

# ミクロ計量経済学入門：

## 第5回 出産の意志決定モデルの推定

北村行伸

一橋大学経済研究所

2006年6月15日

### 概要

連載の第5回はプロビット・モデルやロジット・モデルを使いながら、二項選択モデルが実際にどのように使えるかを示す。具体的には出産の意志決定モデルを考える。また、出産のタイミングが生涯出生児数にどのような影響を与えるかを見る目的でサバイバル分析も行う。

## 1 はじめに

2005年に15-49歳までの女性の年齢別出生率を合計して1人の女性が平均して生むと推定される子供の数である合計特殊出生率は1.25にまで低下した。人口を維持するために必要な出生率が2.07であると言われており、この低下傾向に歯止めがかからなければ、労働力不足や各種人材の不足に陥り、経済の活力が低下するのではないかと考えられている。また、少子化は高齢化の裏返しでもあり、これからますます増えていく高齢者をより少ない現役世代で支えていかねばならず、年金や税制の改革も必至であることは周知の通りである。

今回は、この最も緊急に政策が求められている分野についてミクロ計量経済学で何が出来るかを論じてみたい。具体的には出産の意志決定について分析してみよう。分析手法としてはプロビット、ロジット推定について解説し、出産のタイミングに関してはサバイバル分析を行う。

これまでの出産行動に関する研究の蓄積は人口学、社会学、経済学などを中心に膨大な蓄積がある。手際よいサーベイとしてHotz, Klerman and Willis (1997) や Ermisch(2003) がある<sup>1</sup>。そこでの論点は、(1) 子供の質か数かという問題、(2) 所得水準と出産の関係、(3) 出産による労働の質の低下や逸失賃金、再雇用可能性の問題、(4) 第1子出産のタイミングと第2子、第3子出産の最適間隔の選択、(5) 税制や社会保障、育児支援制度の問題など広範にわた

<sup>1</sup>また研究論文として Ermisch (1989)、Gustafsson (2000)、Happel, Hill and Low (1984)、Willis (1973,1999)、Willis and Rosen (1979) などを挙げておきたい。

る。少子化対策として、日本政府が提案している対策の多くはヨーロッパを中心とした少子化先進国の経験から学んだものである。

日本の実証研究では Ermisch and Ogawa (1994) が先駆的かつ包括的な研究である。そこでは既に女性の高学歴化と就業率の上昇が晩婚化、少子化の原因であることを指摘している。その後、家計経済研究所によって提供されている『消費生活に関するパネル調査』データを利用して樋口・阿部 (1999) や松浦・重野 (2001) が出産行動に関して分析している。樋口・阿部 (1999) では、就業、結婚、出産を包括的に分析しており、出産に関しては、Easterlin(1973) で提示された世代間の相対的経済状況に基づく出産行動仮説である「相対所得仮説」がある程度成り立っていること、高学歴化は初産年齢を引き上げていること、またその時々の労働市場やマクロ経済の状況が出産行動に影響を与えていることが示されている。松浦・重野 (2001) ではサバイバル分析によって、保育所の整備状況や親との同居が出産行動にどのような影響を与えているかを検討し、出産育児の機会費用を低下させるような要因に対しては出産促進的効果があり、機会費用を高める要因に対しては出産抑制的効果があることが示されている。

日本で出産行動について実証研究を行う場合には、『出生動向基本調査』(国立社会保障・人口問題研究所) を用いるのが標準的であろう。著者は現在、このデータの利用許可を得ていないので、残念ながら、今回の講義ではアメリカの『一般社会調査』(General Social Survey: GSS) を用いる。この調査は 1972 年より隔年、あるいは毎年行われてきたクロスセクション調査である。このデータは一般公開されており誰でも使える。著者は Winkelmann and Boes (2006) で利用されているデータセットを利用させてもらった。ここで用いたのは 1974 年から 2002 年までの 4 年毎のデータである。この一般社会調査は人口学的な質問、社会問題一般に関わる質問など 1972-2004 年までに 3260 の質問に対し、延べ 38000 人分の回答を得ている。調査主体は National Opinion Research Center (NORC) である。

本論に入る前に、ここで用いるデータの基本的情報を見ておきたい。現在では高齢者出産も増えてきたが、一般には女性 40 才ではほぼ出生行動を終えていると判断して、表 1 では 40 才以上の女性の生涯出生児数の頻度を度数分布として表している。容易に想像がつくように、アメリカでも平均的な子供の数は 2 人であり、次いで 3 人、1 人と続く。ここで注目したいのは子供を生んでいない 14.45% の女性に関してである<sup>2</sup>。まず、子供を産んでいない女性に焦点を当てて分析してみよう。

<sup>2</sup>日本の『出生動向基本調査』(1940-2002 年、5 年毎調査) で見た場合、2002 年調査 (第 12 回) でも子供を持たない夫婦は 3.4% と極めて低い、アメリカの一般社会調査では独身女性も含んだ上での数字であることに注意されたい。

## 2 出産に関するプロビット・ロジット推定

質的（離散的）従属変数を用いたモデルは、具体的には、車を買うかどうか、あるいは車を所有しているかどうか、住宅を買うかどうか、労働組合に参加するかどうか、結婚するかどうかなどの意思決定問題として扱うことができる。出産も全く同じ枠組みで分析することが可能である<sup>3</sup>。

従属変数  $y_i$  は一般に選択しなければ 0、選択すれば 1 の 2 項選択の形をとることが多い。このような 2 項選択の確率モデルはベルヌーイ確率関数 (Bernoulli probability function) として定式化できる。

$$f(y_i|x_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, \quad y_i = 0, 1$$

