

多変量 AR モデルによる年齢各歳別死亡率 の長期予測*

大 場 保

I 人口の将来推計と将来予測

従来より人口の将来推計においては、その作業の最初に死亡率および出生率についてなんらかの形で未来のシナリオと呼ばれる一連の(仮定値)を設定し、現在の値とシナリオを結ぶために平滑化手法と人口学的数理が適用されてきた。

例えば1994年の国連による将来推計¹⁾においては、出生率に関しては将来の TFR が2.1になるなどとされたし、死亡率に関しては平均寿命が女で87.5年、男で82.5年という値に近づくというシナリオが描かれた。そして、現在の値からシナリオで設定した将来の仮定値に到達するまでの間を、平滑化手法と人口学的な数理により曲線で結ぶ作業が将来推計の中核だったのである。

本研究においては未来のシナリオは用いない。過去から現在までの時系列的变化を統計学的に将来へ延長する。そのための手法として多変量 AR モデル (Multivariate AutoRegressive Model) を採用し、日本の性年齢各歳別年次別死亡率に適用した。これにより、将来推計ではなく将来予測を実現した。

II 多変量 AR モデル

最初にここで用いた多変量 AR モデルについて触れておく。このモデルは、 N 個のデータ長よりなる l 変量の時系列において、 n 番目のデータ (ベクトル) を \mathbf{y}_n と表すとき、次式により与えられる。

$$\mathbf{y}_n = \sum_{i=1}^m \mathbf{A}_i^m \mathbf{y}_{n-i} + \mathbf{v}_n \quad (2.1)$$

ここで、 \mathbf{v}_n は l 次の残差ベクトルであり、 $\mathbf{v}_n \sim N(\mathbf{0}, \mathbf{V}_m)$ である。 m は AR 次数と呼ばれる。 \mathbf{A}_i^m は

1) United Nations, World Population Prospects The 1994 revision, pp.142-151, 1994.

$$A_i^m = \begin{pmatrix} a_i^m(1,1) & a_i^m(1,2) & \cdots & a_i^m(1,l) \\ a_i^m(2,1) & a_i^m(2,2) & \cdots & a_i^m(2,l) \\ \vdots & \vdots & \ddots & \vdots \\ a_i^m(l,1) & a_i^m(l,2) & \cdots & a_i^m(l,l) \end{pmatrix} \quad (2.2)$$

なる $l \times l$ 行列であり、AR 次数が m 次のときの自己回帰係数行列と呼ばれる。 $a_i^m(j, k)$ は、 m 次の AR 次数のうちの i 次の自己回帰係数である。

実データを (式2.1) に代入することで、 A_i^m および v_n の分散共分散行列 V_m を推定し、これらを用いて時系列の長期予測、あるいは補間推定ができる。

長期予測は、(式2.1) における残差項を略し、

$$y_{N+1} = \sum_{i=1}^m A_i^m y_{N+1-i} \quad (2.3)$$

という操作を必要な年数分だけ繰り返せばよい。したがって本研究では、原データに対して適当な変換処理を行った後に (式2.1) を推定し、(式2.2) を繰り返し用いて将来の予測値を得るというステップを経る。

多変量 AR モデルの推定には、レビンソンのアルゴリズムを用いてユール-ウォーカー方程式を解く方法と、ハウスホルダー変換に基づく最小二乗法などがあるが、ここでは、変量ごとに m を変えられる後者を採用した。詳細については北川²⁾を参照されたい。

最小二乗法によるモデル推定プログラムは、北川²⁾を参考に Microsoft FORTRAN Power Station 4.0で作成し、PC-9821 Ap2の Windows NT 3.51上でコンパイルおよび実行した。

Ⅲ 死亡率データの検討

過去の死亡率データに対して多変量 AR モデルを適用するに際して、死亡率の時系列にそのまま適用すればよいのか、あるいは対数値に変換したものがよいのか、さらには対数値の差分がよいのかを検討した。

今回用いた性年齢各歳別年次別の死亡率は、厚生省大臣官房統計情報部の人口動態統計の性年齢別年次別死亡数と、総務庁統計局の国勢調査報告および推計人口の性年齢別人口を用いて1947～1994年について算出した。

1995年は、次の理由で用いていない。例えば1994年における11歳女子の死亡数は56人である。阪神淡路大震災による死者は約6000人にもものぼったため、性別に考えると、1年階級あたりの死者はおおよそ30人程度である。これは、上記年齢の女子の場合、いきなり50%程度も増加したことに相当する。残念ながら、時系列の最終データだけこのような異常値を含むような場合でも問題なく処理できるモデルはほとんどない。今回のモデルも、

2) 北川源四郎, 「FORTRAN-77時系列解析プログラミング」, 岩波書店, 1993.

このような突然の変化に対応できるほどロバストではない。

(i) 死亡率

年齢各歳別年次別死亡率 $M_{a,t}$ は、年齢 a の増加とともに指数関数的な変化がある。また、年次方向にも急激な変化が観察されるが、これが単純な指数関数なのか修正指数曲線のような若干複雑なものかはっきりしない。

(ii) 死亡率の対数値 (図 1-M (ale), 図 1-F (emale))

$\log_{10}(M_{a,t})$ とすると指数関数的な変化はなくなり、年齢方向、年次方向での緩やかな曲線的な増加が見られる。これにモデルを適用した場合曲線的な増減の傾向が結果に反映されにくい難点がある。

図 1-M 年齢別死亡率 (常用対数値, 男)

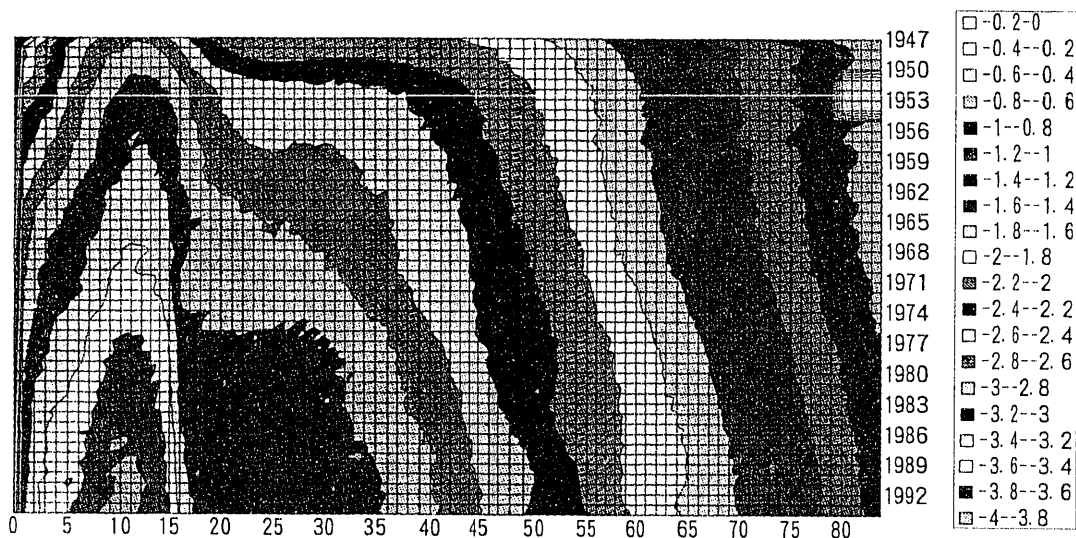
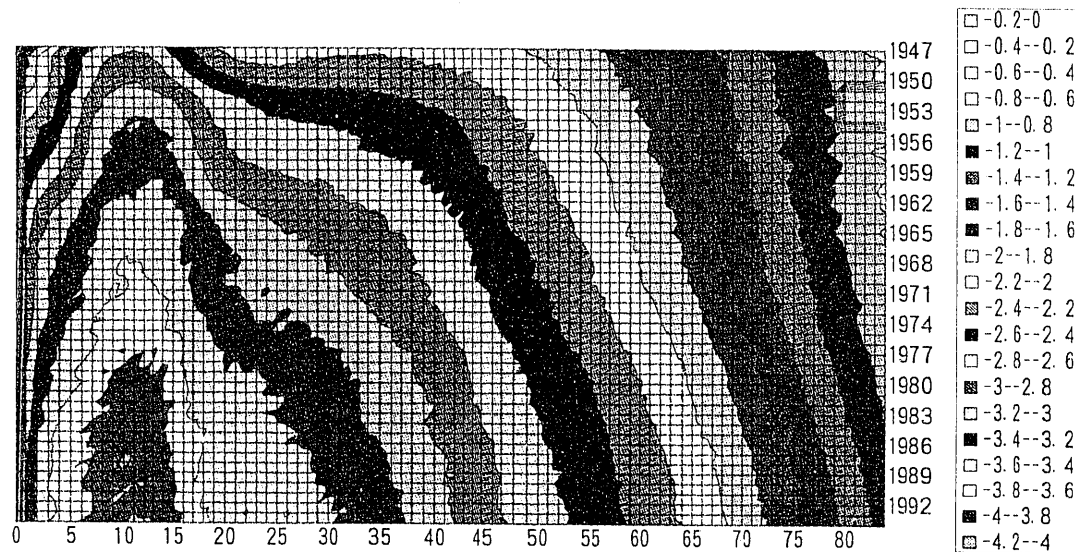


