

# 日本の年齢別出生力パタンのトレンドの分析

中澤 港\*

キーワード：合計出生力，Coale and Trussell モデル，Hadwiger モデル，トレンド

Title: Trends in age-specific fertility in postwar Japan

## Abstract

Continuous reduction of fertility has been a focus of demographic problem in Japan. Most previous studies suggested that the reduction is mainly attributable to the increase of delayed or never married population, and that marital fertility itself is not decreased. Here I applied the Hadwiger's model and Coale and Trussell's model to the data of age-specific fertility and age-specific marital fertility in postwar Japan, and clarified some trends as cubic functions. Future projection based on these functions yielded different results due to the period of used data. When using the data after 1960, the projection suggests gradual increase of fertility.

## 要旨

近年の出生力の低下は日本においては大きな人口問題の一つである。多くの先行研究は、この低下が結婚タイミングの遅れあるいは非婚率の上昇に起因し、有配偶出生力自体は低下していないとしているが、年齢別有配偶出生力のパターンを数学的に詳細に検討したものは少ない。本研究は、戦後の日本における年齢別有配偶出生力パターンについて Coale and Trussell のモデルと Hadwiger のモデルを、年齢別出生力パターンについて Hadwiger のモデルを適用し、そのパラメータの経時的なトレンドを三次関数で近似したものである。Hadwiger のモデルに基づく将来予測によれば、日本の出生力のトレンドは、1950 年以降のデータをすべて使うか、1960 年以降のデータだけを使うかで結果が違っていた。1960 年以降のデータだけを使った場合、今後やや上向くと予想された。

## はじめに

日本は 21 世紀初めには、世界でも最も出生力の低い国の 1 つである。日本における合計出生力は、ベビーブームで一時的に上向いたのを除けば戦後一貫して減り続けてきており、この傾向が労働力供給や経済発展にとって問題であるという指摘が数多くなされている。先行研究によれば、この出生力低下には、主として結婚していない人の割合が増加したことと、平均結婚年齢が遅くなったことによる寄与が大きいとされている。合

---

\* 山口県立大学看護学部

計有配偶出生力は問題を含んだ指標であるが<sup>1)</sup>、仮に計算してみると、1960年以降の合計有配偶出生力は上昇しているし、20歳以上の年齢別有配偶出生力の合計（20歳以上の合計有配偶出生力）に限ってみても横ばいである（図1）。

しかし、合計有配偶出生力という指標は、足しあわされる年齢別有配偶出生力の分母が年齢別既婚女性人口であるため、分母が小さい低年齢層の変化の影響が大きくなりすぎるという問題があり、この相対的な安定が、21世紀も続くのかどうかは不明である。年齢別有配偶出生力のパターンが低年齢層での上昇が大きくなり、高年齢層では低下してきているならば、合計有配偶出生力が安定しているのは見掛けだけかもしれない。本研究は、このことを調べるために、年齢別有配偶出生力と年齢別出生力のパターンに対してパラメタライズドモデルを当てはめ、パラメータの経時的な変化に何らかのトレンドを探索したものである。

## モデル

年齢別有配偶出生力のパターンのパラメタライズドモデルとしては、1978年に発表された Coale and Trussell のモデル（以下 CT1978）<sup>2)</sup> を当てはめた研究がこれまで多くなされているので、本研究でもこれを採用する。式で書くと、

$$F(x) = G(x)F_m(x), \quad F_m(x) = Mn(x)\exp(m\nu(x)) \quad [1]$$

となる。 $x$  が年齢、 $F(x)$  が年齢別出生力、 $G(x)$  が既婚割合、 $F_m(x)$  が年齢別有配偶出生力、 $n(x)$  が自然出生力、 $\nu(x)$  が年齢に依存する自然出生力からの下向きの乖離パターン（ $n(x)$  と  $\nu(x)$  は Coale と Trussell の論文の中に表の形で与えられている）、 $M$  は全体としての出生力水準を表す「スケール」パラメータであり、 $m$  は意図的な出産抑制の程度を示す「コントロール度」パラメータである。このモデルは、多くの集団について年齢別有配偶出生力パターンによく当てはまり、かつパラメータがデータによらず実体的な意味をもっている点が優れている。

なお、ここで注意すべきは、前半の  $F(x) = G(x)F_m(x)$  の部分である。これが成り立つためには、非嫡出生が無視できるくらい小さい必要がある（式が厳密に成り立つためには、 $F_m(x)$  の分子に非嫡出生も含まなくてはならないが、非嫡出生は分母となる女性から生まれていないので論理的におかしい）。日本では非嫡出生は少ないので、適用しても問題ないと考えられるが、他の国・集団に適用する際には、この点について注意すべきである。