ここで  $\pi_i = \pi(x_i)$  は  $\Pr(y_i = 1|x_i)$  の省略形である。経済主体  $i$  が意思決定をする確率を  $\pi_i = \pi(x_i)$  と表せば、従属変数の期待値は  $E(y_i) = 1 \cdot \pi_{it} + 0 \cdot (1 - \pi_{it}) = \pi_{it}$  となり、これは通常、なんらかの変数 ( $x_{it}$ ) で説明される。

$$\pi_i = G(x'_i\beta)$$

確率変数  $\pi_i$  を  $x_i$  の  $l$  番目の要素について偏微分すると次のように表すことができる。

$$\frac{\partial \pi_i}{\partial x_{il}} = g(x'_i\beta)\beta_l$$

ここで  $g(x'_i\beta) = dG(x'_i\beta)/d(x'_i\beta)$ 。確率変数の偏微係数は  $x_i$  の値（要素）によって変動する。すなわち、確率変数自体が非線形関数なので、 $x_i$  のどの要素 ( $l$ ) のどの値の周辺で偏微分するかによってとる値が違ってくるのである。この点については後で触れるが、最小 2 乗法などによる線形モデルとは違い、この種の確率モデルは限界効果が一意には定まらないことに注意する必要がある。

クロスセクションデータを用いた実証研究では  $G(x'_i\beta)$  の定式化としてプロビット・モデルとロジット・モデルがそれぞれ次のように定義されている。

両端で 0 と 1 をとる単調な関数として逆関数  $\Phi^{-1}(x'_i\beta)$  に関する線形モデルを考える。

$$\Phi^{-1}(x'_i\beta) = \frac{x_i - \bar{x}}{\sigma}$$

この累積分布関数は

$$\pi_i = G(x'_i\beta) = \Phi(x'_i\beta) = \int_{-\infty}^{x'_i\beta} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

と表せる。これは標準正規分布の密度関数に他ならず、このようなモデルを**プロビット・モデル**と呼ぶ。具体的には上式を冒頭に定義したベルヌーイ確率関数に代入すればよい。

<sup>3</sup>本節では Winkelmann and Bose (2006、第 4 章)、Cameron and Trivedi (2005、第 14 章) を参照している。

$$f(y_i|x_i) = [\Phi(x'_i\beta)]^{y_i} [1 - \Phi(x'_i\beta)]^{1-y_i}, \quad y_i = 0, 1$$

また同様にしてロジスティック分布の密度関数を考えることもできる。

$$\pi_i = G(x'_i\beta) = \Lambda(x'_i\beta) = \frac{\exp(x'_i\beta)}{1 + \exp(x'_i\beta)}$$

$\Lambda(x'_i\beta)$  はロジスティック分布の累積分布関数を表している。累積分布関数  $\Lambda(x'_i\beta)$  を  $x_i$  に関して微分すると、ロジスティック分布の密度関数が得られる。

$$\frac{d\Lambda(x'_i\beta)}{dx_i} = \Lambda(x'_i\beta)[1 - \Lambda(x'_i\beta)]$$

**ロジット・モデル**は累積分布関数をベルヌーイ確率関数に代入することで得られる。

$$f(y_i|x_i) = [\Lambda(x'_i\beta)]^{y_i} [1 - \Lambda(x'_i\beta)]^{1-y_i}, \quad y_i = 0, 1$$

プロビット・モデルもロジット・モデルも標本  $\{y_{it}\}$  を得る同時確率は次のような尤度関数として表せる。

$$L(x) = \prod_{y_i=1} F(x'_i\beta) \cdot \prod_{y_i=0} (1 - F(x'_i\beta))$$

この式に対して最尤法推定を行うことで一致推定量を得ることができる。標本の大きさが増加するに従って、真のパラメータ値に近づく。不偏性は一般には成り立たない。ここで注意しなければならないことは、プロビット・モデルとロジット・モデルの推定パラメータは直接比較できないということである。直感的に言えば、プロビット・モデルもロジット・モデルもともに非線形の確率分布に従っていて、その曲率が違うので説明変数  $x_i$  の平均値で評価したパラメータを比較することはできないということである<sup>4</sup>。

**限界確率効果** (marginal probability effect: MPE) は次のように定義できる。

$$MPE_{il} = \frac{\partial \pi_i}{\partial x_{il}} = \frac{\partial G(x'_i\beta)}{\partial (x'_i\beta)} \frac{\partial x'_i\beta}{\partial x_{il}} = g(x'_i\beta)\beta_l$$

さらに具体的に、プロビット・モデルの場合の限界効果は次のように表せる。

$$MPE_{il}|_{probit} = \phi(x'_i\beta)\beta_l$$

ロジット・モデルの場合。

$$MPE_{il}|_{logit} = \Lambda(x'_i\beta)[1 - \Lambda(x'_i\beta)]\beta_l$$

<sup>4</sup>極めて大雑把な比較のためには、ロジット推定のパラメータはプロビット推定のパラメータに 1.6 倍したもので近似できると考えられている。

上式から限界効果は  $x_i$  のどの要素 ( $l$ ) で評価されるか、誰 ( $i$ ) を評価しているかで違ってくる。総合的な評価の仕方の一つは限界効果の平均を求めるという考え方で、**平均限界効果**(average marginal probability effect) と呼ばれているものである。

$$AMPE_l = \frac{1}{n} \sum_{i=1}^n g(x_i' \hat{\beta}) \hat{\beta}_l$$

もう一つの評価の仕方は変数  $x$  の期待値  $E(x)$  で評価するというのである。これを**期待限界効果** (expected marginal probability effect) と呼ぼう。

$$EMPE = g(\bar{x}' \hat{\beta}) \hat{\beta}_l$$

$AMPE$  と  $EMPE$  は  $g(x' \beta)$  が非線形であり一致しないことに注意されたい。

ロジット・モデルの場合、確率比である**オッズ比**(odds ratio) で説明変数の効果を評価することができる。

$$\frac{P(y_i = 1 | x_i)}{P(y_i = 0 | x_i)} = \frac{\pi_i}{1 - \pi_i} = \exp(x_i' \beta) \Rightarrow \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = x_i' \beta$$

