

(x:年齢, t:年次, $\varepsilon_{x,t} \sim N(0, \sigma_\varepsilon^2)$)

ここで a_x は通時的に平均的な年齢階級の死亡水準を示すパラメータであり、 b_x は、年次に特有の死亡水準 k_t に対する各年齢階級の死亡率の弾力性を示すパラメータである。なお、式(1)の右辺には推計されるべき係数のみ並んでいることに注意しなければならない。それ故単純に最小二乗法などを適用して推計することはできず、従って b_x, k_t は SVD を用いて推計される。また、 k_t は 1900 年以降のアメリカのデータによれば直線的に低下しており (p. 662)、具体的には以下のようなランダムウォークとして推計される。

$$k_t = k_{t-1} - 0.365 + 5.24flu + e_t \quad (2)$$

(0.069) (0.461) $R^2 = 0.995$

* ()内は t 値を表す

但し flu は 1918 年頃から猛威を振るい、特異な死亡状況を生じた、いわゆる「スペイン・インフルエンザ」の影響を統御するためのダミー変数である。Tuljapkar et al.(2000)はこのリー・カーター・モデルを G7 諸国の死亡率データに適用し、あてはまりが良いことを実証している。

2.2 flexible model

以下、本研究の直接の先行研究である Wilmoth et al. (2012) についてレビューする。Wilmoth et al. (2012) は限定的な情報から各国の多様な年齢パターンを効率的に表現できるモデル (flexible two-dimensional model, flexible model) を開発した。なお、Wilmoth et al.(2012) は以下の(a)~(e)の論理展開で議論を進めている³：

(a) モデルの提示と各パラメータの説明(p.8)

³ 本稿 3.1 節は(a)、3.2 節は(b)に対応し、(c), (d)は本稿 3.3 節に対応する。(e)は、4.1 節および、市区町村別生命表への応用に関する 4.2 節に対応している。

- (b) 推計ツールとしての利用法(p.11)
 - (c) モデルにより示される年齢パターンの例示(p.12, Figure3 および p.13, Figure4)
 - (d) 様々な年齢階級の実績値とモデルの関係とパラメータ k の効果(pp.14~16 Figure5, 6a, 6b)
 - (e) モデルの正確性(Figure7 および pp.20~23)
- そこで、以下、このような論理展開に沿ってレビューしていく。

(a) モデルの提示と各パラメータの説明(p.8)

一般に発展途上国のセンサスは先進諸国のものに比べて、精度が低く、入手可能かつ信頼できる人口動態の情報が限定的である。しかし、Wilmoth et al. (2012)によれば、発展途上地域も含め、ほとんどすべての国において、乳幼児死亡確率、つまり 0~4 歳の死亡確率(${}_5q_0$) は入手可能である。flexible model はそのような事実に着目し、以下の式(3)によって全年齢の死亡率を推計するものである：

$$\log m_x = a_x + b_x h + c_x h^2 + v_x k \quad (3)$$

但し $h = \log {}_5q_0$

パラメータ h は $\log {}_5q_0$ であり、これが全体的な死亡水準の高低を代表する。これに対して、パラメータ k は典型的な死亡の年齢パターンからの乖離を表すパラメータであり、 $k > 0$ であることは、実績の成人(15~59 歳)死亡率が、 h のみから理論的に予測されるものよりも高いことを示すことになる((c)で後述する)。この点について、もう少し詳しく説明しよう。

特に式(3)において、右辺第 3 項の h^2 の項以下を無視して、 $\log m_x$ と h との間の線形関係

$$\log m_x = a_x + b_x h \quad (3-1)$$

によってもかなりの程度的確に死亡率を推計できるが、各年齢階級の実績の $\log m_x$ を h に対しプロットすると、若年層では下に凸、高齢層では上に凸な曲線を描くように $\log m_x$ が分布しており、その意味で、

$$\log m_x = a_x + b_x h + c_x h^2 \quad (3-2)$$

という 2 次関数による推計のほうが式(3-1)よりも推計誤差は小さくなる。

上の式(3-1)、(3-2)は式(3)において $k = 0$ とした場合のものであるが、式(3)における v_x は実績の $\log m_x$ の、式(3-2)の推計値からの乖離のうち、年齢によって説明される部分を捉

え、 k はその乖離の方向と大きさを捉えることになる。

さらに、式(3-2)による $\log m_x$ の推計値と実績値との乖離について、特異値分解 (Singular Value Decomposition, SVD) を施して、その左特異ベクトル v_x を導出する。

(b) 推計ツールとしての利用法(p.11)

$h (= \log {}_5q_0)$ は、HMD に収録された 719 の生命表から入手する。さて、いま入手可能なデータが h のみであったとすると、式(3)において $k = 0$ と仮定した推計式、すなわち対数 2 次関数による式(3-2)を最小二乗法によって推計することとなる。15~59 歳の死亡確率 ${}_{45}q_{15}$ が入手可能であれば、 ${}_{45}q_{15}$ の実績値を再現するような k を数値解析的に求めることとなり、 ${}_{45}q_{15}$ が入手可能な場合においてはこの方法がより推計誤差を小さくする方法となる。

(c) モデルにより示される年齢パターンの例示(p.12, Figure3 および p.13, Figure4)

(a) で述べた、パラメータ h, k の効果については、p.12, Figure3 においてシミュレーションによって確認されているが、それによれば、パラメータ h は全体的な死亡水準の高低を規定するのに対して、パラメータ k は典型的な死亡の年齢パターンからの乖離を表すパラメータであり、 $k > 0$ であることは、実績の成人 (15~59 歳) 死亡率が、 $h = \log {}_5q_0$ のみから理論的に予測されるものよりも高いことを示すことになるから、死亡率曲線の形状を規定するパラメータであるといえる。また、Wilmoth et al. (2012) では、男女別に① ${}_5q_0 = 0.05, k = -4, \dots, 0, \dots, 4$ とした場合の e_0 の理論値及び m_x 曲線の形状 ② $k = 0, e_0 = 38, \dots, 78$ とした場合の ${}_5q_0$ の理論値及び m_x 曲線の形状 ③ $e_0 = 60, {}_5q_0 = 0.06, \dots, 0.14$ とした場合の k の理論値及び m_x 曲線の形状 をコンピューターでシミュレーションし、 $k \in (-4, 4)$ で多様な m_x 曲線を再現できることを確認した。

(d) 様々な年齢階級の実績値とモデルの関係とパラメータ k の効果(pp.14~16 Figure5, 6a, 6b)

${}_{45}q_{15}$ の実績値を、 ${}_5q_0$ の実績値に対してプロットし、これを式(3)で $k \in (-2, 2)$ の範囲で動かした時の m_x の理論値から得られる ${}_{45}q_{15}$ の理論値と比較してみると (Figure5)、実績の $({}_{45}q_{15}, {}_5q_0)$ のペアのほとんどすべては、上の ${}_{45}q_{15}$ の理論値の範囲に入っていることが確認できた。このことは、 k を適切に推計すればいかなる $({}_{45}q_{15}, {}_5q_0)$ のペアをも再現できるのみならず、いかなる実績の $({}_{45}q_{15}, {}_5q_0)$ のペアに対しても一意的に k を導出できることを示唆している。さらに、各年齢階級の死亡率の実績値を、 ${}_5q_0$ の実績値に対してプロットすると、一部の例外を除き、 $k \in (-2, 2)$ の範囲で k を可変とすることによって、ほとんどすべての実績値をカバーできることが分かった (Figure6a, 6b)。

(e)モデルの正確性(Figure7 および pp.20~23)

推計の精度は、平均寿命(e_0), 0~1 歳時死亡確率(${}_1q_0$), 15~44 歳時死亡確率(${}_{45}q_{15}$)および 60~79 歳時死亡確率(${}_{20}q_{60}$)について、推計誤差を平均平方 2 乗誤差(Root-mean-squared errors, RMSE) ;

$$(RMSE) \equiv \sqrt{\frac{1}{n} \sum_{i=1}^n (\theta_i - \hat{\theta}_i)^2} \quad (4)$$

(但し $\theta_i = e_0, {}_1q_0, {}_{45}q_{15}, {}_{20}q_{60}$ (いずれも実績値) であり、 $\hat{\theta}_i$ はそれらの理論値である)

で測って評価した。その結果、flexible model の精度は Coale-Demeny のモデル生命表や UN-lifetable よりも高く、Brass-logit モデルと同程度であったが、Brass logit model では分析の前に l_5 と l_{60} との間の関係式を予測する必要があるのに対して、flexible model では少数のパラメータから直接的に全年齢の死亡率を推計できることに長所があるとしている。

なお、式(3)の推計において、1940~1944 年のフィンランドなどが極端な外れ値として存在し、この外れ値に影響されて死亡率を過大評価してしまうことになる。そのような状況でも最良線形不偏推定量(Best Linear Unbiased Estimator, BLUE)を得るため、Wilmoth et al. (2012)では Tukey の bi-weighted estimation を行っている。

ただし、この柔軟性の高いモデルでも当てはめが困難なのが近年の日本の死亡率であり、このモデル生命表を日本に適用するには問題がある。

2.3 わが国における死亡モデル研究—石井(2008)・石井(2013)

すでに、石井(2008)は、わが国の 1970 年代以降の死亡動向に関して、死亡水準の低下ではなく、死亡の遅延が生じて死亡率曲線が高年齢にシフトするという年齢シフトを考慮して (p. 32)、年齢シフトモデルを開発した。これは、Bongaats(2005)のシフティング・ロジスティック・モデルを用いてリー・カーター・モデルを改善したもので(石井 2008)、まず、死力を以下の 3 パラメータロジスティックモデルとして当てはめた(Bongaats 2005, 石井 2008) :

$$\mu(x, t) = \frac{\alpha(t)e^{\beta x}}{1 + \alpha(t)e^{\beta x}} + \gamma(t) \quad (5)$$

さらに、年齢シフト量を

$$S(t) \equiv -\frac{\ln(\alpha(t))}{\beta} \quad (6)$$

として求めた。その上で先の年齢シフト量 $S(t)$ と β_t から得られる年齢変換 \mathcal{A} と、リー・カーター・モデルの推計過程 \mathcal{L} とを用いて、

$$\mu_{x,t} \xrightarrow{\mathcal{A}} \widetilde{\mu}_{z,t} \xrightarrow{\mathcal{L}} \widehat{\mu}_{z,t} \xrightarrow{\mathcal{A}^{-1}} \widehat{\mu}_{x,t}$$

