

論文(投稿)

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

坂本 和靖

(財団法人 家計経済研究所研究員)

本稿の目的は、「消費生活に関するパネル調査」(1993～2003年)を用いて、サンプル脱落の特徴を把握すること、またサンプル脱落が与える推計へのバイアスに関する分析を行うことにある。本稿での分析結果は、以下の2点にまとめることができる。第一に、脱落傾向がある対象者の特徴を捕捉するべく、当該年度(t 年度)の脱落ダミー(脱落=1, 継続回答=0)を被説明変数、前年度($t-1$ 年度)調査の情報を説明変数としたプロビット分析を行った。その結果、無配偶者、有配偶者に見られた共通の特徴として、結婚予定者、新婚者などのライフイベント前後における脱落傾向が強く見られた。また経済的特徴としては、無配偶者では、本人の収入が低いほど、収入変化額が負に大きいほど、さらに、借り入れ負担感が大きい者ほど脱落する傾向が見られた。逆に、有配偶者では、本人、夫の収入変化額が正に大きいほど、脱落する傾向となった。第二に、結婚選択関数における脱落による推計バイアスを確認するべく、Inverse Probability Weighting法による調整を行ったモデルと何の調整も行わなかったモデルとの係数の比較を行った。Hausman Testを行ったところ、両モデルの係数間には、統計的に有意な差異が見られ、また係数が過小評価されていたことが確認された。

目次

- I はじめに
- II 脱落の過程と特性について
- III JPSCにおける脱落の現状
- IV 結婚選択と脱落
- V 分析結果
- VI むすび

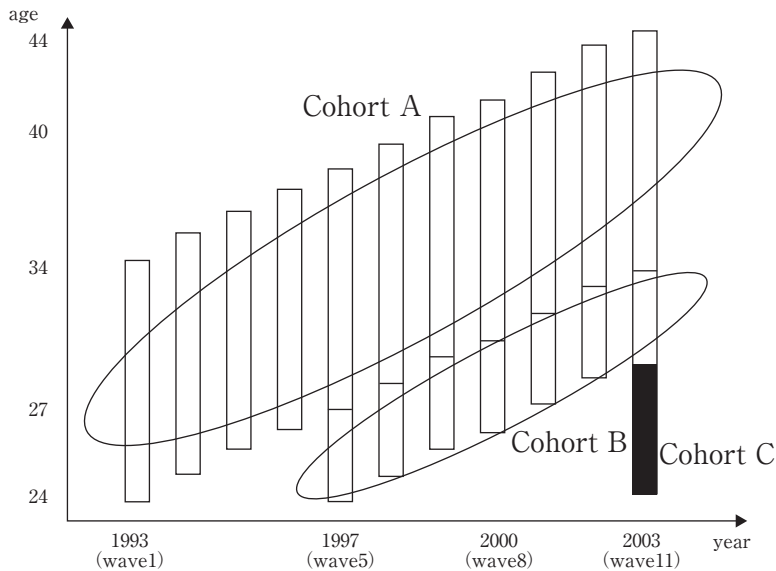
I はじめに

財団法人家計経済研究所では、1993年より、「消費生活に関するパネル調査」(The Japanese Panel Survey of Consumers, 以下JPSC)という、個人(調査開始時に20～30歳代の女性)を対象とした追跡調査を行っている(図1)。公開データという特性から、多くの経済学者、社会学者を中心とした研究者の手により、政策的インプリケーシ

ョンに富んだ研究が数多く発表されており、わが国の家計データの中でも、非常に重要な調査データとなりつつある(樋口・岩田1999, 樋口・太田・家計経済研究所2004)。しかしながら、本データにおいても、パネル調査一般が抱える問題が存在することは否めない。それは、同一個人を調査対象としているため、途中時点で回答を拒否された場合、一般的な調査(クロスセクションデータ)のように、代わりに他の対象者(予備サンプル)に調査を行うことができない点である(サンプル脱落の発生)。特に、内生変数によってサンプルが脱落する場合、そのデータによって得られた推計結果は深刻なバイアスをもつ可能性が高くなるため、注意を払わなければならない。

本稿の目的は、JPSCを用いて、脱落したサンプルの特徴を把握し、かつ脱落により結婚選択関数の推計結果にバイアスが生じるかどうか、

図1 消費生活に関するパネル調査における
調査年度と調査対象年齢



Inverse Probability Weighting 法による調整をしたモデルと未調整のモデルとの推計結果を比較し、脱落による推計バイアスの影響を考察することにある。

次節以降では、IIにおいて、一般的な脱落過程とその諸特徴について論じ、IIIでは、JPSCにおける脱落の現状を探るべく、第一に、継続回答者と途中脱落者の（諸変数における）平均値比較を、第二に、脱落ダミー（脱落=1、継続回答=0）を被説明変数とした、プロビット分析を行った。最後に、IV、Vでは、Inverse Probability Weighting 法で調整したモデルと未調整モデルを比較し、結婚選択関数推計における脱落によるバイアスについて考察した。Hausman Testを用いて、両モデルの係数間の比較を行ったところ、統計的に有意な差異が見られ、またいくつかの係数が過小評価されていたことが確認された。

II 脱落の過程と特性について

1 脱落過程

本節では、脱落のメカニズムについて簡単に紹

介したい。脱落には大きく分けて、3種類のパターンがある。

以下では、 a_i は脱落 (attrition) するかどうかの指標を、 ϕ は未知パラメータを示すものとする。

また、 $a_i = j$ は j 時点となった時に脱落し、 $j-1$ 時点の変数 (y_1, \dots, y_{j-1}) は観測可能だが、 j 時点以降の変数 (y_j, \dots, y_T) は観測不可能であることを意味する。

i) Missing Completely At Random (MCAR)

脱落が完全にランダムな場合、過去 ($j-1$ 時以前)、また脱落が発生する時点、それ以降 (j 時以降) の変数にも依存しないため、

$$\Pr(a_i = j | y_{i1}, \dots, y_{iT}; \phi) = \Pr(a_i = j | \phi) \quad (1)$$

for all y_{i1}, \dots, y_{iT} ,

と示すことができる。この場合、脱落による推計バイアスは無視することが可能となる。

ii) Missing At Random (MAR)

MARでは、脱落は観測可能な変数 y_1, \dots, y_{j-1} へのみ依存する。

$$Pr(a_i = j | y_{i1}, \dots, y_{iT}; \phi) = Pr(a_i = j | y_{i1}, \dots, y_{ij-1}; \phi) \text{ for all } y_{i1}, \dots, y_{iT} \quad (2)$$

つまり、 j 時点で脱落するかどうかは、 $j-1$ 以前の観測値から推測可能ともいいかえることができる。後述するプロビット分析による脱落関数の推計はこの考え方に相当する。

iii) Non-ignorable, Missing At No Random (MANR)

脱落が、脱落時点以降の値 (y_{j+1}, \dots, y_T) などの観測不可能な値に依存している場合、観測済みデータ ($j-1$ 時点) だけでは、脱落値 (j 時点) を推測することは困難となる。

$$Pr(a_i = j | y_{i1}, \dots, y_{iT}; \phi) = Pr(a_i = j | y_{ij}, y_{ij+1}, \dots, y_{iT}; \phi) \text{ for all } y_{i1}, \dots, y_{iT} \quad (3)$$

