

2-3. 離散時間ロジットモデルにおける競合イベントの取り扱い

あるイベントの生起によって、他のイベントの生起リスクがなくなる場合、2つのイベントは競合するイベント (competing events) であるという。例えば、死因別死亡率の分析において、「癌による死亡」と「心臓病による死亡」は相互に競合するイベント (mutually competing events) である。また、相互にではなく、一方のみが他方の競合するイベントとなることもある。例えば、結婚は婚前妊娠にとって競合するイベントである。しかし、婚前妊娠が生起しても、結婚のリスクはなくならないため (むしろ増大する)、婚前妊娠は結婚に競合するイベントではない。

パネル調査における脱落は、あらゆるイベントにとって競合するイベントである。なぜならば、脱落が生じることによって対象とするイベントの生起リスクが観測できなくなるためである。また、イベントの生起によって、少なくともリスク期間における脱落の発生リスクは消失する。そのため、脱落は常に対象とするイベントと相互に競合するイベントであるといえる。

山口 (2002b) によれば、離散時間モデルにおける競合するイベントの取り扱いには次の3つの方法がある。1) 競合する他のすべてのイベントをその生起時点でセンサリングとして扱う、2) 競合するイベントを従属変数とする離散時間多項ロジットモデル (discrete-time multi-nominal logit model) を適用する、そして3) 競合するイベントを従属変数とする離散時間ネステッド・ロジットモデル (以下、ネステッド・ロジットモデル) を行う。以下に山口 (2002b) を参照しつつ、どのような場合に各方法を使用すべきなのかについて解説する。なお、以下では相互に競合するイベント A とイベント B があるとする。

競合するイベントを右センサリングとして扱うという第1の方法は、最も一般的に用いられる手法である。しかし、離散時間モデルにおいてこの方法が妥当であるのは、イベント A とイベント B のハザード確率 $P_A(t)$ と $P_B(t)$ の積が無視できるほど小さい場合のみである。連続時間を仮定するモデルにおいては、競合するイベントの同時モデルにおいて、各イベントが独立に起こるという条件が成立する場合、競合するイベントをセンサリングとして扱うことが可能である。この条件が成立するには、競合するイベントがいずれも起こらない確率が各イベントの生存確率の積となる必要がある ($S_{A+B}(t) = S_A(t) \times S_B(t)$)。しかし、離散時間モデルにおいては、時点 t においてイベント A も B も起こらない確率は、 $1 - P_A(t) - P_B(t)$ であり、 $(1 - P_A(t)) \times (1 - P_B(t))$ とはならない。したがって、離散時間モデルでは、競合イベントが独立である条件の $(1 - P_A(t)) \times (1 - P_B(t))$ に対して、 $P_A(t) \times P_B(t)$ 分だけ誤差が生じることとなる。そのため、イベント A か B、あるいは双方の生起確率が著しく小さく、 $P_A(t) \times P_B(t)$ が無視できるほど小さい場合に限り、他の競合イベントをセンサリングとして扱うことが妥当となる。

$P_A(t) \times P_B(t)$ が無視できるほど小さくない場合、第2の方法である多項ロジットモデルによる競合リスク分析が検討される。この方法では、前項において解説した人・期間別データに対して、多項ロジットモデルを適用し、競合する各イベントのハザード確率の同時推定

を行う (Allison 1982)。ただし、多項ロジットモデルでは IIA (Independence from Irrelevant Alternatives) の仮定を前提としている。IIA の仮定とは、いかなる 2 つの確率の比も他の確率の大きさによる影響を受けないことをいう。 $P_A(t)$ と $P_B(t)$ がともに起こらない確率を $P_C(t)$ ($=1-P_A(t)-P_B(t)$) とすると、IIA が成立するとき、以下の関係が成り立つ。

- ① $P_A(t)/P_C(t)$ が $P_B(t)$ に依存しない
- ② $P_B(t)/P_C(t)$ が $P_A(t)$ に依存しない

① の関係が成立する時、イベント B が起こらないという条件の下でイベント A の生起確率が、イベント B の生起確率から独立である (A は B から条件付きで独立)。また、② の関係が成立する時、イベント A が起こらないという条件の下でイベント B の生起確率が、イベント A の生起確率から独立である (B は A から条件付きで独立)。IIA が成立する時、条件付きでイベント A とイベント B の決定要因が独立と考えられるため、離散時間多項ロジットモデルを適用することができる。

競合するイベントの条件付き生起確率に IIA が成立するか否かをより直接的に検証し、かつ IIA が成り立たない場合でも偏りなくパラメータを推定する方法が、第 3 の選択肢であるネステッド・ロジットモデルである。したがって、初婚要因の離散時間モデルにおいては、脱落をセンサリングとして扱う第 1 の方法とネステッド・ロジットモデルを用いる第 3 の方法を比較することによって、パラメータ推定におけるバイアスの大きさについて検討することが可能となる。また、ネステッド・ロジットモデルでは脱落と初婚の非観察要因 (shared unmeasured risk factors)⁹ に相関があるか否かを統計的に検定することができる (Hill, et. al. 1993)。IIA が成立する場合、この相関は 0 となる。そのため、ネステッド・ロジットモデルによる分析を通して、第 2 の方法である離散時間多項ロジットモデルの適用が妥当か否かを検討することもできる。次節ではネステッド・ロジットモデルの概要について述べる。

3. 離散時間ネステッド・ロジットモデルの概要

3-1. 概要

離散時間ネステッド・ロジットモデルとは、McFadden (1981) が多項ロジットモデルの拡張として導いたネステッド・ロジットモデルを Hill 等 (1993) が離散時間モデルに応用したものである。その要諦は、競合イベントの同時分析において、各イベントの誤差項に部分的な相関を許容することで、多項ロジットモデルにおける IIA の仮定を緩和することにある。以下に、Hill 等 (1993) や山口 (2002b) を参照しつつ、その概要について述べる。

⁹ 連続時間を仮定したハザードモデルにおける観察されない異質性 (unobserved heterogeneity) と同義である。

m 個の競合するイベントがある場合に、個人 i が t 時においてどのイベントを経験するのは、各イベントの潜在的な生起傾向 (state propensity index) によって決定されている。この潜在的な生起傾向は、直接には観察できない連続量 (latent variable) で、確率のような固定範囲をもたないとする。その場合、個人 i の t 時における潜在的なイベント生起傾向 S_{tmi} は以下の (3) 式によって表すことができる。

$$\begin{aligned}
 S_{t0i} &= \beta_0^* X_{t0i} + \varepsilon_{t0i} \\
 S_{t1i} &= \beta_1^* X_{t1i} + \varepsilon_{t1i} \\
 &\vdots \\
 S_{tmi} &= \beta_m^* X_{tmi} + \varepsilon_{tmi}
 \end{aligned}
 \tag{3}$$

S_{tmi} は、説明変数の分散によって説明される部分 $\beta_m^* X_{tmi}$ と誤差分 ε_{tmi} とに分けられる。式

(3) では、個人 i は S_{tmi} が最も高いイベントを経験すると仮定する。離散時間ハザードモデルにおいては、リスク開始時点においてイベント未経験の状態である S_{t0i} の値が最も高いと仮定される。他の潜在的イベント生起傾向 S_{tmi} がこれを超えるまで、いずれのイベントも生起しない。しかし、誤差項による攪乱もしくは共変量 X_{tmi} の値の変化によって、 S_{tmi} が S_{t0i} を超えると最も潜在的イベント生起傾向が高いイベントが生起する。