という手順により死亡率をモデリングした⁴。

石井(2013)は上述の石井(2008)を発展させ、線形差分(LD)モデルを開発した。これは、対数死力の逆関数が年齢の線形関数になるようなモデルである。

いま、 $X = [0, +\infty)$ を年齢、 $T = (-\infty, +\infty)$ は時間を表す変数とする。年齢 $x \in X$ 、時刻 $t \in T$ における死力関数 $\mu_{x,t}$ を用いて、

$$y = \lambda_{x,t} = \log \mu_{x,t} \in Y \quad (7)$$

で対数死力を表すことにする。対数死亡率曲面 (the log mortality surface) を

$$S = \{(x, t, y) | y = \lambda_{x,t}\} \quad (8)$$

で定義し、

$$v_{y,t} = \lambda_{x,t}^{-1} \quad (9)$$

なる逆関数 $v_{x,t}$ とその時間微分

$$\tau_{y,t} = \frac{dv_{y,t}}{dt} \quad (10)$$

を定義すると、 y_0 を固定したときに対数死亡率曲線上に描かれる曲線 C_{y_0} は、

$$C_{y_0} = (v_{y_0,t}, t, y_0) \quad (11)$$

と書けて、これを時間微分すると、

$$(\tau_{y_0,t_0}, 1, 0) \quad (12)$$

⁴ 日本人口学会第 62 回大会で示された年齢変換アプローチについての説明を援用する (石井(2010))

という接ベクトル場を考えることができる。各時点・各年齢における死亡率改善の方向は、上のような接ベクトル場を用いて表現でき、このような枠組みの中で LD モデルは水平方向への死亡率のシフトを表現するモデルとして見通せるようになる。また、 $\tau_{y,t}$ が年齢の線形関数となることは、高齢層での死亡率を最も的確に表現するとされている Kannisto モデルも持つ性質である。以上のような背景から、石井(2013)では若年層にはリー・カーター・モデルを適用し、高齢層では LD モデルによってモデリングを行い、その両者を、ウェイトを用いて線形結合して全年齢のモデルとしたものである。

2.4 HMD と JMD⁵

本研究では、1975 年以降の JMD の都道府県別生命表に対して、まず flexible model を適用するので、JMD と、その基となった HMD についてレビューしておく。

Human Mortality Database(HMD)は、カリフォルニア大学バークレー校の John Wilmoth をプロジェクトリーダーとして作成された。すでに同氏の Berkeley Mortality Database(BMD)は 1997 年に開始され、HMD の直接的な先行研究となったが、これはフランス・日本・スウェーデン・アメリカの 4 か国のみのデータを提供したものであった。一方、1993 年に Kannisto と Thatcher によって創設された “Kannisto-Thatcher Database on Old Age Mortality”(KTD)は 30 か国のデータを収集し、とりわけ 80 歳以上の高齢死亡率推定に関して先駆的業績となった。HMD は BMD の形式を踏襲しながら、KTD の高齢死亡率推定の先進的方法を活用して 30~40 の国と地域のデータを収録することを目指して展開されたプロジェクトであるといえる⁶。

HMD では、入力データから最終的な生命表を作成するまでの工程を、すべての国や地域、あらゆる時代に適用できるような統一的方法として、方法手順書(methods protocol)をまとめている(Wilmoth et al. 2007)。HMD はそうした統一的方法のもと、男女ともに 110 歳以上を開放年齢として統一的に整理されたデータであり、国際的な死亡動向の比較調査に最適化されたデータであるといえる。

「日本版死亡データベース(JMD)」はこの HMD の方法手順書を詳細にレビューしたうえで、国際比較可能性を阻害しない範囲内で、いくつかの修正を行ったものである(石井 2015)例えば、JMD では、 a_0 について、(HMD では Coale-Demeny のモデル生命表の値を用いているが)それぞれ男性・女性について、 m_0 が小さい場合、実績に基づく以下の関係式を用いた：

⁵ 石井(2015), pp. 5-18 を参考にまとめた。

⁶ HMD のウェブサイト(www.mortality.org) (U.C. Berkeley)もしくは www.humanmortality.de (Max Planck Institute for Demographic Research, Germany) に、歴史(History)が記されている

男性：

$$a_0 = \begin{cases} 0.330 & (m_0 \geq 0.107) \\ 0.045 + 2.684m_0 & (0.0612 \leq m_0 < 0.107) \\ 0.132 + 1.264m_0 & (0.00869 \leq m_0 < 0.0612) \\ 0.242 - 11.373m_0 & (m_0 < 0.00869) \end{cases}$$

女性：

$$a_0 = \begin{cases} 0.330 & (m_0 \geq 0.107) \\ 0.045 + 2.800m_0 & (0.0557 \leq m_0 < 0.107) \\ 0.152 + 1.015m_0 & (0.000637 \leq m_0 < 0.0557) \\ 0.239 - 12.537m_0 & (m_0 < 0.00637) \end{cases}$$

また、JMD では都道府県別生命表も用意しているが、死亡パターンの安定性を保つため、5歳×5年、10歳×10年の生命表に限って提供している(石井 2015)。

3. データと方法

3.1 修正モデル

本研究では、以下の式(14)によってモデル生命表(以下「修正モデル」という)を作成する：

$$\log m_x = a_x + b_x h + c_x h^2 + v_x k + u_x r \quad (13)$$

(但し、 $h = \log_5 q_0$ である)

修正モデルは、以下のような問題意識のもとに開発された。すなわち、Wilmoth et al. (2012)は、非常に柔軟性の高いモデルであったが、このモデルでも日本の高齢死亡率推計において推計誤差が大きく、モデリングが困難であることが確認された(後述)。これは、石井(2008)においても指摘されているように、近年の日本の死亡率改善は、高齢死亡率改善によるところが大きく、高齢死亡率改善がまだ大きくない他の国々のデータによるモデリングでは、高齢死亡率を過大評価してしまうためであると考えられる。以上のような理由により、Wilmoth et al. (2012)に代わる、日本の死亡動向をより正確に表現するモデル生命表として、本稿では修正モデルを提案するものである。

式(13)の係数 a_x, b_x, c_x, v_x は、Wilmoth et al.(2012)によってHMDの719の生命表から得られたものをそのまま援用する。また、式(13)の右辺第5項は、後述する式(14)による死亡率の推計誤差に対して、特異値分解(SVD)を施し、その第一特異値のみを抽出したもので、 u_x はその左特異ベクトルとして導出され、推計誤差の、年齢パターンを示す。同じく右辺

第4項のパラメータ r は、右特異ベクトルとして得られ、年次や地域などによる推計誤差の方向とその大きさを表す。これは、Wilmoth et al.(2012)におけるパラメータ k に対応するものである。

3.2 修正モデルの作成

最初に、flexible modelによる、以下の式(3) (再掲載)をJMDの都道府県生命表に適用する：

$$\log m_x = a_x + b_x h + c_x h^2 + v_x k \quad (3)$$

(但し $h = \log_5 q_0$ である)

但し、係数 a_x, b_x, c_x, v_x は、Wilmoth et al.(2012)によってHMDの719の生命表から得られたものをそのまま援用する。

図1,2は、それぞれ女性、男性について、2013~2017年の対数死亡率 $\log m_x$ のJMDから得られた実績値、 $k=0$ の場合(mx_0)と k を可変としたとき(mx_f)のflexible modelから得られる推計値を示したものである⁷。なお1~47の数字は都道府県番号を示している。さらに、図3,4は、それぞれ女性、男性について、 $k=0$ の場合(d_lmx_0)と k を可変としたとき(d_lmx_f)の、flexible modelから得られる推計値の2013~2017年の対数死亡率 $\log m_x$ の実績値に対する誤差を

$$\begin{aligned} (d_lmx_f) &= \log(mx_f) - \log m_x \\ (d_lmx_0) &= \log(mx_0) - \log m_x \\ &\dots\dots\dots (14) \end{aligned}$$

として示したものである。これらを見ると、 k を可変とした場合のflexible modelでは、15~59歳の成人死亡率の推計において、 $k=0$ の場合よりも誤差が小さくなっており、また同じ年齢区間では極めて0に近い値を取っていることが分かる。Wilmoth et.al(2012)では、flexible modelの推計式におけるパラメータ v_x の値が15~59歳の年齢階級では大きくなることが指摘されているが、この年齢区間では k の値が推計の精度に強く有効に作用しているためであると考えられる (Wilmoth et al. 2012)。

一方、高齢層では、 $k=0$ においても k を可変とした場合においても同様に、推計誤差が大きくなっている。これは、上述の理由の逆で、この年齢層では推計式におけるパラメータ v_x の値が極めて0に近く、 k の効果が極めて小さくなることに起因するものと考えられる。

さて、このような推計誤差を平均2乗誤差(Mean-Squared Error,MSE)：

⁷ 本研究では、1978~1982, 1983~1987,..., 2013~2017の8つの年次について同様の分析を行った結果、いずれの年次においても同様の結果が得られたため、本稿ではそのうち、直近の2013~2017年についてのみ結果を示すこととした。

$$\frac{1}{n} \sum_{i=1}^n (\theta_i - \hat{\theta}_i)^2 \quad (15)$$

で測定して 65 歳時平均余命 e_{65} に対してプロットしたのが図 5（女性）・図 6（男性）である。これを見ると MSE は e_{65} が大きくなると増加し、高齢死亡率の改善にともなって flexible model の推計誤差が大きくなる傾向が見て取れる。

このように、65 歳以上の高齢層で flexible model の推計誤差が大きくなる傾向が確認されたことから、本研究では、この推計誤差に対して SVD を施し、第一特異値に対応する箇所のみ抽出して推計誤差のモデリングを試みる。具体的には、

$$X \equiv \{65 \sim 69, 70 \sim 74, \dots, 110 +\}$$

を（65 歳以上の）年齢の集合とし、

$$P \equiv \{1, \dots, 47\}$$

を都道府県の集合とする。さらに、

$$Y \equiv \{1978 - 1982, 1983 - 1987, \dots, 2013 - 2017\}$$

を年次の集合とする。そして、

$$a \equiv \dim(X) = 10, b \equiv \dim(P) \times \dim(Y) = 47 \times 8 = 376$$

と表し、 $U \in R^{a \times a}$ を左特異ベクトル、 $R \in R^{b \times a}$ を右特異ベクトルとし、

$$U \equiv (U_1 \quad \dots \quad U_{10}), \quad R \equiv (R_1, \dots, R_{10}), \quad \Sigma \equiv \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_{10} \end{pmatrix} \in R^{a \times a} \quad (16)$$

（但し $d_i (\forall i \in \{1, \dots, 10\})$ は特異値）

としたとき、65 歳以上での (k を可変とした場合の) flexible model の誤差を $b \times a$ の行列に収納したものを d_lmx_f と表すことにする。すると、以下のように d_lmx_f を特異値分解することができる：

$$(d_lmx_f) = U \Sigma R^T \quad (17)$$

そこで、(17)式の右辺の第一特異値に対応する項のみを抽出して、それを推計誤差のモデルの行列 d_lmx_MDL と表すことにすると、

$$(d_lmx_MDL) = U_1 d_1 R_1^T \quad (18)$$

$$\left(\text{但し } U_1 = \begin{bmatrix} u_{65-69} \\ u_{70-74} \\ \vdots \\ u_{110+} \end{bmatrix} \text{ である} \right)$$