2 先行研究

続いて、脱落するサンプルには具体的にどのような特性がみられるのか。以下では、海外のパネルデータを用いた先行研究について触れたい。

Panel Study of Income Dynamics (PSID) を用いた包括的な脱落に関する研究論文である、Fitzgerald, Gottschalk, and Moffitt (1998) では、継続回答者と途中脱落者との基本統計量の比較、またプロビット分析から脱落者の特徴を以下のように挙げている。

第一に、脱落者は、賃金が低い、就学年数が短い、無配偶である、賃貸住宅に居住しているなどの社会経済的な傾向がある。

第二に、脱落者は、賃金の分散が大きい、結婚や離婚の変化が多い、また地理的移動が多いなどの生活環境が不安定である。特に賃金の下落、離婚や死別などのイベントがあるほど脱落する傾向がある。

また、National Longitudinal Survey of Youth (NLSY) を用いた、MaCurdy, Mroz, and Gritz (1998) では、無就学者、無就業者、また賃金が高いものほど (特に、女性では賃金分布の上方に位置するサンプル) 脱落する結果が得られている。

ただし、これら、両パネルデータからの結果と

JPSC の結果を直接比較するのは難しい。何故なら、JPSC では調査対象者が女性に限定されており、また年齢層が 20 歳代から 30 歳代であるという点から、脱落の特性も多少特異なものと考えられる。次節以降では、Fitzgerald, Gottschalk, and Moffitt (1998) に倣い、MAR の考え方に則り、 $j-1$ 以前の観測値から、脱落者の特徴について考察していきたい。

III JPSC における脱落の現状

現時点 (wave11, 2003 年) までの JPSC のサンプルの累積脱落率は、それぞれ 42.2% (Cohort A¹⁾, 1993~2003 年の 11 年間), 35.4% (Cohort B²⁾, 1997~2003 年の 7 年間) となっている (表 1)。Cohort 別の各年脱落率 (t 時点の脱落者/ $t-1$ 時点の回収数 $\times 100$) を見ると、Cohort A では前年比の 5% 近くだったものが、徐々に 3% へと収束している。これに対して、Cohort B では、第 2 年度に 11.6% と大幅に脱落した以降は、だいたい 6% 程度に落ち着いている。この JPSC での結果と比べて、PSID では、調査開始から 11 年間 (1968~1977 年) の累積脱落率は 32.2% と JPSC より小さい値をしめしている³⁾。

III では、3つの手法を用いて、脱落者の特性について考察している。第一に、途中脱落者と継続回答者との間における、観測可能前年の個人的属性、経済的属性などの平均値比較、第二に、脱落ダミー (t 年度, 脱落 = 1, 継続回答 = 0) を被説明変数、前年 ($t-1$ 年度) の個人の家族属性、経済的属性などの情報を説明変数としたプロビット分析、そして、第三に、調査拒否者の拒否理由について記されている「『消費生活に関するパネル調査』回収結果」という、記述的な資料を基に、脱落理由の分類を行った。また、以下では、Cohort A, B に分割して、脱落要因に関する分析を行っている。これは、調査方法は同じであっても、調査年時が異なるため、母集団に多少の違いが生じる恐れがあり、脱落要因の解析にあたる III では慎重を期するために分割している。

表1 脱落率

調査年	CohortA				CohortB			
	回収数	脱落数	脱落率 ¹⁾	累積脱落率	回収数	脱落数	脱落率 ¹⁾	累積脱落率
1993	1,500	—	—	—	—	—	—	—
1994	1,422	78	5.2%	5.2%	—	—	—	—
1995	1,342	80	5.6	10.8	—	—	—	—
1996	1,298	44	3.3	14.1	—	—	—	—
1997	1,255	43	3.3	17.4	500	—	—	—
1998	1,196	59	4.7	22.1	442	58	11.6%	11.6%
1999	1,137	59	4.9	27.1	412	30	6.8	17.6
2000	1,102	35	3.1	30.1	386	26	6.3	22.8
2001	1,059	43	3.9	34.0	366	20	5.2	26.8
2002	1,032	27	2.5	36.6	343	23	6.3	31.4
2003	974	58	5.6	42.2	323	20	5.8	35.4

注：1994～2002年までは復活サンプルを含む。

1) 脱落数/前年の回収数×100

1 平均値比較

1 個人的属性

CohortA, Bの両方において、途中脱落者のほうが統計的に有意に年齢が若いことが確認できた(表2)⁴⁾。

また、CohortA, B両方において、無配偶者の内、途中脱落者の方が結婚予定者割合が高い(いかにすれば、結婚予定者は脱落する確率が高い)という結果となった。これは、本調査期間中に結婚したサンプルの61.0%(346人中211人)が転居を経験していることからわかるように、結婚予定というのが、転居のpredictorとなっており、転居による影響が大きいことが考えられる。また、途中脱落者の方が新婚者割合が高く、結婚後まもない新婚者において、翌年脱落する確率が有意に高い結果が得られている。後述する「『消費生活に関するパネル調査』回収結果」では、結婚以降、「ご主人に断られる」「嫁の立場があるためと拒否」など家族による反対、あるいは家族を意識した拒否が少なからず発生しており、結婚は単なる転居理由ではないことがうかがわれる。特にJPSCでは、調査対象者本人への質問だけではなく、夫、その親などに関する質問項目も含まれており、それらについて、回答することができないことによる回答拒否がみられる。

また、CohortA, CohortBの双方において、Lepkowski and Couper (2002)などの出産以降

に脱落するという結果とは異なり、CohortBでは、継続回答者のほうが出産割合が高い結果が得られた。これは、出産を機会に退職し、家事育児のため在宅時間が長くなるといった、日本における出産退職慣行などの独特な就業環境と関連している可能性がある。

最後に、居住に関する属性別にみると、居住都市規模別では、途中脱落者のほうが、「大都市」⁵⁾居住割合が高く、「その他の市」居住割合が低い傾向が見られ、都市に住んでいる者ほど脱落する傾向が見られた。また住居形態では、継続回答者の方が持ち家率が高いことが確認された。

2 経済的属性

CohortA, Bともに、有配偶では、途中脱落者のほうが平均年収が高い傾向が見られるが、統計的に有意な値ではなかった(表3)。逆に無配偶では、途中脱落者のほうが平均年収が低く、Fitzgerald, Gottschalk, and Moffitt (1998)に合致した結果が得られている。前年比の収入変化額は、有配偶、無配偶ともに途中脱落者の平均変化額は負となり、所得が減少する場合、脱落傾向が強まることがいえる。

また、夫の年収は、CohortAでは、本人(有配偶)同様、脱落者の年収のほうが正に大きく、逆にCohortBでは夫の年収は継続回答者の方が正に大きい結果が得られているが、双方とも統計的に有意な値とはならなかった⁶⁾。夫の所得の変化額でも、CohortAでは継続回答者ほど正に大き

表2 継続回答者と途中脱落者比較 (個人的属性)

	CohortA		CohortB	
	継続回答者	途中脱落者	継続回答者	途中脱落者
本人年齢 (歳)	33.09	32.61**	27.76	27.37**
サンプル数	11,807	536	2,273	177
本人就学年数 (年)	13.12	13.13	13.29	13.26
サンプル数	11,807	536	2,273	177
労働時間 (平日, 分)	258.78	285.79**	314.79	357.49**
サンプル数	11,558	523	2,229	171
家事時間 (平日, 分)	356.38	317.01***	281.57	213.01***
サンプル数	11,559	522	2,228	171

注: 1) ***: 1%水準, **: 5%水準, *: 10%水準, #: 15%水準で統計的に有意 (t 検定)