簡略化のため、ここで競合するイベントが 2 つであるとする。誤差項 ε_{tmi} に極値分布を仮定すると、これがイベント間で独立である場合に IIA が成立し、離散時間多項ロジットモデルを得る。しかし、 ε_{t0i} は他の 2 つから独立であるが、 ε_{t1i} と ε_{t2i} の間に相関がある場合、ネステッド・ロジットモデルを得る¹⁰。

ここで注目すべきは、多項ロジットの成立要件である IIA は、競合イベントの同時分析における誤差項、すなわち非観察要因が、各イベント間で独立であるときに成立するということである。初婚と脱落について考えてみると、これは非常に強い仮定であるといえる。なぜならば、パネル調査においては、結婚はそれ自体が脱落の要因となるためである (坂本 2006)。女性にとって結婚は転居を伴うことが多い。そのため、結婚直後のサンプル捕捉が困難となる。また、結婚により夫や夫の家族による調査拒否、またそれを忌避することによる本人からの調査拒否などが発生する (坂本 2006)。分析において用いる「21 世紀成年者縦断調査」においても同様に、少なくとも一定割合の脱落は結婚によって生起しているものと思われる。その結果、結婚と脱落の生起傾向は類似したものとなり、非観察要因についても共通の傾向をもつ可能性が高いのである。

ネステッド・ロジットモデルでは、非類似係数 (index of dissimilarity) ρ を説明変数の

¹⁰ ネステッド・ロジットモデルの数式的展開については、Hill ら (1993) や山口 (2002b) を参照のこと。

係数と同時に推定する。 ε_1 と ε_2 の相関係数は $1-\rho^2$ として表される。したがって、 ρ が1の時は競合イベントの非観察要因には相関がない、つまり IIA を仮定できることを意味する。また、 ρ の標準誤差もモデルで計算されるため、 ρ が1と統計的に有意に異なるのかの検証も行うことができる(山口 2002b)。この ρ の解釈を通して、競合するイベントの非独立性の存在やその強さについて検証することができる。また、説明変数の係数は、 ρ すなわち、競合するイベント間における非観察要因の相関、を補正した上で得られた値となる。モデルで ρ を統制することは、競合するイベントの生起過程に条件付き独立を留保した状態を統計学的に作り出すことに等しい。そのため、ネステッド・ロジットモデルにおける係数は、競合するイベントが起こらなかった場合に、説明変数が当該イベントのハザード確率に与える効果を表す。

3-2. 適用手順

初婚と脱落を競合イベントとして取り扱う場合を例として、Hillら(1993)や山口(2002b)によって示されているネステッド・ロジットモデルの適用手順を以下に示す。

(1) はじめに、通常の離散時間ロジットモデルと同様に人・期間別データを作成する。また、従属変数 $Y(t)$ はイベントが生起していなければ0、初婚が生起する場合は1、そして脱落が生起する場合は2となるようにコーディングする。

(2) (1)で作成した人・期間別データより、初婚もしくは脱落を経験したサンプル($Y(t)$ が1もしくは2のケース)のみを取り出し、初婚対脱落を対比としたロジットモデルを行う。ここでの分析は、初婚か脱落が生起したとして、それが脱落ではなく初婚である確率を推定するモデルとなる。

(3) (2)で得られた回帰係数をもとにして、以下の値を算出する。

$$z_1(t) = \log[1 + \exp(-\sum_k b_k x_k(t))]$$

$$z_2(t) = \log[1 + \exp(\sum_k b_k x_k(t))]$$

この時、 $\sum_k b_k x_k(t)$ は(2)のモデルで得られた各レコードごとの予測値を表す。 $z_1(t)$ と $z_2(t)$ の値を(1)で作成した人・期間別データの各レコードに対して計算して、各レコードに対して変数として付帯する。さらに、このデータに初婚か脱落が生起した場合に1、いずれも生起せずに未婚のままである場合に0をとる新しい変数 $Y^*(t)$ を作成して追加する。

(4) (3)で作成した人・期間別データを用いて、従属変数を $Y^*(t)$ とする離散時間ロジッ

ト分析を行う。ただし、この時 (3) で作成した $z_1(t)$ もしくは $z_2(t)$ の一方を説明変数としてモデルに追加する。 $z_1(t)$ を追加した場合には、脱落を経験せずに結婚するという競合イベントへの決定要因 β_{1k} を得る。一方、 $z_2(t)$ を追加した場合は、結婚せずに脱落するという競合イベントへの決定要因 β_{2k} を得る。なお、(2) と (4) では異なる説明変数をもつことも可能である (Hill et. al. 1993)。この時、 z_1 と z_2 の係数として算出されるのが ρ の推定値である。 ρ は z_1 と z_2 のどちらを用いても全く同じ値を示し、理論的には 0 から 1 までの値をとる。初婚と脱落の観察されない異質性 (誤差項) の相関係数は、 $1 - \rho^2$ によって与えられる。また、この時 ρ の標準誤差は ρ が 1 である、つまり初婚と脱落の相関係数が 0 であるという帰無仮説を検定している。

3-3. 利用における留意点

ネステッド・ロジットモデルの適用においてはいくつか留意する点がある。第 1 に、このモデルは連続時間を仮定する競合ハザードモデルには適用できない。離散時間モデルによる同時的な競合ハザード分析では、人・期間別データに対する多項ロジット、多項プロビット、あるいは多項 complementary log-log といったモデルを適用する。これらのモデルでは、いずれも回帰式に何らかのパラメトリックな分布をもつ誤差項 ε を仮定している。そのため、競合イベントのそれぞれについて誤差項、すなわち非観察要因の影響力に関する値を個人別に算出し、その相関係数を推定することができる。しかし、通常の連続時間ハザードモデルでは誤差項が仮定されない (Allison 1995)。連続時間モデルでは、誤差項の分布はベースライン・ハザード関数によって部分的に表現されるに留まり、明示的に定義されないためである (Blossfeld and Rowher 2002)。そのため、通常の連続時間モデルでは、競合するイベントごとに誤差項を算出することができず、したがって、その相関係数を推定することもできない¹¹。

次に、ネステッド・ロジットモデルでは、競合するイベントの非観察要因の相関は、リスク期間を通じて一定と仮定されている (Hill et. al. 1993)。したがって、非観察要因がリスク期間を通じて、初婚と脱落に異なる影響を与える場合、この仮定が成立しない。例えば、調査の初期においては結婚を契機として脱落するサンプルが多いが、調査回が進むに

¹¹ 連続時間ハザードモデルにおいて、パラメトリックな分布をもつ誤差項を仮定するモデルは、いくつか紹介されている (Blossfeld and Rowher 2002)。なかでも、誤差項の分布についてのパラメトリックな仮定が少ない Heckman と Singer (1984) によるものが代表的である。近年において、Heckman と Singer (1984) のモデルを拡張した分析手法が注目を集めている。例えば、Lillard と Waite (1993) は、誤差項を仮定した連続モデルを用いて、ハザード率の同時回帰モデルを行っている。このモデルでは、依存関係にある複数のハザード率を対象として、観察できない異質性の相関を統制した上で、各ハザード率のパラメーター推定を行うことが可能である。この方法では、対象とするハザード率は競合するイベントに限らず、同時推定を行うことが可能である。そのため、あるイベントが他のイベントに対してもつ選択バイアス (内生性 Endogeneity) を統制した上で、因果関係を検証できるという利点をもつ (山口 2002a)。しかし、その適用には aML (Lillard and Panis 2003) という特殊な統計ソフトが必要となる。なお、aML は以下のサイトから無料でダウンロードして使用することができる。