また、説明変数  $x$  の限界的な変化 (factor change) に対してオッズ比がどのように変化するかを考えると次のようになる。

$$\frac{\exp(x_i' \beta + \Delta x_{il} \beta_l)}{\exp(x_i' \beta)} = \exp(\Delta x_{il} \beta_l)$$

もし変数  $x$  の変化が 1 単位 ( $\Delta x_{il} = 1$ ) であれば、上式は  $\exp(\beta_l)$  と表し、これを**オッズ比**と呼ぶこともある。

説明変数  $x$  の限界的な変化に対するオッズ比の**相対的变化** (relative change) は次のように表せる。

$$\frac{\exp(x_i' \beta + \Delta x_{il} \beta_l) - \exp(x_i' \beta)}{\exp(x_i' \beta)} = \exp(\Delta x_{il} \beta_l) - 1$$

これはオッズ比の変化率を表示したものであり、こちらの表示が使われることもある。

具体的に出産の意志決定をモデル化してみよう<sup>5</sup>。ここでは、子供を産まない場合を 1、産んだ場合を 0 とする二項選択モデルを考える。用いるデータは先に紹介したアメリカの『一般社会調査』である。具体的には、従属変数は子供有無ダミー (childless)、説明変数には時間トレンド (linear time trend)<sup>6</sup>、

<sup>5</sup>Winkelmann and Bose (2006, pp.111-112, pp.119-121) を参照しているが、推計は独自に行っている。

<sup>6</sup>時間トレンドは 1974=0、1978=4、1982=8、...2002=28 と 4 年毎に 4 ずつ増える設定にしている。

教育年数 (years of education)、白人ダミー (white)<sup>7</sup>、兄弟姉妹数 (number of siblings)<sup>8</sup>を用いた。

推定結果は表 2 に掲載されている。推定方法は線形確率モデルに基づく OLS 推定、プロビット推定、ロジット推定の 3 種類である。既に論じたように、推定結果のパラメータを直接比較することは出来ないが、符号条件は OLS 推定の定数項を除いて一致している。また、係数の有意性もほぼ同じ結果になっている。白人ダミーと時間トレンドはそれほど有意な説明力は持っていないのに対して、教育年数は有意に正に効いており、教育水準の高まりが出生を遅らせるあるいは阻害する方向に働いていることがわかる。推計モデルでは従属変数は子供の有無で表現してあるが、生涯出生児数と教育年数の関係を図示したものが図 1 である。この図から明らかなように、教育年数の増加は出生児数を低下させていることがわかる。とりわけ大学進学をすると教育年数は 16 年を超え、その場合は平均で出生児数は 2 人を割り込んでいる<sup>9</sup>。次に兄弟姉妹数の係数は負であり、出生を増加させる傾向にあることが示されている。しかし、脚注 8 でも述べたように、兄弟姉妹数と出生児数の関係は単調な線形関係ではない。これを確認するためには図 2 を参照していただきたい。確かに、兄弟姉妹数が 10 人ぐらいまでは単調に出生児数は増加するが、その水準を超えると出生児数の分散が大きくなり、かつ平均では出生児数は低下しはじめている。ちなみに、日本もかつては大家族制で兄弟姉妹数が 10 人いる人も稀ではなかったが、Ermisch and Ogawa (1994) の日本の実証結果では兄弟姉妹数は出生に関して有意な説明力は持っていないことが報告されている。

表 3 は表 2 と同じ推定モデルに関して限界効果を求めたものである。プロビット推定に関しては限界効果をロジット推定に関してはオッズ比を載せてある。プロビットの限界効果は説明変数の平均値を用いて評価した値で、教育年数が 1 年増えると、子供を産まない確率が 0.0071% 増えるということの意味している。もちろんこれは平均教育年数 (約 12 年) 周辺で評価した値であり、実際に考える際には、高卒に人が大学に行くことによって、生涯子供を産まない確率が何パーセント増加するかという計算をしなければあまり意味はない。それをすると 0.03% 増えることがわかる。表 1 で見たように生涯子供を産まない人のシェアは 14.5% 程度であり、教育年数が子供を産まな

<sup>7</sup>白人ダミーは白人であれば 1、それ以外の人種であれば 0 を割り振ってある。

<sup>8</sup>兄弟姉妹の数が出生の意思決定に影響を与えるかどうかということは、複数の意味を持っている。(1) その家系が多産系かどうかのコントロール変数になる。しかし、サンプルの中には兄弟姉妹数が 35 人という人もおり、それはむしろ逆効果であることがわかる。すなわちある水準 (10 人程度) を超えると、むしろ平均的には子供数は低下してくる。(2)(1) の点に関連しているかもしれないが、宗教的なバックグラウンドを反映している可能性もある。これには、避妊を認めない立場や一夫多妻制を認める立場などが含まれているかもしれない。(3) また時間の経過とともに核家族化が進んでおり、大家族制をとっていたのはかなり古い世代であり、その人たちも大家族を再生産していたと考えれば、サンプルを年代別に分割して見る必要がある。(4) より現実的には、近くに兄弟姉妹がいれば、出産や育児の際に互いに助け合うことができるという外部性があるかもしれない。

<sup>9</sup>このサンプルには独身のサンプルも入っているので平均が 2 人を割りこんでいるが、結婚している人でみれば 2 人は超えている。

い確率に与える影響は全体としてみれば少ないとはいえる。ロジット推定のオッズ比では、教育年数が1年延びることで子供を産まないという確率と産むという確率の比は1.0651であり、相対変化でみると産まない確率が相対的に6.51%増加することを意味している。