2) 上段: %, 下段: 各項目該当者/継続回答 (途中脱落) 者全体

	CohortA		CohortB	
	継続回答者	途中脱落者	継続回答者	途中脱落者
中学校卒	4.9 579/11,807	5.97 32/536	5.68 129/2,273	7.34 13/177
高校卒	44.12 5,209/11,807	42.16 226/536	38 866/2,273	40 71/177
専門学校・短大卒	39.26 4,636/11,807	38.81 208/536	39.6 900/2,273	32.2* 57/177
大学・大学院卒	11.71 1,383/11,807	13.06 70/536	16.63 378/2,273	20.34 36/177
有配偶	77.12 9,105/11,807	70.9*** 380/536	53.81 1,223/2,273	42.94*** 76/177
結婚予定あり (無配偶)	6.07 164/2,702	14.74*** 23/156	7.33 77/1,050	13.86*** 14/101
新婚 (有配偶)	2.42 220/9,105	6.32*** 24/380	7.93 97/1,223	6.58 5/76
出産	6.17 729/11,078	5.04 27/536	8.40 191/2,273	4.52* 8/177
持ち家	66.11 7,796/11,792	62.87* 337/536	55.07 1,250/2,270	52.27 92/176
大都市	23.68 2,796/11,807	27.05* 145/536	27.67 629/2,273	23.73 42/177
その他の市	56.26 6,643/11,807	52.05* 279/536	57.02 1,296/2,273	61.02 108/177
町村	19.86 2,345/11,807	20.90 112/536	15.31 348/2,273	14.12 25/177
同居	47.51 5,610/11,807	47.20 253/536	51.25 1,165/2,273	48.02 85/177
夫婦関係満足度	34.04 3,099/9,105	32.63 124/380	44.24 541/1,223	40.79 31/76

注: ***: 1%水準, **: 5%水準, *: 10%水準, #: 15%水準で統計的に有意 (χ^2 検定)

表3 継続回答者と途中脱落者比較（経済的属性）

	CohortA		CohortB	
	継続回答者	途中脱落者	継続回答者	途中脱落者
本人年収（万円，有配偶）	95.91	102.59	97.16	121.89
サンプル数	8,357	339	1,105	65
本人年収（万円，無配偶）	289.45	253.49***	250.56	240.06
サンプル数	2,629	151	1,023	96
夫年収（万円）	541.99	547.19	444.33	437.38
サンプル数	8,440	347	1,130	65
本人年収変化（万円，有配偶）	-0.79	-11.15***	-15.22	-4.04
サンプル数	7,010	279	884	46
本人年収変化（万円，無配偶）	10.19	-14.19***	12.67	-19.81***
サンプル数	2,135	113	754	59
夫年収変化（万円）	13.50	8.78	12.40	28.80
サンプル数	6,987	269	829	45

注：1）***：1%水準，**：5%水準，*：10%水準，#：15%水準で統計的に有意（t検定）

2）上段：%，下段：各項目該当者／継続回答（途中脱落）者全体

	CohortA		CohortB	
	継続回答者	途中脱落者	継続回答者	途中脱落者
正規就業	28.06	33.77***	38.63	42.94
	3,313/11,807	181/536	878/2,273	76/177
パート・嘱託	23.00	18.84***	22.57	23.00
	2,700/11,807	101/536	513/2,273	41/177
専業主婦・学生・無職	40.00	38.25	33.88	30.00
	4,746/11,807	205/536	770/2,273	53/177
借り入れ負担	5.94	5.41	5.72	9.04*
	701/11,807	29/536	130/2,273	16/177

く、逆に CohortB では脱落者ほど正に大きい相反する結果が得られたが、ここでも双方とも統計的には有意でなく、安定的な結果が得られなかった。さらに、借り入れ返済負担感の有無⁷⁾についてみると、CohortB において、脱落者のほうが負担感を持つ者の割合が有意に高いという結果が得られた。

本人の就業状態では、CohortA において、途中脱落者のほうが正規就業として働いている者の割合が高く、継続回答者のほうはパート・嘱託として働いている者の割合が高い結果が得られた⁸⁾。

また、JPSC の質問内容が本人だけでなく、子ども、夫、両親などに関する項目も含まれているため、家族内での人間関係が円滑でない、正確な回答が得られなかったり、回答拒否される可能性が高い。これを測るために、心理的な特徴である、「夫婦関係の満足度」（「非常に満足している」、 「まあまあ満足している」= 1、それ以外= 0）が、脱

落に影響しているか確認したところ、CohortA、B 両方とも、継続回答者のほうが夫婦関係に満足している割合が高いが、その差は統計的に有意とはならなかった。

2 プロビット分析からみた脱落特性

本節では、脱落ダミー（ $a_{it} = 1$ は脱落を、 $a_{it} = 0$ は継続回答をさす）を被説明変数、観測可能な前年の個人の家族属性、経済的属性などを説明変数 x_{it-1} としたプロビット分析を行っている。

1 無配偶

すべてのケースで有意であったのは、結婚予定ダミーであった（表4）。繰り返しになるが、これは、先の脱落理由の節でも示したように、結婚を機に脱落するサンプルが多いことを裏づけている。この原因として、JPSC には、家計内容、配偶者やその親に関する、非常に詳細な質問項目があるために、先述したように、配偶者によって忌避さ

表4 脱落関数（無配偶）

	CohortA		CohortA		CohortB		CohortB	
	係数	頑健的標準誤差	係数	頑健的標準誤差	係数	頑健的標準誤差	係数	頑健的標準誤差
年齢	-0.017	0.012	-0.021	0.012*	-0.007	0.028	0.021	0.036
大都市	0.004	0.101	0.023	0.100	-0.138	0.133	-0.045	0.159
(その他の市)								
町村	0.008	0.135	0.030	0.139	0.004	0.182	0.126	0.224
親との同居	-0.005	0.119	0.012	0.119	-0.021	0.130	-0.167	0.155
中学校卒	0.047	0.243	0.294	0.222	-0.311	0.317	-0.314	0.395
高校卒	0.016	0.135	0.055	0.133	-0.157	0.179	-0.133	0.224
(専門・専修学校卒)								
短大卒	0.164	0.139	0.179	0.139	-0.282	0.183	-0.192	0.230
大学・大学院卒	0.063	0.148	0.045	0.149	-0.099	0.182	-0.083	0.229
本人の年取（対数値）	-0.214	0.066***			0.018	0.086		
本人の年取（階差）	—		-0.001	0.000***			-0.002	0.001***
借入負担感あり	-0.050	0.290	-0.255	0.316	0.754	0.252***	-0.085	0.481
結婚予定	0.487	0.135***	0.535	0.132***	0.318	0.181*	0.354	0.218*
正規就業	0.154	0.113	-0.001	0.097	-0.031	0.137	0.060	0.143
定数項	-0.156	0.533	-1.158	0.428***	-1.125	0.796	-1.878	0.999*
サンプル数		2,210		2,248		1,093		813
Wald χ^2		30.23		32.92		15.9		11.24
Prob > χ^2		0.0026		0.001		0.1959		0.5083
擬似決定係数		0.0312		0.0356		0.024		0.0289
Log pseudo-likelihood		-429.2935		-432.0887		-310.4609		-205.4557