<http://www.applied-ml.com/index.html>

つれて結婚以外の事由による脱落が増えるという場合には、非観察要因の相関がリスク期間を通じて一定であることを仮定できない。この仮定が成立しない場合、時間依存性共変量のパラメーターや ρ の推定値にバイアスが生じる (Hill et. al. 1993)。しかし、非観察要因の相関がリスク期間を通じて変化する場合においても、時間固定共変量の係数についてはバイアスが少なく、比較的安定的に推定されることが指摘されている (Hill et. al. 1993)。

さらに、ネステッド・ロジットモデルでは、誤差項に負の相関を仮定することができないという制約がある (山口 2002b)。例えば、婚前同棲の解消について競合するイベントが結婚と別離である場合、非観察要因 (例えば、性格の相性) は結婚に対しては正の効果をもち、別離に対しては負の効果をもつことが十分に起こりえる。しかし、 ρ は理論上、 $0 < \rho \leq 1$ の範囲の値を取るため、非観察要因の相関係数 $1 - \rho^2$ は正であることが仮定されている¹²。したがって、非観察要因の相関が負である競合イベントは、ネステッド・ロジットモデルでは分析することができない¹³。

また、ネステッド・ロジットモデルにおける推定上の問題として、2段階推定においては、パラメーター推定値の標準誤差が平均してやや小さめに推定される可能性が指摘されている (山口 2002b)。これは1段階目のパラメーター推定値には実際には誤差があるにもかかわらず、2段階推定では定数として扱うことから生じる。しかし、通常はこのバイアスは有意度に影響を与えない程度であるため、それほど問題とはならない。

最後に、 ρ が統計的に有意に1と異なる場合 ($\rho \neq 1$)、回帰係数 β_k をオッズ比として解釈することができないという制約がある (山口 2002b)。そのため、本稿における分析ではオッズ比ではなく、係数を用いて解釈を行う。離散時間モデルにおける係数は、ハザード確率のログオッズに対する共変量の影響を表しており、共変量 X_k における1単位の増加はイベント生起のログオッズを β_k 分増加 (減少) させると解釈することができる。

4. データと分析対象

分析には、厚生労働省が2002年11月より実施している「21世紀成年者縦断調査」(以後、成年者調査)の第1回から第4回(パネル1からパネル4)までの個票データを使用した。成年者調査は、男女の結婚、出産、就業等の実態及び意識の経年変化の状況を把握することを目的として実施されているパネル調査である。同調査は、全国より無作為に抽出された20-34歳(平成14年10月末日現在)の男女及びその配偶者35,448人を対象としており、第1回調査では対象サンプルの83.4%にあたる29,566人より回答を得ている。調査は毎年11月に行われており、前回調査の回答者を対象として継続的に回答を得ている。

¹² しかし、実際の分析においては、 ρ の推定値が $0 < \rho \leq 1$ の範囲を超えることが起こりうる。 ρ の推定値が統計的に有意で、1より大きいか、0より小さい場合には2段階推定の妥当性に問題があると考えられ、その結果は信頼できない(山口 2002b)。

¹³ このような場合は、Vermunt (1997)によって提案されている離散時間多項ロジットモデルに潜在クラスを導入する方法が推奨される。

調査方法は調査員による留置き法を基本とし、第1回調査以降に転出した者に対しては郵送法が併用されている¹⁴。成年者調査はわが国における既存のパネル調査の中でも最大規模の調査の1つである。

分析対象は、第1回調査において未婚である女性である¹⁵。未婚女性のうち、調査期間を通じて性別や年齢の回答に不備がある者、ならびに分析に用いる説明変数に外れ値や欠損値がある者を除外したところ、分析の対象となるサンプル数は7,588人であった。

分析では、初婚ならびに脱落を競合するイベントとして取り扱い、第1回調査における未婚女性が結婚するか、調査から脱落するか、未婚のまま第4回調査を向かえるまでをリスク期間とした離散時間ロジットモデルならびに、その応用的分析手法であるネステッド・ロジットモデルによるイベントヒストリー分析を行う。

なお、パネルデータに対するイベントヒストリー分析では、かなりの割合で左打ち切り(left-truncation)が発生することに留意する必要がある。左打ち切りとは、観察期間より前にイベント生起のリスクが開始していることをいう(Guo 1993, Allison 1995)。本調査においては、20-34歳の女性を分析対象としている。女性の法定結婚年齢は16歳であるため、実際には同調査におけるすべての未婚女性が左打ち切りのケースといえる。結婚リスクの高いサンプルは調査開始までの間に結婚してしまう可能性が高いため、左打ち切りのケースはイベント発生リスクの低いケースに限定される可能性が指摘されている(Allison 1995)。しかし、同調査においては第1回調査のサンプルは無作為に抽出されていることから、調査開始時における未婚女性にはこのようなセレクション効果は生じておらず、無作為に選択されていると仮定できる。また、Guo (1993)はイベントのリスク開始時点が明らかである場合、条件付き尤度法(conditional likelihood approach)を用いることで左打ち切りによる偏りなく推定が行えることを示している¹⁶。本分析においては、最も早く初婚が観察される年齢である20歳を初婚リスクの開始時点と仮定して、条件付き尤度法によるパラメーター推定を行う¹⁷。

¹⁴ 留置き法では、調査員があらかじめ配布した調査票に被調査者が自ら記入し、密封したものを後日調査員が回収する留置き法によって回収されている。また、第1回調査以降に転出した者については、厚生労働省から郵送された調査票に被調査者が自ら記入し、郵送により厚生労働省に提出する方法を用いている(厚生労働省大臣官房統計情報部 2005)。

¹⁵ 「21世紀成年者縦断調査」では結婚歴に関する項目がないため、未婚者と既婚の独身者を区別することができない。そのため、本稿では子どものいない独身女性を未婚女性と定義する。「2000年国勢調査」(総務省統計局, 2001)によると、20-34歳の男女のうち、独身者に占める離死別者の割合は男が1.63%、女が4.41%である。本分析では、子どもがいる独身者を分析より除外しているため、分析サンプルにおける既婚者の割合はこれと同程度かそれ以下と推測される。

¹⁶ Guo (1993)の条件付き尤度法による分析手法は、山口(2001a)において、区間調査法によるイベントヒストリー分析として示されている。

¹⁷ ただし、今回の分析では調査開始から3年間に生起した初婚と脱落のみを取り扱う。そのため、各年齢における初婚経験は対応する出生コーホートに限定して生起している。例えば、34歳以降の結婚を経験できるのは、調査開始時に34歳であった1968年出生コーホートの女性のみである。同様にその他の年齢においても、最大で3つの単年コーホートにしか結婚が経験されない。そのため、ハザード確率の年齢効果はコーホート効果と分離することができない。このことは年齢の係数が、普遍的な効果ではなく、コーホートに特有の効果である可能性があることを意味する。

これらの分析を通して、1) 初婚と脱落の生起過程にはどの程度の相関があるのか、2) 脱落を右センサーとして扱い、通常の離散時間ロジットモデルによる分析を行った場合、初婚要因のパラメーター推定にはどの程度のバイアスが生じるのかを明らかにする。