図 1-F 年齢別死亡率 (常用対数値, 女)



(iii) 死亡率の対数値の年次方向での差分 (図2-M, F)

年次方向での差分 $\log_{10}(M_{a,t+1}) - \log_{10}(M_{a,t})$ では、年次ごとの特徴を示す水平方向の尾根と谷がみられ、また、コーホート方向 (斜め方向) での尾根や谷がみられる。

図2-M 年次方向での死亡率差分 (男)

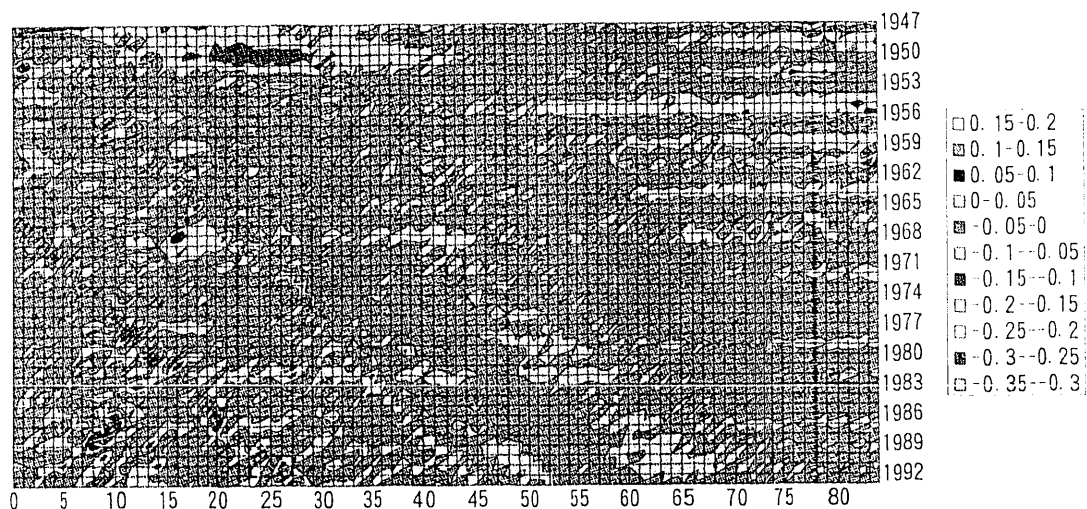
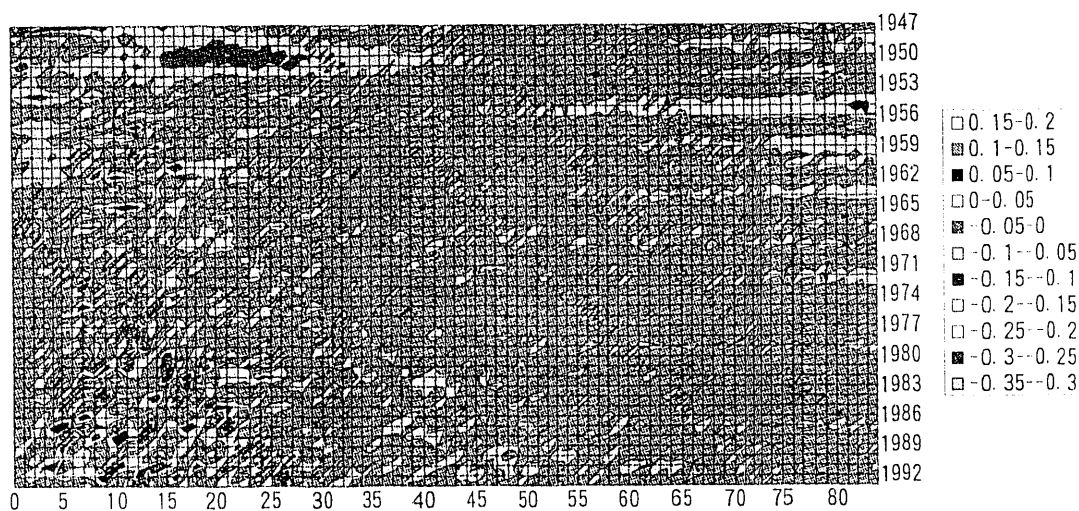


図2-F 年次方向での死亡率差分 (女)



(iv) 死亡率の対数値の年齢方向での差分 (図3-M, F)

年次の場合と対照的に、 $\log_{10}(M_{a+1,t}) - \log_{10}(M_{a,t})$ では縦方向の尾根と谷、加えてコーホート方向の尾根と谷がみられる。また、男子15 (~21) 歳においては1968年以降突出した「塀」がみられる。ただし、この年齢範囲の15歳という下の値はかなりはっきりしているが、上の21歳というのはかなり曖昧である。

男子に塀が見られる一方で、女子にはそのようなはっきりした変化は見られない。しかし、1968年以降、15歳の差分は上下にランダム変動をしつつも増加傾向にある。

図3-M 年齢方向での死亡率の差分（男）

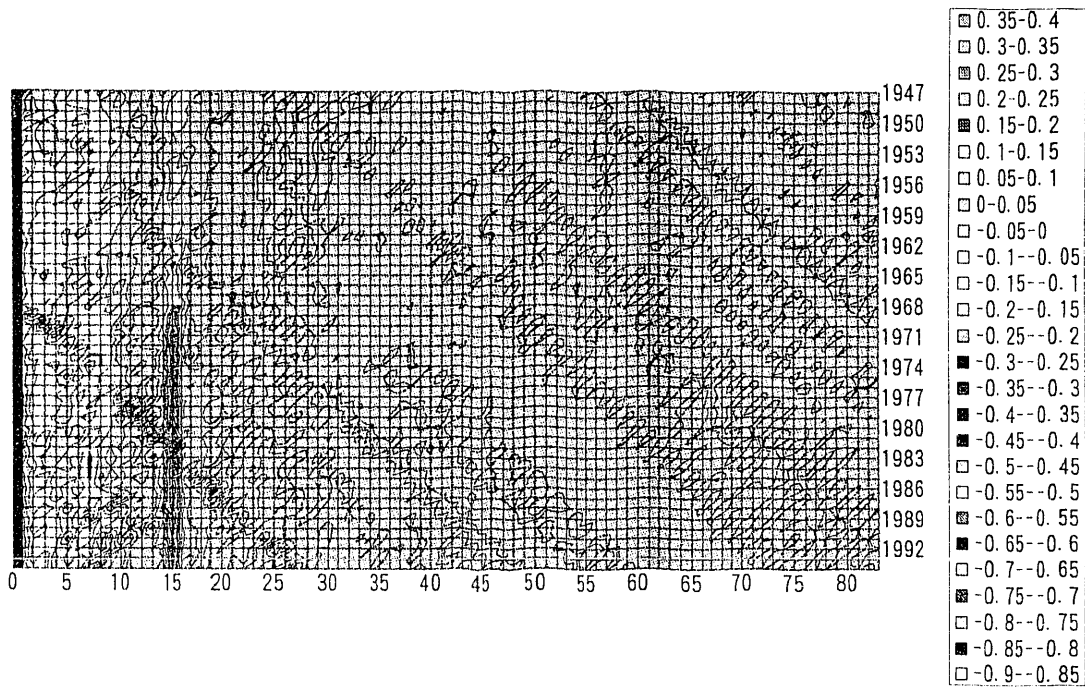
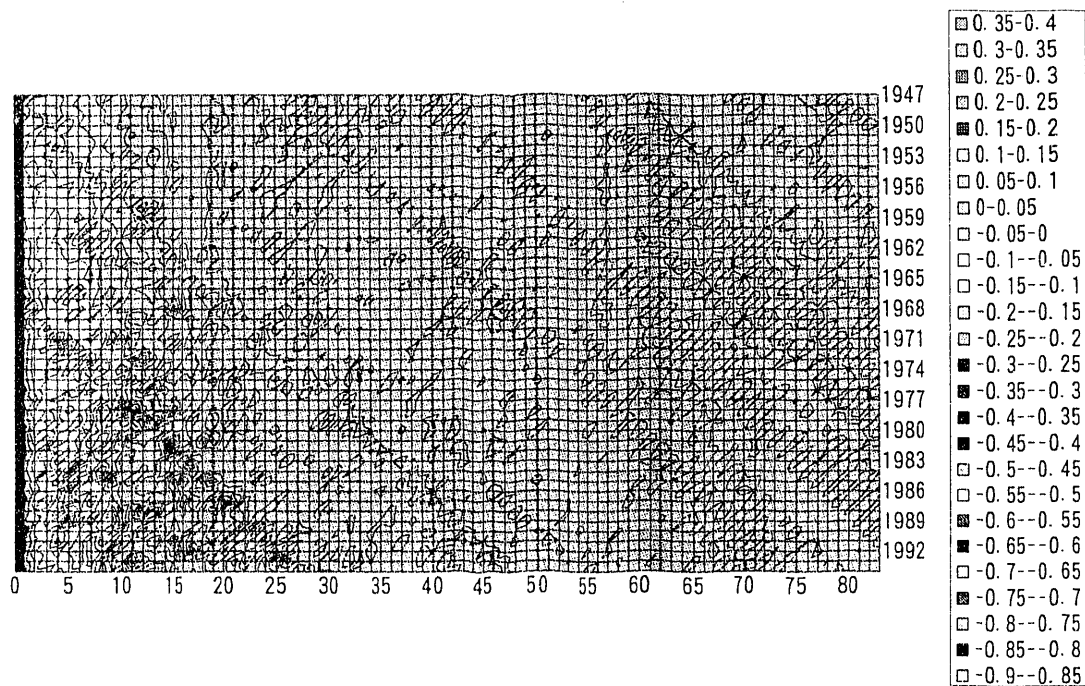