とできる。そこで、

$$u_x = \begin{bmatrix} u_0 \\ u_{1-4} \\ u_{5-9} \\ \vdots \\ u_{60-64} \\ U_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ u_{65-69} \\ u_{70-74} \\ \vdots \\ u_{110+} \end{bmatrix} \in R^{24} \quad (19)$$

を男女別に算出して、

$$\log m_x = a_x + b_x h + c_x h^2 + v_x k + u_x r \quad (13)$$

$$\left(\text{但し } h = \log {}_5q_0 \text{ である} \right)$$

により、修正モデルを推計する。

但し、(13)式の u_x をそのまま使って推計すると、対数死亡率曲線が 65 歳時で不連続になってしまう。その不連続性は男性については無視可能なほど小さいが、女性では顕著に現れてくることから、本研究では、女性について、

$$u_{50-54} = \frac{1}{4} \cdot u_{65-69},$$

$$u_{55-59} = \frac{2}{4} \cdot u_{65-69},$$

$$u_{60-64} = \frac{3}{4} \cdot u_{65-69}$$

$$\dots\dots\dots (20)$$

と加重平均してモデルの推計を行った。

上述のようにして作成された修正モデルを利用して、任意の死亡率を推定するためには、flexible model において 15~59 歳の死亡確率 ${}_{45}q_{15}$ の実績値を再現するような k を導出した

のと同様の方法によって、対象となる集団の65歳平均余命 e_{65} の実績値を再現するように r を男女別に導出する。

なお、表1は、JMDの都道府県別生命表から得られる修正モデルの u_x を、Wilmoth et al.(2012)のTable3で示されたパラメータ (a_x, b_x, c_x, v_x) とともにまとめたものである。以降この表1の係数を用いて、市区町村別生命表に修正モデルを適用する。

3.3 パラメータ r の効果

図7は、Wilmoth et al.(2012)のTable3に対応し、推計式(13)におけるパラメータ h, k を、 $k = -1.0, {}_5q_0 = 0.01$ として、 $r = -2.10, \dots -3.00$ と r を動かして、パラメータ r の、モデル生命表に対する効果を例示したものである。但し、推計式(14)の係数は女性のものを使用している。 r の値を小さく（絶対値を大きく）することにより、高齢層での死亡率曲線が改善していることが分かる。 k のように、パラメータ r は、flexible modelでは十分に表現できない日本の高齢死亡率改善を表現するパラメータであると解釈することができるのである。

図8・9は、それぞれ女性・男性について、修正モデルから計算される65歳時平均余命(e_{65})の実績値を、修正モデルの推計によって得られたパラメータ r に対してプロットしたものであるflexible modelにおけるパラメータ k と同様に、65歳平均余命が長くなるとパラメータ r の値は小さくなり、逆に65歳平均余命が短い場合には、パラメータ r の値は大きくなる。以上より、修正モデルにおけるパラメータ r は、適切に推計することにより、生命表から得られる様々な値の65歳時平均余命を再現できることがわかり、パラメータ r を、65歳時平均余命(e_{65})の実績値を再現するように推計することは妥当であるといえる。

4. 結果と考察

4.1 修正モデルにおける推計の正確性

図10・11は、それぞれ、55歳以上の女性・男性について、修正モデル・flexible model・実績値をプロットしたものである。これらの図から、修正モデル(mx_MDL)では、高齢層での死亡パターンをより正確に表現できていることが分かる。図12・13はそれぞれ全年齢の女性・男性について、推計式(13)から得られる推計値をflexible model・実績値と比較したものである。65歳以上の高齢層での死亡率推計における誤差が推計式(13)では改善されたことから、全年齢の死亡率推計もあてはまりが良くなっていることが分かる。

4.2 修正モデルの市区町村別生命表への応用とモデルの有効性

今回得られた修正モデルの有効性を示すためには、パラメータの推計に用いた上記の（センサス中心、5年×5歳間隔の）JMDの都道府県生命表以外の生命表に対して、修正モデルを当てはめ、その精度を検証する必要がある。

今、2015年の市区町村別生命表はわからなかったとし、2010年の市区町村別生命表と、2010、2015年の全国JMDが既知であったとする。このとき、当該市区町村生命表から得られる2010年の $5q_0$ 、 $15q_{45}$ 、 e_{65} の、その全国値に対する比率を計算し、2015年の全国と同 $5q_0$ 、 $15q_{45}$ 、 e_{65} それぞれにこの比率を掛けることによって、2015年の当該市区町村の $5q_0$ 、 $15q_{45}$ 、 e_{65} を求めることができる。すると、これらから、本研究のモデル生命表を作成することができ、これと2015年の実績値を比較し、精度を検証することができる。

市区町村別生命表の開放区間は95歳以上であり、本研究のモデル生命表の開放区間は110歳以上であるので、本研究のモデル生命表について、まず

$${}_{\infty}d_{95} = \sum_{i=0}^4 {}_{99+5i}d_{95+5i} + {}_{\infty}d_{110} \quad (21)$$

を求め、

$${}_{\infty}m_{95} = \frac{{}_{\infty}d_{95}}{T_{95}} \quad (22)$$

とすることによって、本研究のモデル生命表と実績値を比較検討することができる。

今回、検討対象とした市区町村は、埼玉県和光市と長野県天龍村である。前者は老年人口割合が最も低い市区町村の一つであり、後者は老年人口割合が最も高い地域であることから選んだものである⁸。それぞれ2015年の女性について、図14、15の結果を得た。和光市の場合に比べて、天龍村の若年層での推計誤差が大きくなるのは、同村では高齢化がかなり進行しているために、若年層が少なく、この年齢層でのデータが不安定なことや、人口規模が小さいことなどが理由として考えられる。但し、若年層が少数であることから、この年齢層における推計誤差は将来人口推計においてそれほど深刻な誤差を生み出すことはないと考えられる。また、いずれの自治体においても、65歳以上の高齢死亡パターンを適切に再現できていることが分かる。このことから、今回の修正モデルは、市区町村レベルの将来人口推計にも有効であることが示唆される。

⁸ 平成27(2015)年度『国勢調査報告』によれば、埼玉県和光市の人口総数は80,826人、老年人口割合は16.83%であり、全国では8番目に低い。一方、長野県天龍村の人口総数は1,365人で、老年人口割合は59.05%であり、全国で2番目に高い。

5. 議論

・「アジア型人口学モデル」の構築に向けての貢献

本研究の修正モデルは、わが国における死亡状況をよりの確に表現可能なモデル生命表であるといえるが、今後このモデルをさらに発展させることにより、真に「日本型モデル生命表」と呼びうるモデル生命表が開発できるものと期待される。この「日本型モデル生命表」によって、これまで以上に正確な将来人口推計も可能となるであろうが、これが社会保障をはじめとする政策決定のための有効な提言を生み出していくものと期待される。

さらに、「日本型モデル生命表」が開発された後には、それを広くアジア、特に東アジア中国文化圏に適用可能かについても検討したい。既にアイリーン・トイバーは、国民の同質性や高い教育水準など、中国文化圏に共通な特質が1950年代以降の東アジア諸国の急速な出生力低下の文化的要因であると指摘しているが(Taeuber 1966)、死亡率改善についても日本型モデル生命表が他の東アジア諸国に良く当てはまるようであれば、いわば「アジア型人口学モデル」の構築にもつなげることが可能であると考えられる。これは、それまで欧米の経験から帰納的に構築されていた人口学において、画期的な成果となろう。

・歴史人口学および日本経済史への応用可能性

加えて、「日本型モデル生命表」の開発は、日本の歴史人口学研究において、より日本の死亡特性に適合した推定を可能にするという効果も期待できる。例えば、フランスの歴史人口学者レイ・アンリは「教区簿冊」に記されたヨーロッパ各地の教区民の出生・死亡・結婚に関する年次記録を用いる調査方法を開拓したが、これに対応して、日本では「宗門人別改帳」（「宗門改帳」）から「ベーシック・データ・シート」（「BDS」）と呼ばれる、出生・死亡・移動などの人口動態を記録した表を作成し、調査する手法が速水融によって開発された（速水 2020）。「日本型モデル生命表」が開発できれば、このような先行研究を踏まえ、近世以前の人口をより正確に把握することができる可能性があり、広く経済史研究の発展にも寄与し得ると考えられる。

・今後の課題

本研究で推計した修正モデルにおいて、SVD から求めた u_x をそのまま用いると、女性の死亡率については顕著な不連続性が現れることが確認されたため、本稿では加重平均の方法によってこの不連続性を修正した。しかしながら、特異値分解を施す年齢階級を例えば50歳以上などとすることによって、このような問題に対応することなども考えられる。

また、本研究では Wilmoth et al.(2012)の flexible model の推計で用いられた係数 (a_x, b_x, c_x, v_x) をそのまま用いたが、JMD の都道府県生命表から直接これらの係数を推計する方法も考えられる。

さらに、本稿では修正モデルの有効性を検証するため、市区町村別生命表から2つの自治体を例として適用したが、今後はより多くの自治体の生命表に対して本モデルを適用することなどにより、自治体における将来人口推計への有効性を示していくことも課題である。

最後に、前節で言及したように、本研究は、将来的に「アジア型モデル生命表」・「アジア型人口学モデル」の開発・確立を視野に入れて行われた。それ故、今後は先述のような有効性の検証とともに、これを韓国・台湾などといった東アジア諸国の死亡パターン分析に応用し、必要に応じて地域の実情にあったモデルに修正していくことも課題である。

6 結論

本研究を通して、まず JMD の都道府県生命表に Wilmoth et al.(2012)の flexible model をそのまま適用すると、65 歳以上の高齢層で、死亡率を過大に推計してしまうこと、またそうした推計誤差は、65 歳時平均余命の伸長とともに大きくなる傾向が見取れたが、これは近年、わが国の高齢死亡率改善が急速に進展してきている事実を考慮すると無視できない事実である。

そこで、本研究では flexible model の k を可変とした場合の 65 歳以上の年齢階級における推計誤差自体に SVD を施し、そのうち、第一特異値と第一特異値に対応する左右特異ベクトルのみによってこの推計誤差を表現し、その結果を flexible model の推計値から減ずることによって推計の改善を試みたところ、死亡率推計を改善することが観察された。