	CohortA&B		CohortA&B	
	係数	頑健的標準誤差	係数	頑健的標準誤差
年齢	-0.029	0.009***	-0.026	0.010
大都市	-0.055	0.080	-0.002	0.084
(その他の市)				
町村	-0.019	0.106	0.030	0.115
親との同居	-0.053	0.085	-0.073	0.091
中学校卒	-0.049	0.188	0.143	0.193
高校卒	-0.013	0.105	0.032	0.113
(専門・専修学校卒)				
短大卒	0.022	0.111	0.087	0.119
大学・大学院卒	0.052	0.113	0.056	0.123
本人の年取（対数値）	-0.133	0.053**		
本人の年取（階差）			-0.001	0.000***
借入負担感あり	0.338	0.175*	-0.227	0.265
結婚予定	0.409	0.108***	0.467	0.112***
正規就業	0.067	0.087	-0.001	0.080
定数項	-0.003	0.372	-0.854	0.338***
サンプル数		3,303		3,061
Wald χ^2		43.17		39.6
Prob > χ^2		0.000		0.0001
擬似決定係数		0.0258		0.0308
Log pseudo-likelihood		-748.4698		-641.8872

注：***：1%水準，**：5%水準，*：10%水準，#：15%水準で統計的に有意

れるのみならず、回答者自身の忌避がみられた。無論、結婚を機に転居したので、「転居先まで来ないで欲しい」などの回答も（『消費生活に関するパネル調査』回収結果）に存在するため、結婚予定は転居の predictor という考えられる。また、本人年取は（脱落確率に対して）負に有

意となっており、本人年取が低いサンプルほど、脱落する傾向が高いことが確認できる。また年取の変化額 ($x_{t-1} - x_{t-2}$) も負に有意となっており、所得が減少するほど、脱落する傾向にある。また、CohortB, A & B では、借入れ負担感には正に有意となり、経済的に逼迫している家計ほど、脱

表5 脱落関数（有配偶）

	CohortA		CohortA		CohortB		CohortB	
	係数	頑健的標準誤差	係数	頑健的標準誤差	係数	頑健的標準誤差	係数	頑健的標準誤差
年齢	-0.008	0.009	0.009	0.008	-0.043	0.041	-0.055	0.048
大都市 (その他の市)	0.063	0.088	0.061	0.069	-0.071	0.186	-0.196	0.176
町村	0.175	0.084**	0.059	0.077	-0.088	0.227	-0.349	0.223 [#]
親との同居	-0.136	0.075*	-0.162	0.065**	0.065	0.185	0.047	0.167
中学校卒	0.494	0.176***	-0.025	0.161	—	—	0.195	0.386
高校卒 (専門・専修学校卒)	0.147	0.093	0.015	0.077	0.024	0.223	0.419	0.236*
短大卒	-0.002	0.113	0.013	0.090	-0.106	0.280	0.390	0.265
大学・大学院卒	0.107	0.127	0.033	0.107	0.334	0.273	0.322	0.305
本人年取 (対数値)	-0.006	0.034			0.105	0.074		
夫年取 (対数値)	0.070	0.084			-0.009	0.165		
本人年取 (階差)			2.3 E-04	1.5 E-04 [#]			0.001	0.001
夫年取 (階差)			1.6 E-04	1.0 E-04 [#]			0.000	0.000
借入負担感あり	0.119	0.123	0.008	0.108	0.314	0.283	-0.565	0.400
新婚	0.324	0.164**	—	—	-0.323	0.271		
出産	-0.072	0.149	-0.046	0.108	0.158	0.227	-0.056	0.196
正規就業	0.198	0.087**	0.227	0.072***	-0.129	0.201	0.265	0.184 [#]
定数項	-2.076	0.555***	-2.119	0.272***	-0.786	1.547	-0.320	1.437
サンプル数		4,735		6,912		597		823
Wald χ^2		29.74		25.54		10.26		17.82
Prob > χ^2		0.0083		0.0196		0.6725		0.1646
擬似決定係数		0.0184		0.0118		0.0267		0.0405
Log pseudo-likelihood		-785.5569		-1,072.375		-135.0202		-153.5007

	CohortA&B		CohortA&B	
	係数	頑健的標準誤差	係数	頑健的標準誤差
年齢	-0.014	0.008*	0.001	0.007
大都市 (その他の市)	0.033	0.079	0.029	0.064
町村	0.140	0.078*	0.012	0.072
親との同居	-0.115	0.070 [#]	-0.142	0.060**
中学校卒	0.334	0.170**	-0.004	0.147
高校卒 (専門・専修学校卒)	0.129	0.087 [#]	0.059	0.073
短大卒	-0.018	0.105	0.045	0.085
大学・大学院卒	0.155	0.114	0.050	0.101
本人年取 (対数値)	0.015	0.031		
夫年取 (対数値)	0.045	0.077		
本人年取 (階差)			2.4 E-04	1.4 E-04*
夫年取 (階差)			1.5 E-04	1.1 E-04 [#]
借入負担感あり	0.134	0.112	-0.046	0.104
新婚	0.146	0.141	-0.045	0.094
出産	-0.003	0.123	0.236	0.066***
正規就業	0.124	0.080 [#]	-1.865	0.244***
定数項	-1.739	0.497***		
サンプル数		5,356		7,735
Wald χ^2		25.49		25.11
Prob > χ^2		0.0301		0.0223
擬似決定係数		0.0142		0.0103
Log pseudo-likelihood		-929.4101		-1233.678

注：***：1%水準，**：5%水準，*：10%水準，#：15%水準で統計的に有意

落の可能性が高いことを示している。

2 有配偶

CohortAにおいて、新婚ダミーは、脱落確率に対して正に有意な結果となり、ここでも結婚が脱落に対して、影響を持つことが確認された(表5)。

また、親との同居ダミーは負に有意となり、親と同居している場合、脱落確率が小さい。これは、JPSCでは「留置調査法」という調査方法を採用しており、調査票の受け渡しのためには、誰かが在宅している必要があるため、親と同居している場合、本人が不在であっても、その親に調査票を受け渡しできることがその要因と考えられる。

経済的属性では、夫・本人の年収は、ここでは安定的な結果が得られなかったが、それらの変化額は増加するほど、脱落していく傾向がみられた。また本人の正規就業ダミーが、脱落確率に対して正の影響を与えていることから、勤めに出ており、在宅時間が短い場合は、調査票を受け取れないため、脱落確率が高いことが挙げられる⁹⁾。

上記の分析においては、脱落以前に観測される情報だけに限定されるため(MAR)、調査対象者が、前年度(t-1年度)の調査回答から当該年度(t年度)に調査を拒否するまでの間の出来事に関する情報は得られていない。調査直前のt-1年度からt年度の間に発生したイベントが発生し、それが原因となって脱落する可能性は高い(MANR)。

こうした情報を補完するべく、JPSCでは、「『消費生活に関するパネル調査』回収結果」¹⁰⁾(以下、回収結果)において、t年度での脱落者の脱落理由の内容についてみることができる¹¹⁾。ただ本来「回収結果」資料は、脱落理由の情報を収集することを目的としていなかったため、脱落理由がわからないものが半分を占めており、限定的な資料といわざるをえないが、観察可能なデータからでは把握することができないt-1年度からt年度の情報を得ることができる意味では、本節で扱う意義は十分にあると思われる。

3 「『消費生活に関するパネル調査』回収結果」からみる脱落理由

ここでは、「回収結果」11カ年分を用いて、脱落原因に関する考察を行いたい。ここでは、村上(2003)にならい、調査経過期間別に分類し、調査の時点において、脱落理由が変化するか注意を払った。