5. 記述統計

5-1. 初婚と脱落の推移

表 1 初婚と脱落の発生頻度

		未婚→			合計
		未婚	結婚	脱落	
パネル1→2	n	6110	316	1162	7588
(2002年-2003年)*	%	80.5	4.2	15.3	100.0
パネル2→3	n	4926	304	880	6110
(2003年-2004年)	%	80.6	5.0	14.4	100.0
パネル3→4	n	4069	283	574	4926
(2004年-2005年)	%	82.6	5.8	11.7	100.0
合計	n	15105	903	2616	18624
(2002年-2005年)	%	81.1	4.9	14.1	100.0
*: 各年とも11月時点		$\chi^2(4)=84.02, p<0.01$			

表 1 は、各調査間における結婚と脱落の発生頻度を表している。パネル 1 において未婚であった女性のうち、4.2%にあたる 316 人がパネル 2 までの間に結婚している。一方、パネル 2 に回答しなかった脱落ケースは 1,162 人と全体の 15.3%に達している。同様に、パネル 2 の時点で未婚であった女性のうち、パネル 3 までの間に結婚したケースは 304 人 (5.0%) であり、脱落は 880 ケース (14.4%) であった。パネル 3 からパネル 4 においては、結婚が 283 ケース (5.8%)、脱落が 574 ケース (11.7%) となっている。

未婚女性に対する結婚と脱落の発生頻度をみると、調査回が進むにつれて、結婚が増加し、脱落が減少する傾向がみられる。調査回が進む毎に結婚が増加しているのは、より結婚が起きやすい 20 歳代後半へと回答者の年齢がシフトしていることを反映しているものと思われる。また、脱落については、調査に非協力的な回答者は早い段階で脱落するため、調査が回を重ねるにつれて、調査に協力的な回答者が残ることとなる。そのため、あとの調査回ほど脱落が少ないものと思われる。

減少傾向にあるとはいえ、調査からの脱落率が各年で 10%を超えていることに留意する必要がある。結婚と脱落の各生起確率のうち、どちらかあるいは一方が無視できるほど小さい場合には、それぞれのイベントが独立に生起していると仮定して、一方のイベントの生起を他方のイベントのセンサーされたケースとして取り扱うことが可能である (山口 2002b)。しかし、ここでは脱落の生起確率が比較的大きく、また初婚の生起確率も無視できるほどには小さくはない。そのため、脱落をセンサーとして扱い、初婚をイベントとす

る通常の離散時間ロジットモデルではパラメーター推定にバイアスが生じる可能性が高い。したがって、ここでは脱落を初婚と競合するイベントとして取り扱い、多項ロジットあるいはネステッド・ロジットモデルによって、競合イベントのパラメーターを同時推定する手法が望ましいことが示唆される。

5-2. 説明変数の定義と度数

分析に用いた共変量（説明変数）の定義ならびに度数分布を表 2 に示した。各変数はわが国における結婚の先行研究（Tsuya and Mason 1995, Ono 2003, Raymo 2003a, Raymo 2003b, Raymo and Ono 2007, 福田 2007b）を参照しつつ、同調査において使用可能な変数を選択した。

モデルでは、初婚のベースライン・ハザードは年齢の関数として表される。年齢は 20-25 歳、25-30 歳、30-37 歳までの 3 つに区分され、各区間で初婚のハザード率が線形に上昇あるいは下降すると仮定するスプライン関数によって近似する¹⁸。他の共変量はベースライン・ハザードを比例的に増減させる効果をもつことを仮定している。いわゆる比例オッズモデルである。

共変量の多くが、時間と共に値が変化することを許容する時間依存性共変量である。説明変数（原因）の従属変数（結果）に対する時間的先行を留保するため、時間依存性共変量は前年度調査で得られた値を使用した。また、リスク期間を通じて一定の値もつ時間固定共変量として、居住都道府県の SMAM (singulate mean age at marriage)¹⁹と結婚意欲を用いた²⁰。両変数共に、サンプル確定時あるいは第 1 回調査で得られた値を使用しているため、イベント生起に対する時間的先行は留保されている。

なお、休日家事時間、週当たり労働時間、そして年間勤労所得の各変数には、無視できないほど高い割合で欠損値がみられた。これらの欠損値を分析より除外するとサンプル数が大きく減少し、パラメーターの推定上望ましくない。そのため、これらの変数については、欠損値に平均値を代入して分析に含めた。また、各変数について欠損値ダミー変数を作成して、欠損値を持つケースには 1 を、そうでないケースには 0 を付した。例えば、週当たり労働時間が欠損値であるケースについては、週当たり労働時間が 1 時間以上あるサンプルを用いて算出した平均値（性、パネル別）を代入し、労働時間不明ダミーを 1 とした²¹。この処置を前述の各変数に行うことにより、いずれかの変数に欠損値であるケースに

¹⁸ ベースラインハザードの形状がそれほど複雑ではない場合、各歳別のダミー変数を使用する場合と比べて、スプライン関数では自由度が小さくて済むため、モデルの節約性が高い。

¹⁹ SMAM は国勢調査の年齢別未婚率から計算する結婚年齢であり、次式により計算する。

$$SMAM = (\sum Cx \cdot 50 \cdot S) / (1 - S) \quad Cx: \text{年齢別未婚率}, S: \text{生涯未婚率}$$

²⁰ これらの変数を時間独立共変量として用いたのは、1) 今回提供を受けたデータでは、回答者の居住地域はサンプル確定時の居住都道府県しか分からないこと、ならびに 2) 結婚意欲に関する質問が第 1 回調査と第 4 回調査でしか得られないこと、による。