JMD の都道府県別生命表から得られたパラメータ u_x を固定し、 e_{65} を一致させるように r を推定することで、年齢別死亡パターンのモデル化を行った。パラメータ r は、flexible model では十分に表現できない日本の高齢死亡率改善を表現するパラメータであると解釈することができるのである。次に、この修正モデルの有効性を確認するため、市区町村別生命表に修正モデルを適用した。市区町村別生命表からは、老年人口割合が低い埼玉県和光市と、老年人口割合が高い長野県天龍村の2つの自治体の、2015 年女性の生命表を抽出した。その結果、修正モデルは 2015 年の当該市区町村の実績値を的確に再現できており、市区町村レベルの将来人口推計にも応用可能であることが示唆された。

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, 64-3, pp. 28~44.
- 石井太(2010)『死亡率推計モデルにおける年齢変換アプローチの応用』, 日本人口学会第62回大会, お茶の水女子大学.
- 石井太(2013)『死亡率曲線の自由な方向への変化を表現する数理モデルとわが国の将来生命表への応用』, 人口問題研究, 69-3, pp. 3~26.
- 石井太(2015)『日本版死亡データベースの構築に関する研究』, 人口問題研究, 71-1, pp. 3~27.
- 速水融(2020)『歴史人口学事始め—記録と記憶の九〇年』, ちくま新書.
- Bongaarts, John (2005), "Long-Range Trends in Adult Mortality: Model and Projection Methods", *Demography*, Volume42, Number1, pp.23-49.
- Lee, Ronald D. and Lawrence R. Carter (1992), "Modeling and Forecasting U.S. Mortality", *Journal of the American Statistical Association*, Vol.87, No.419, pp.659-671.
- Ohshansky, S. and A. Ault (1986), "The Fourth Stage of the Epidemiologic Transition: The Age of Delayed Degenerative Diseases", *The Milbank Quarterly*, Vol64, No.3, pp.355-391.
- Omran, A.(1971), "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change", *The Milbank Memorial Fund Quarterly*, Vol.49, No.4, 509-538.
- Tuljapurkar, Shripad, Nan Li and Carl Boe (2000), "A Universal Pattern of Mortality Decline in the G7 Countries", *Nature*, Vol.415, pp.789-792.
- Wilmoth J.R., K. Andreev, D. Jdanov, and D.A. Gleijeses(2007), "Method Protocol for the Human Mortality Database", available at, <http://www.mortality.org/public/Docs/MethodsProtocol.pdf> .
- Wilmoth, John, Sarah Zureick, Vladimir Canudas, Mie Inoue, and Cheryl Sawyer(2012) "A flexible two-dimensional mortality model for use in indirect estimation", *Population Studies*, Vol.66, No.1, pp.1-28.

図表

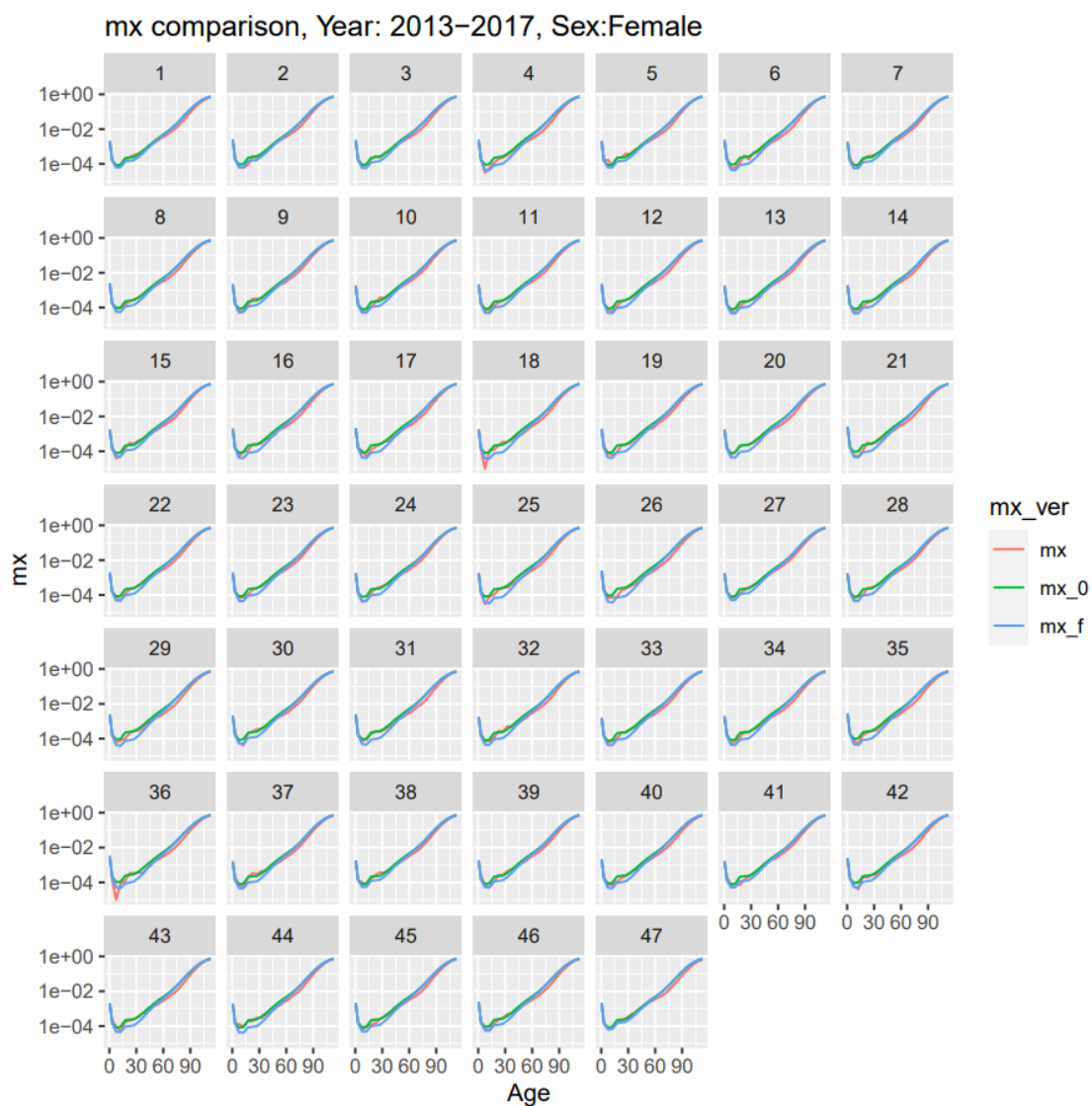


図1 : flexible two-dimensional model による対数死亡率($\log m_x$)と JMD から得られた実績値(2013~2017年, 女性)。” m_{x_0} ” は flexible model において $k=0$ とした場合の推計結果を表し、 m_{x_f} は k を可変とした場合の推計結果を示す。1~47 の数字は都道府県番号を示す。

(出所) 筆者推計

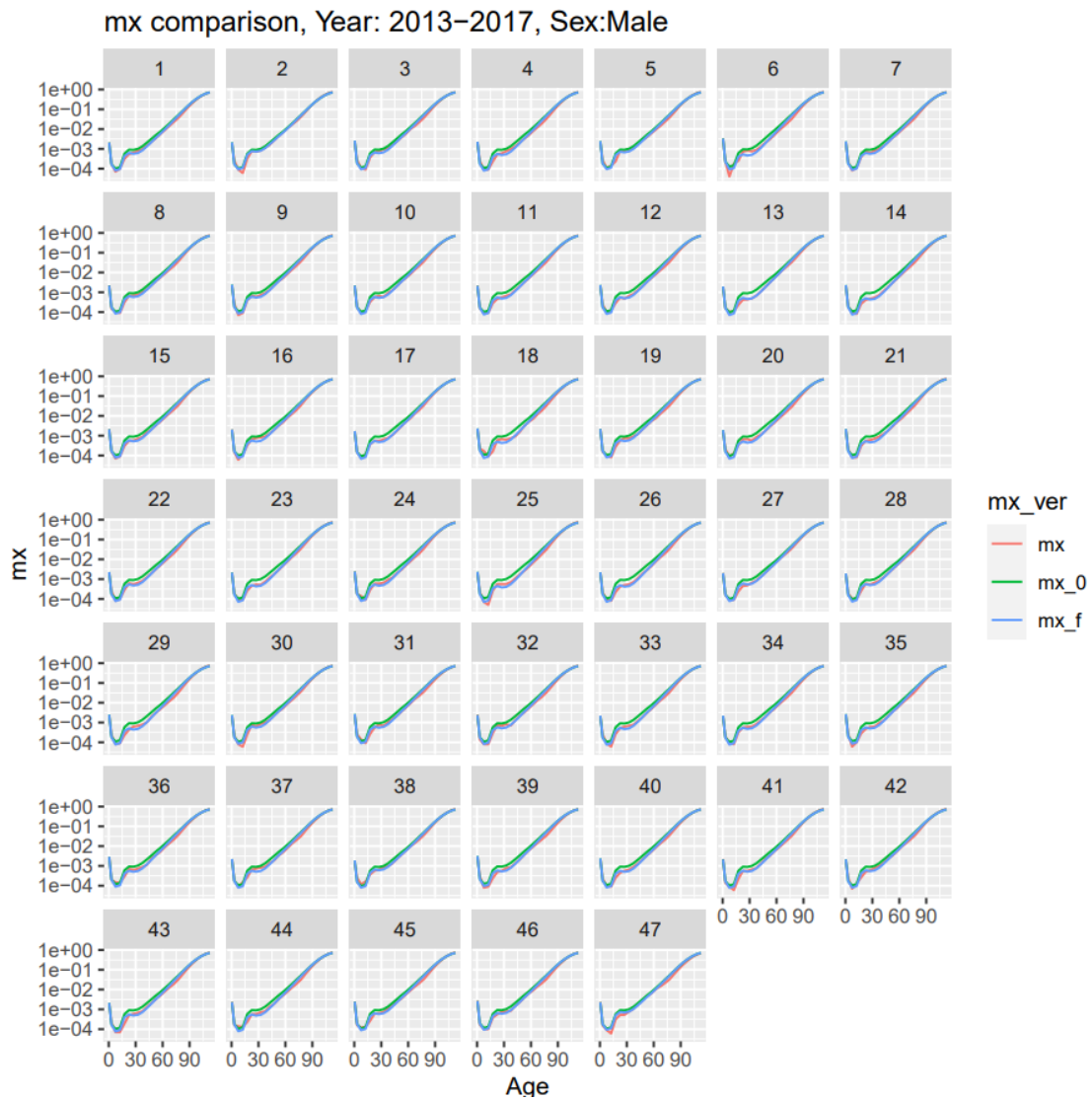


図 2 : flexible two-dimensional model による対数死亡率($\log m_x$)と JMD から得られた実績値(2013~2017 年, 男性)。” mx_0” は flexible model において $k=0$ とした場合の推計結果を表し、mx_f は k を可変とした場合の推計結果を示す。1~47 の数字は都道府県番号を示す。