実際には「理由」が複数にわたるものも多いため(例:「転居、仕事を始め、育児も大変であるから」など)、ここでは一番初めに記載されている理由(この場合、「転居」と分類する)を最も重要な理由として仮定し、分類を行った(表6)¹²⁾。また、脱落理由の分類が多岐にわたり、各セルにおちるサンプル数が限定されてしまうため、ここでは、配偶別の分類を行っていない。

1 CohortA

wave 1→4期間では、「(詳細不明)を除いて」理由の中で最も大きな割合を占めているのは「多忙」¹³⁾、続いて多かったのが「転居・転居先不明」であった。またこれらに加え、「結婚」「出産・育児」などライフイベントを理由とする脱落サンプルが目立っている。これは、本データの調査対象者の多くが20歳代後半から30歳代前半の「適齢期」に差し掛かっていることがその要因と考えられる。

wave 4→7期間では、「転居+転居先不明」が12.7%と(「詳細不明」を除いて)最も大きな割合を占め、「出産・育児」(7.9%)、「(長期)不在」(6.1%)が続いている。調査初期で脱落理由の大きな割合を占めていた「多忙」(4.8%)が大きく落ち込んでいる。

wave 7→11期間になると、「結婚」、「出産・育児」を理由とするサンプルが減少し、かわりに「離婚・別居」を理由とする脱落サンプルが徐々に増加し、また「入院中・体調不良・病気」の割合が増加する。ライフステージの移行、年齢の変化により、脱落理由が変わっていくことが確認できる。

11年間通した合計でみると、CohortAでは(「詳細不明」を除いて)「多忙」(11.7%)、「転居+転居先不明」(9.7%)が上位を占めている。

表6 「『消費生活に関するパネル調査』回収結果」からみる脱落理由

(単位：%)

	CohortA			合計
	wave1→wave4	wave4→wave7	wave7→wave11	
本人死亡	0.0	0.6	1.7	0.8
転居+転居先不明	8.4	12.7	8.1	9.7
(長期)不在	6.3	6.1	1.2	4.5
入院中+体調不良・病気	2.1	3.0	5.2	3.4
多忙	19.5	4.8	9.8	11.7
記入が大変、興味がない	2.1	2.4	2.9	2.5
プライバシーに立ち入っている	3.2	0.6	0.6	1.5
結婚	3.2	3.6	2.9	3.2
出産・育児	6.3	7.9	1.7	5.3
離婚・別居	2.1	0.0	4.0	2.1
就職・転職	2.6	3.0	1.7	2.5
家族の反対	5.8	4.8	0.6	3.8
家族の病気・不幸	6.3	1.8	2.3	3.6
詳細不明	31.6	45.5	52.6	42.8
分類不能	0.5	3.0	4.6	2.7
	100.0	100.0	100.0	100.0
脱落数(復活サンプル含む, 人)	190	165	173	528

	CohortB		合計
	wave5→wave8	wave8→wave11	
本人死亡	0.0	0.0	0.0
転居+転居先不明	11.5	12.3	11.8
(長期)不在	6.9	6.2	6.7
入院中+体調不良・病気	5.4	3.1	4.6
多忙	16.2	6.2	12.8
記入が大変、興味がない	0.8	3.1	1.5
プライバシーに立ち入っている	1.5	0.0	1.0
結婚	6.2	9.2	7.2
出産・育児	10.8	1.5	7.7
離婚・別居	1.5	1.5	1.5
就職・転職	0.0	1.5	0.5
家族の反対	3.1	4.6	3.6
家族の病気・不幸	0.8	0.0	0.5
詳細不明	32.3	49.2	37.9
分類不能	3.1	1.5	2.6
	100.0	100.0	100.0
脱落数(復活サンプル含む, 人)	130	65	195

2 CohortB

wave 5→wave 8期間では、CohortAと同様に、「多忙」(16.2%)、「転居+転居先不明」(11.5%)などを理由とした脱落が大きな割合を占めている。それと変わらないくらいにライフイベント(「結婚」(6.2%)、「出産・育児」(10.8%))を理由とするサンプルの割合が大きい。これは、CohortAと比べると、調査対象者が相対的に若い年齢のサンプルに限定されているため、結婚、出産などの

ライフイベントによる影響が強いものと考えられる。

wave 8→wave11期間では、「詳細不明」を除いて「転居+転居先不明」(12.3%)が最も大きな割合を占め、「結婚」(9.2%)、「(長期)不在」(6.2%)が後に続いている。7年間通した合計で見ると、CohortBでは、「多忙」(12.8%)、「転居+転居先不明」(11.8%)の割合がCohortA同様に高い。「結婚」(7.2%)、「出産・育児」(7.7%)

とライフイベントを理由とする脱落も目立っている。

JPSC における、脱落理由を大別すると、①「転居+転居先不明」、②「多忙」、③「結婚」「出産・育児」「離婚・別居」などのライフイベントとなっていることが確認された。

IV 結婚選択と脱落

JPSC における調査対象者の調査開始年度の年齢は、それぞれ満 24~34 歳 (CohortA)、満 24~27 歳 (CohortB) と、比較的若い年齢層女性を対象としている。このように若い年齢層に限定することにより、結婚や出産などの世帯変動、それに伴う多様なライフスタイルを観察することを調査の狙いとしている (永井 2000)。しかし、その反面で、ライフイベントの頻発に伴い、脱落者が発生するという問題も孕んでいることが、前述した分析結果から確認された。

本節では、脱落が推計に与えるバイアスの問題について考察していきたい。特にここでは、結婚選択関数の推計に与える、脱落の推計バイアスを考察することを目的とする。

これら欠損値による推計問題を克服するため、これまでさまざまな方法が開発されてきた。大別すると、①Complete Case Analysis、②Imputation Analysis、③Available Case Analysis の 3 つに分けることができる¹⁴⁾。

経済学において、推計バイアスを考慮した推計を行う場合、よく用いられるのが、Heckman 推計である (Heckman 1979)。しかしながら、脱落によるバイアスを推計する際に、Heckman 推計の場合、以下のような問題があることが明らかとなっている。

第一に、Wooldridge (2002) によれば、Heckman 推計では、行動方程式には現れない、選択方程式に対して影響を持つ外生変数 (除外変数) を必要としているが、パネルデータでの回答拒否や脱落の場合においては、いくつかの外生変数が観察できないという事実は、除外変数を見つけることを困難にしており、Heckman 推計を難しいものにしてしている¹⁵⁾。

第二に、Wooldridge (1997)、Terza (1998) では、Heckman 推計は、非線形推計の場合、特定の推計方法 (exponential regression) でしか、用いることができないことが示されている。

本稿では、代替的な方法として、脱落による selection がある場合にも、consistent な非線形推計ができる、Inverse Probability Weighting 法 (以下、IPW 法) を用い、脱落による推計バイアスの調整を行った (Fitzgerald, Gottschalk, and Moffitt 1998, Wooldridge 2002)¹⁶⁾。この手法は、脱落過程を MAR (あるいは MCAR) と仮定しており、IPW の算出方法は、 $t-1$ 時点以前の観察可能な情報を説明変数とし、継続回答ダミー ($c_{it} = 1$ は継続回答を、 $c_{it} = 0$ は脱落を示す) を被説明変数としたプロビット (ないしはロジット) 分析から得られた結果から、(各 wave ごとに) 推計した継続回答確率 (\hat{p}_{it}) の逆数を導き、最後に、目的としている推計関数に算出した IPW を weight づける方法である¹⁷⁾。