²¹ 休日家事時間については、0 時間のケースも含めて平均値を算出した。

ついても、分析に使用できるようにした。なお、欠損値ダミー変数は対応する変数の平均値をリファレンスとする係数を表す。

表2 共変量の定義と度数分布

変数名	定義と度数分布
年齢 (t)	1)20-24歳、2)25-30歳、3)30-37歳に区分し、スプライン関数として表現。 1) 7546 [40.5%], 2) 6882 [37.0%], 3) 4196 [22.5%]
年次 (t)	1)2002-03年、2)2003-04年、3)2004-05年 1) 7588 [40.7%], 2) 6110 [32.8%], 3) 4926 [26.5%]
学歴 (t)	1)中学卒、2)高校卒、3)専門学校卒、4)短大・高専卒、5)大学・大学院卒 1) 420 [2.3%], 2) 6620 [35.6%], 3) 3417 [18.4%], 4) 4359 [23.4%], 5) 3808 [20.5%]
職業 (t)	1)大企業雇用、2)中小企業雇用、3)専門・技術職、4)自営業主・家族従業者・会社役員、5)非正規雇用(パート、アルバイト、派遣、契約社員)、6)無職、7)学生、8)不明 1) 2036 [10.9%], 2) 3336 [17.9%], 3) 2426 [13.0%], 4) 480 [2.6%], 5) 4904 [26.3%], 6) 1585 [8.5%], 7) 2024 [10.9%], 8) 1833 [9.8%]
親との同別居 (t)	1)親と別居、2)両親と同居、3)片親と同居、4)不明 1) 2084 [11.2%], 2) 12585 [67.6%], 3) 2021 [10.9%], 4) 1934 [10.4%]
居住都道府県のSMAM	サンプル確定時(2001年6月)に対象者が居住していた都道府県のSMAMの値。2000年と2005年の都道府県別SMAMを按分して算出した調査開始年(2002年)の値を用いた。 平均: 28.9, 標準偏差: 0.64, 最小値: 27.7, 最大値: 30.4
健康状態 (t)	1)定期的通院あり、2)入院あり、3)いずれもなし、4)不明 1) 1722 [9.3%], 2) 498 [2.7%], 3) 15776 [84.7%], 4) 628 [3.4%]
休日家事時間 (t)	休日における家事時間。平均より4標準偏差以上大きい値は外れ値として分析より除外。 平均: 1.57, 標準偏差: 1.43, 最小値: 0歳, 最大値: 12
休日家事時間不明ダミー (t)	休日における家事時間が不明なケースを表すダミー変数。この変数が1である場合、休日家事時間には平均値を代入。 0) 13500 [72.5%], 1) 5124 [27.5%]
労働時間(週) (t)	1週間の労働時間。平均より4標準偏差以上大きい値は外れ値として分析より除外。 平均: 33.06, 標準偏差: 18.17, 最小値: 0, 最大値: 95
労働時間(週)不明ダミー (t)	1週間の労働時間が不明なケースを表すダミー変数。この変数が1である場合、労働時間には平均値を代入。 0) 17577 [94.4%], 1) 1047 [5.6%]
年間勤労所得(十万円) (t)	調査年度の前年における勤労所得。平均より4標準偏差以上大きい値は外れ値として分析より除外。 平均: 205.87, 標準偏差: 109.31, 最小値: 0, 最大値: 900
年間勤労所得不明ダミー (t)	年間勤労所得が不明なケースを表すダミー変数。この変数が1である場合、勤労所得には平均値を代入。 0) 14602 [78.4%], 1) 4022 [21.6%]
結婚意欲	第1回調査における結婚意欲。1)絶対したくない、2)あまりしたくない、3)どちらとも言えない、4)なるべくしたい、5)絶対したい 1) 359 [1.9%], 2) 1246 [6.7%], 3) 4208 [22.6%], 4) 6731 [36.1%], 5) 6080 [32.7%]

(t): 時間依存性説明変数 (time-varying covariates)

*: 各共変量は前年もしくは初年度調査の値を使用した。

6. 分析結果

分析では、脱落を通常の右センサリングとして扱った場合の離散時間ロジットモデルと初婚と脱落を競合イベントとして扱ったネステッド・ロジットモデルの2つの分析を行い、パラメーター推定の結果を比較した。比較の結果を表3に示す。なお、参考までに離散時間多項ロジットモデルを用いた初婚と脱落の競合ハザード分析の結果を付表1に示した。

6-1. 離散時間ロジットモデルの結果

表3の第1列は、脱落を右センサリングとして扱った離散時間ロジットモデルの結果を示している。ベースライン・ログオッズについてみると、初婚のハザード確率は20-25歳において急激に上昇し、その後25-30歳の年齢においても緩やかな上昇を続ける。しかし、30歳以降になるとその推移は下降に転じており、初婚のハザード確率は典型的なベル型の生起パターンを示している。年次については、最終年次において初婚のハザード確率が上昇している。

先行研究(Tsuya and Mason 1995, Raymo 2003a, Raymo and Ono 2007)においては、主要な説明変数となっている学歴および職業については、統計的に有意な変数がほぼ皆無であるとの結果を得た。また、パラサイト・シングル仮説(山田 1999)の検証(Raymo 2003b)などで繰り返し用いられてきた親との同別居についても初婚の生起に対する影響は認められない。

居住都道府県のSMAMについては、有意な負の効果が認められる。未婚率が高いという地域的要因が、個人の結婚行動にも作用して、結婚のハザード確率が低い傾向がある。週当たりの労働時間については、2次関数による表現が最も良くデータに適合した。係数を実数値に当てはめてみると、週34時間までは労働時間が長いほど初婚のハザード確率が上昇する傾向がみられた。しかし、それ以上労働時間が長くなると、初婚のハザード確率は低下する。やがて、労働時間が週70時間を越えると、労働時間が0時間の女性よりも初婚のハザード確率は低くなっている。

女性の経済的自立と結婚選択の観点から注目される所得については、強い正の効果が認められた。この結果は、高収入の女性ほど結婚しにくくなるとする先行研究(Ono 2003, 福田 2007b)とは異なるものであり、この係数がバイアスによる値なのか否かは分析の含意を左右する重要な問題である。

最後に、結婚意欲については、結婚を強く望む女性ほど現実に結婚しやすいことが示されている。しかし、結婚を「あまりしたくない」、もしくは「絶対したくない」と回答した女性の結婚ハザードは、「どちらとも言えない」と回答した女性と比べて有意に異なるようである。以上の結果は、ネステッド・ロジットモデルではどのように異なるのだろうか。

表3 離散時間ロジットモデルならびに離散時間ネステッド・ロジットモデルによる初婚ハザード確率の決定要因に関する分析結果

	離散時間ロジット		ネステッド・ロジット	
	(1) 結婚ハザード β	(2) 結婚(対:脱落) β_i	(3) 結婚ハザード β_1	(4) 脱落ハザード β_2
年齢スプライン				
20-25歳	0.192 ***	0.152 **	0.118 ***	0.014
25-30歳	0.104 ***	0.081 ***	0.095 ***	0.040 ***
30-37歳	-0.127 ***	-0.082 **	-0.103 ***	-0.047 ***
年次(対:2002-03年)				
2003-04年	0.100	0.113	0.121 **	0.044
2004-05年	0.176 **	0.382 ***	0.115	-0.145 **
教育水準(対:高校卒)				
中学卒	-0.125	-0.107	0.032	0.105
専門学校卒	0.089	0.130	0.234 ***	0.145 **
短大・高専卒	0.104	0.207 *	0.189 ***	0.048
大学・大学院卒	0.103	0.166	0.316 ***	0.203 ***
職業(対:中小企業雇用)				
大企業雇用	-0.162	-0.260 *	-0.155 *	0.022
専門・技術職	0.061	0.042	-0.021	-0.050
自営・家従・会社役員	0.199	0.034	0.210 *	0.187
非正規雇用	-0.036	-0.106	-0.010	0.063
無職	0.369	0.247	0.356 ***	0.188
学生	-0.692 ***	-0.602 **	-0.406 **	0.004
不明	-0.002	-0.082	0.012	0.068
親との同別居 (対:両親と同居)				
親と別居	-0.127	-0.969 ***	0.346 **	1.006 ***
片親と同居	0.071	-0.068	0.140 **	0.187 ***
不明	0.142	-0.309 **	0.310 ***	0.520 ***
居住都道府県のSMAM	-0.308 ***	-0.485 ***	-0.187 **	0.143 ***
健康状態(対:いずれもなし)				
定期的通院あり	-0.052	0.112	-0.124 *	-0.201 ***
入院あり	0.159	-0.068	0.247 **	0.294 ***
不明	0.343 *	0.415 *	0.254 **	-0.028
休日家事時間	0.029	0.035	0.024	0.000
休日家事時間不明ダミー	-0.062	-0.351 ***	0.060	0.299 ***
労働時間	0.021 **	0.026 **	0.017 ***	-4.00E-04
労働時間の2乗	-3.00E-04 **	-3.00E-04 **	-3.00E-04 ***	0.00E+00
労働時間不明ダミー	0.030	-0.009	0.016	0.023
年間勤労所得(十万円)	0.013 ***	0.010 **	0.009 ***	0.002
年間勤労所得不明ダミー	-0.297 ***	-0.530 ***	-0.132	0.229 ***
結婚意欲 (対:どちらとも言えない)				
絶対したくない	-0.995 *	-0.786	-0.696 ***	-0.160
あまりしたくない	-0.254	-0.286	-0.210 **	-0.015
なるべくしたい	0.559 ***	0.563 ***	0.367 ***	-0.016
絶対したい	1.119 ***	1.057 ***	0.871 ***	0.151 *
z1			0.681 ***	
z2				0.681 ***
定数	-0.084	8.092 ***	-1.180	-6.725 ***
person-year数	18624	3519	18624	18624
カイ2乗値	437.670	491.072	512.905	512.905
自由度	35	35	36	36