(出所) 筆者推計

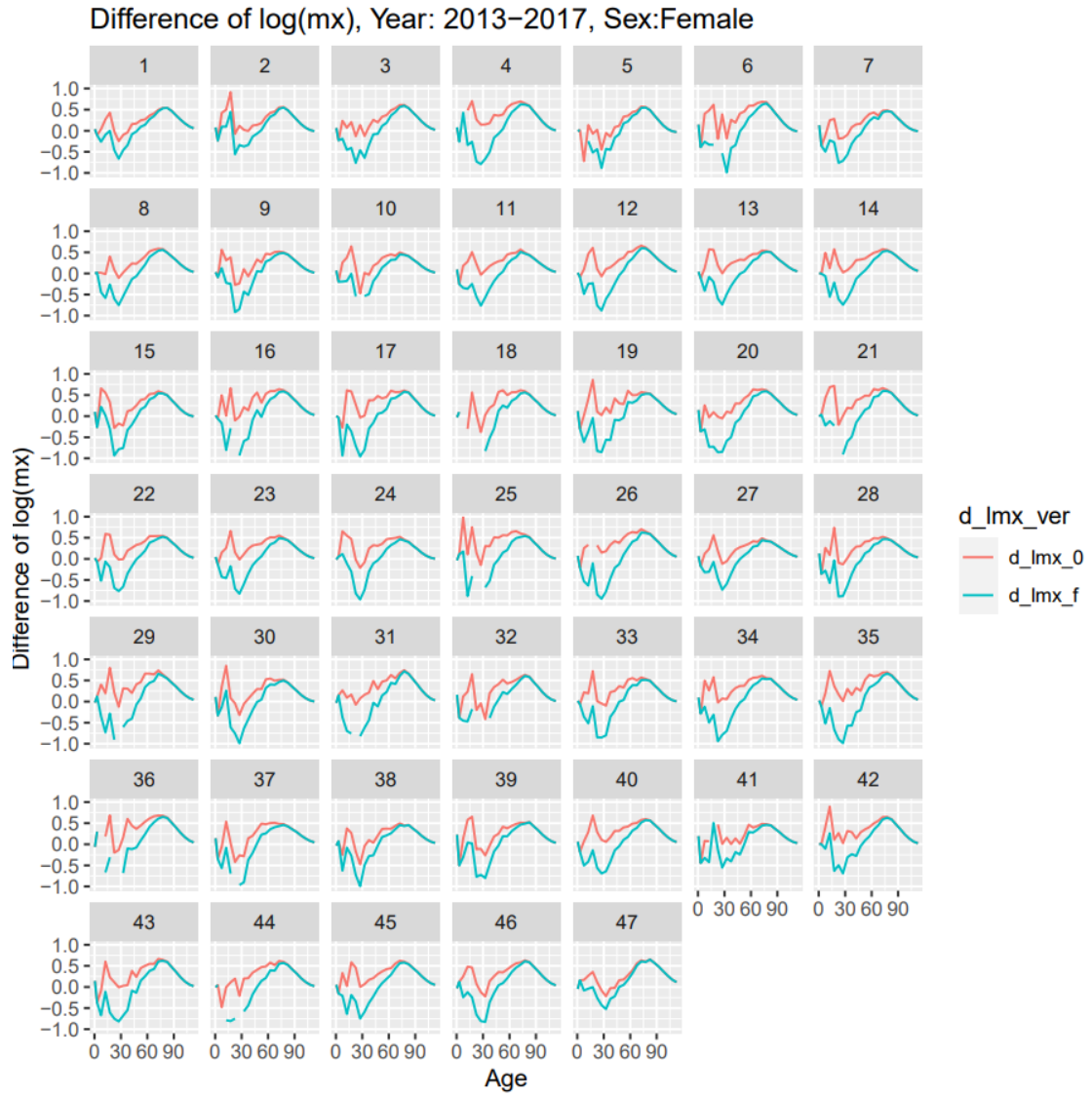


図3 $k=0$ のときの死亡率の誤差(d_{lmx_0})と k を可変とした場合の誤差(d_{lmx_f}) (女性, 2013~2017 年), 横軸に年齢、縦軸に対数死亡率を取っている。1~47 の数字は都道府県番号を示す。

(出所) 筆者推計

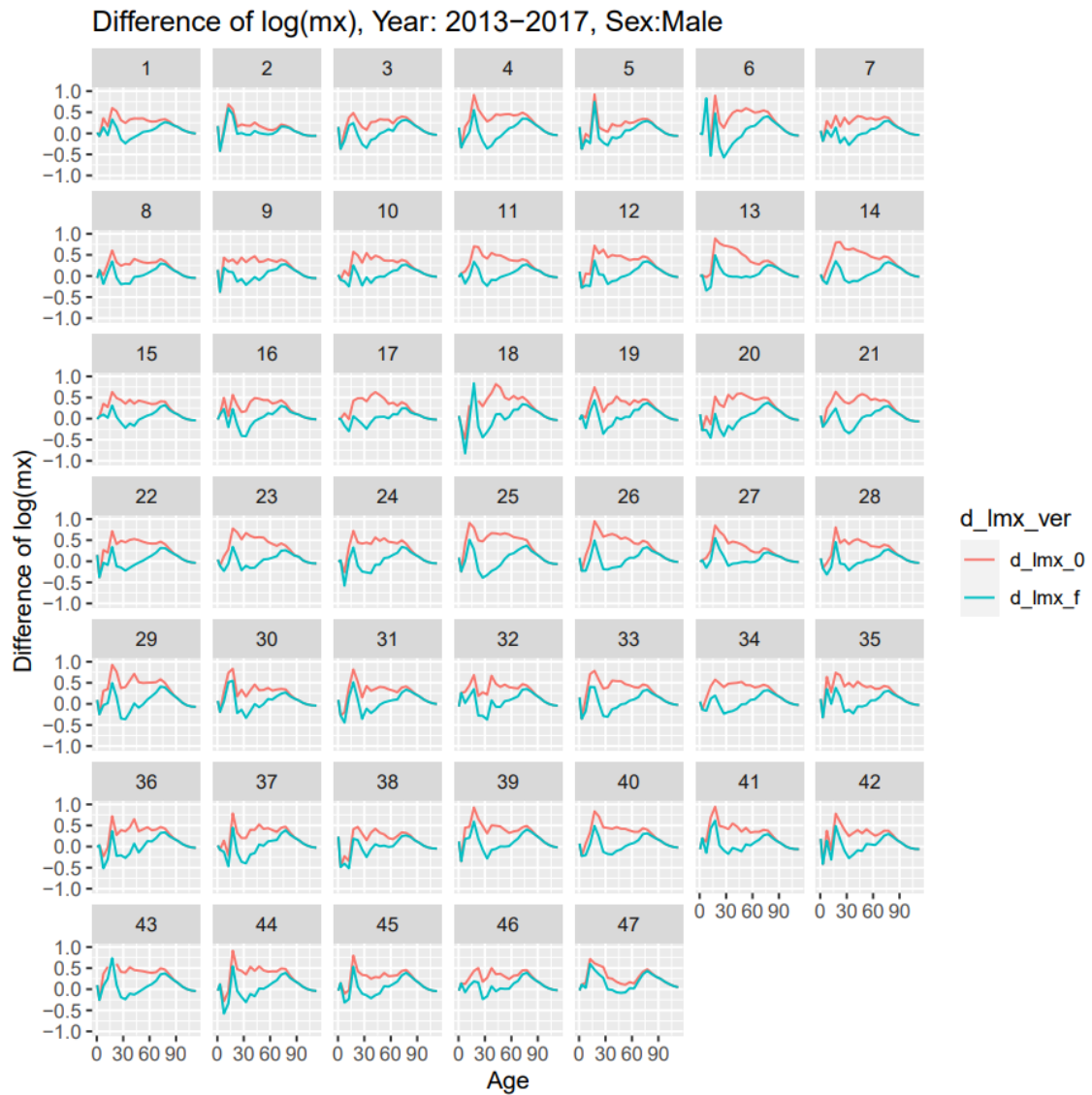


図4 $k=0$ のときの死亡率の誤差(d_{lmx_0})と k を可変とした場合の誤差(d_{lmx_f}) (男性, 2013~2017 年), 横軸に年齢、縦軸に対数死亡率を取っている。1~47 の数字は都道府県番号を示す。