一方、年齢別出生力のパターンを示すパラメタライズドモデルは、これまで数多く提案されている。Hoem らは、1962年から1971年までのデンマークの年齢別出生力に対して、最小二乗法によって、三次スプライン関数、Hadwiger のモデル、1974年に発表された Coale and Trussell のモデル（以下 CT1974）、ガンマ分布、ベータ分布、2種類の多項式、2種類の Brass のリレーショナルモデル、Gompertz 曲線を当てはめた結果、三次スプライン関数をもっとも良く適合し、CT1974 とガンマ分布、Hadwiger と続き、他の関数はあまり当てはまりが良くないことを示している<sup>3)</sup>。ベータ分布やガンマ分布は、とくに若い年齢層の出生パターンに合わなかった。ガンマ分布を変形したものはデンマークのデータによく適合するが、パラメータに人口学的意味がない（たんに数式として適合するだけであり、その実体としての意味を解釈することができない）。パラメータの経時的トレンドを分析する以上、そのパラメータには人口学的な意味がなくてはいけない。その意味で、三次スプライン関数は、いくら当てはまりが良くても、本研究には使えない。CT1974 は当てはまりが良く、パラメータに人口学的な意味もあるが、関数の形が複雑なためにデータによっては収束しにくい場合があり、本研究では採用しなかった。

Hadwiger のモデルは、そのままでも 1994 年のオーストリア、デンマーク、フランス、スウェーデンなどのデータにはよく適合するし、二峰性であるためにそのままでは適合度が高くない 1994 年の英国やアイルランドのデータにも、2 つの部分集団を考えてやればよく適合することも示されている<sup>4)</sup>。英国やアイルランドで出生力の年齢分布が二峰性になっているのは、これら二国では若い年齢層での非嫡出生がかなり多く、それが嫡出生とは質的に違った出生パターンをもっているからと推論されている。本研究では、日本人のデータを対象にするため、非嫡出生は 1 % にも満たないので、Hadwiger のモデルをそのまま当てはめてよいと思われた。

$$F(x) = \frac{ab}{c} \left(\frac{c}{x}\right)^{3/2} \exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\} \quad [2]$$

[2] 式が Hadwiger のモデルである。年齢別出生力  $F(x)$  を 3 つのパラメータ  $a, b, c$  で表す、比較的単純な関数である。3 つのパラメータには、もともとは人口学的な意味はないが、Chandra らは、ヨーロッパのデータについて、 $a$  が合計出生力 (TFR; Total Fertility Rate) とよく相関し、 $b$  が年齢別出生力のピークの高さと相関し、 $c$  が平均出産年齢と相関することを報告している<sup>4)</sup>。

## データ

資料としては、1950 年から 1995 年までの人口動態統計と国勢調査報告を用いた。年齢別出生率 (ASFR) は年齢各歳別総出生数を年齢各歳別女性人数で割って求めた。年齢別有配偶出生率 (ASMFR) は、原則として年齢各歳別嫡出生数を年齢各歳別既婚女性人数で割って求めたが、1970 年だけは年齢各歳別嫡出生数が掲載されていなかったため、年齢各歳別非嫡出生数を年齢各歳別総出生数から引いて求めた年齢各歳別嫡出生数を計算に用いた。前述のように、CT1978 を厳密に成り立たせることよりも、年齢別有配偶出生率の論理的整合性を重視したためである。

## 解析方法

CT1978 は未知のパラメータが 2 つ、Hadwiger のモデルでは未知のパラメータが 3 つあるので、残差二乗和が最小になるようなパラメータを探索しなくてはならない。関数の最小化はいくつかの統計パッケージでも可能だが、処理の量が多いと煩雑になるので、Nelder=Meade の滑降シンプレックス法として知られるアルゴリズム<sup>5)</sup> を使って、自動的に最も適合度が良いパラメータを見つけるプログラムを開発した。このプログラムは web ページで公開し、ダウンロードできるようにする予定である。

統計ソフト R (R-1.6.1, <http://www.r-project.org/>) を用いて、得られたパラメータを代入した関数の形をデータ点とともにプロットした。さらに、横軸に年次をとってパラメータ推定値をプロットしてみると、どのパラメータも三次曲線で近似できそうだったので、R の線型回帰関数により、西暦年とその 2 乗、3 乗を独立変数とし、各パラメータを従属変数とする回帰分析を行った。その結果、この三次曲線の当てはめが 90 % 以上の説明力をもつことがわかったので、2010 年までこの関数を外挿し、各パラメータの変化と TFR の変化を予想した。

## 結果と考察

年齢別有配偶出生力に対して CT1978 と Hadwiger のモデルを当てはめた結果を図 2 に示す。図中の  $M$  と  $m$  は偏差平方和の平方根  $E$  が最小になるように推定された CT1978 のパラメータ（最小のときの  $E$  が括弧内に示してあり、モデルの曲線が実線で示してある）で、 $a$  と  $b$  と  $c$  は偏差平方和の平方根  $E$  が最小になるように推定された Hadwiger のモデルのパラメータ（そのときの  $E$  が括弧内に示してあり、モデルの曲線が破線で示してある）である。全体に適合は悪くないが、Hadwiger モデルのほうが CT1978 よりも適合度が良いようにみえる。1985 年以降の適合度が低くなっているのは、20 歳未満の値がおかしいためである。日本の法律では女性は 16 歳以上でないと結婚できないので、15 歳の嫡出出生はゼロであるべきであり、15 歳で低いのは当然である。17 歳から 20 歳がモデルより高いのは、妊娠したから結婚するという、いわゆる「できちゃった婚」のケースが多いからと思われる。もし結婚可能な年齢に法的下限がなければ、15 歳や 16 歳でも有配偶出生力は高くなると思われ、その場合は Hadwiger モデルの適合度は上がるであろう（パラメータもまったく違うものになるはずである）。