* p<.1; ** p<.05; *** p<.01

6.2. ネステッド・ロジットモデルの結果

表3の第2列から第4列では、ネステッド・ロジットモデルによる分析結果を示している。第2列目では、1段階目の推定である初婚対脱落のロジットモデルのパラメーターを表している。1段階目の推定モデルでは、イベントヒストリー分析ではなく、初婚か脱落が生じたレコードのみを用いて、脱落を0、結婚を1とするロジットモデルを行った。

第2列目のモデルの予測値を用いて、2段階目の推定を行ったのが第3列目と第4列目のモデルである。第3列目では、脱落が起きなかったと仮定した場合における初婚ハザード確率の決定要因を示している。一方、第4列目では、結婚が起きなかったと仮定した場合における未婚者の脱落ハザード確率の決定要因が示されている。両モデルの z_1 と z_2 は同じ値を示しており、1以下の正の数となっていることから、ネステッド・ロジットモデルが成功裏になされたことが確認できる(山口 2002c)。また、 z_1 と z_2 の値は1%水準で統計的に有意であることから、初婚と脱落の生起過程には、共通の非観察要因が存在し、独立とは見なせないことが明らかである。なお、2つのモデルの誤差項の相関は $\rho = 0.536$ ($=1 \cdot 0.681^2$)と高い値を示している。これらのことからIIAの仮定は成立せず、離散時間多項ロジットよりもネステッド・ロジットモデルが適したモデルであることが明らかである。

初婚の決定要因における離散時間ロジットモデルとネステッド・ロジットモデルの違いを検証するため、表3の第1列目と第3列目のパラメーターを比較してみよう。年齢の効果については、第3列目では20歳代前半の上昇傾向と30歳以降の下降傾向が、第1列目と比べて緩やかになっている。年次については、最終年次ではなく、2年目にあたる2003年から2004年にかけて最も初婚のハザード確率が高かったことが示されている。第1列目では、脱落のハザード確率(第4列目)が最終年次で低かったために、初婚のハザード確率が相対的に過大に推定されていたことが示唆される。

第1列目において統計的有意性がみられなかった学歴と職業による初婚ハザード確率の差異は、第3列目のモデルではより明確に現れている。第3列目では、専門学校卒以上の学歴において、高校卒以下よりも初婚ハザード確率が高い。また、職業については、中小企業雇用の女性に比べて、大企業雇用の女性ほど結婚しにくく、自営や家族従業・会社役員的女性ほど結婚しやすい傾向が弱いながらも認められる($p < .10$)。また、学歴や労働時間、そして所得を統制した上での結果であるが、無職の女性ほど結婚しやすい傾向も1%水準で有意となっている。

親との同別居についても、ネステッド・ロジットモデルでは統計的に有意な要因へと転じている。親と別居している女性あるいは片親と同居している女性は、両親と同居している女性よりも初婚ハザード確率が高い。第4列目をみると、脱落のハザード確率にも同様の傾向がみられる。したがって、初婚と脱落の決定要因が相互に打ち消しあった結果、第1列目ではこれらの変数による影響が統計的に有意とならなかったものと思われる。

居住都道府県のSMAMについては、第1列目と比べて弱まっている。SMAMが高い地

域では脱落のハザード確率が高い。そのため、SMAM が高い地域における初婚ハザード確率が過小に推定されていたことが分かる。

結婚意欲についてみると、第 1 列目では結婚意欲が低い場合の負の効果が不明瞭であった。しかし、第 3 列目では結婚意欲別の初婚ハザード確率の差異がより明瞭に表れており、結婚意欲に準じて初婚のハザード確率も増減する様子がみてとれる。また、結婚意欲が最も高い層において、脱落のハザード確率が高い傾向が弱いながらも認められる。このことから、成年者調査において結婚と脱落との間に相関があることが示唆される。

最後に、労働時間と勤労所得についてみると、これらの変数は脱落のハザード確率には影響せず、初婚ハザード確率にのみ統計的に有意な影響を与えている。このことは第 2 列における初婚対脱落のロジット分析においても明らかである。その結果、これらの変数については第 1 列目と第 3 列目の結果には大きな差はみられない。したがって、モデルの他の変数による影響が一定である場合、高学歴ならびに高所得の女性ほど初婚しやすいという結論が得られる。この結果は、1990 年代のデータを用いた先行研究 (Ono 2003, 福田 2007b) とは異なるものであり、2000 年以降におけるわが国の女性の結婚行動が、1990 年代までと比べて大きく変化している可能性を示唆する。また、経済的に自立している女性ほど結婚しやすいという傾向は、アメリカや北欧における結婚行動と類似するものである (Ono 2003, 福田 2007a)。このような女性の結婚行動における変化の要因や背景、ならびにこうした変化が示唆する社会的、人口学的意義の解明は、今後の結婚分析の重要な課題となるであろう。

以上の結果より、脱落を右センサリングとする通常の離散時間ロジットモデルでは、係数の推定値にかなり大きなバイアスが生じており、分析の妥当性が大きく歪められる恐れがあることが明らかとなった。脱落はパネル調査において不可避の問題である。そのため、パネルデータを用いたイベントヒストリー分析では、その取り扱いに細心の注意が必要である。ネステッド・ロジットモデルは、脱落による影響を考慮したパラメーター推定を行える点において、非常に有用であることが示された。また、その適用も容易であることから、今後は同様の分析において積極的に活用されるべきである。

7. まとめ

本稿では、「21 世紀成年者縦断調査」の第 1 回から第 4 回までの個票データを用いて、脱落の取り扱いが、イベントヒストリー分析におけるパラメーター推定に与える影響について検証した。分析では、離散時間ロジットモデル (Allison 1982) ならびにその拡張である離散時間ネステッド・ロジットモデル (Hill et. al. 1993) を用いて、初婚ハザード確率の要因分析を行い、両者のパラメーター推定値を比較した。

分析の結果、脱落を右センサリングとして取り扱う離散時間ロジットモデルでは、係数の推定において重大なバイアスが生じており、分析の妥当性が著しく損なわれる恐れがあ

ることが明らかとなった。その理由としては、同調査における初婚と脱落が独立には生起しない競合イベントであることが挙げられる。初婚と脱落は相互に競合するイベントである。なぜならば、脱落が生起することで初婚の生起リスクは消失し、初婚が生起することで初婚のリスク期間における脱落のリスクが消失するためである。しかし、この2つのイベントは独立ではない。なぜならば、初婚と脱落には共通の非観察要因が存在するためである。言い換えるならば、初婚と脱落の同時モデルにおける両者の誤差項には、かなり高い相関 ($r=0.536$) が認められる。