(出所) 筆者推計

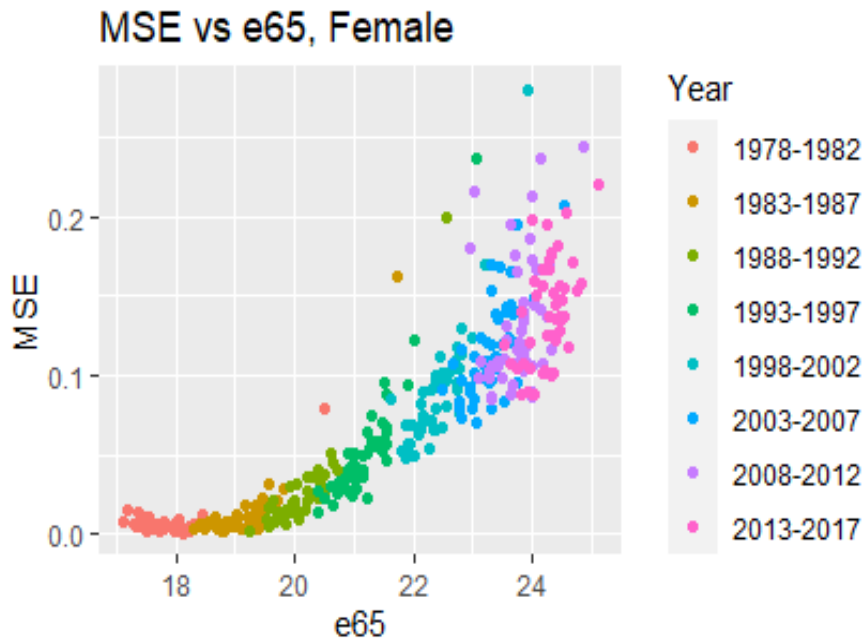


図5 平均2乗誤差(Mean Squared Error, MSE)と65歳時平均余命 e_{65} の関係(女性, 47都道府県)。

(出所) 筆者推計

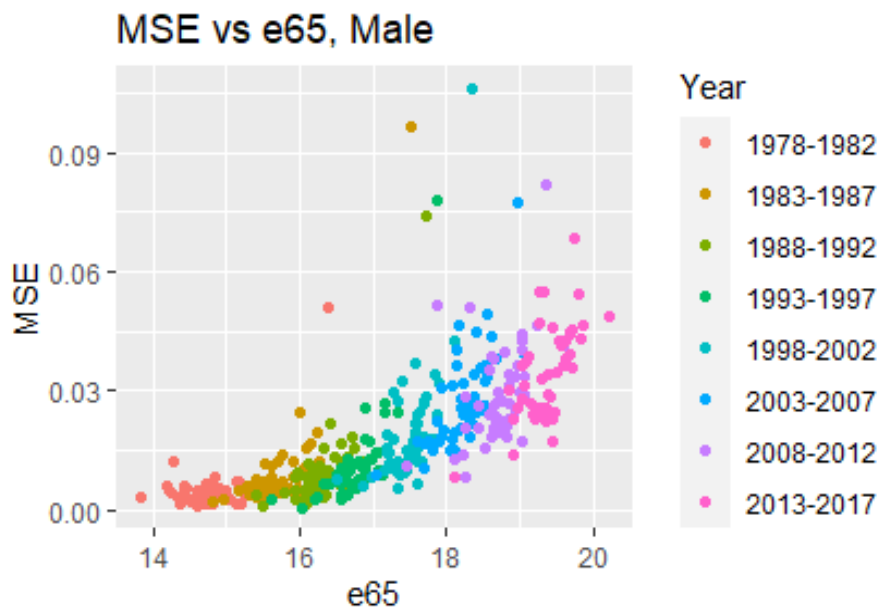


図6 平均2乗誤差(Mean Squared Error, MSE)と65歳時平均余命 e_{65} の関係(男性, 47都道府県)。

(出所) 筆者推計

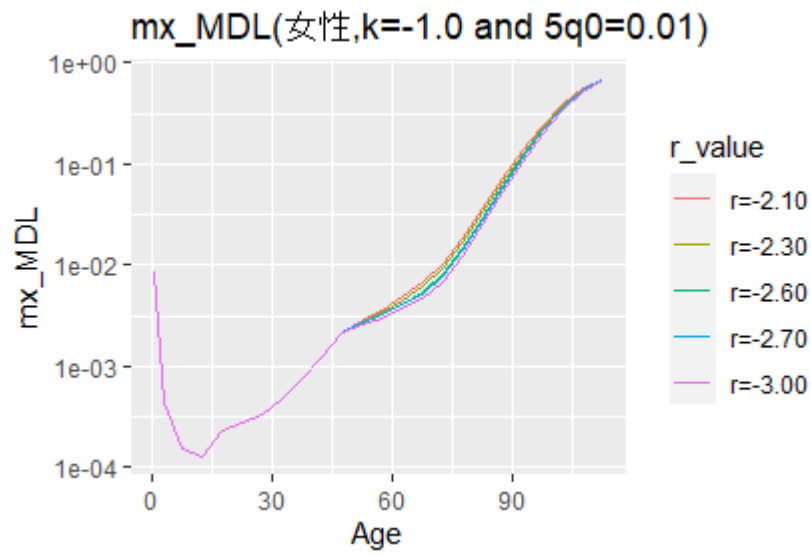


図 7: 修正モデル(mx_MDL)のシミュレーション($r=2.10, \dots, 3.00$) (女性)。0~4 歳時死亡確率 ${}_5q_0$ は、 ${}_5q_0 = 0.01$ とし、 $k = -1.0$ で固定した。
 (出所) 筆者推計

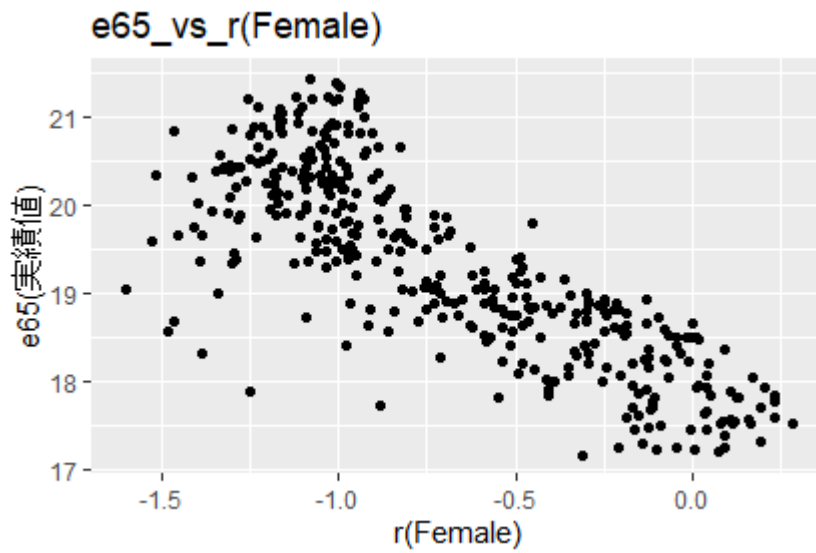


図 8: 65 歳時平均余命 (e_{65}) の実績値と r の関係 (女性)
 (出所) 筆者推計

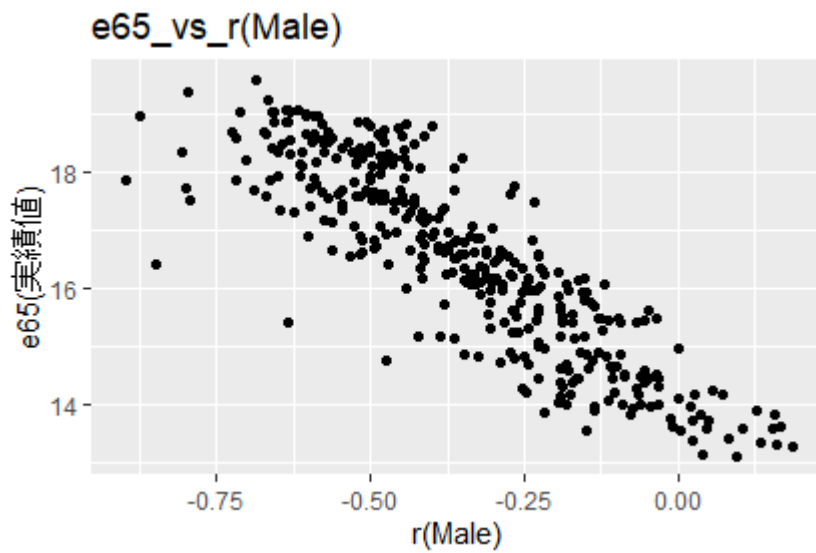


図 9: 65 歳平均余命 (e_{65}) の実績値と r の関係 (男性)
 (出所) 筆者推計

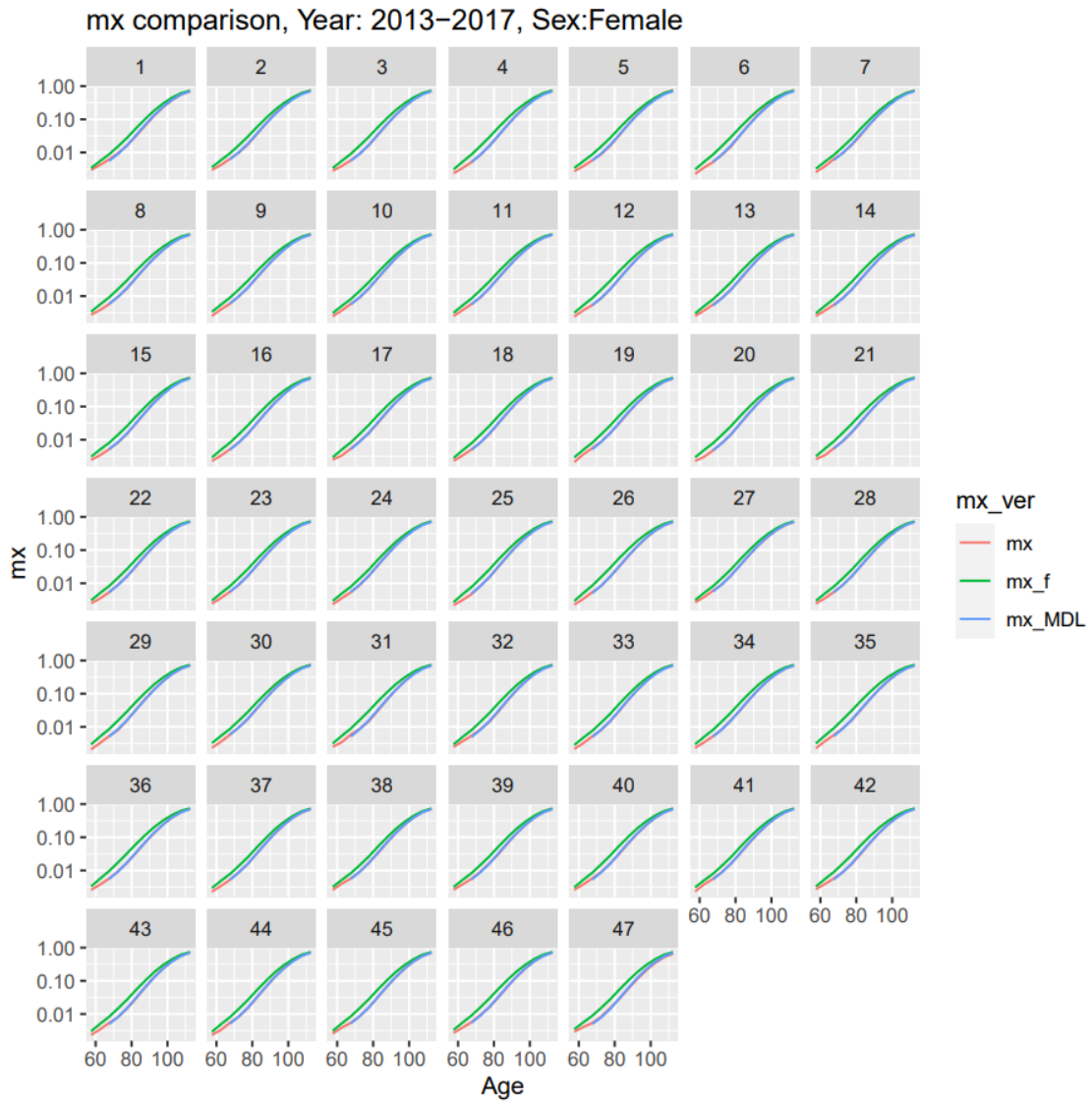


図 10：対数死亡率 $\log m_x$ の比較(2013~2017年，女性，55歳以上)。” m_x ”はJMDから得られた実績値、” m_{x_f} ”は k を可変とした場合のflexible modelから得られた推計値、” m_{x_MDL} ”は修正モデルによる推計結果。