図3-F 年齢方向での死亡率の差分（女）



(v) 死亡率の対数値のコーホート方向での差分（図4-M, F）

年次の場合や年齢の場合と異なり、 $\log_{10}(M_{a+1,t+1}) - \log_{10}(M_{a,t})$ ではコーホート方向の尾根と谷がみられない一方で、年次方向および年齢方向での尾根と谷がみられる。また、男子15（～21）歳においては年齢方向の時と同様に1968年以降突出した「塀」がみられる。

図4-M コーホート方向での死亡率差分（男）

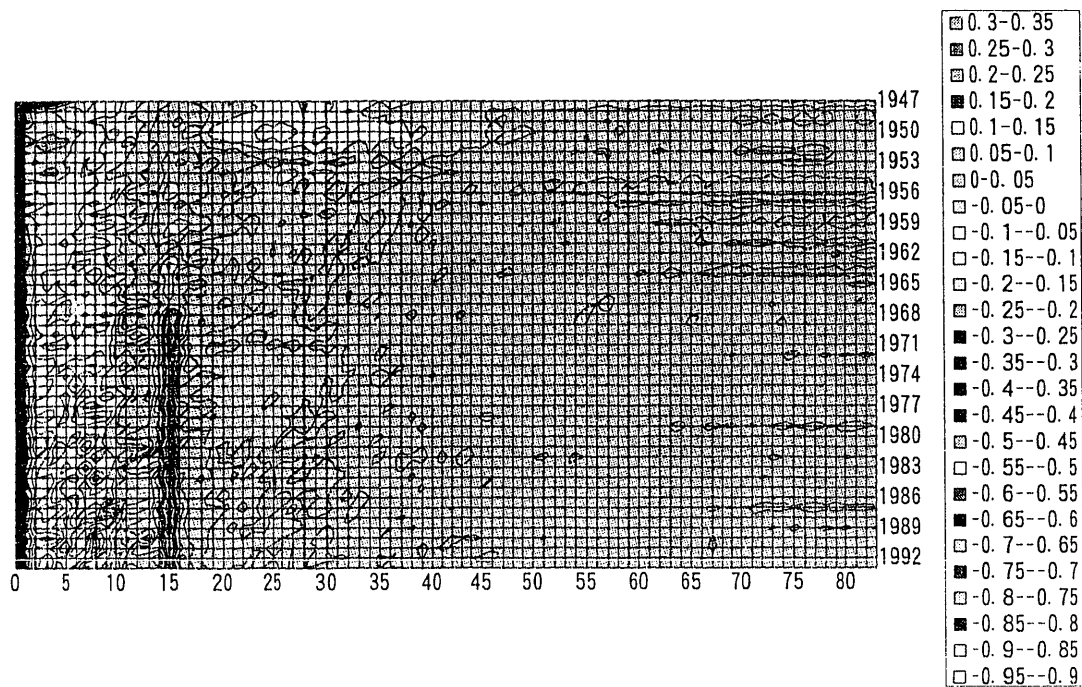
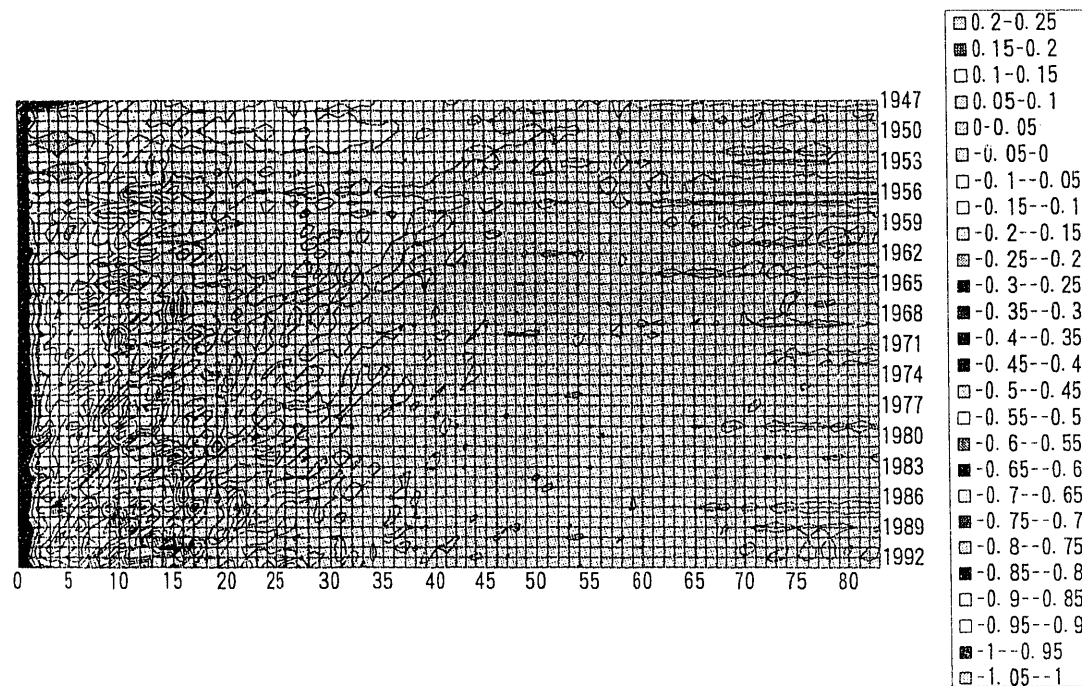


図4-F コーホート方向での死亡率差分（女）



コーホート差分は年齢方向の差分も含んでいるので、15歳における1968年の男女の変化については、年齢方向の差分の場合と事情はほとんど同じである。

以上のものを比較して、時系列としては(Ⅲ), (Ⅳ), (Ⅴ)の各差分がかなり直線的な変化を見せていることから、これらの中から選ぶこととした。

コーホート方向の差分には、年齢及び年次方向の差分に有るコーホート方向での尾根や谷がなく、最も安定した変化をみせているので、これを採用することとした。年次方向および年齢方向のものにはコーホート方向の尾根や谷が存在し、これがデータの分散を増加させる。

コーホート差分を採用するとして、時系列長をどれだけにするか、これも検討した。なるべく長くしたい一方で、戦後まもない混乱期は、戦争で疲弊した人々が活力を取り戻すにつれて急激な死亡率の低下が見られ、これはある種の異常事態でもある。実際、1947年からの10年間ほどはコーホート方向の差分の極めて急激な低下が観察される。そこで、この時期をなるべく外し、かつ安定した長期予測が得られるだけの時系列長として、時系列長を36年間（1958～1993年）とした。後述するように、 $((m=)3+1) \times ((l=)8+1)$ が比較的安定していたためである。