初婚と脱落が独立ではないことは、2つのイベントの生起過程が類似していることを意味する。財団法人家計経済研究所が実施している「消費生活に関するパネル調査」では、結婚自体が脱落の主要な要因となっていることが示されている(坂本 2006)。坂本(2006)は、女性にとって結婚は転居を伴うことが多いこと。また、夫や夫の家族による調査拒否、あるいはそれを忌避することによる本人からの調査拒否などによって、結婚を契機に脱落するケースが少なからず発生していることを指摘している。この場合、脱落と初婚の発生パターンや決定要因は、多くの点で共通した要素をもつこととなる。今回の分析結果は、成年者調査においても初婚と脱落との間に同様の関係がみられることを示唆する。

ネステッド・ロジットモデルを用いた競合リスク分析では、初婚と脱落の非観察要因の相関を統制した上で、パラメーター推定を行うことが可能である。また、通常のロジットモデルによる推定が可能であることから、汎用的な統計パッケージを用いて分析を行うことができる。パネル調査における脱落が不可避である以上、脱落を競合リスクとするネステッド・ロジットモデルは、不偏的なパラメーターを得る方法として極めて簡便かつ有用であり、パネルデータを用いたイベントヒストリー分析では必ず検討すべき分析手法であるといえる。

成年者調査において、初婚と脱落が独立ではない一因として、転居者へのフォローアップが十分ではないことが挙げられる。成年者調査では、調査員による訪問留置き法によって調査票を回収している。しかし、第1回調査以降に転出したサンプルに対しては、郵送法による調査に切替えている。一般に、郵送法による回答率は、留置き調査によるそれと比べて低いことが知られており、郵送法への切り替えが転居者の調査継続の低さに繋がっているものと思われる。転居による脱落は、結婚のみならず、就職や出産、離婚、そして住宅取得といった主要なライフイベントと相関をもつため、成年期におけるライフコース分析をより困難なものとする可能性がある。この点において、ネステッド・ロジットモデルの利用は、あくまで2次的な対処であり、長期的にみた場合、脱落により生じる諸問題の解決には繋がらない。したがって、「21世紀成年者縦断調査」のデータの代表性を維持し、様々な政策的課題に対応した分析を行っていく上では、転居者に対するフォローアップを充実させて、移動による脱落を減少させていくことが肝要である。

付表1 離散時間多項ロジットモデルによる初婚と脱落の競合ハザード分析

	結婚ハザード β_{m1}	脱落ハザード β_{m2}
年齢スプライン		
20-25歳	0.193 ***	-0.001
25-30歳	0.108 ***	0.033 **
30-37歳	-0.132 ***	-0.035 *
調査年度(対:1年目)		
2年目	0.108	0.050
3年目	0.155 *	-0.162 ***
教育水準(対:高校)		
中学	-0.106	0.139
専門学校	0.115	0.181 ***
短大・高専	0.114	0.065
大学・大学院	0.140	0.258 ***
職業(対:中小企業雇用)		
大企業雇用	-0.159	0.025
専門・技術職	0.050	-0.079
自営・家従・会社役員	0.224	0.187
非正規雇用	-0.026	0.072
無職	0.397 *	0.188
学生	-0.687 ***	0.038
不明	0.007	0.075
親との同別居 (対:両親と同居)		
親と別居	0.068	1.079 ***
片親と同居	0.097	0.203 ***
不明	0.235 *	0.547 ***
居住都道府県のSMAM	-0.282 ***	0.172 ***
健康状態(対:いずれもなし)		
定期的通院あり	-0.080	-0.215 ***
入院あり	0.210	0.303 **
不明	0.335 *	-0.055
休日家事時間	0.029	-0.002
休日家事時間不明ダミー	-0.015	0.324 ***
労働時間	0.021 **	-0.001
労働時間の2乗	-3.00E-04 **	0.00E+00
労働時間不明ダミー	0.033	0.018
年間勤労所得(十万円)	0.013 ***	0.001
年間勤労所得不明ダミー	-0.253 **	0.262 ***
結婚意欲 (対:どちらとも言えない)		
絶対したくない	-1.016 *	-0.144
あまりしたくない	-0.255	-0.010
なるべくしたい	0.550 ***	-0.063
絶対したい	1.129 ***	0.072
定数	-0.765	-7.264 ***
person-year数	18624	
カイ2乗値	1011.195	
自由度	70	

* p<.1; ** p<.05; *** p<.01

参考文献

- Alderman, H., J. R. Behrman, H. Kohler, J. A. Maluccio, S. C. Watkins, 2001, "Attrition in Longitudinal Household Survey Data", *Demographic Research*, 5(4), 80-123.
- Allison, Paul D., 1982. "Discrete-Time Methods for the Analysis of Event Histories", *Sociological Methodology*, 13: 61-98.
- Allison, Paul D., 1995. *Survival Analysis Using The SAS System: A Practical Guide*. Cary: SAS Institute Inc.
- Blossfeld, Hans-Peter and Gots Rohwer, 2002. *Techniques of Event History Modeling second edition: New Approaches to Causal Analysis*. New Jersey: Lawrence Elbaum Associations, Inc., Publishers.
- Falaris, E. M., 2003, "The Effect of Survey Attrition in Longitudinal Surveys: Evidence from Peru, Cote d'Ivoire and Vietnam", *Journal of Development Economics*, 70, 133-57.
- Fitzgerald, J., P. Gottschalk, and R. Moffitt, 1998, "An Analysis of Sample Attrition in Panel Data", *The Journal of Human Resources*, 33(2), 251-99.
- Guo, Guang, 1993. "Event-History Analysis for Left-Truncated Data" *Sociological Methodology* 23:217-43.
- Heckman, J. and B. Singer, 1984. "A Methods For Minimizing the Impact of Distributional Assumptions in Econometric Models For Duration Data" *Econometrica* 52(2): 271-320.
- Hill, Daniel H., William G. Axinn, and Arland Thornton, 1993. "Competing Hazards with Shared Unmeasured Risk Factors" *Sociological Methodology* 23 : 245-77.
- Lepkowski, J.M., and M.P. Couper, 2002. "Nonresponse in Second Wave of Longitudinal Household Surveys" , in *Survey Non-response*, edited by Eltice Groves, John Willey and Sons, pp. 259-72.
- Lillard, Lee A. and Linda J. Waite, 1993. "A Joint Model of Marital Childbearing and Marital Disruption " *Demography* 30:653-81.
- Lillard, Lee A. and Constantijn W. A. Panis, 1998. "Panel Attrition from the PSID" *The Journal of Human Resources* 33(2), 437-57.
- Lillard, Lee A. and Constantijn W. A. Panis, 2003. *aML User's Guide and Reference Manual version 2*. California: Econ Ware.
- Macfadden, D., 1981. "Economic Models of Probabilistic Choice" in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. M. Manski and D. McFadden, Cambridge, Mass: MIT Press.
- Ono, Hiromi, 2003. "Women's Economic Standing, Marriage Timing, and Cross-National Contexts of Gender" *Journal of Marriage and Family* 65:275-86.
- Raymo, James M., 2003a. "Educational Attainment and the Transition to First Marriage Among Japanese Women" *Demography* 40: 83-103.
- Raymo, James M., 2003b. "Premarital Living Arrangements and the Transition to First Marriage in Japan" *Journal of Marriage and Family* 65:302-15.
- Raymo James M. and Ono Hiromi, 2007. "Coresidence With Parents, Women's Economic Resources, and the Transition to Marriage" *Journal of Family Issues* 28: 653-81.
- Tsuya Noriko O. and Karen Oppenheim Mason, 1995. "Changing Gender Roles and Below-Replacement Fertility in Japan" in *Gender and Family Change in Industrialized Countries*, edited by K. Oppenheim Mason and A-M. Jensen. Oxford: Clarendon Press, pp. 139-67.