(出所) 筆者推計

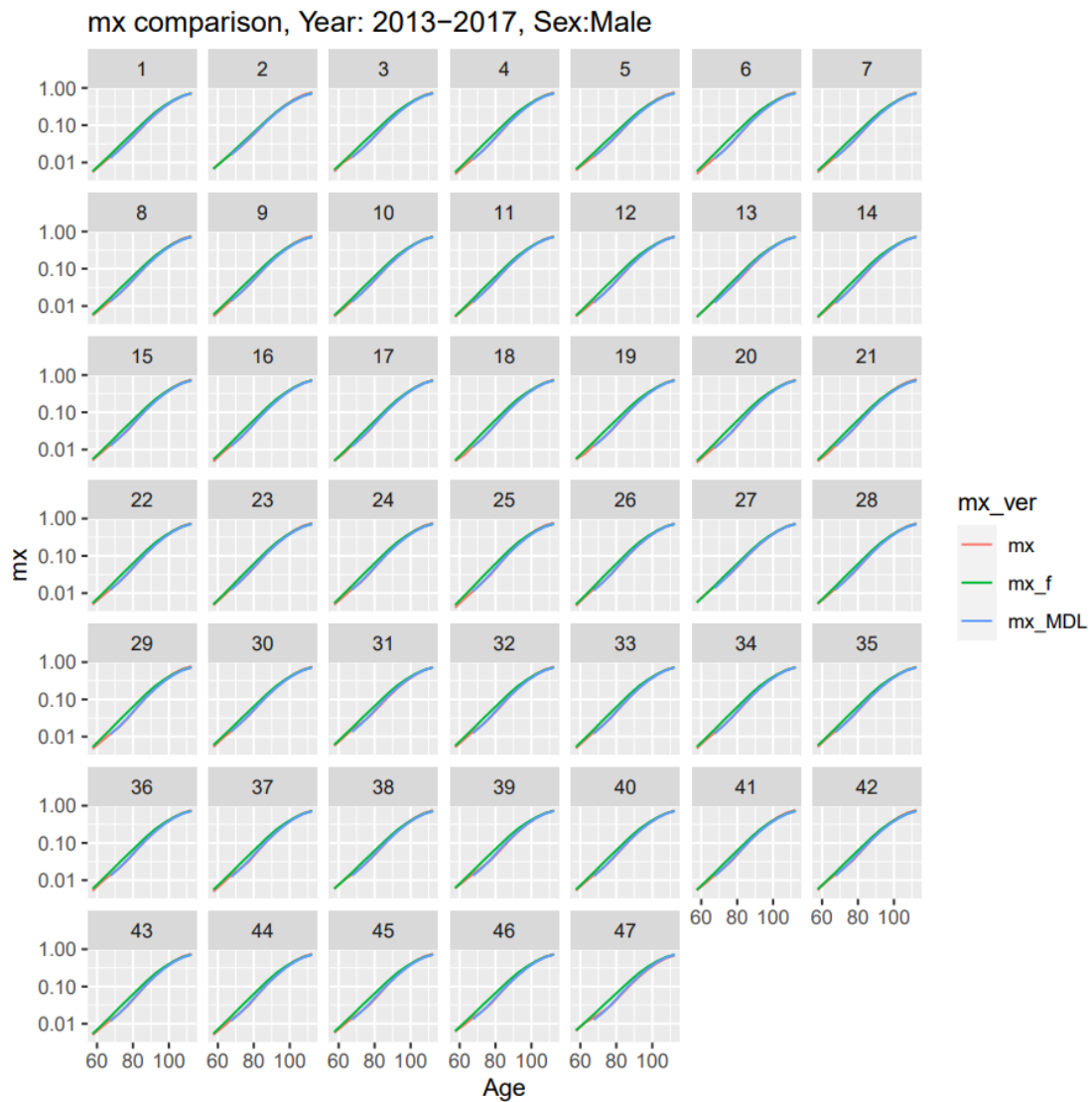


図 11：対数死亡率 $\log m_x$ の比較(2013~2017年，女性 55 歳以上)。”mx”は JMD から得られた実績値、”mx_f”はkを可変とした場合の flexible model から得られた推計値、”mx_MDL”は修正モデルによる推計結果。

(出所) 筆者推計

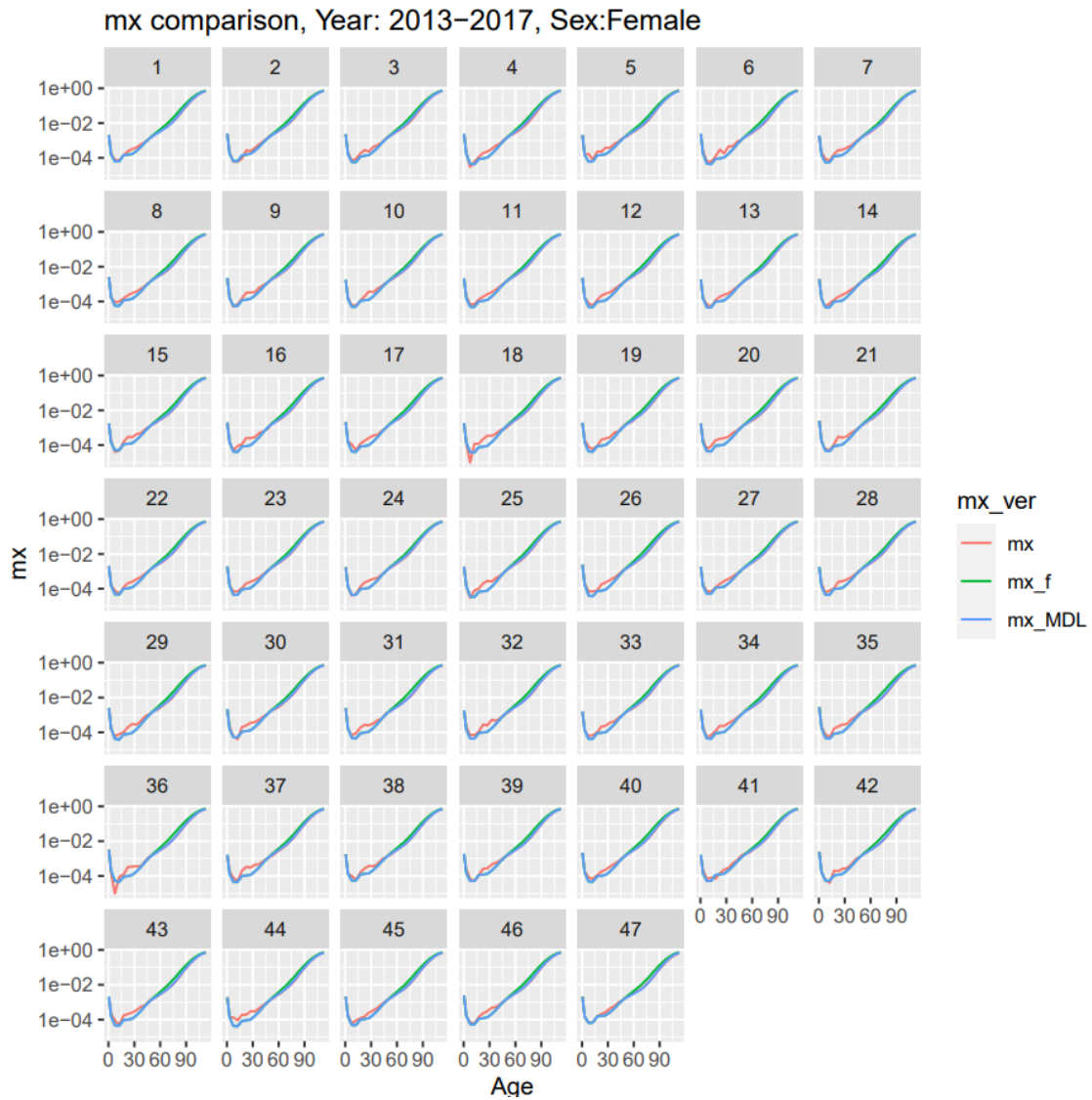


図 12：対数死亡率 $\log m_x$ の比較(2013~2017年，女性，全年齢)。”mx”はJMDから得られた実績値、”mx_f”はkを可変とした場合のflexible modelから得られた推計値、”mx_MDL”は修正モデルによる推計結果。