ここでは、 y_{it} は行動方程式の被説明変数にあたる変数、 x_{it} は説明変数にあたる変数、 z_{it} は調査初年度における y_{it} 、 x_{it} を含む情報となっている。推計式 ($F(x)$ は推計関数、ここでは結婚選択関数) は以下のように処理されている。

$$\sum_{i=1}^N \sum_{t=1}^T (c_{it}/\hat{p}_{it}) \text{Log}F(x) \quad (4)$$

IPW 法には大きく分けて、2 種類の推計方法がある。第一の方法では、継続回答関数を推計する際に、調査初年度における情報は、各 wave において、調査対象者が継続回答するかどうかに対して有効な変数であると仮定している (Wooldridge 2001 p.588, Wooldridge 2003)。いわば、調査初年度における情報が、その後、調査対象者が継続回答するかどうかに対して有効な変数であることを示している。本稿では、これを IPW (Wooldridge), $1/\hat{p}_{it}$ とする。

$$\hat{p}_{it} = P(c_{it} = 1 | y_{it}, x_{it}, z_{it}) = P(c_{it} = 1 | y_{it-1}, x_{it-1}, z_{it}) \quad t = 2, \dots, T \quad (5)$$

また、Robins, Rotnitzky and Zhao (1995) では、Wooldridge とは異なり、調査前年度に回答していた調査対象者の情報を重視し、かつ各

wave の継続回答率 (p_{it}) は、前年度以前のものを掛け込んだものを用いている。本稿では、これを IPW (Robins, Rotnitzky, and Zhao), $1/\hat{p}_{itR}$ とする。

$$\hat{p}_{it} = P(c_{it} = 1 | y_{it}, x_{it}) = P(c_{it} = 1 | y_{it-1}, x_{it-1}, c_{it-1} = 1) \quad t = 2, \dots, T \quad (6)$$

$$\hat{p}_{itR} = \hat{p}_{i2} \times \hat{p}_{i3} \times \dots \times \hat{p}_{iT} \quad (7)$$

本稿では、脱落と関連している変数 (結婚) を被説明変数として用いた推計における、推計バイアスの分析を行いたい。脱落による結婚選択関数推計への影響は、以下のように考えられる。

まず、結婚選択 (被説明変数) に与える説明変数の効果の過小評価が挙げられる。分析対象から、結婚したサンプルが一部脱落していることで、未婚サンプルに偏るといふ、サンプルセレクション・バイアスが発生している。これにより、結婚選択に対して影響のある変数の効果が過小に評価されてしまっている恐れがある。

例えば、正社員として働くことが結婚選択の確率を高めているとすると、結婚したサンプルが脱落しやすい傾向がある場合、正社員として働くという効果の一部が、脱落により捕捉されず、正社員であることが結婚選択に与える影響は過小推定されていることとなる。また、サンプル脱落により、観測可能なデータ上では、正社員として働きつづける人は結婚しない性向が高い人々に偏ってしまい、結婚選択と正社員であることとの間には、正ではなく、負の関係が発生してしまい、正社員であることが結婚選択に与える正の影響は相殺されてしまい、過小推定となってしまう。

事例でも挙げている、正規就業やパート・アルバイトなどの職業形態の区別は、結婚・出産に対して影響をもっていることが先行研究においても示されており、特に、学卒後のまもない職業として、パート・アルバイトなどのフリーター経験者と比べて、正規就業経験者は結婚選択確率が高い (酒井・樋口 2005¹⁸⁾)。初職選択がその後の結婚、出産などの人生に多大な影響を与えることを示している。

こうしたセレクション・バイアスを調整するた

めに、本稿では、上記の 2 つの IPW 法を用いた推計を行う。

また、結婚選択関数の説明変数として、パラサイトシングル、イースターリン仮説などで挙げられている、親の経済力が子どもの結婚選択に与える影響を見るべく、親の年取と子どもの年取の比や親からの小遣いや仕送りなどの移転の有無を用いている。

ここでは、双方の Cohort において、脱落に強く影響を与えている「結婚」に焦点を絞る、(サンプルの過少を補うべく) 両 Cohort をプーリングしたデータを用いて、V の推計バイアスの分析を行う。なお、分析に用いるサブサンプルは、前年に未婚状態であった者とし、またここでは親子間の年取の比をみるため、親と同居しているサンプルに限定している¹⁹⁾。

また、IPW を推計するために、継続回答関数推計で用いた説明変数は、IPW (Wooldridge) では、調査初年度における、本人の年齢、都市規模ダミー (13 大都市、その他の市 = レファレンス、町村)、親との同居、本人学歴 (中学校卒、高校卒、専門学校卒 = レファレンス、短期大学卒、大学・大学院卒)、本人の年取 (対数値)、結婚意欲、正規就業、借り入れ負担感ありを用いて、各年ごとにプロビット推計を行い、継続回答確率を推計した。一方、IPW (Robins, Rotnitzky, and Zhao) では、前年度における、IPW (Wooldridge) と同じ諸変数を用いた。

V 分析結果

IPW による調整を施したモデルと何の調整も施していない ORIGINAL モデルの係数の差がないことを帰無仮説とした、Hausman Test を行ったところ、ORIGINAL モデルと IPW (Wooldridge) モデルとの比較においては、帰無仮説が棄却され ($\chi^2 = 37.90$)、係数の差が有意であるとの結果が得られた (表 8)。

また同様に、ORIGINAL モデルと IPW (Robins, Rotnitzky, and Zhao) モデル間でも、Hausman Test を行ったところ、ここでも、係数に有意な差が見られる結果となった ($\chi^2 = 60.85$)。

表7 基本統計量

変数	サンプル数	平均値	標準偏差	最小値	最大値
結婚ダミー	1,387	0.119	0.324	0	1
年齢	1,387	28.959	3.359	24	42
13 大都市	1,387	0.317	0.466	0	1
その他の市 町村	1,387	0.567	0.496	0	1
13 大都市	1,387	0.116	0.320	0	1
初職／自営・家族従業者	1,387	0.011	0.103	0	1
初職／正規就業 (500人以上, 官公庁)	1,387	0.387	0.487	0	1
初職／正規就業 (500人未満)	1,387	0.474	0.500	0	1
初職／パート・嘱託	1,387	0.102	0.303	0	1
初職／その他	1,387	0.001	0.027	0	1
中学校卒	1,387	0.016	0.125	0	1
高校卒	1,387	0.329	0.470	0	1
専門学校卒	1,387	0.162	0.369	0	1
短期大学卒	1,387	0.267	0.443	0	1
大学・大学院卒	1,387	0.224	0.417	0	1
年収 (本人)	1,387	292.228	125.887	7	1,120
年収 (親家族)	1,387	702.249	490.406	0	2,740
年収 (親家族) / 年収 (本人)	1,387	3.643	7.391	0.0	131.0
親からの小遣い・仕送りダミー	1,387	0.177	0.382	0	1

諸変数の限界効果の違いについてみると、(初職がパート・アルバイト, その他である, フリーター経験者と比べて) 正規就業 (500人以上, 官公庁) の限界効果はそれぞれ, 0.261 (ORIGINAL), 0.263 (IPW (Wooldridge)), 0.293 (IPW (Robins, Rotnitzky, and Zhao)) と, ORIGINAL の限界効果に比べて, IPW で処理した限界効果が高くなっており, 想定通り, 推計値が過小推定されていたことが確認できた。