### 3 初産に関するサバイバル分析

前節では教育年数が出産の意思決定に負の効果を持っていることを見たが、それが単に出産の先延ばしで、最終的に結婚した人が2-3人の子供を産めば問題はないのではないかという議論もありうる。そこで本節では、出生のタイミングが教育年数とどのように関連しているのか、また、生涯出生児数にどのような影響を与えるのかを調べてみたい。

ある意思決定やライフイベント、例えば、学歴、就業、結婚、出産などのタイミングを分析するには、サバイバル分析の手法を用いることが有益である。サバイバル分析をするためには特にパネルデータは必要はなく、クロスセクションデータにライフ・イベントに関する情報が含まれていれば十分である。むしろ、パネルデータでは同一個人データを追っているためライフイベントの情報はサンプル数に限定されてしまうが、繰り返し行われるクロスセクションデータのサンプルから情報を集めれば、かなり多量なライフイベント情報が蓄積されることになるので、サバイバル分析には繰り返しクロスセクションデータを用いる方が望ましいといえる。

以下では、サバイバル分析の基本的概念について概説しておきたい<sup>10</sup>。

ある主体にある事象（ここでは初産）が起こるまでの時間を確率関数の形で表したものを、**生存率関数**(survival rate function) あるいは**生存時間関数**(survivorship function) と呼び、ある事象が起こる時間を  $T$  とすると、生存確率は  $Pr(T > t) = S(t)$  と表すことができる。また、ある事象が起こる時間は個々の主体により変動し、それが  $T$  より短い場合の確率密度関数  $f(x)$  と累積分布関数  $F(x)$  を次のように定義する。

$$F(t) = Pr(T \leq t) = \int_0^t f(s) ds$$

とすると、生存確率は次のように表せる。

$$S(t) = Pr(T > t) = 1 - F(t)$$

生存確率密度関数は1から0への単調減少関数である。**平均生存時間**は次のようになる。

<sup>10</sup>本節は Winkelmann and Bose (2006 第8章)、(Cameron and Trivedi (2005、第17-18章)を参照している。

$$E(T) = \int_0^\infty (1 - F(u))du = \int_0^\infty S(u)du$$

すなわち、平均生存時間は生存確率密度関数の面積と等しい。

**ハザード関数**は  $t$  期まで生存する（出産しない）という条件の下で次の期に事象が起こる確率を表している。

$$\lambda(t) = f(t)/S(t) = -d \ln(S(t))/dt$$

ハザード関数と生存確率密度関数の関係は次のように表される。

$$S(t) = \exp\left(-\int_0^t \lambda(u)du\right)$$

**累積ハザード関数**は次のように定義される。

$$\Lambda(t) = \int_0^t \lambda(s)ds = -\ln S(t)$$

これまでは時間  $t$  は連続数として扱ってきたが、実証上用いている時間は離散数であり、関数型も離散型で考える必要がある。しかし、基本的な考え方は同じなので、ここでは簡便化のために連続数として扱っている。

実証研究でパラメトリックにハザード関数を推定するためには、その関数型を特定化する必要がある。ハザード率  $\gamma$  が一定の場合には**指数分布**(exponential duration distribution)<sup>11</sup>が用いられる。ハザード率  $\gamma$  が一定でない場合には**ワイブル分布**(Weibull distribution)<sup>12</sup>、**ゴンペルツ分布** (Gomperts distribution)、**対数正規分布** (Log-normal distribution)、**対数ロジスティク分布**(Log-logistic distribution)、**ガンマ分布**(Gamma)などが用いられる<sup>13</sup>。

これらのハザード関数を分類するとき、ハザード関数が各主体に共通の**ベースラインハザード**(baseline hazard :  $\lambda_0(t)$ ) と各主体に独自の**相対ハザード**(relative hazards :  $r(x'_i\beta)$ ) の積で表されるような場合を**比例ハザードモデル** (proportional hazards (PH) models) と呼び、次のように定義する。

$$\lambda(t|x'_i\beta) = r(x'_i\beta)\lambda_0(t)$$

このモデルに従うハザード関数には**指数分布**、**ワイブル分布**、**ゴンペルツ分布**が含まれる<sup>14</sup>。

これに対して、**対数正規分布**、**対数ロジスティク分布**、**ガンマ分布**は**加速ハザードモデル**(accelerated failure time (AFT) models) に分類されている。

<sup>11</sup>指数分布関数の場合、生存確率関数は  $S(t) = \exp(-\gamma t)$  となり、密度関数  $f(t) = -S'(t) = \gamma \exp(-\gamma t)$ 、累積ハザード関数は  $\Lambda(t) = -\ln S(t) = \gamma t$  である。

<sup>12</sup>ワイブル分布の場合、 $\lambda(t) = \gamma \alpha t^{\alpha-1}$  と定義されている。これは比例ハザードモデルの特殊型である。

<sup>13</sup>これらの関数の詳細に関しては、例えば、Cameron and Trivedi (2005, Table 17.5) を参照。計量ソフト STATA ではこれらの関数を自由に選ぶことができる。

<sup>14</sup>比例ハザードモデルの詳細については中村 (2001) を参照。

ここでは、生存時間の対数の変化が説明変数  $x_i$  の変化によって説明できると考え、次のようにモデル化する。

$$\ln t = x_i' \beta + u_i$$

ここから、

$$t = \exp(x_i' \beta) u_i \Leftrightarrow u_i = t \exp(-x_i' \beta)$$

ハザード関数は次のように表せる。

$$\lambda(t | x_i' \beta) = \lambda_0(t \exp(-x_i' \beta)) \exp(x_i' \beta)$$

加速ハザードモデルとはベースラインハザードが時間とともに加速していくことを想定したモデルを意味している。ここで誤差項  $u_i$  の分布を正規分布と仮定すれば、ハザード関数は**対数正規分布モデル**になる。誤差項がロジスティック分布に従っていれば、ハザード関数は**対数ロジスティック分布モデル**になる。同様に誤差項がガンマ分布に従っていれば、**ガンマ分布モデル**になる。