(出所) 筆者推計

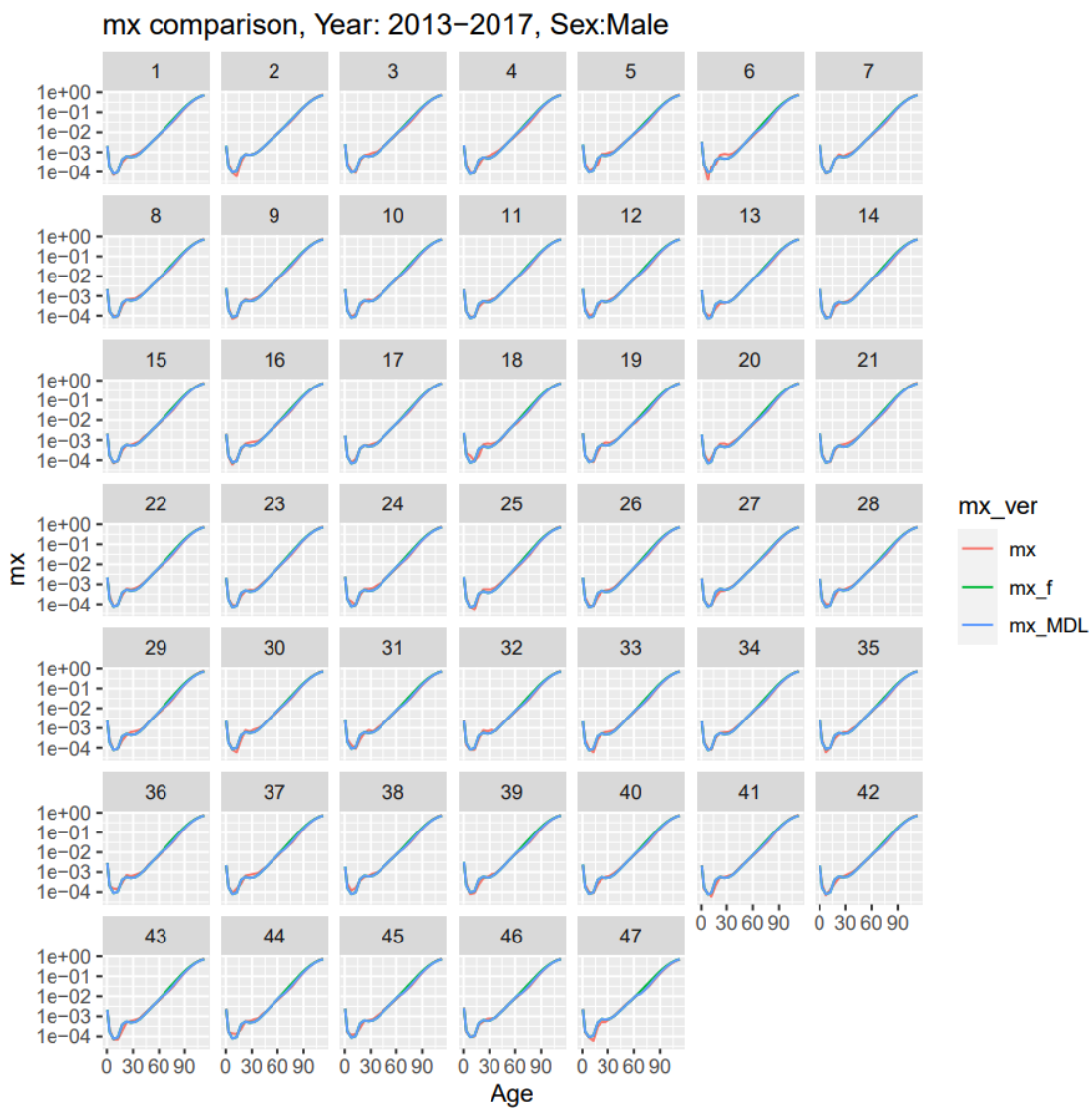


図 13：対数死亡率 $\log m_x$ の比較(2013~2017年，女性、全年齢)。” mx ”はJMDから得られた実績値、” mx_f ”は k を可変とした場合のflexible modelから得られた推計値、” mx_{MDL} ”は修正モデルによる推計結果。

(出所) 筆者推計

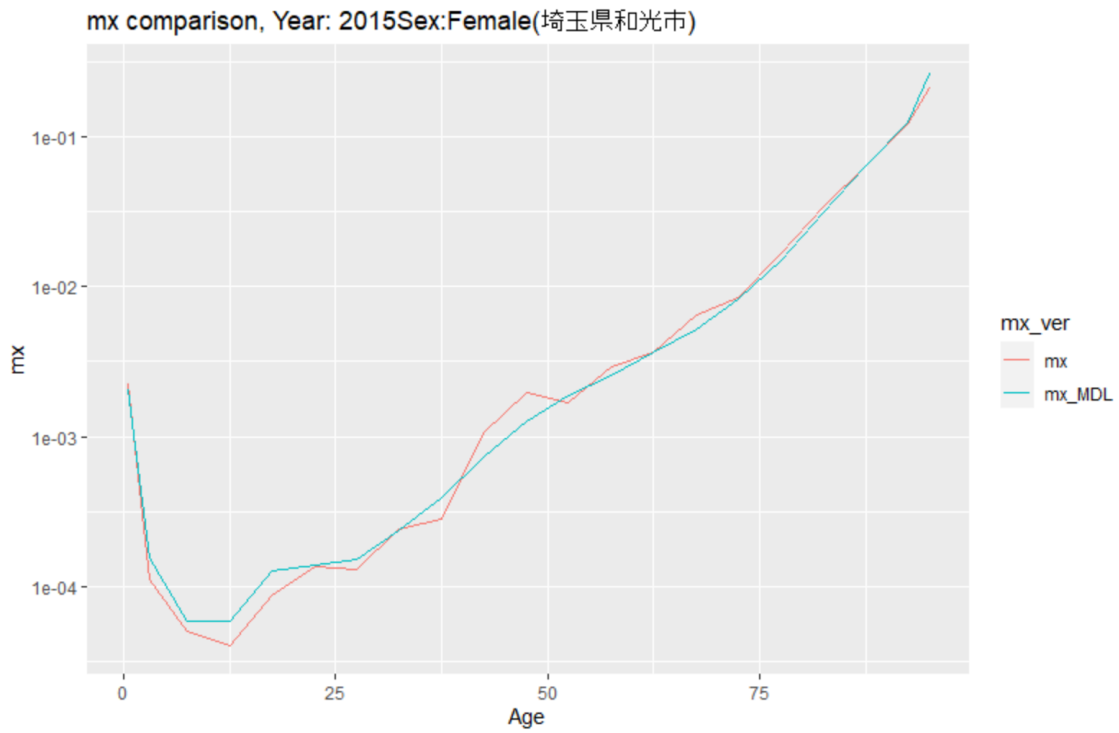


図 14:埼玉県和光市（女性,2015 年）における対数死亡率 $\log m_x$ の比較。”mx”は市区町村別生命表から得られる実績値、”mx_MDL”は修正モデルから得られる推計値を表す。

（出所）筆者推計

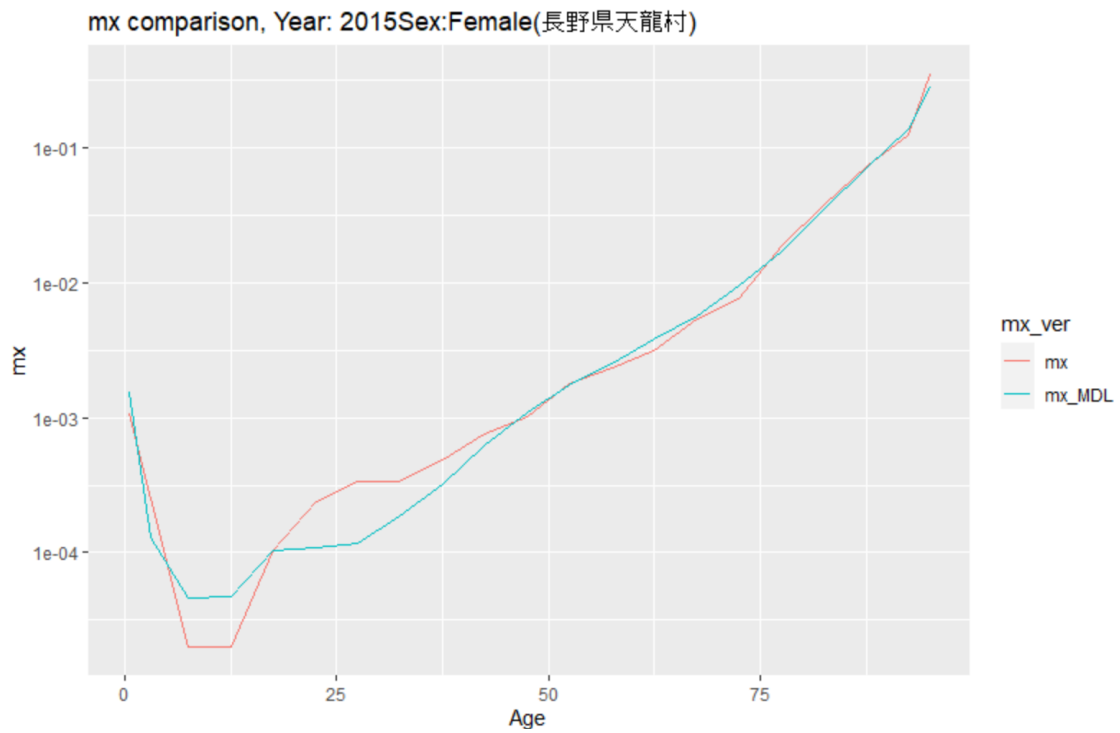


図 15:長野県天龍村（女性, 2015 年）における対数死亡率 $\log m_x$ の比較。“mx”は市区町村別生命表から得られる実績値、“mx_MDL”は修正モデルから得られる推計値を表す。

（出所）筆者推計

表1：修正モデルのパラメータ（男性・女性）

	女性					男性				
年齢	a_x	b_x	c_x	v_x	u_x	a_x	b_x	c_x	v_x	u_x
0	-0.66	0.77	-0.03	0.00	0.00	-0.51	0.82	-0.02	0.00	0.00
1-4	-	-	-	-	0.00	-	-	-	-	0.00
5-9	-2.56	1.79	0.11	0.28	0.00	-0.30	1.53	0.08	0.17	0.00
10-14	-3.24	1.67	0.11	0.34	0.00	-3.96	1.24	0.06	0.17	0.00
15-19	-3.11	1.58	0.11	0.40	0.00	-3.94	1.04	0.08	0.22	0.00
20-24	-2.98	1.51	0.10	0.41	0.00	-3.42	1.17	0.09	0.30	0.00
25-29	-3.02	1.37	0.08	0.39	0.00	-3.42	1.14	0.09	0.36	0.00
30-34	-3.02	1.29	0.08	0.34	0.00	-3.44	1.07	0.08	0.38	0.00
35-39	-3.15	1.11	0.06	0.29	0.00	-3.42	0.96	0.07	0.38	0.00
40-44	-3.27	0.93	0.05	0.22	0.00	-3.38	0.83	0.06	0.35	0.00
45-49	-3.52	0.66	0.03	0.18	0.00	-3.45	0.60	0.04	0.31	0.00
50-54	-3.41	0.56	0.02	0.14	0.10	-3.42	0.40	0.01	0.26	0.00
55-59	-3.26	0.45	0.01	0.12	0.20	-3.41	0.18	0.01	0.20	0.00
60-64	-2.89	0.40	0.00	0.08	0.29	-3.14	0.09	-0.02	0.16	0.00
65-69	-2.66	0.26	-0.01	0.06	0.39	-2.86	0.02	-0.03	0.12	0.42
70-74	-2.29	0.18	-0.02	0.03	0.45	-2.41	0.04	-0.02	0.09	0.53
75-79	-2.04	0.05	-0.04	0.01	0.46	2.04	0.01	-0.03	0.05	0.49
80-84	-1.73	-0.01	-0.03	0.00	0.41	-1.65	0.01	-0.02	0.03	0.86
85-89	-1.45	-0.05	-0.03	0.00	0.35	-1.32	-0.02	-0.02	0.01	0.29
90-94	-1.16	-0.06	-0.03	0.00	0.28	-1.04	-0.03	-0.02	0.00	0.24
95-99	-0.87	-0.06	-0.02	0.00	0.20	-0.73	-0.02	-0.01	0.00	0.14
100-104	-0.63	-0.05	-0.01	0.00	0.13	-0.50	-0.01	-0.01	0.00	0.08
105-109	-0.43	-0.03	-0.01	0.00	0.08	-0.33	0.00	-0.00	0.00	0.04
110+	-0.30	-0.02	-0.00	0.00	0.04	-0.22	0.00	-0.00	0.00	0.01

注：小数点以下第3位を四捨五入した。1～4歳の係数(a_x, b_x, c_x, v_x)については、 ${}_5q_0$ が推計式の説明変数(h)であるため、係数はない。

(出所)：係数(a_x, b_x, c_x, v_x)については Wilmoth et al.(2012), p.10, Table3。 u_x については筆者推計。