



家計調査のパネル化について : 世帯照合の方法

宇南山, 卓

(Citation)

国民経済雑誌, 204(3):51-64

(Issue Date)

2011-09

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD01)

<https://doi.org/10.24546/81008361>

(URL)

<https://hdl.handle.net/20.500.14094/81008361>



家計調査のパネル化について： 世帯照合の方法

宇 南 山 卓

国民経済雑誌 第 204 巻 第 3 号 抜刷

平 成 23 年 9 月

家計調査のパネル化について： 世帯照合の方法

宇 南 山 卓

本稿の目的は、総務省統計局の公表する「家計調査」の個票データを用いてパネルデータを作成する手順を示すことである。家計調査は同一世帯を6ヶ月追跡して調査しているパネルデータであるが、同一世帯の異なる調査月での結果を識別する符号が付与されていないためパネルデータとして扱うことが困難であった。ここでは、家計調査において同一月内での世帯の識別に用いられる符号を改善することで、パネルデータを作成するための照合キーを構築する手順を示した。

キーワード 家計調査、パネルデータ、世帯識別

1 は じ め に

ここでは、家計調査の個票データを利用することを前提に、パネルデータとして利用するために必要な、同一世帯の各月レコードの照合方法を構築する。多くの先行研究で、「家計調査」は月次のクロスセクション統計として分析されているが、実際には各世帯が6ヶ月間調査され、一定比率がサンプル替えされるローテーション方式のパネル (Rotating Panel) である。6ヶ月の短期とはいえ、パネルデータとして利用できれば同一世帯の行動が追跡でき、情報量は格段に増える。

しかし、家計調査の世帯識別符号は、同一世帯を異なる月で追跡することを前提に設計されていないため、そのままではパネルデータを構築するための符号として利用できない。家計調査の個票データでは世帯識別符号として「市町村番号」、「単位区符号」、「世帯番号」、「一連世帯番号」が利用できるが、これらの符号は世帯と完全には対応しておらず、同一世帯を完全にはマッチさせることはできず、しかも同一でない世帯のレコードをマッチさせる誤りを発生させる。

この問題に対し、家計調査をパネルデータとして利用した最初の試みである林 (1986) では、これらの識別符号に「年間収入」も照合キーに加えることでパネルを構築する試みをしている。¹⁾ この方法は、異なる世帯をマッチしてしまう可能性を下げることはできるが、家計調査の世帯の識別符号の持つ問題を解決しておらず、依然として不適切な照合方法である。

にもかかわらず、その後の研究でも基本的にこの林（1986）の方法が用いられている（清水谷・堀，2005；Hori and Shimizutani, 2005a, 2005b, 2006；Shimizutani, 2006；美添・荒木，2006²⁾）。林（1986）の方法による照合の失敗は家計行動と構造的に関連しており，サンプル数の減少という情報の損失にとどまらず，サンプルセクションバイアスを引き起こし，家計調査のパネルデータとしての利用価値を低下させる。

ここでは，家計調査の世帯の識別符号について概説して問題の所在を明確にし，その上で，先行研究の方法での問題点を完全に解決した新たな照合方法を提示する。さらに，提示された新たな方法と先行研究の方法を比較することで，先行研究でとられた方法の問題の大きさを評価した。

本稿の以下の構成は次の通りである。まず，第2節では家計調査の標本抽出手順と世帯符号の概要を説明した。第3節では，先行研究の方法を批判的に再検討することで，公式の世帯符号の問題点を指摘した。つづく第4節で新たな照合方法のアルゴリズムを提示した。また，実際のデータを用いて，先行研究の方法による照合結果と比較した。最後の第5節はまとめである。

2 家計調査の識別符号の概要

家計調査の個票データは，通常，月別に別ファイルとして提供される³⁾。各世帯は月ごとに別のレコードとして記録されており，世帯ごとのパネルデータを構築するには異なる月のファイルを統合し，プールしたレコードを照合して世帯ごとに関連付けしなければならない。しかし，異なる月で共通した世帯識別のための符号は存在しない。それが，家計調査をパネルデータとして扱う際の最大の問題である。

そこで，まず本節では家計調査の各月における世帯の識別方法について概観する。その上で，複数の月をプールした場合における世帯識別方法の問題点を指摘する。

家計調査の月次ファイルの各レコードには3つの世帯識別に関する符号：

- 市町村番号（5桁）
- 単位区符号（3桁）
- 世帯番号（3桁：1999年6月までは2桁）

が付与されている。家計調査を単月のクロスセクションデータとして見た場合，これらの符号によって世帯をユニークに識別することができる。言い換えれば，同じ月のサンプルの中でこれら3つが完全に一致するレコードは存在しない⁴⁾。

各符号は家計調査の層化3段階抽出法による標本抽出に対応している。その意味で，問題

の所在を明らかにするには、家計調査の標本抽出を示すことが必要である。その上で、各符号の役割・位置づけを述べる。

家計調査の標本抽出は以下の手順による：

- ① 全国の市町村から規模別に確率比例方式で調査市町村を決定する。
- ② 選定された市町村の全域を一定のルールに従ってブロックと呼ばれる区域に区分し、調査対象のブロックを市町村の規模に応じた数だけ抽出する⁵⁾。その抽出されたブロック内を調査世帯数が75以上になるように「単位区」とよばれる区域に分割し、2単位区ずつ調査対象として選定する。
- ③ 選定された単位区で、調査員が実地踏査によって単位区内全世帯の名簿を作成し、世帯ごとに番号を振る。その番号をもとに各単位区6世帯ずつ原則として6ヶ月間調査される。
- ④ 6ヶ月の調査終了後に、世帯名簿の修正によって最新の世帯情報に更新した上で、さらに6世帯が選定される。