年齢別出生力に対して Hadwiger のモデルを当てはめた結果を 3 に示す。こちらは RMSE が 0.002 から 0.012（1960 年以降に限れば 0.006 未満）と、有配偶出生力への当てはめよりも適合度が 1 桁良かった。Chandra らが北欧諸国についてよく適合することを示したのと同様、Hadwiger のモデルは戦後日本の年齢別出生パターンについてもよく当てはまることが示された。これは、日本人口が、英国やアイルランドとは異なり、非嫡出出生の割合がきわめて小さいことから、当然予想された結果である。

Chandra らが示したのと同様、戦後日本のデータでも、パラメータ  $a$  は TFR との高い相関 ( $R^2 = 0.999$ ) を示し、パラメータ  $c$  は平均出産年齢との高い相関 ( $R^2 = 0.96$ ) を示した。1950 年から 1995 年までの各パラメータの変化をプロットしてみると (4)、上述した通り三次曲線で近似できそうにみえたので、三次曲線を当てはめてみたところ、悪くない当てはまり ( $R^2 \geq 0.91$ ) が得られた。この曲線によって 2010 年までの各パラメータの値を予想し (5)、それを式 [2] に代入して合計して TFR を求めたところ、2000 年に 0.63、2005 年に -0.34 という、明らかに正しくない値になった (6)。これは、図 5 の左上のパネルに示すように  $a$  が負になってしまうためである。1950 年から 1995 年までのデータを使った場合の  $a$  の変化のトレンドに対する三次曲線の当てはめで得られた重相関係数の二乗は 0.91 であり、他のパラメータに比べて低かったことから、 $a$  についてだけ 1960 年以降のデータだけで当てはめをやり直すと、重相関係数の二乗は 0.95 とかなり改善した。この場合は、2005 年以降も  $a$  は正であり続け、その値を代入して得られる TFR の予測値は、2000 年、2005 年、2010 年にそれぞれ、1.37、1.44、1.61 となった。もっとも、データ点の選び方が恣意的であり (1955 年を入れると重相関係数の二乗があまり改善されず、1965 年以降だけにしても 1960 年以降の場合よりもあまり改善されないため、という理由はあるが)、トレンドの当てはめが三次曲線であることに必然性がまったくないので、この予測に意味があると言い切ることはできない。

最後に、CT1978 を有配偶出生力パターンに当てはめて得られた  $M$  と  $m$  のトレンド (7) を示す。どちらも三次関数がかかなり良く当てはまるが、1970 年の値だけは大きく外れている。第二次ベビーブームの最中であつたために、突発的に出生力が上昇したのではないかと考えられる。三次関数からの予想によれば、有配偶出生力はスケールとコントロール度の両方が今後上昇することになる。この予想は、ピークの高さは上がるけれども、高齢での抑制も強くなることを意味するので、大雑把に見れば、短期間に集中して子どもを産むようになることを意味する。しかし、三次関数の意味付けがない以上、この結果からそこまで解釈してよいのかという点については慎重に判断すべきであろう。

本研究は、トレンドを調べる際の恣意性を免れていない点が本質的な弱点であり、それゆえに予測のためにトレンドを外挿することを正当化する根拠はない。従って、この予測が長期的に有効であるとは考えられない。しかし、トレンドの適合度が高いことから、少なくとも短期的には成り立つ可能性が高いと考えられる。

## 付録：用語説明

本論文では人口学の専門用語がいくつも出てくるので、重要なものについて簡単に解説しておく。

- 1) 有配偶出生力：Marital Fertility Rate の訳語である。即ち、年齢別既婚女性人口を分母とし、母の年齢別の嫡出出生数（結婚しているカップルからの出生数）を分子とした値を年齢別有配偶出生力という。それを全年齢層について合計したものを合計有配偶出生力 (Total Marital Fertility Rate; TMFR) という。
- 2) 合計出生力：Total Fertility Rate (通常 TFR と略される) の訳語である。年齢別女性人口を分母とし、母の年齢別出生数を分子とした値を年齢別出生力といい、それを全年齢層について合計したものを合計出生力という。
- 3) 自然出生力：Natural Fertility の訳語である。避妊や人工妊娠中絶などの意図的な出産抑制がない場合の出生力（女性人口あたりの出産数）をいう。対義語は抑制された出生力 (Controlled Fertility) である。
- 4) パラメタライズドモデル：ここでは、比較的少数のマクロな変数（パラメータ）を使ってシステムのふるまいを示すように構築された数理モデルをさす。

## 引用文献

- 1) 廣嶋清志: 出生率低下をどのようにとらえるか? - 年齢別有配偶出生率の問題性 - 数理社会学会『理論と方法』, 16: 163-183, 2001.
- 2) Coale AJ, Trussell TJ: Technical Note: Finding the two parameters that specify a model schedule of marital fertility. *Population Index*, 44: 203-213, 1978.
- 3) Hoem JM et al.: Experiments in modelling recent Danish fertility curves. *Demography*, 18: 231-244, 1981.
- 4) Chandola T, Coleman DA, Hiorns RW: Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Population Studies*, 53: 317-329, 1999.
- 5) Nelder JA, Mead R: A simplex method for function minimization. *The Computer Journal*, 7: 308-313, 1965.