これまではハザード関数自体には誤差が含まれることは想定していなかったが、実証上は観察不可能な主体の**異質性**(unobserved heterogeneity あるいは frailty) があることで、説明変数  $x_i$  だけでは説明できないハザードが起こることは知られている。その場合は対数ハザード関数は  $\ln \lambda(t_i | x_i' \beta, \varepsilon_i)$  と定義される。さらに、誤差項が**相加的**に入ってくれば次のような定式化ができる。

$$\ln \lambda(t_i | x_i' \beta, \varepsilon_i) = x_i' \beta + \varepsilon_i$$

誤差項が**相乗的**に入れば次のように表せる。

$$\ln \lambda(t_i | x_i' \beta, v_i) = \exp(x_i' \beta) v_i$$

ここで  $v_i = \exp(\varepsilon_i)$ 。このような場合、誤差項は**相乗的異質性** (multiplicative heterogeneity) を示しているという。誤差項  $v$  の密度関数を  $g(v_i | x_i)$  と定義し、この関数型を決める必要がある。実証的には**ガンマ分布**が用いられている。

$$g(v_i | x_i) = \frac{\gamma^\theta}{\Gamma(\theta)} v_i^{\theta-1} e^{-\gamma v_i}$$

ここで  $\Gamma(\theta) = \int_0^\infty z^{\theta-1} e^{-z} dz$ . 誤差項の期待値を 1 に正規化すると ( $E(v_i | x_i) = \theta/\gamma = 1$ )、 $\theta = \gamma$  となり、整理するとハザード関数は次のように表せる。

$$\lambda(t_i | x_i' \beta) = \exp(x_i' \beta) [1 + \theta^{-1} \exp(x_i' \beta) t_i]^{-1}$$

ここで  $\theta^{-1}$  はガンマ分布で表された誤差項  $v$  の分散を表すパラメータであり、 $\theta^{-1} > 0$  ではハザード率は  $t_i$  の減少関数となる。

本節では初産をイベントとし、出産した人はサンプルから外れ、出産していない人が残っていくという想定の子バヤル分析を行う<sup>15</sup>。現実的には、出産をイベントとして考えると、出産を繰り返す人はサンプルに再び入ってくるはずである。そのようなタイプのデータを扱う方法もあるが、ここでは初産に焦点を当てることで、生涯に一度あるか無いかの繰り返しのないイベントとして扱っている。

子バヤル分析では所謂、生命表 (lifetable) を作り、生存確率を計算する。ここでは前節と同じアメリカの『一般社会調査』のデータを用いているが、2002年のデータに限定しているため、総サンプル数は1517人で、そのうち出産した人は1154人である。アメリカのデータでは初産年齢は9-42歳までに分布しており20代で初産を経験する確率が最も高くなり、30歳で出産していない人は17%まで減ることがわかる<sup>16</sup>。

先ほど見たハザードモデルのうち、加速ハザードモデル (AFT) 系のモデルを選んで推定した結果が表4に報告されている<sup>17</sup>。表2-3で用いた変数以外に新しく使われている変数の定義は次の4つである。親の一方か両方が移民であるダミー変数 (immigrant)、16歳の時に家計所得が平均より低かったかどうかの貧困ダミー変数 (low income at age 16)、16歳の時に人口5万人以上の都市に住んでいたかどうかの都市ダミー変数 (lived in city at age 16)。

推定パラメータは対数正規推定も対数正規 (誤差項ガンマ分布) 推定もほぼ同じであり安定しているが、モデル全体の当てはまりは、誤差項にガンマ分布を当てはめて相乗的異質性に対応したモデルが良さそうである。初産の年齢を有意に高める効果があるのは教育年数、白人ダミー、移民ダミーであり、初産を低める効果は兄弟姉妹数、貧困ダミー、都市ダミーであるが、貧困ダミーを除いて有意ではない。とりわけ、教育年数は際だって強い年齢引き上げ効果を持っている。図3は教育年数と初産年齢の関係をプロットしたものである。教育年数が12年ぐらいまでは分散は大きい平均して初産年齢は20歳程度で安定しているが、教育年数が12歳を超えると初産年齢は単調に増加していく。

しかし、教育年数が初産年齢に強い効果を持つようになったのは最近のことであるという議論がある。それを確認するために、サンプルを生年に応じてコーホート別に表4と同じモデルを推定した。表5はコーホート別の教育年数のパラメータだけを取り上げて比較したものである。表5によると年配のコーホート (1901-1920 生まれ、1921-1940 生まれ) と比べると若いコーホート (1941-1960 年生まれ、1961-1983 年生まれ) の係数は絶対値も2倍以上に上昇し、有意性も上昇している。すなわち、高学歴化は戦後の現象であり、そ

<sup>15</sup> Winkelmann and Bose (2006, pp.255-259, pp.275-277) を参照しているが、サンプル、推定方法は独自に選択している。

<sup>16</sup> 1994年から2002年までの出産経験者総数3312人に拡張すると、30歳時点でまだ出産していなかった人は7%程度まで低下する。

<sup>17</sup> 指数分布、ワイブル分布、ゴンベルツ分布などを当てはめたハザードモデルも推定したが、表4に載せてある対数正規分布の場合と比べると、モデルの当てはまりは悪かった。