また、コーホート方向の差分だけでは0歳の将来死亡率および将来生まれるコーホートの死亡率は予測できないので、0歳に関しては年次方向の差分を採用することとした。こちらの時系列長であるが、1歳以上の全年齢を対象とするコーホート差分の場合と事情が異なり、戦後まもない頃の大きな変化は見られない1948～93年とした。

IV 将来予測値作成のプロセス

本研究での将来の予測値を求めた手順を以下に示す。

(i) 差分の算出

性別年齢各歳別年次別死亡率の対数値 $\log_{10}(M_{a,t})$ より、その年次方向の差分 $\log_{10}(M_{a,t+1}) - \log_{10}(M_{a,t})$ とコーホート方向の差分 $\log_{10}(M_{a+1,t+1}) - \log_{10}(M_{a,t})$ を求めた。

(ii) 異常値の補正

上述のように、男子15歳近辺には1968年以降に異常な差分値の突出がある。このような階段状のレベルの変化は、今回のモデルがガウス分布を前提としているだけに、単純なARモデルでは追従できない。そこで、それ以前と平均値を比べて、これより突出分を算出し、1968年以降の値から差し引いて、ここでの階段状の変化を補正した。

このような場合、時系列を2つに分割してそれぞれの部分にARモデルを適用する方法として局所定常ARモデルがあるが、それではもともと短い時系列のデータ長がさらに短くなることと作業全体が複雑化することから、今回は用いないこととした。

(iii) リニアトレンドの除去

ARモデルは平均値＝一定である必要があるので、各時系列データに最小二乗法で回帰直線を当てはめ、それとの差分に変換した。

(iv) 正規化

ARモデルでは正規分布を前提としているので、(iii)で求めた値を標準偏差で除して正規化した。これで大数の法則により概ね標準正規分布に従う時系列が得られる。

(v) 同時に予測計算する時系列の組み合わせ

実際のデータにおいては時系列のデータ長に限りがあるため、同時に扱える時系列の数にも限りがある。今回用いたモデルには、データ長 N 、変量数 l 、AR 次数 m の間に次のような制限がある。

$$(l+1) \times (m+1) \leq N \quad (4.1)$$

一方、例えば中学生の年齢においては近年は死亡数が非常に少なく、女子の場合で年間の死亡数が50程度という年齢もあり、ランダム変動の影響を強く受けている。そのため、この年齢層だけで予測を行おうとしても、ランダム変動が大きすぎて実用にならなかった。

そこで、死亡数が比較的多い他の年齢層の時系列と組み合わせで予測を行った。ただし、よりよい組み合わせを追求することが本研究の主旨ではないので、重回帰分析におけるステップワイズ法のような複雑な方法は用いず、次のような単純な方法で行った。

まず0歳から84歳までの時系列を3または4に分割した。3に分割した場合なら0-27, 28-55, 56-84である。そしてこれらから一つずつ抜き出して、順に並べた。{56, 0, 28, 57, 1, 29, ...} となる。

(vi) 多変量 AR モデルによる予測

(v)の時系列に対して多変量 AR モデルによる予測計算を行う。ここでは1958~1993年の死亡率差分に対して、その将来予測値の分散が最小になるような A^m を求め、これを用いて1994年以降2035年までの長期予測値を求めた。ただし、最後の5年間は、移動平均を用いた平滑化の際に利用するためであり、最終的に死亡率として求めたのは2030年までである。また、AR 次数 m と変量数 l は、 l を2から11まで変化させてみて、予測誤差分散が最も小さく、予測値の振動発散がほとんどみられなかった $m=3, l=8$ とした。

前項で得られた年齢ごとの時系列順序 {56, 0, 28, 57, 1, 29, ...} に対して、左側から8変量ずつ順次組み合わせ、対象を1変量ずつ右にシフトさせながら、予測区間の分散が最小になる場合を順次求めた。

(vii) 逆正規化およびリニアトレンドの付加

(iv)および(iii)の手順の逆を行い、データを復元した。

(viii) 塀の復元

男子15歳の1968年以降の値から差し引いておいた値を再び加え、突出した塀を復元した。

(ix) 平滑化

AR モデルによる予測値は上下変動を伴うが、予測している値が差分だけに一つの大きな変動が後々まで強い影響を与えてしまう。そこで、年次方向および年齢方向の移動平均を求めて平滑化した。ただし、年齢方向の平滑化は、上下動の比較的少ない20歳以上に限った。

(x) 差分から率を求める

時間方向の差分より0歳の将来予測死亡率(対数値)を求め、次いでコーホート方向の

差分より1～84歳の将来予測死亡率（対数値）を求めた。この指数をとって将来予測死亡率を求めた。

(xi) 85歳以上の補外

得られた予測死亡率に対して、非線形最小二乗法³⁾を用いてコーホート方向にメイカム曲線 ($M_{a,t_0+a} = \alpha + \beta e^{ra}$) を当てはめ、85歳以上の高年齢部分を補外した。実績値についても平均寿命を求める必要性から補外を行ったが、データ長の問題からコーホート方向での補外は困難だったため、年齢方向の補外 ($M_{a,t} = \alpha + \beta e^{ra}$) である。

V 結果と考察

得られた結果について、死亡率の差分、死亡確率、サバイバルカーブ、平均寿命、平均寿命の男女差の順に見ていき、加えて考察も行う。

1. 死亡率の差分の推移

死亡率の差分は、多変量 AR モデルで直接求められた値であるので、どのような結果が得られたかを少し丁寧にみていく。図5-M, Fに0, 5, 10, 15, 30, 50, 78歳についての差分の推移を示す。ただし、0歳は年次方向の差分、他はコーホート方向の差分であり、1994年以降が今回求めた予測値である。

最初に目に付くのは、男子15歳における差分の異常な高さであろう。すでに述べたように、実績値は1968年から急激に上昇し、以後ほぼ一定の高さを保っている。よって予測値も、若干低下気味とはいえ今後ともその高さを保ち続けるものとなっている。

15歳の女子ではどうかというと、実績値が、男子ほどはっきりした階段状のレベル変化を見せておらず、上下動を繰り返しつつも徐々に増加しつつある。予測値も今後は次第に上昇し、2030年には男子並の0.2という値に近づいていく。

次に、低年齢の死亡率に大きな影響を与える0歳の差分をみると、男女ともに過去から現在までマイナスであるが微増傾向にあり、これが2030年頃まで続く結果となった。

5歳では、男女ともに過去から現在までマイナスの値であったが、同時に一貫して増加傾向にあり、今後はプラスに転じて増加が続く。10歳では、男子では緩やかな上下動をしつつも増加し、女子では大きくゆれながらもやはり増加していく。10歳では死亡数自体がかなり少ないこと、原データ自体がかなり大きな変動を伴って推移してきたことから、AR モデルによる予測値も若干の上下動を示す結果となっている。30歳では、今後とも緩やかに増加、50, 78歳ではほとんど横這いのままである。

以上は、特定の年齢に限ってみてきたが、次にすべての年齢における差分の変化を、1950年から10年ごとの特定の年次についてみる（図6-M, F）。

これを見てまず気が付くことは、1950年はほぼ全年齢にわたってマイナスの値であった

3) 岸田孝一、藤井良治、「FORTRAN-80版 マイコン統計手法」、秋葉出版、pp.67-75.