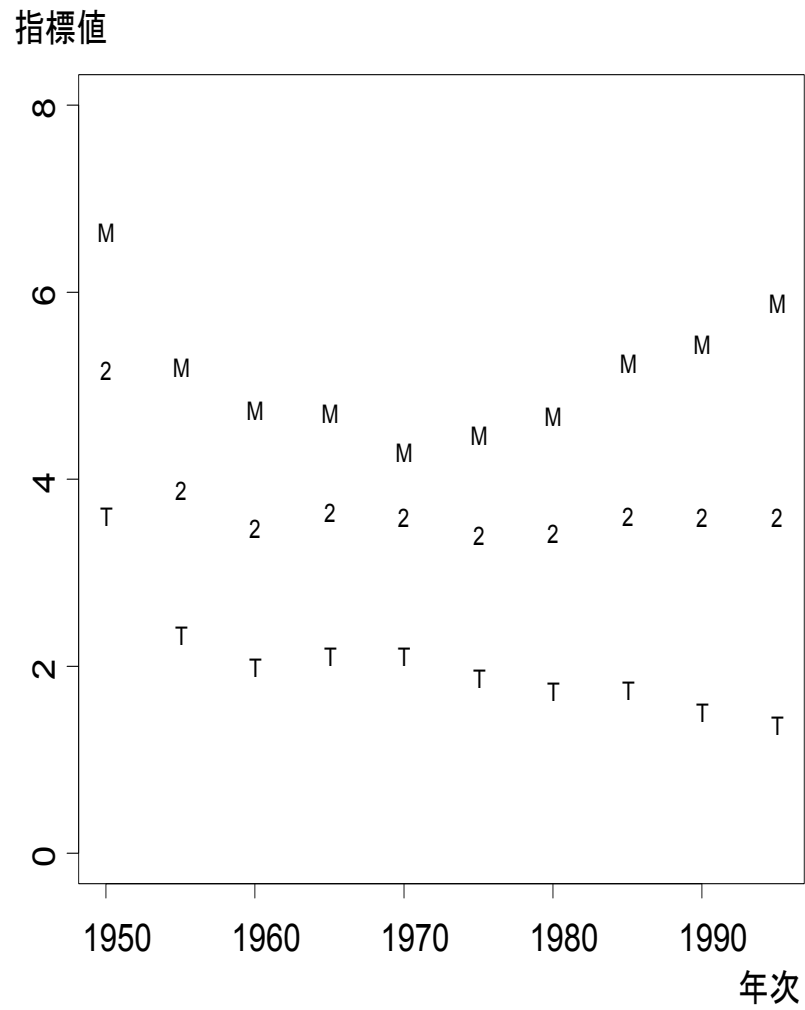


図 1: 日本における出生力指標の年次推移 (M: 合計有配偶出生力, 2: 20 歳以上合計有配偶出生力, T: 合計出生力)