- Vermunt, J. K., 1997. *Loglinear Models of Event Histories*. Thousand Oaks, CA: Sage.
- Wooldridge, J., 2002. "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification" *CeMMAP Working Paper CWP11/02*.
- 厚生労働省大臣官房統計情報部, 2005, 『第2回 21世紀成年者縦断調査（国民の生活に関する継続調査）結果の概況』。
- 坂本和靖, 2006, 「サンプル脱落に関する分析：「消費生活に関するパネル調査」を用いた脱落の規定要因と推計バイアスの検証」, 『日本労働研究雑誌』, 第551号, 55-70ページ。
- 津谷典子, 2002, 「イベント・ヒストリー分析」, 日本人口学会編, 『人口大事典』, 428-31ページ, 培風館。
- 福田節也, 2007a, 「ジェンダーシステムと女性の結婚選択（1）：ヨーロッパ諸国における「女性の経済的自立仮説」の検証」, 『季刊家計経済研究』, No. 76, 45-53ページ。
- 福田節也, 2007b, 「ジェンダーシステムと女性の結婚選択（1）：日本における「女性の経済的自立仮説」の検証」, 『季刊家計経済研究』, No. 76, 54-62ページ。
- 山口一男, 2001a, 「イベントヒストリー分析（2）」, 『統計』, 2001年10月号, 70-75ページ。
- 山口一男, 2001b, 「イベントヒストリー分析（4）」, 『統計』, 2001年12月号, 73-78ページ。
- 山口一男, 2002a, 「イベントヒストリー分析（13）」, 『統計』, 2002年9月号, 60-65ページ。
- 山口一男, 2002b, 「イベントヒストリー分析（14）」, 『統計』, 2002年10月号, 66-71ページ。
- 山口一男, 2002c, 「イベントヒストリー分析（最終回）」, 『統計』, 2002年11月号, 55-60ページ。
- 山田昌弘, 1999, 『パラサイト・シングル時代』, 筑摩書房。

(2) 1990年代以降における第1子出生の傾向と要因： 第2回成年者縦断調査を用いた分析

福田 節也

1. はじめに

近年の人口学研究においては、夫婦出生力の低下が指摘されている（佐々井 1998, 岩澤 2002, 金子 2004）。1990年代以降、夫婦が結婚してから第1子をもつまでの期間が長期化しており、第2子以降をもたない夫婦が増加していることがその一因である。第1子出生タイミングの遅れならびに第2子出生確率の低下は、晩婚化や女性の高学歴化といった従来の要因では説明できないことが明らかとなっている（佐々井 1998, 金子 2004）。これまでわが国においては、結婚すれば夫婦は平均して2人以上の子どもを生むという状態が持続してきた。しかし、近年における夫婦出生力の低下は、このような2子規範の崩壊を示唆するものである。したがって、このような夫婦の産み方の変化が、どのような構造の下に生起しているのかを明らかにすることは、わが国の今後の出生力変動を見極める上で重要な課題である。

本稿では、厚生労働省大臣官房統計情報部によって実施されている「21世紀成年者縦断調査」の第2回調査のデータを用いて、わが国の1990年代以降における第1子出生の傾向と要因に関する分析を行う。第1子出生がどの年齢で、あるいは結婚からどのくらい経過した時点で生起しているのかは、個人の出生力を規定する上で重要な要素である。また、第1子出生のタイミングは、婚前妊娠や晩婚化といった結婚行動における変化をもっとも敏感に反映しているため、近年における出生パターンの変化を理解するうえで多くの情報を与えるものと思われる。本稿では、晩婚化、高学歴化、雇用の非正規化、若者の失業、そして有配偶女性の雇用労働力化といった変化が、1990年代以降における第1子出生のタイミングにどのような影響を与えているのかについて知見を得る。

2. データ

本稿において用いるデータは、前出の「21世紀成年者縦断調査」の第2回調査（2003年11月5日実施）の個票データである。同調査は、2002年10月末時点で20-34歳であった全国の男女（及びその配偶者）を対象とし、そのうち、第1回調査において協力を得られた者等（及びその配偶者）を調査客体としている（厚生労働省大臣官房統計情報部 2005）。同調査は、同一個人を対象とした追跡調査であり、パネル調査と呼ばれる。パネル調査を利用することの利点として、同一個人の属性の時間的変化を特定できること、また横断調査に比べてより直接的に標本のもつ異質性を統制し、説明要因と非説明変数との因果関係

を考察できることなどが挙げられる（樋口他 2004）。しかし、このようなパネル調査の利点を享受するには、データの十分な蓄積が必須であり、長い時間と調査継続のための莫大な費用を要する。それでは、データが蓄積するまでの間、パネルデータを用いて有効な分析を行うことはできないのであろうか。答えは否である。本稿においては、パネル調査において得ている事歴情報（event history data）を用いた分析事例を示したい。

同調査においては、調査対象者のライフコースに関する様々な事歴情報を得ている。事歴情報とは、結婚、出生、就業などのライフイベントについて、それぞれのイベントを対象者が「いつ」経験したのかを表す情報である。こうした事歴情報は、イベントを経験した年齢や年月によって得られる。このように対象者のイベント経験履歴について情報を集めている調査を遡及的調査（retrospective survey）という。「21世紀成年者縦断調査」の第2回調査においては、対象者の現在の配偶者との同居開始年月、すべての子どもについての出生年月、最終学校卒業あるいは中退の年月、そして過去13回までの就業の開始および終了の年月を得ている。これらの時点に関する情報を生かすことにより、「時間」の概念を反映したモデルを構築し、結婚や出生の規定要因に関する縦断的分析（longitudinal analysis）を行うことが可能である。以下に今回の分析において用いた分析手法であるイベントヒストリー分析（event history analysis）の概略と分析モデルについて述べる¹。

なお、今回の分析では同調査のうち第1回から第3回調査までの個票データ使用することが可能であったが、1) 卒業・中退及び就業に関する履歴情報は第2回調査以降において得られること、そして2) 脱落の発生により、第2回調査のほうが第3回調査よりもサンプル数が多いことの2つの理由により、第2回調査のデータのみを用いた。

3. 分析手法とモデル

3-1. 分析手法

本稿における分析は、イベントヒストリー分析によって行う。イベントヒストリー分析とは、あるイベントの発生パターンとその要因に関する分析手法の総称である。別名、生存分析（survival analysis）ともいわれる。

イベントヒストリー分析では、リスク人口（population at risk）におけるイベント発生確率である「ハザード率（hazard rate）」を分析の対象とする。リスク人口とは、イベントを経験する可能性がある人口を指す。例えば、離婚をイベントとして分析を行う場合、離婚のリスク人口は有配偶の男女であり、未婚者や死別者、離別者はリスク人口に含まれない。ハザード率は、より正確には「時間 t に至るまでの期間に、当該イベントが起こらなかったという条件のもとでの、時間 t におけるイベント発生の瞬間確率（instantaneous rate）」（津谷

¹ なお、Blossfeld and Rowher（2002）は、このような事歴情報を用いた因果分析における注意点として、1) 価値観や態度などの変数を遡及的に扱うことができない、2) 記憶違いによる誤った情報の収集の可能性、3) 生存者のみを対象とすることによるサンプリング・バイアスなどを挙げている。