図5-M 死亡率差分の推移（男）

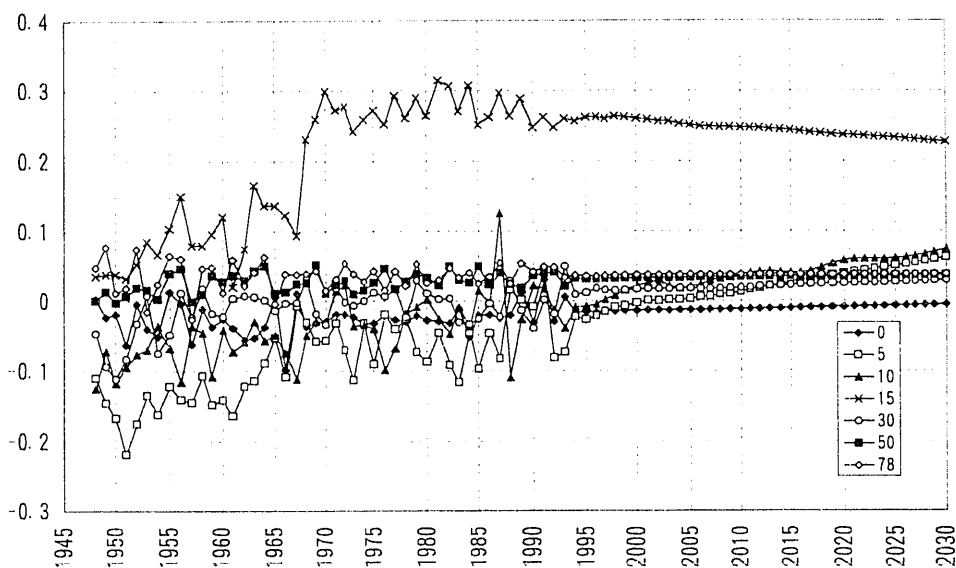
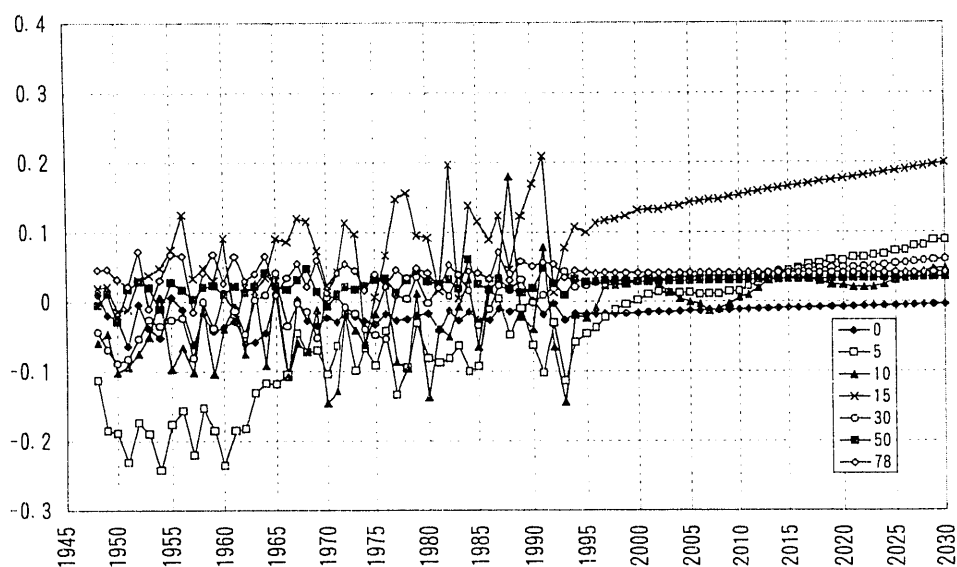


図5-F 死亡率差分の推移（女）



ものが、年を経るに従い次第に増加し、多くの年齢でプラスの値になっていくという点であろう。特に男子で5歳前後、10～12歳で大きく増加、女子では5歳と15歳で大きく増加する。さらに25歳以上では男女ともに暫増し、とりわけ女子の24～49歳の働き盛りの年齢における増加が大きいことがわかる。

なぜ女子の働き盛り年齢層でこれほど増加するという予測結果になったかを見るため、女子20～49歳の5歳階級ごとのコーホート差分の推移を図7に示す。5歳ごとの合計値で示してあるが、25～49歳の各年齢階級で、過去から現在まで一貫した増加が観察され、これを将来に延長した結果として、今後の差分の増加となっているのが分かる。

図6-M 年齢別死亡率差分, 10年ごと, 男

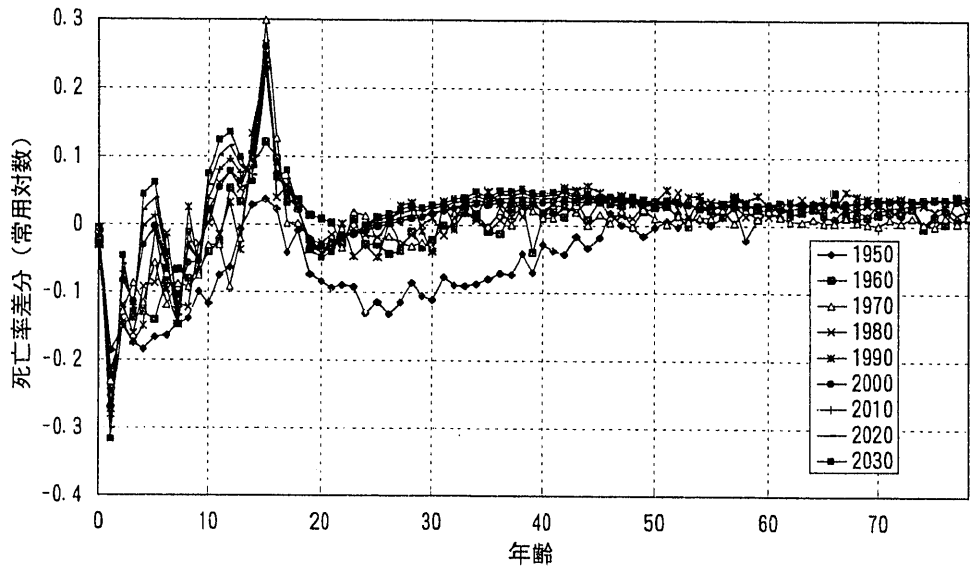
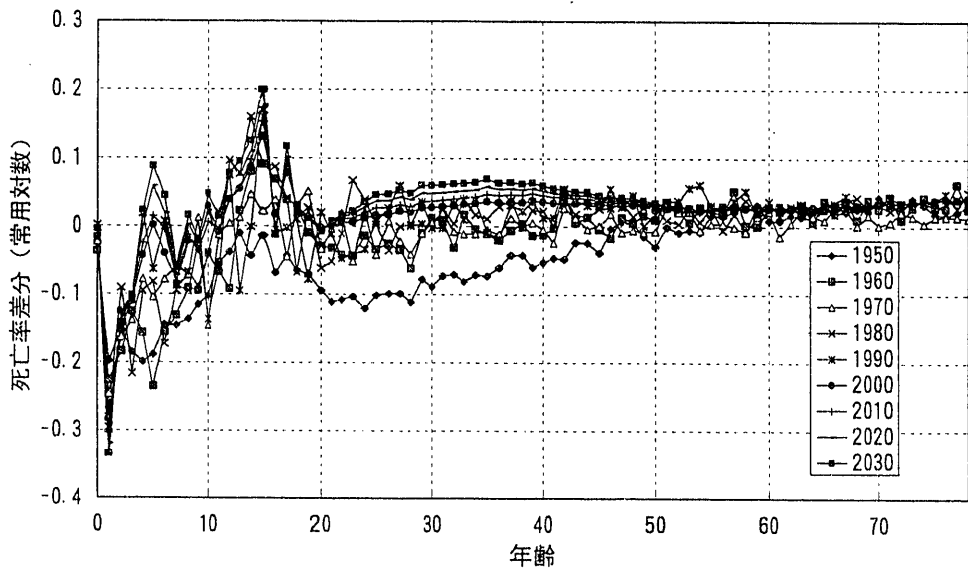


図6-F 年齢別死亡率差分, 10年ごと, 女



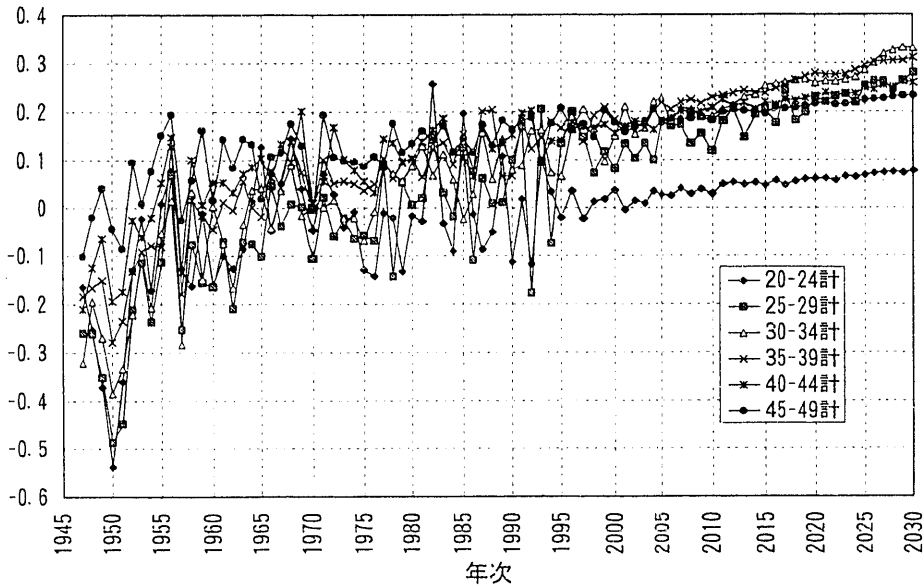
なぜ女子の差分が戦後ほぼ一貫して増加傾向にあるのかと言う点は、死因別あるいは職業別に詳しく調べてみないと分からない。女子の社会進出と関係があるのかないのか大変興味深い点である。

2. 死亡確率の推移

死亡率の差分より得られた死亡率、これを Chang⁴⁾の方法により死亡確率に変換した結果を図8-M, Fに示す。これをみると、1980年までは50歳以下の年齢における死亡率

4) Chang, C. L., The Life Table and Its Applications, Robert Krieger Publishing Co., 1984.