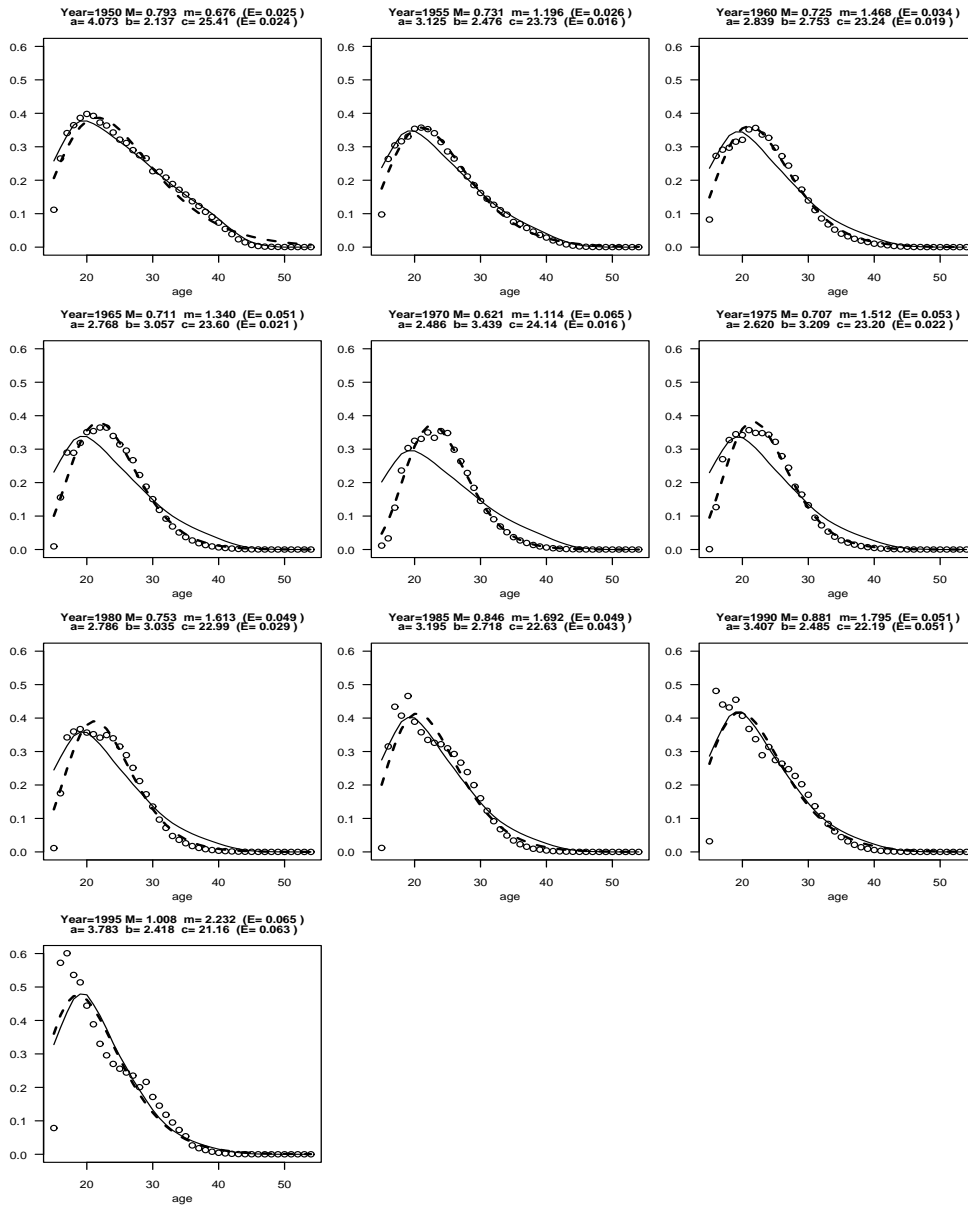


図 2: 戦後日本の年齢別有配偶出生力データ ( 印) に Coale and Trussell (1978) のモデル ( 実線) と Hadwiger のモデル ( 点線) を当てはめた結果