の結果としての初産年齢の上昇も戦後、特に 1960 年以後生まれ世代の現象であることがわかった。

初産をハザードとして推計したモデルでは、その後何人の子供を産んだかは分からないので、1994-2002 年のデータを用いて初産年齢と生涯出生児数との関係をプロットしたのが図 4 である。この図が意味していることは重要である。すなわち、生涯出生児数を 2 人以上にすることが、人口減少をくいとめる最低限の条件であるとするれば、初産年齢は遅くとも 30 歳あるいはそれよりも若いことが望ましいことを示している。このことは例えば女性の学歴が上昇し、就業が一般化したとしても、結婚や出産は 30 歳前に経験しておくことが望ましいということになる<sup>18</sup>。女性にとっては学業の終了と就業、結婚、出産が 20 代に集中してあわただしいかもしれないが、出産可能年齢の限度が 40 代後半にあり、これは科学技術の進歩によっても、それほど大幅には延長させることができないとするれば、体力的にも余裕のある 20 代での第 1 子出産が望ましいと言えそうである。

## 4 おわりに

今回は連載の第 5 回ということで、ミクロ計量経済学の中心的分析手法であるプロビット分析、ロジット分析を中心に解説し、経済学では比較的新しい手法であるサバイバル分析についても紹介した。

今回の講義は出産行動を題材にしたが、晩婚化・少子化の主因は女性の高学歴化と就業機会の増大であるとされている。もちろん、女性の高学歴化、高所得化あるいは男女間格差の解消などは何ら否定すべきものではなく、むしろ歓迎すべきことである。しかし、その結果として、人間として当たり前の行動である結婚や出産、育児がしにくくなったり、過度の負担が女性にかかるということであれば、それは仕事の仕方や、出産育児の負担の軽減を図るような仕組みを社会として作らなければならないことを意味している。

現在、政府は次々と少子化対策を打ち出しているが、総花的な政策が多く、ミクロ計量経済学を用いた限界的なインセンティブの分析は不十分である。例えば、第 3 子を生みやすくするためのインセンティブは初産のインセンティブとは違うことは Heckman and Walker (1990) でも明らかにされている。これは精緻なミクロ計量経済学の分析を行って始めて明らかになることであり、日本でもこの種の分析がさらに進められることが危急の課題であろう。また、今回の講義でも指摘したが、初産のタイミングと出産総数の関係を見ると、与えられた出産可能期間の中で、労働機会の喪失や人的資本の劣化を阻止しつつ、いかに最適な間隔で出産を続けられるか、またそのためには初産は何

<sup>18</sup> 日本の場合、2005 年の『人口動態統計』によると、女性の初産平均年齢は 29.1 歳にまで上昇してきている。

歳で経験するのが望ましいかなどの分析も必要である。これらのタイミングの問題も政策提言にはほとんど反映されていないが重要な論点である。

## 参考文献

- [1] 北村行伸 (2005) 『パネルデータ分析』、岩波書店
- [2] 中村剛 (2001) 『Cox 比例ハザードモデル』、朝倉書店
- [3] 樋口美雄 (1996) 『労働経済学』、東洋経済新報社
- [4] 樋口美雄、岩田正美 (編) (1999) 『パネルデータからみた現代女性 結婚・出産・就業・消費・貯蓄』、東洋経済新報社
- [5] 樋口美雄、阿部正浩 (1999) 「経済変動と女性の結婚・出産・就業のタイミング」、樋口・岩田 (1999)、第 1 章、pp.25-65.
- [6] 松浦克己、滋野由紀子 (2001) 『女性の選択と家計貯蓄』、日本評論社
- [7] Cameron, A.C. and Trivedi, P.K.(2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.
- [8] Easterlin, Richard A.(1973) “Relative Economic Status and the American Fertility Swing”, in Sheldon, E.B. (ed) *Family Economic Behavior: Problems and Prospects*, Lippincott.
- [9] Ermisch, John F.(1989) “European Women’s Employment and Fertility Again”, *Jornal of Population Economics*, 3., pp.3-18.
- [10] Ermisch, John F.(2003) *An Economic Analysis of the Family*, Princeton University Press.
- [11] Ermisch, John and Ogawa, Naohiro.(1994) “Age at Motherhood in Japan”, *Journal of Population Economics*, 7, pp.393-420.
- [12] Gustafsson, Siv.(2000) “Optimal Age at Motherhood. Theoretical and Empirical Considerations on Postponement of Maternity in Europe”, *Journal of Population Economics*, 14, pp.225-247.
- [13] Happel, S.K., Hill, J. K. and Low, S.A.(1984) “An Economic Analysis of the Timing of Childbirth”, *Population Studies*, 38(2), pp.299-311.
- [14] Heckman, James J.and Walker, James R.(1990) “The Third Birth in Sweden”, *Journal of Population Economics*, 3, pp.235-275.

- 
- [15] Hotz, V. Joseph, Klerman, Jacob Alex., and Willis, Roberts J.(1997) “The Economics of Fertility in Developed Countries”, in Rosenzweig, M.R. and Stark, O.(eds) *Handbook of Population and Family Economics*, Elsevier, vol. IA, Chapter 7, pp.275-347.
- [16] Singer, Judith D. and Willett, John B.(2003) *Applied Longitudinal Data Analysis*, Oxford University Press.
- [17] Willis, Robert J.(1973) “A New Approach to the Economic Theory of Fertility Behavior”, *Journal of Political Economy*, 81(2), Part 2, pp.S14-S64.
- [18] Willis, Robert J.(1999) “A Theory of Out-of-Wedlock Childbearing”, *Journal of Political Economy*, 107(6), Part 2, pp.S33-S64.
- [19] Willis, Robert J. and Rosen, Sherwin. (1979) “Education and Self-Selection”, *Journal of Political Economy*, 87(5), Part 2. pp.S7-S36.
- [20] Winkelmann, Rainer and Boes, Stefan.(2005) *Analysis of Microdata*, Springer.
- [21] Wooldridge, Jeffrey. M.(2003) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press