図7 女子20-49歳の5歳階級ごとのコーホート差分（常用対数値）の推移
（1947-1993は実績値，1994-2030は予測値）



の減少が顕著であるが、2000年にかけて次第に減少幅が小さくなり、それ以降はほとんど減少しないかむしろ増加する。一方で、65歳以上の死亡率は男女ともにほぼ一貫して低下していく。

女子では高齢者の死亡率減少がみられる一方で、25～50歳での死亡率の顕著な増加がみられる結果となった。これは差分レベルでのこれまでの一貫した増加傾向が反映された結果といえるが、はたしてどこまで伸び続けるかについてはさらに詳細な研究が必要であろう。

また、コーホート方向での差分から将来値を積み上げているため、全体的にコーホート方向での尾根や谷がみられる。これは過去の実績値においても顕著に観察されるものであるが、予測死亡率ではそれより若干多い。

3. サバイバルカーブと平均寿命

サバイバルカーブ（図9-M, F）の大きな特徴としては、今後は高齢域での曲線が右へシフトすることによる寿命の伸びがみられることであろう。これは65歳以上での死亡率の一貫した低下によるものである。

この結果は、寿命に限界があるという説と整合しないが、近年の実績値は高齢域が右方向にシフトする様相を呈しており、これとは整合している。一方で、男子は2020年以降の60～70歳における下方向のシフトが見られる。女子ではさらに顕著で2010以降25～70歳の領域で下方向へのシフトがみられる。

サバイバルカーブより平均寿命を求めると（図10）、男子では2030年の81.9歳まで増加が続く。一方、女子に関しては2026年の88.8歳をピークに増加が止まり、以後はゆるやかな減少に転じている。

図 8 - M 死亡確率の推移 (男)

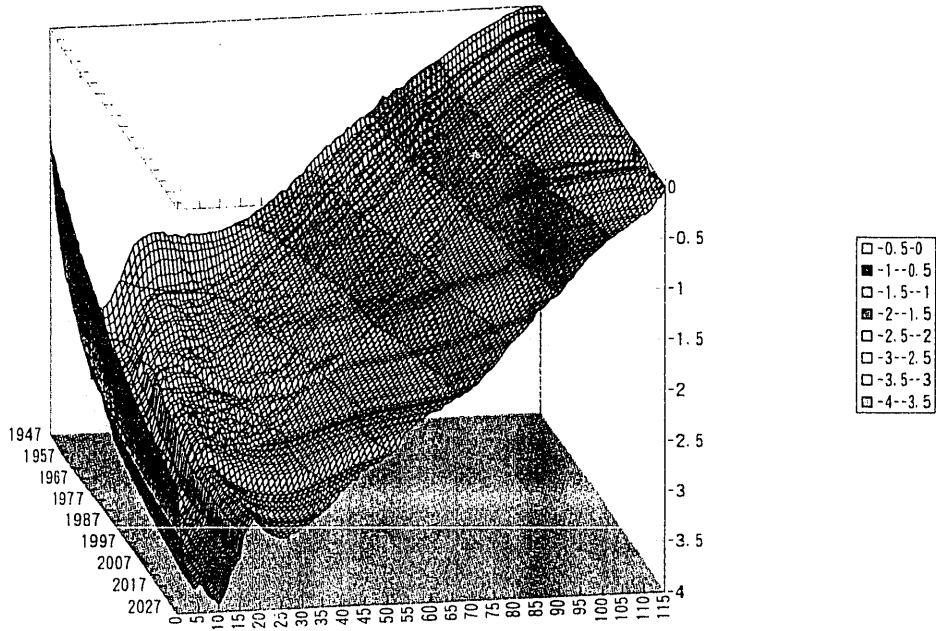


図 8 - F 死亡確率の推移 (女)