また, 同様に正規就業 (500人未満) の限界効果は, それぞれ 0.040 (ORIGINAL), 0.044 (IPW (Wooldridge)), 0.044 (IPW (Robins, Rotnitzky, and Zhao)), またさらに, 自営業・家族従業者の限界効果は, 0.033 (ORIGINAL), 0.035 (IPW (Wooldridge)), 0.038 (IPW (Robins, Rotnitzky, and Zhao)) と他の初職変数においても, 推計値の過小推定が確認できた。

その他にも, 高校卒ダミー, 大学・大学院卒ダミーにおいても, 推計値の過小推定が確認できた。

VI むすび

本稿では, JPSC における脱落者の特性の捕捉, また脱落による推計バイアスが結婚選択関数にど

のような影響を与えているのかを確認するべく, IPW 法による調整を行ったモデルと何の調整も行わなかったモデルとの係数の比較を行った。

II では, 一般的な脱落過程と脱落者の特徴に関する簡単な紹介を行い, III では, JPSC における脱落者の特徴を, i) 継続回答者と途中脱落者間の諸変数の平均値比較, ii) 脱落ダミーを被説明変数, 前年の情報を説明変数としたプロビット分析, iii) t 年度における回答拒否者の脱落理由がわかる「回収結果」の脱落理由の分類, と3つの方法で分析した。その結果, 無配偶者, 有配偶者に見られた共通の特徴として, 結婚予定者, 新婚者などのライフイベント前後における脱落傾向が強く見られた。特に, 無配偶者での結婚予定ダミーはほとんどのケースにおいて, (脱落確率に対して) 有意に正であるという結果が得られた。また, 経済的要因に目をむけると, 無配偶者では, 本人の年収が低いほど, 収入変化額が減少するほど, 脱落する結果が得られた。有配偶者では, 逆に, 本人 (妻), 夫の双方と年収の変化額が正に大きいほど, 脱落する傾向が見られた。

またさらに, 「回収結果」から脱落理由を分類した結果, 「多忙」「転居+転居先不明」が多くを占めており, 若い年齢層に限定すると, 「結婚」

表 8 結婚選択関数

	ORIGINAL			IPW (Wooldridge)			IPW (Robins, Rotnitzky, and Zhao)		
	限界効果	頑健的 標準誤差	z	限界効果	頑健的 標準誤差	z	限界効果	頑健的 標準誤差	z
年齢	-0.010	0.003	-3.31***	-0.010	0.003	-3.30***	-0.010	0.003	-3.26***
13大都市 (町村)	-0.032	0.018	-1.72*	-0.033	0.018	-1.72*	-0.028	0.019	-1.38
その他の市	0.075	0.034	2.53**	0.079	0.035	2.62***	0.074	0.035	2.42**
初職／正規就業 (500人以上, 官公庁)	0.261	0.116	2.93***	0.263	0.118	2.90***	0.293	0.121	3.12***
初職／正規就業 (500人未満) (初職／非正規就業)	0.040	0.031	1.34	0.044	0.032	1.42	0.044	0.033	1.36
初職／自営・家族従業者	0.033	0.029	1.16	0.035	0.029	1.19	0.038	0.031	1.24
中学校卒	-0.088	0.022	-1.95*	-0.091	0.023	-1.98**	-0.093	0.023	-1.96**
高校卒	-0.050	0.021	-2.19**	-0.052	0.022	-2.24**	-0.053	0.022	-2.23**
(専門学校卒)									
短期大学卒	-0.034	0.022	-1.45	-0.035	0.023	-1.47	-0.034	0.023	-1.39
大学・大学院卒	-0.048	0.022	-2.00**	-0.050	0.022	-2.05**	-0.052	0.023	-2.05**
所得(親家族)／所得(本人)	-0.002	0.002	-1.02	-0.002	0.002	-1.02	-0.002	0.002	-0.96
親からの小遣い・仕送り	0.013	0.024	0.56	0.016	0.025	0.65	0.020	0.027	0.77
1994年	-0.010725	0.036	-0.29	-0.003	0.038	-0.07	0.028	0.113	0.26
1995年	0.012	0.040	0.31	1.14E-02	0.040	0.29	4.31E-02	0.120	0.39
1996年	-0.055	0.029	-1.57*	-0.058	0.029	-1.63*	-0.031	0.085	-0.33
1997年	-4.92E-02	0.029	-1.48	-0.052	0.030	-1.53*	-0.025	0.091	-0.26
1998年	-0.032	0.032	-0.91	-0.033	0.032	-0.94	0.003	0.103	0.03
1999年	-0.041	0.030	-1.20	-0.043	0.030	-1.24	-1.65E-02	0.094	-0.17
(2000年)									
2001年	—			—			0.042	0.121	0.38
2002年	-0.023	0.082	-0.26	-0.025	0.083	-0.27	—		
定数項	0.554※	0.553	1.00	0.557※	0.555	1.00	0.393※	0.771	0.51
サンプル数		1,387			1,387			1,387	
Wald χ^2		63.930			66.470			63.040	
Prob > χ^2		0.000			0.000			0.000	
擬似決定係数		0.054			0.057			0.052	
Log pseudo-likelihood		-478.860			-485.774			-492.226	
Hausman Test									
H0: difference in coefficients not systematic									
$\chi^2 = (b-B)'(V(b-B))^{-1}(b-B)$		---			37.90			60.85	
Prob > $\chi^2 =$		---			0.0061			0.0000	

注：1) ***：1%水準，**：5%水準，*：10%水準，#：15%水準で統計的に有意
2) ※定数項のみ係数値を示している

「出産」などのライフイベント発生によるものが多かった。また、特に結婚というイベントは、転居の predictor となるだけでなく、夫、夫の家族による調査拒否、それを忌避して自ら拒否するケースが発生しており、JPSC における脱落と非常に密接な関係にあると考えられる。

その結果を受けて、IVでは、結婚選択推計に対して、脱落が与える推計バイアスの問題について考察した。IPW法 (Wooldridge 2001 2003, Robins, Rotnitzky, and Zhao 1995) による調整を施した推計結果によると、初職がパート・アルバイトなどのフリーターであった者と比べて、初職が正規就業であった者、自営・家族従業者であった者の、

(結婚選択確率に対する) 限界効果は、何の処理も施していない ORIGINAL モデルに対して、IPWを用いた2つのモデルは、(正に) 大きくなっていることが確認され、かつ (Hausman Test によれば) それらの係数の差は有意であることが確認された。

以上の分析結果から、結婚したサンプルが一部脱落していることで、未婚サンプルに偏るという、サンプルセレクション・バイアスが発生していること、また結婚に影響を与えている変数の結果が過小となることが確認できた。

* 論文作成にあたり、川口大司、栗田巨相、伴金美、町北朋洋

各氏、一橋大学 21 世紀 COE プログラム『社会科学の統計分析拠点構築』若手研究会参加者より、また『日本労働研究雑誌』の匿名レフェリーの方に、有益なコメントならびにアドバイスを賜った。記して感謝を申し上げたい。いうまでもなく、本稿の誤りのすべては筆者に帰す。