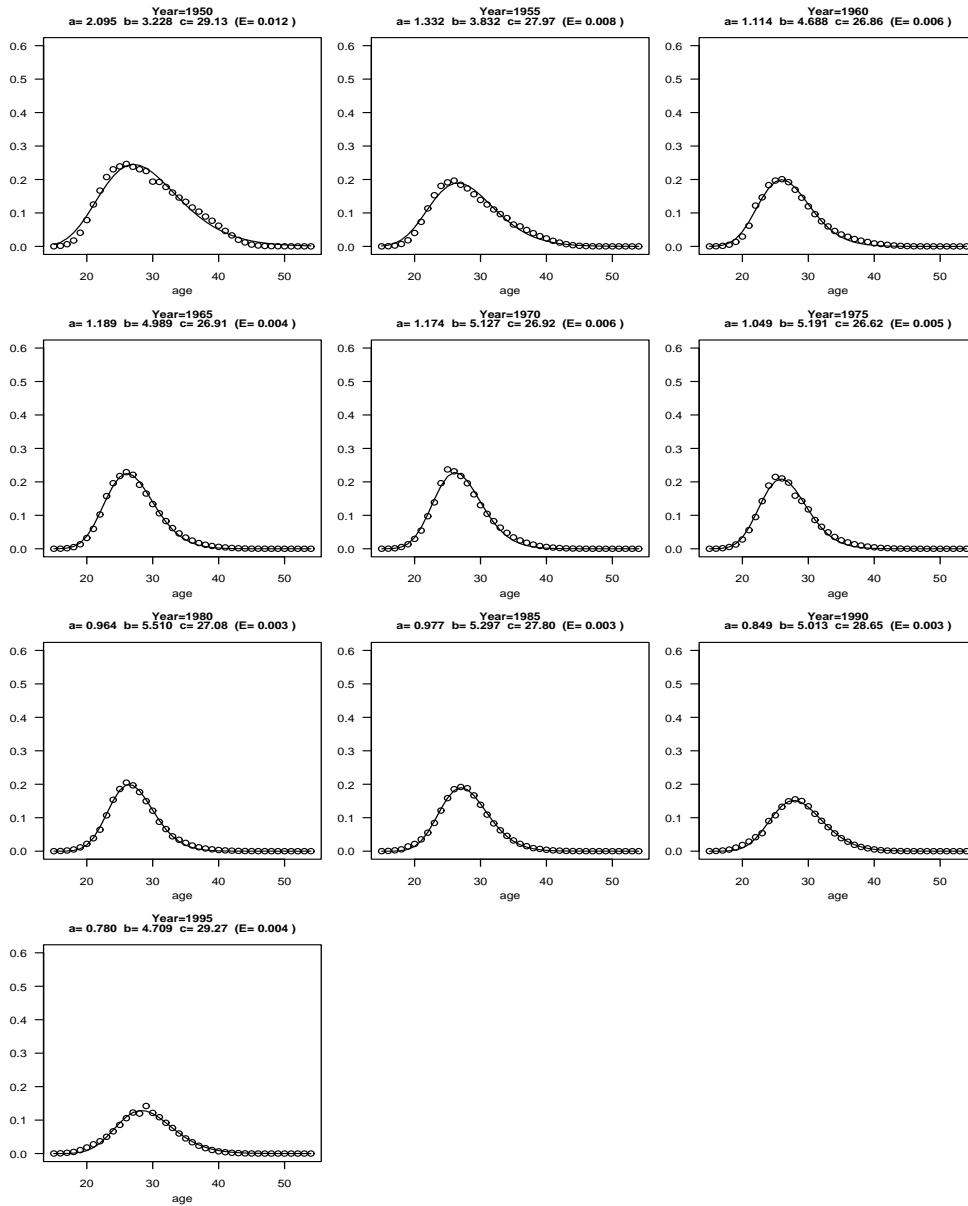


図 3: 戦後日本の年齢別出生力データ ( 印 ) に Hadwiger のモデル ( 実線 ) を当てはめた結果



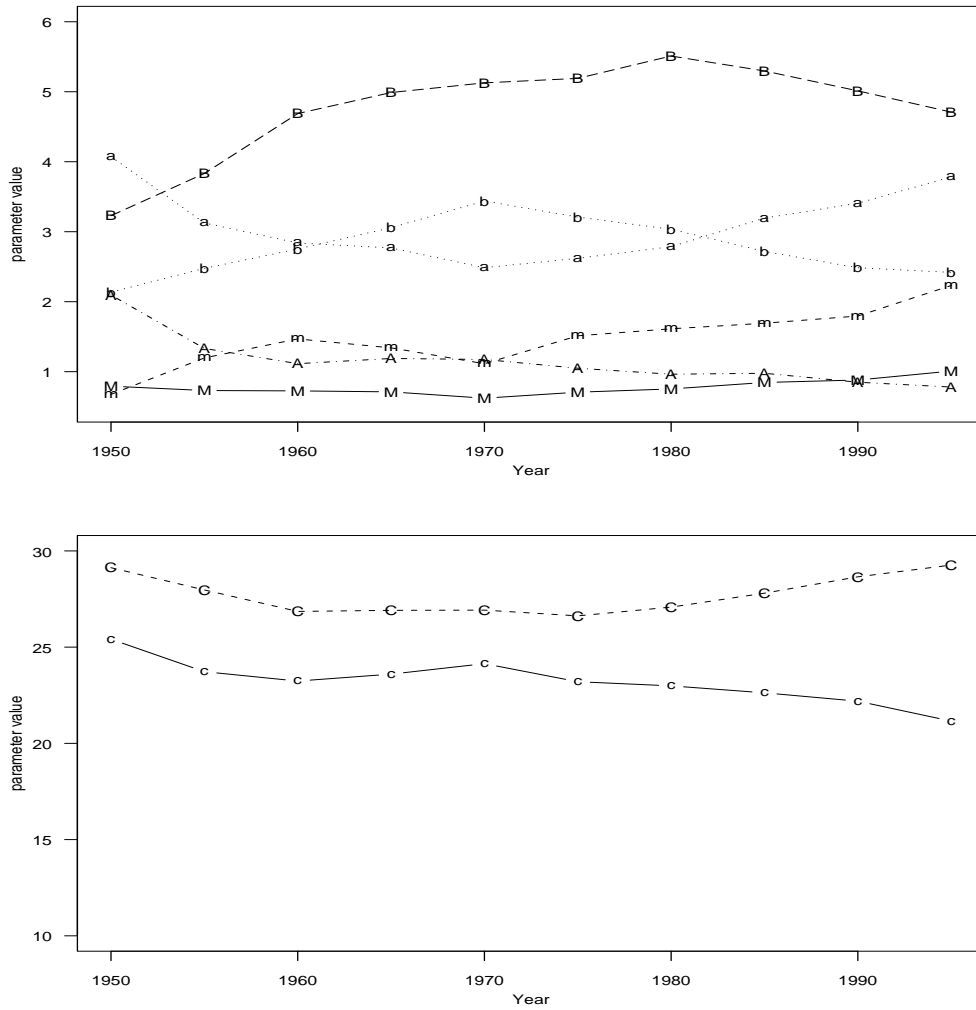


図 4: 戦後日本のデータに当てはめた, Coale and Trussell (1978) のモデルと Hadwiger モデルの各パラメータの年次推移 (a, b, c は Hadwiger モデルを年齢別有配偶出生力に当てはめたもの, M と m は Coale and Trussell (1978) のモデルを年齢別有配偶出生力に当てはめたもの, A, B, C は Hadwiger モデルを年齢別出生力に当てはめたものの a, b, c に相当する)