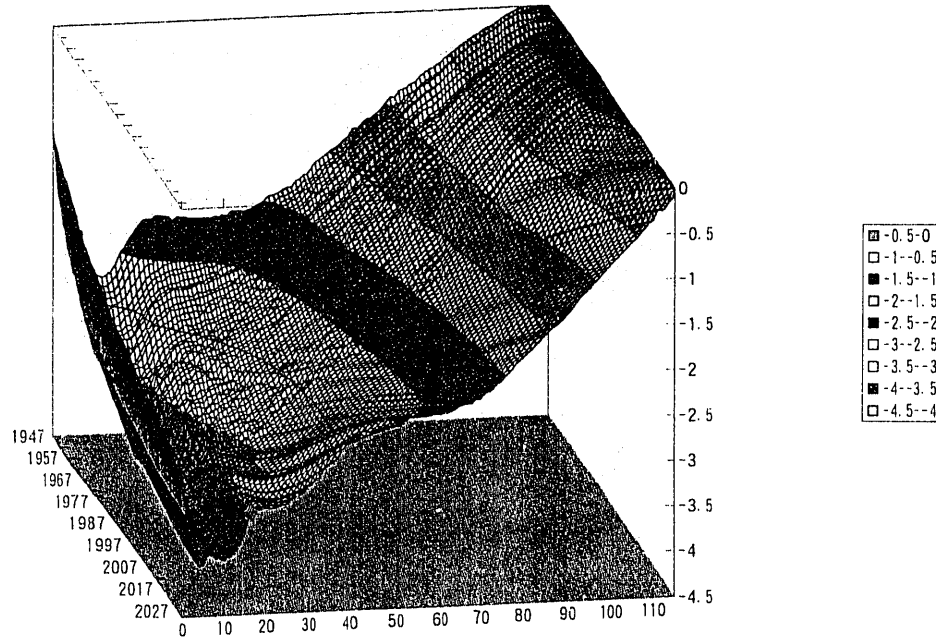


図9-M サバイバルカーブの推移（男）

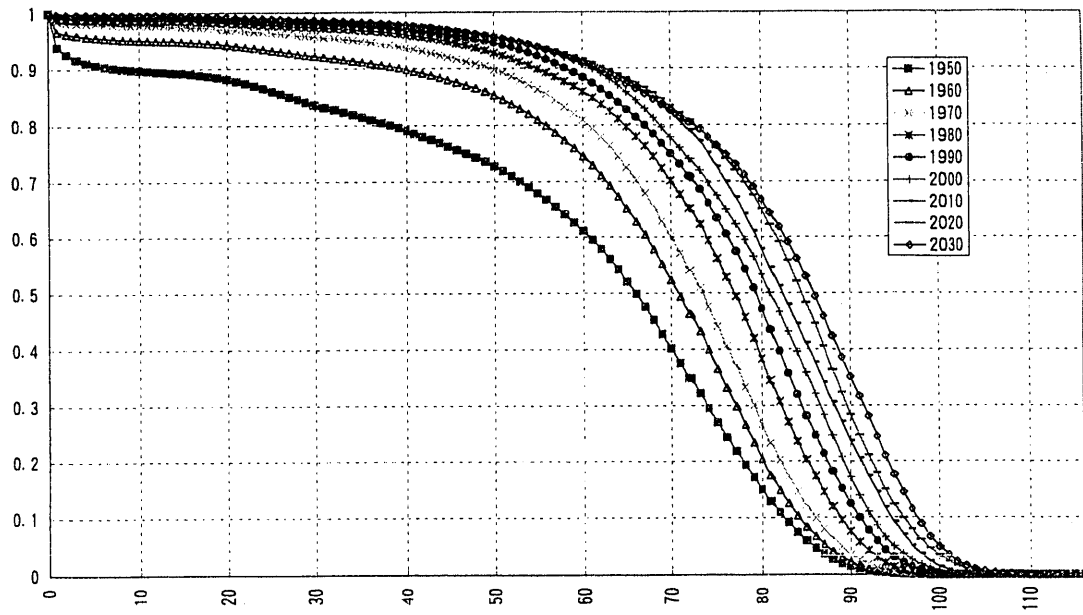
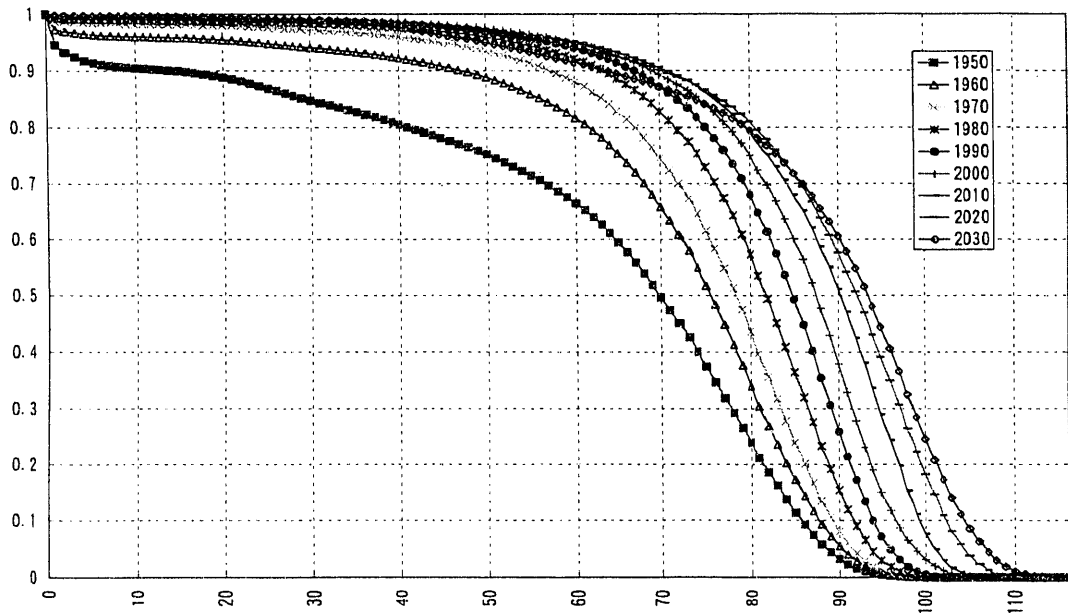


図9-F サバイバルカーブの推移（女）



平均寿命の男女差（図11）は、2016年の7.6年まで広がり続けるが、以後縮まる傾向を見せる。

このように、平均寿命は男女ともに5年以上は伸びると思われるが、これが高齢者における死亡率の一層の低下によるものであり、その一方で働き盛りの年齢層における死亡率低下にブレーキが掛かるということも見逃せない。本研究の結果からは、高齢化は従来想定されていたよりもさらに深刻な事態になることが予測される。

図10 平均寿命の推移

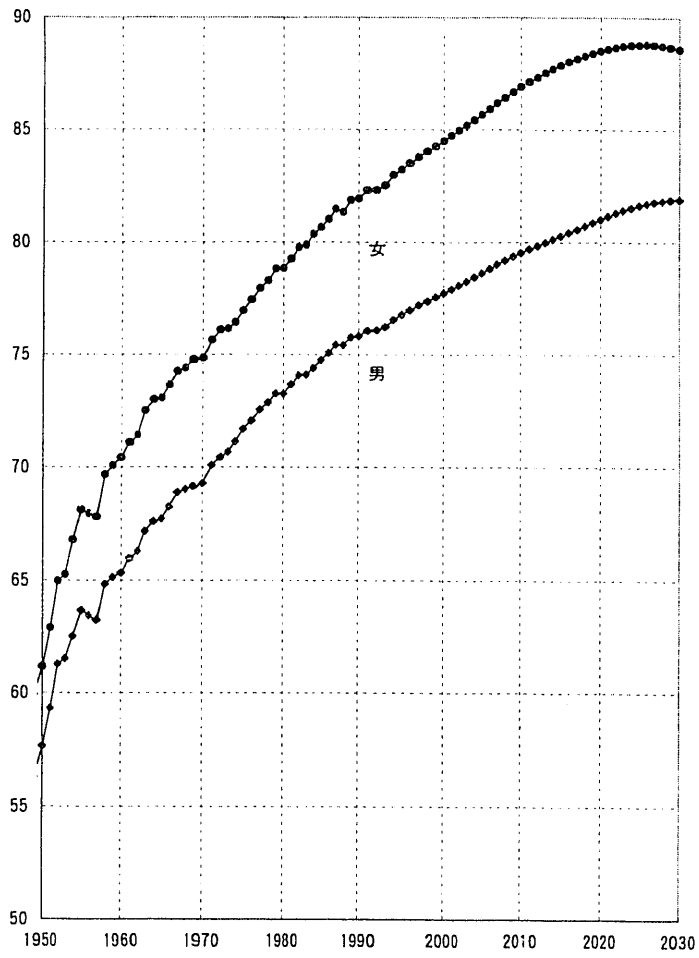
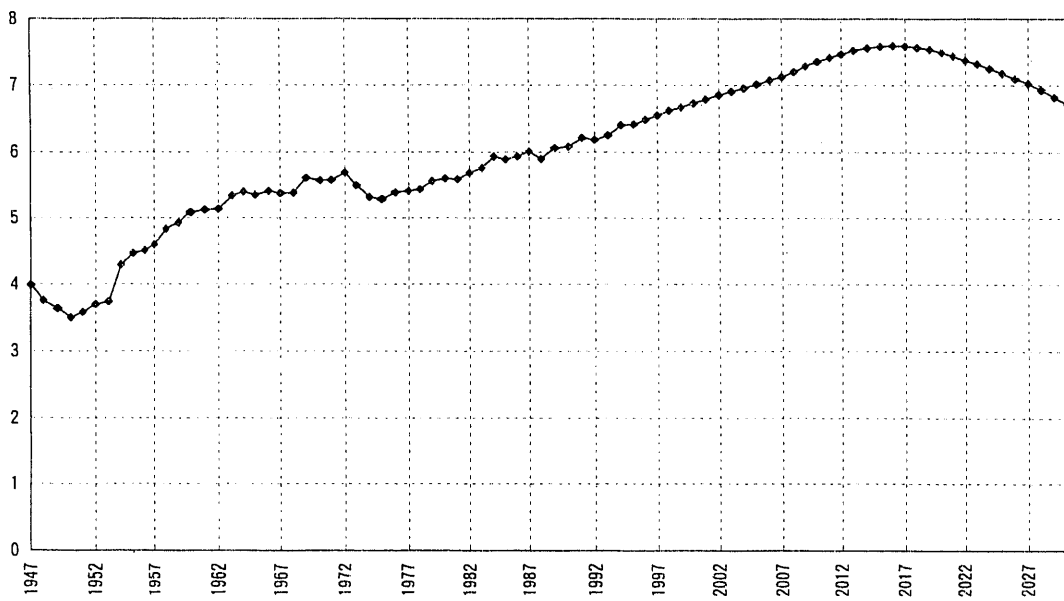


図11 平均寿命の男女差の推移



4. 多変量 AR モデルの利点

年齢別年次別データの予測における多変量 AR モデルの利点を挙げれば次の 4 点であろう。

(i) 複数の時系列を同時に扱えること。

名前が示す通り、複数の時系列の相互の関係を含んだ推定が可能である。従来の推計で用いられてきた方式は、年齢別死亡率曲線のカーブ自体はどの年齢も同一とし、曲線全体を設定した平均寿命を実現するよう上下に平行移動して作成されることが多かった。むしろこの場合は、各年齢ごとの時系列的变化は一切考慮されない。多変量 AR モデルでは、各年齢ごとの複数の時系列を同時に扱うことにより、お互いの関係のある程度保ったまま将来に向けて予測が可能となる。

(ii) 対象とする時系列は、ガウス分布にしたがうならば何でもよいこと。

人口学的な時系列とそうでないものを同時に扱える。したがって、天候の指標、経済学的変数、社会学的変数など、時系列でデータが取ればなんでも構わない。人口学の数理と関係のない変数も同時に扱うことが可能である。ある程度の時系列長を取ることができれば、大数の法則と正規化の操作によりほぼガウス分布に従う時系列を得ることができる。