- 1) 1993 年時に 24~34 歳であった女性 (1993 年より調査継続)。
- 2) 1997 年時に 24~27 歳であった女性 (1997 年より調査継続)。
- 3) 他のパネル調査, the Dutch Transportation Panel では、7 年 (1984~1990 年, 年 2 回) でオリジナルサンプルの約 38% が脱落している (Ridder (1992))。
- 4) また、途中脱落者の年齢は (脱落時点で) 止まってしまうのに対して、継続回答者は回答を続けていく限り、年齢が上昇していくことを考慮し、調査期間ごとで分割して比較を行った場合、CohortA の調査開始前半 (第 1 年度~第 4 年度) のみ差が有意となった (調査期間別の結果は図示していない)。
- 5) 札幌市, 仙台市, さいたま市, 千葉市, 東京都区, 横浜市, 川崎市, 名古屋市, 京都市, 大阪市, 神戸市, 広島市, 北九州市, 福岡市。
- 6) しかし、CohortA の場合、調査期間ごとにみると、wave1 から wave3 では、496 万 7300 円 (継続回答者), 562 万 7200 円 (途中脱落者), wave4 から wave7 では、552 万 7500 円 (継続回答者), 623 万 5000 円 (途中脱落者) と調査期間前半、中盤において、途中脱落者の方が統計的に有意に高い結果が得られている (調査期間別の結果は図示していない)。
- 7) ここでは、ローン返済について、「苦しい」「かなり負担である」と回答した者を「借入れ負担感あり = 1」、それ以外の者を「借入れ負担感なし = 0」としている。
- 8) これを配偶別にみると、有配偶においてのみ、正規就業割合が 16.95% (継続回答者), 22.11% (途中脱落者), パート・嘱託割合が、24.03% (継続回答者), 20.26% (途中脱落者) とそれぞれ統計的に有意に差が見られたが、無配偶では差がみられなかった (配偶別結果は図示していない)。
- 9) 前節と同様、転居 + 転居先不明などの移動性向の影響をはかるために、ここ 3 年の間に転居したかどうかというダミー変数を用いたが、有意な結果が得られず、また収入だけでなく、収入に対する支出の割合をみることで、家計としての余裕の有無で回答継続に違いがあるか、さらに先行研究を参考に、本人、夫の転職、離職、就職などの労働移動も用いたが、有意な結果が得られなかった。
- 10) 調査員が捕捉できる範囲 (調査者からの電話での拒否連絡があった、直接訪問した際に、拒否理由を述べられた) で、脱落者の脱落理由についてまとめている資料である。なお、これは内部資料とされており、公開されていない。
- 11) 村上 (2003) では、これを用いた分析が行われている。
- 12) 「wave 2 → wave 3」, 「wave 3 → wave 4」, 「wave 4 → wave 5」に関しては、「事前拒否連絡」サンプルの脱落理由については捕捉されていない。
- 13) 「多忙」といっても、ここで分類されたものは、「理由」が明記されておらず、その具体的な内容は千差万別であると思われる。仕事を理由とするもの、家事を理由とするもの、育児を理由とするものなど他の分類項目と重なる可能性も否定できない。
- 14) 脱落によるバイアス克服の方法については、岩崎 (2002), Verbeke and Molenberghs (1997) に詳しい。
- 15) また、一般的な Heckman 推計の批判は、Puhani (2000) において述べられている。
- 16) Wooldridge (2002) では、parametric model を用いて、

継続回答率を推計しているが、Hirano, Imbens, and Ridder (2000) は nonparametric model (Propensity Score) を用いて推計している。

- 17) STATA8.0 では、「pweight」というオプションを用いることで weight づけることができる。
- 18) 「第 1 回慶應家計パネル調査」を用いたサバイバル分析から、学卒後 1 年から 2 年の間、フリーター経験者は正規就業経験者よりも結婚年齢が高い。
- 19) 同居している場合のみ、正確な親の年取を捕捉できる。

参考文献

- 岩崎学 (2002) 『不完全データの統計解析』エコノミスト社。
- 酒井正, 樋口美雄 (2005) 「フリーターのその後——就業・所得・結婚・出産」『日本労働研究雑誌』No. 535, 29-41。
- 永井暁子 (2000) 「消費生活に関するパネル調査」佐藤博樹, 石田浩, 池田謙一編『社会調査の公開データ 2 次分析への招待』第 II 部 3, 12 章, 東京大学出版会。
- 樋口美雄, 岩田正美編 (1999) 『パネルデータから見た現代女性』東洋経済新報社。
- 樋口美雄, 太田清, 家計経済研究所編 (2004) 『女性たちの平成不況』日本経済新聞社。
- 村上あかね (2003) 「なぜ脱落したのか——『消費生活に関するパネル調査』における脱落サンプル分析」財団法人家計経済研究所編『家計・仕事・暮らしと女性の現在——消費生活に関するパネル調査 第 10 年度』国立出版局, 第 3 部第 1 章。
- Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998) “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics”, *The Journal of Human Resources*, 33, 251-299.
- Heckman, J. (1979) “Sample Selection Bias as a Specification Error”, *Econometrica*, 47, 153-162.
- Kalton, G., J. Lepkowski, G. Montanari, D. Maligalig (1990) “Characteristics of Second Wave Nonrespondents in a Panel Survey”, *Proceedings of the American Statistical Association, Survey Research Methods Section*, 462-467.
- Hirano, K., G. Imbens, and G. Ridder (2000) Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”, USC Center for Law, Economics & Organization Research Paper No.C02-13.
- Lepkowski, J., and M. Couper (2002) “Nonresponse in Second Wave of Longitudinal Household Surveys”, Groves, Dillman, Eltinge, and Little, *Survey Nonresponse*, John Wiley & Sons, 259-272.
- MaCurdy, T., T. Mroz, and M. Gritz (1998) “An Evaluation of the National Longitudinal Survey on Youth”, *The Journal of Human Resource*, 33, 345-436.
- Neumark, D., and D. Kawaguchi (2004) “Attrition Bias in Economic Relationships Estimated with Matched CPS Files”, *Journal of Economic and Social Measurement*, Vol. 29, No. 4, pp.445-472.
- Puhani, P. (2000) “The Heckman Correction for Sample Selection and Its Critique”, *Journal of Economic Surveys*, 14, 53-68.
- Ridder, G., (1992) “An Empirical Evaluation of Some Models for Nonrandom Attrition in Panel Data”, *Structural Change and Economic Dynamics*, 3, 337-355.
- Robins, J., A. Rotnitzky, and L. Zhao, (1995), “Analysis of

- Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data.” *Journal of the American Statistical Association*, 90, 106-121.
- Terza, J. (1998) “Estimating Count Data Model with Endogenous Switching; Sample Selection and Endogenous Treatment Effects” *Journal of Econometrics*, 84, pp.129-154.
- Verbeke, G., and G. Molenberghs (1997) *Linear Mixed Models in Practice: A SAS-Oriented Approach*, Springer (邦訳 松山裕, 山口拓洋訳『医学統計のための線形混合モデル—SASによるアプローチ』サイエンティスト社).
- Wooldridge, J. (1997) “Quasi-Likelihood Methods for Count Data” Pesaran M., Schmidt, P. (eds) *Handbook of Applied Econometrics* vol.2, pp.352-406, Blackwell, Oxford.
- Wooldridge, J., (2001) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- Wooldridge, J., (2002) “Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification”, CeMMAP Working Paper CWP11/02.
- Wooldridge, J., (2003) “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity”, Centre for Microdata Methods and Practice, Institute for Fiscal Studies in Its Series CeMMAP Working Papers, CWP18/02.
- 〈投稿受付 2005 年 2 月 14 日, 採択決定 2005 年 12 月 9 日〉

<p>さかもと・かずやす 財団法人家計経済研究所研究員。主な著作に『stataで計量経済分析入門』（共著, ミネルヴァ書房, 近刊)。労働経済学専攻。</p>