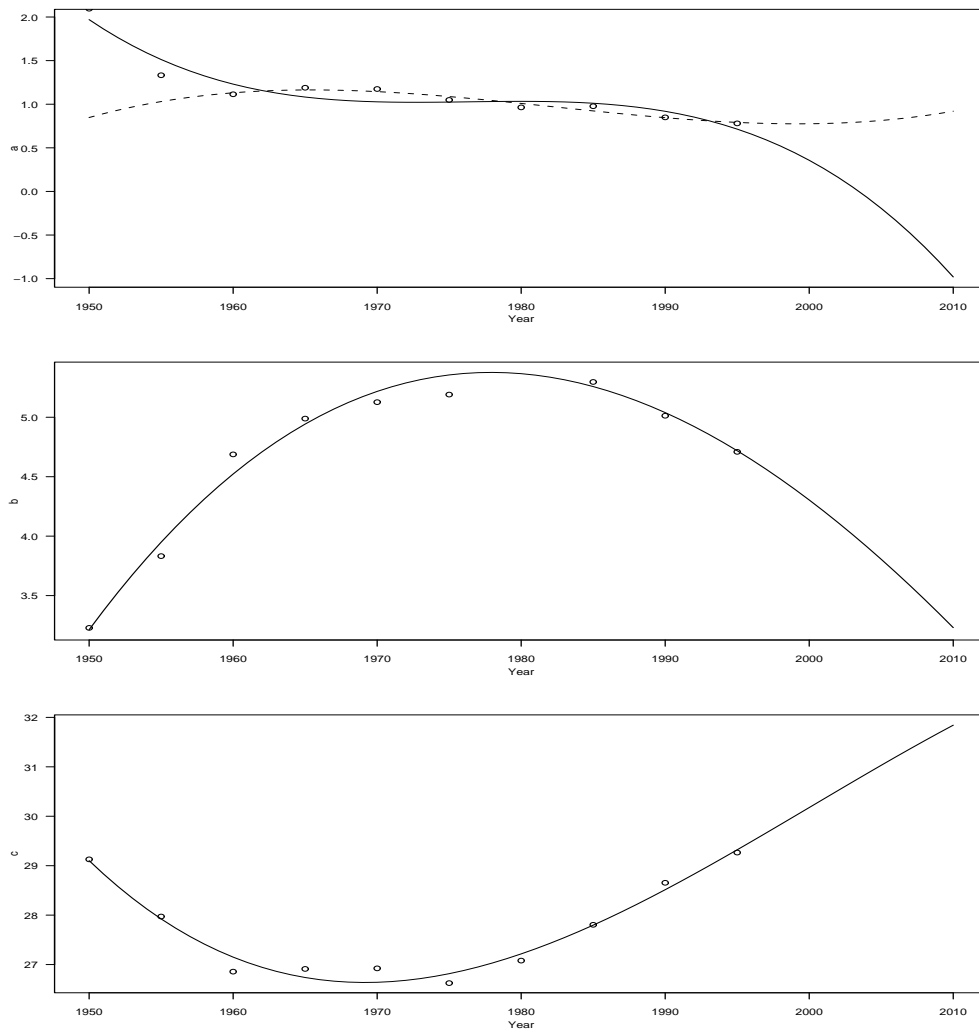


図 5: Hadwiger モデルを年齢別出生力に当てはめて得られるパラメータ  $a$ ,  $b$ ,  $c$  について, 三次曲線近似による 2000 年から 2010 年までに予想される変化