(iii) 時系列の将来予測および補間推定が可能なこと。

例えばデータに欠損値があった場合でも、周囲のデータから欠落値を推定することが可能である。この操作は直線をつないでしのごことが一般的だが、これではその部分だけに特定の分布を仮定していることにはかならないため、対象とする時系列が直線でなければ予測が歪んでしまう。AR モデルで補間を行えば、AIC (Akaike Information Criterion) を最小とすることでより最適な推定値を得ることが可能である。これにより、例えば途中の特定年次が欠落しているようなデータでも、その部分を補間することで、前後の実データすべてを推定に使うことが可能となる。

(iv) 予測値の分散が求められること。

今回はそれなりに安定した予測死亡率を求めることを第一としたため本稿には含めていないが、 v_n の分散共分散行列 V_m を求めているのであるから、例えば 90% 信頼限界の範囲を示すことも可能である。実際の手順としては、Chang³⁾ の方法に従って死亡率の分散 → 死亡確率の分散 → 平均寿命の分散と順次変換していく必要があり、これについては別の機会に示したい。

5. 原データの変動

多変量 AR モデルによる長期予測を死亡率データに適用してみて分かったことは、死亡率自体の一見偶然に見える変動が、決して少なくない事である。このため、 m の値を 3 より増やすと、将来予測値の振動発散が激しくなった。

今後は、死亡率の変化と因果関係のある時系列と同時に推定を行い、死亡率の時系列に含まれる他の時系列からの影響による分散を分離することにより、死亡率の予測誤差分散の減少を導き出すことで精度をさらに高めることが課題の一つと言える。というのも、年

次差分とコーホート差分に関しては、高齢域において、年次ごとにかなり協調的に上下動をしていることから、平均気温などとの関連が強く想定されるためである。

他の時系列を用いて分散を減少させる手法は、単独時系列における AR モデルや従来の人口学的手法ではでき得なかったことであり、さらなる研究が期待される。

死亡率の差分を吟味した段階においても、従来あまり問題にされなかったことが明らかとなった。それは男15歳における1968年以降の死亡率のコーホート方向および年齢方向の差分の塀である。概算によれば、男15歳では、1967年以前に比べて1970年以降では55%も増加している。これは自動車事故の増加によるものである。

実は、今回示した図の解像度では見ても分かりにくいので吟味の節では触れなかったが、同様の物がもう一つ見つかった。57歳には男女ともに1947~1970年にかけて年齢方向およびコーホート方向での死亡率差分の小さな溝が存在する。

これらの溝や塀がなんらかの政策的変化によりもたらされたのであれば、先に述べた補間のテクニックを応用することで、政策的変化の影響を死亡率あるいは平均寿命という尺度で算出できるし、その政策を変更することで平均寿命がどれだけ伸びるかを示すこともできよう。

死亡率差分の塀があることは分かったが、逆の観点から言えることがもう一つある。それは、全死因の死亡率においてはなんらかの政策的影響によって生じたと思われる死亡率差分の改善は、残念ながら現段階では全然見あたらないということである。なぜなら、上述の男子若年者における自動車事故の急増のごときの逆の現象は全く観察されないからである。

6. 死因別死亡率への適用の可能性

今回用いたデータは全死因のものであった。できることなら死因別の死亡率に対して本法を適用することが望ましいが、それに際しては次の2つの問題点がある。

まず第1の問題は、死因別のデータには ICD の改訂に伴う断絶があることである。死因別の死亡数自体が ICD の変更に伴い大きく変わった場合が少なくない。そのため、ICD の改訂による変化を補正する手段を開発する必要がある。この点は多変量 AR モデル固有のものではなく、いかなる方法を用いるにせよ死因別の死亡数を時系列的に扱う場合には必ず生じる深刻な問題であるが、現状では適当な解決策は存在しない。

第2の問題は、若年齢においては全死因でさえ死亡数が少ない事があるのに、死因別にした場合は死亡数が0である場合も珍しくない点である。これから死亡率を求めても、偶然変動が多すぎて使用に耐えない。したがって、いくつかの年齢をまとめて扱うという処理が必要になる。この処理は移動平均を求めているのと同じなので、単純な AR モデルではなく ARMA (AutoRegressive Moving Average) モデルあるいは ARIMA (AutoRegressive Integrated Moving Average) モデルを適用する必要があるかもしれない。いずれのモデルにせよ、AR モデルに比べて定式化も計算も格段に複雑化せざるを得ない。

7. 最小二乗法による計算についての評価

今回はわざわざユール-ウォーカー法に比べて計算の面倒な最小二乗法によってモデル推定を行った。その利点は、変量毎に m を変えられると言う点であった。そこでこの点について利点があったか否か考察しておく。

(式4.1) より、最初に同時にまとめて扱う変量の組み合わせを決めた時点で l が定まり、同時に m の最大値は決まってしまう。つまりプログラムで m を変化させると言っても単に小さくするだけであった。そして m を小さくすると、多くの場合自己回帰係数は y_n が n の増加とともに振動発散してしまうような値となり、収束しないことが多かった。

この点は AR モデルの宿命としての側面もある。つまり、得られる自己回帰係数行列が、 $N \rightarrow N+1$ とするにしたがい y_{N+1} が安定した値となる保証がないのである。場合によっては $y_{n+1} = r \cdot y_n$ ($r > 1$) という具合になることもある。

したがって、今回のような場合においては無理に最小二乗法を採用するまでもなく、レビンソンのアルゴリズムに従ってユール-ウォーカー方程式を解く方法で十分だったかもしれない。今後、人口以外の時系列を組み込んで予測を行うなどのときに、変量の組み合わせ方法とともに再び吟味してみたい。

III まとめ

本研究は、性年齢各歳別死亡率の差分に対して多変量 AR モデルを適用し、これにより性年齢各歳別年次別死亡率の2030年までの長期予測を実現した。得られた平均寿命の推移、平均寿命の男女差の推移、サバイバルカーブの推移は過去から現在までのトレンドを延長するものであり、最終的な評価は時を経てみないと分からないが、まずまず良好な予測値が得られたと思われる。

これによると、平均寿命は男子の場合で2030年の81.9歳まで増加、女子の場合は2026年の88.8歳まで増加した後、減少に転じる。また、サバイバルカーブは、80歳以上で右へシフトする様相を見せ、50歳以下では下にシフトする傾向が見られた。

さらに特徴的なのは23~49歳女子の死亡率である。今後は一貫して増加していくことが予測された。

Long-term Prediction of Age-specific Mortality with Multivariate Autoregressive Model

Tamotsu Ooba

Age and sex-specific mortality rates from 1995 to 2030 in Japan were predicted with multivariate autoregressive (MAR) model. This model is of the form,

$$\mathbf{y}_n = \sum_{i=1}^m \mathbf{A}_i^m \mathbf{y}_{n-i} + \mathbf{v}_n,$$

where \mathbf{y}_n is a vector of time series, \mathbf{v}_n is a vector of a variance of order l and $\mathbf{v}_n \sim N(0, \mathbf{V}_m)$, and l is the number of time series those were considered simultaneously, and \mathbf{A}_i^m is an AR coefficient matrix of order m .

The MAR model was estimated by the least square method with the Householder reduction with the mortality rates during the period from 1947 to 1994. The age (0 to 84 years) and sex-specific mortality rates were log-transformed and their three types of differential, by age, by year, and by cohort, were calculated. These three differentials and the log-transformed mortality rates were compared and the differential by cohort was selected for the forecast. The differential by year was used supplementary because the differential by cohort does not induce the mortality rate at age 0.

Each differential for cohort was converted into the normalized form. Then, \mathbf{A}_i^m was estimated. Long-term prediction of the differentials were made by applying \mathbf{A}_i^m to the differentials repeatedly. The predicted values were reverse transformed into the original form. The death probabilities in the ages 85 years and over were extrapolated toward the direction of the cohort with the nonlinear least square method for the Gompertz-Makeham curve.

The result suggested that the male life expectancy will reach 81.9 years in 2030, and the female life expectancy will reach 88.8 years in 2016 and thereafter decrease slowly. The estimated growth of the life expectancy is contributed with further reduction of mortality rates in elder ages but not younger ages. On the other hand, the result showed that the declining speed of the female mortality for 20 to 65 years old will be slower, and then, their mortality will increase.