表1 40才以上の女性の出生児数の分布

出生児数	頻度	%
0人	744	14.45
1人	706	13.71
2人	1,368	26.56
3人	1,002	19.46
4人	593	11.51
5人	309	6.00
6人	190	3.69
7人	89	1.73
8人以上	149	2.89
合計	5,150	100

(出典) Winkelmann and Bose (2006), Table 1.1

表2 出産に関するプロビット・ロジット推定

従属変数: childless	OLS推定		プロビット推定		ロジット推定	
	係数	t値	係数	z値	係数	z値
説明変数:						
linear time trend	0.0006	1.12	0.0027	1.11	0.0050	1.10
years of education	0.0076	4.03	0.0314	3.89	0.0630	3.97
white	0.0142	1.08	0.0626	0.99	0.1287	1.07
number of siblings	-0.0024	-1.53	-0.0117	-1.60	-0.0211	-1.50
constant	0.0397	1.37	-1.5030	-11.40	-2.6764	-10.15
観察値	5150		5150		5150	
F(4,5145)	8.33					
R-squared	0.0079					
Log-likelihood			-2107.11		-2105.96	
Wald chi2(4)			30.82		30.91	
Pseudo R squared			0.0093		0.0098	

(出典)Winkelmann and Bose (2006) Table 4.1

表3 出産に関する限界効果

従属変数: childless	プロビット限界効果		ロジット限界効果	
	p	$\delta F / \delta x$	オッズ比	z値
説明変数:				
linear time trend		0.00062	1.00499	1.10
years of education		0.00707	1.06508	3.97
white		0.01377	1.13737	1.07
number of siblings		-0.00264	0.97914	-1.50
観察値		5150	5150	
Wald chi2(4)		30.82	30.91	
Pseudo R squared		0.0093	0.0098	
Log-likelihood		-2107.11	-2105.96	

表4 初産のサバイバル分析

被説明変数: age at first birth	対数正規推定		対数正規推定(ガンマ分布)	
	係数	z値	係数	z値
説明変数:				
years of education	0.0256	12.33	0.0274	13.25
number of siblings	-0.0017	-0.88	-0.0014	-0.76
white	0.1047	7.82	0.0990	7.95
immigrant	0.0532	3.09	0.0552	3.39
low income at age 1	-0.0252	-1.83	-0.0340	-2.59
lived in city at age 1	-0.0027	-0.24	-0.0074	-0.70
constant	2.6939	83.05	2.6458	82.42
sigma ( $\sigma$ )	0.1860		0.1502	
theta ( $\theta^{-1}$ )			0.2961	
観察値	1154		1154	
Log-likelihood	303.319		325.495	
LR chi2(6)	247.55		275.04	
LR chi2 test for theta=0:			44.35	

表5 コーホート別初産のサバイバル分析による教育年数の効果

従属変数: age at first birth	対数正規推定		対数正規推定(ガンマ分布)	
	係数	z値	係数	z値
説明変数: years of education				
cohort 1901-1920	0.014	3.18	0.016	3.47
cohort 1921-1940	0.019	8.11	0.020	9.76
cohort 1941-1960	0.034	16.95	0.035	18.23
cohort 1961-1983	0.039	16.40	0.039	16.82