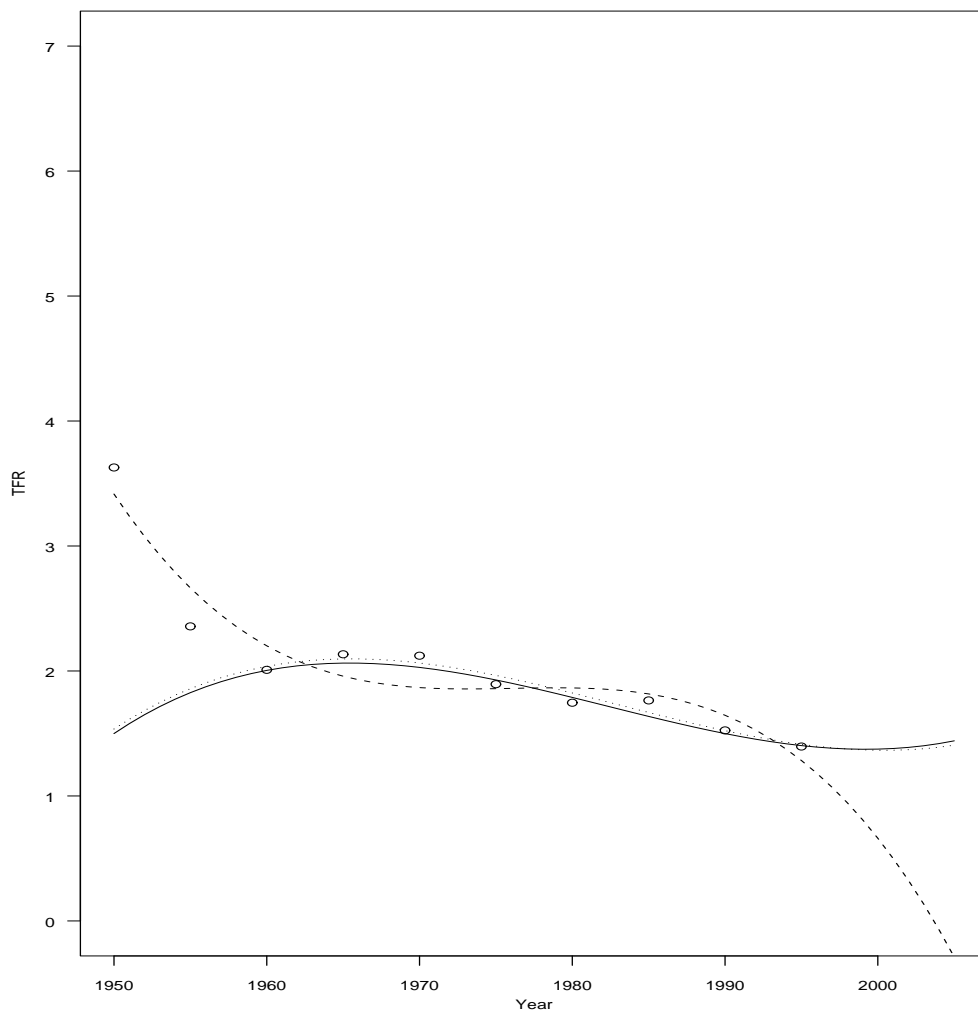


図 6: Hadwiger モデルを年齢別出生力に当てはめて得られるパラメータのトレンド (a は 1960 年以降のデータのみ, b と c は 1950 年から 1995 年までのデータすべてを使って推計) から推計される 2010 年までの TFR 予測値 (実線), TFR そのものの値に三次曲線を当てはめたもの (破線), 1960 年以降の TFR の値に三次曲線を当てはめたもの (点線)

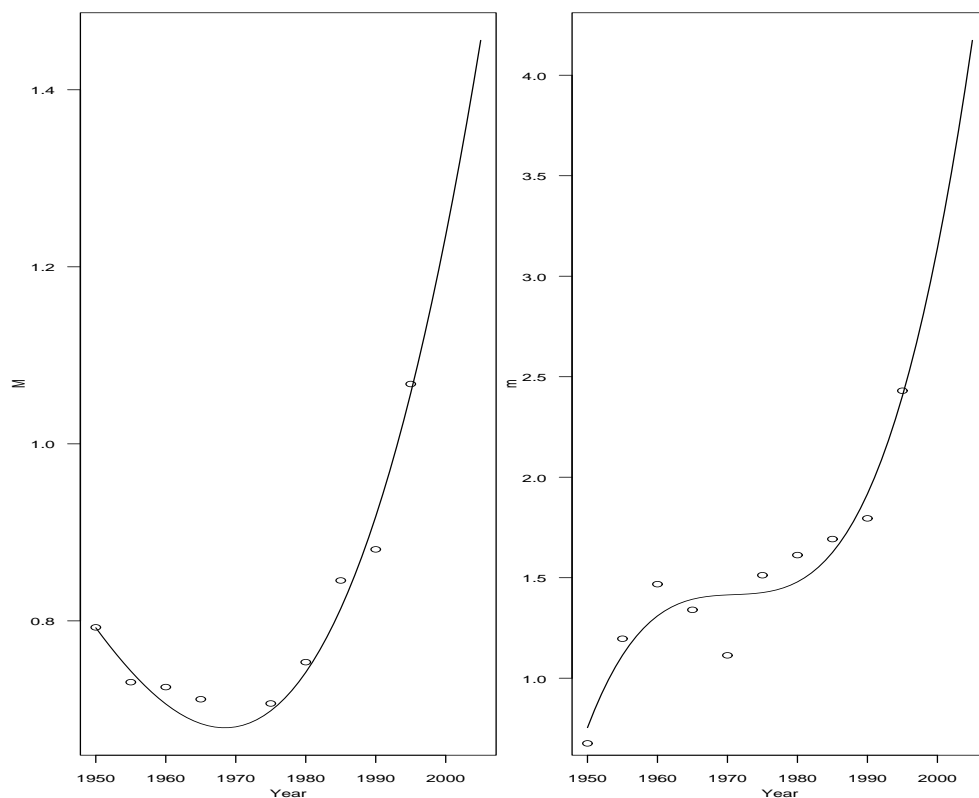


図 7: Coale and Trussell (1978) のモデルを年齢別有配偶出生力にあてはめて得られるパラメータの年次推移への三次曲線の当てはめ (左のパネルが「スケール」 $M$ , 右のパネルが「コントロール度」 $m$  を示す)