図1 教育年数と出生児総数

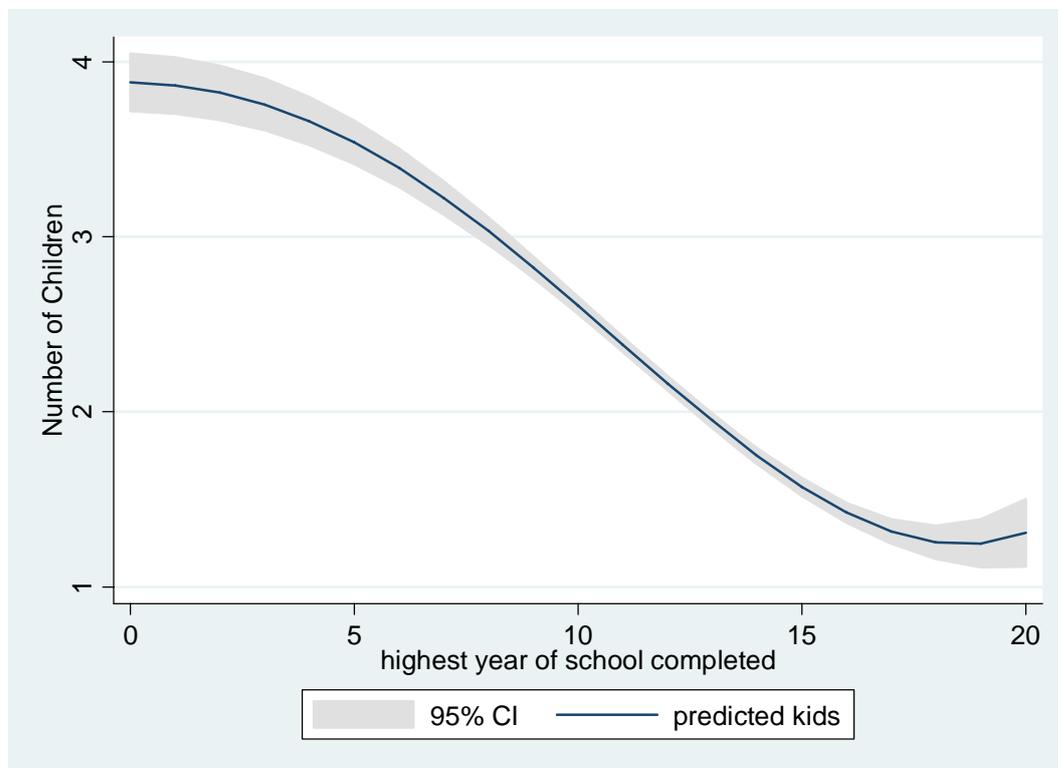


図2 兄弟姉妹数と出生児数

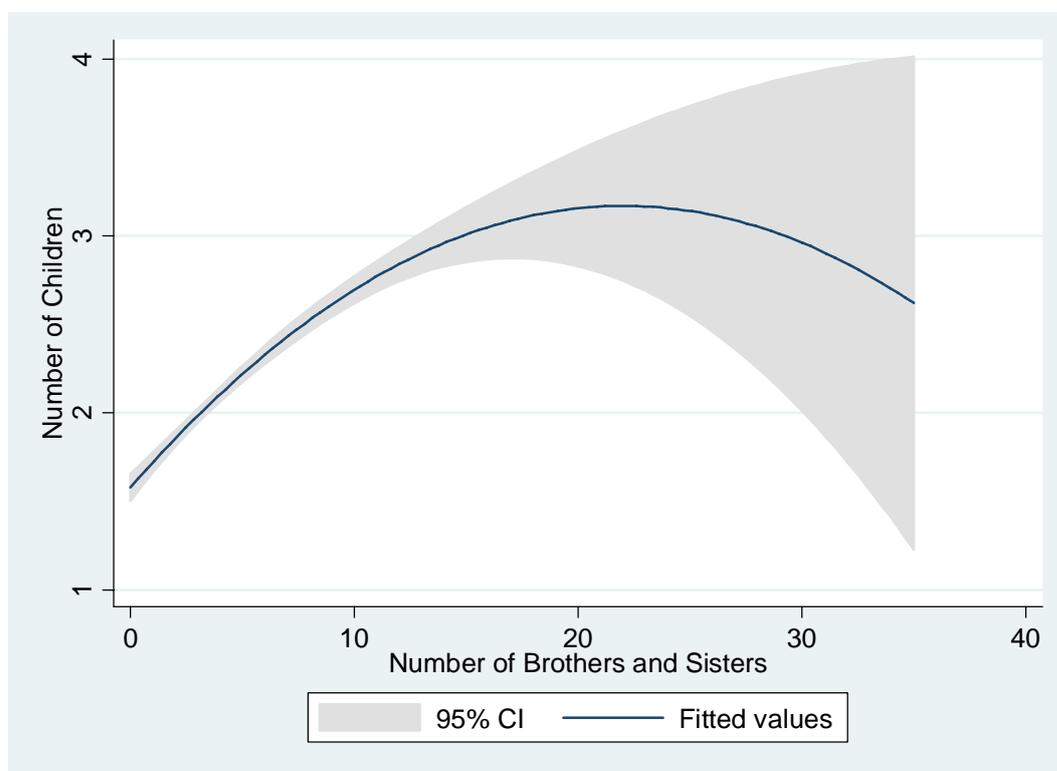


図3 教育年数と初産年齢

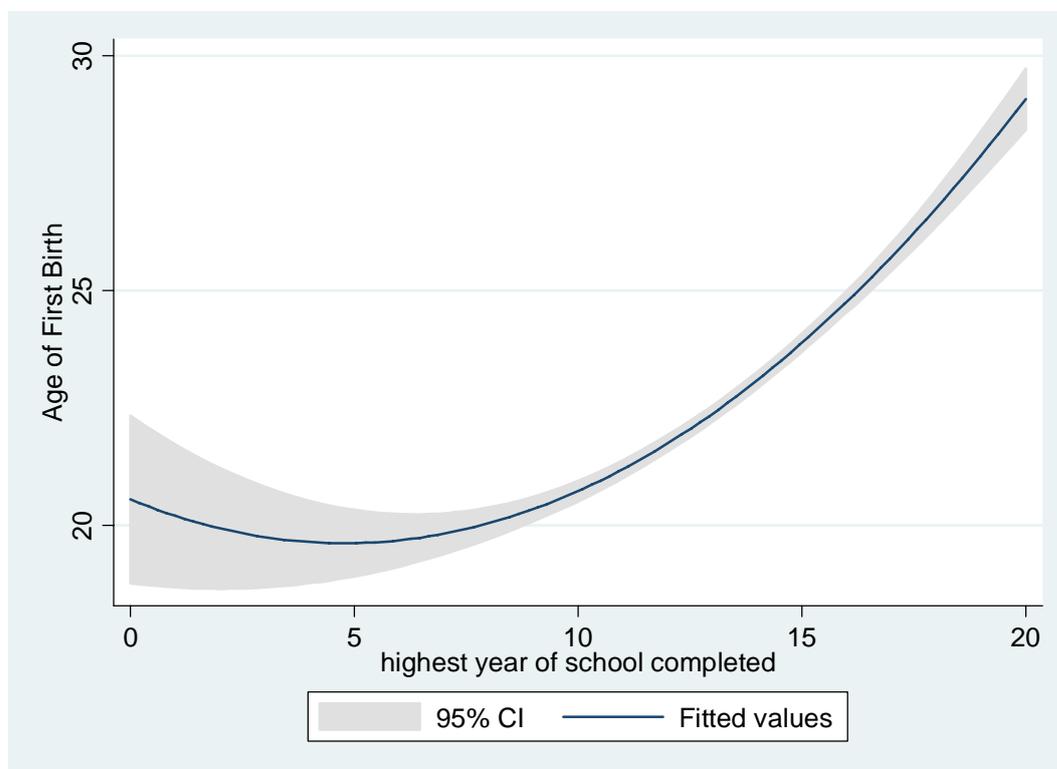


図 4 初産年齢と出生児総数