このうち、最初の市町村の抽出に用いられるのが、市町村番号である。これは総務省（2001年までは自治省）が「地方公共団体コード」として付与している市町村固有の符号である。調査対象かどうかにかかわらず市町村は市町村番号を持っており、事後的にも市町村番号によって調査対象市町村をユニークに同定できる⁶⁾。

ブロック・単位区については家計調査の調査時に独自に符号を付していると考えられるが、ブロックを選定して単位区を抽出する際に用いられる符号は公表されていないため詳細は不明である。公表されている単位区符号は、単位区を特定するために事後的に付与されたものである。最初の2桁がその単位区の調査が開始された月を表し、最後の1桁は同一市町村内の単位区に対し1から順に便宜的に割り振ったものである。そのため、単位区符号は地理的な情報とは結びついておらず、研究者は市町村内のどの地域（ブロック・単位区）が調査されたかを知ることはできない。また、調査主体である総務省統計局も単位区符号からだけでは、地理的な対応を復元することはできない。

世帯番号も、標本抽出の後に割り振られた便宜的な符号であり、世帯固有の符号ではない。3桁の数字はそれぞれ最初の1桁が農林漁家世帯であるかどうかの識別、2桁目が勤労者世帯かどうかの識別符合、3桁目が当該単位区内での世帯に対し割り振られた番号であり、農林漁家、勤労者世帯、それ以外の世帯、それぞれについて1から順につけられている⁷⁾。

抽出手順ごとに符号がついていることから、これら3つの識別符号、市町村番号・単位区符号・世帯番号によって、各月のレコードはユニークに識別できる。しかし、以下で述べる

理由により、複数の調査時点で世帯をユニークに識別することはできない。つまり、同一の世帯が異なる番号を持つ可能性と、異なる世帯であっても同一の世帯番号を持つ可能性がある。そのため、複数の月のデータをプールしてしまうと、世帯識別符号はパネルデータ構築の照合キーにはできない。

3 世帯識別符号の変化とその照合

3.1 先行研究におけるパネルデータの構築方法

前節で述べたように、家計調査の世帯識別符号は同一月内で世帯をユニークに識別することを目的としているため、パネルデータの照合キーとしては問題がある。それに対し林(1986)では、「一連世帯番号」および「年間収入」を照合キーに加えることで、パネルデータにおける世帯の識別を試みている。ここでは林(1986)が追加的に使用した変数によって、どのように影響を受けるかを論ずる。

林(1986)では次の5つの変数を照合キーとして使用して、異なる月のレコードのうち上の変数が一致するレコードを「同一世帯」とみなしている。

- 市町村番号(5桁)
- 単位区符号(3桁)
- 世帯番号(3桁:1999年6月までは2桁)
- 世帯一連番号(3桁:1999年6月までは2桁)
- 年間収入(万円単位)

一連世帯番号とは、選定された単位区において作成される世帯名簿に記載される番号であり、世帯区分(勤労者世帯・非勤労者世帯)ごとに1から順につけられている。同一名簿内では異なる世帯では必ず異なっており、実際に抽出されるかどうかによらず割り当てられる世帯固有の番号である。家計調査の個票データでは、市町村番号・単位区符号・世帯番号に一連世帯番号を加えた4つの変数が世帯識別のために提供される全ての情報である。⁸⁾

世帯は世帯番号によって識別されており、各月内では一連世帯番号を照合キーとしても追加的な情報は無い。しかし、名簿が更新されない限り、異なる月でも一連世帯番号は不変であるため、同一世帯を識別する手がかりとなる。その意味では、異なる月のレコードを照合するには欠かせない変数である。

さらに、林(1986)では、年間収入も照合キーとして利用している。年間収入とは、調査開始以前の12ヶ月の収入を年間収入調査票によって調査したものであり、世帯ごとに1回のみ調査する項目であり、原理的には同一世帯は同一の値をとるはずの変数である。年間収入

を照合キーとして用いる理由について林（1986）ではその根拠を述べていないが、美添・荒木（2006）では「論理チェック」としている。

年間収入は、世帯識別符号だけでは同一世帯を識別できないのであれば、追加的な情報を持つ。実際、年間収入を照合キーに加えると照合の結果は大きく変化をする。しかし、調査主体が体系的に割り当てた符号ではなく、本来は照合の目的に適したものではない。にもかかわらず、ほとんどの先行研究で照合キーとして採用されていることから、ここでは年間収入を用いる意義を検討する。

以下では、この林（1986）による照合方法の問題点を、異なる世帯が同一世帯とみなされる誤りと同一の世帯のレコードを照合できない誤りの2つに分類して論じる。特に、同一世帯であっても照合されない誤りは、経済学的に重要なバイアスを生む可能性がある。

3.2 異なる世帯を同一世帯とみなす誤り

家計調査の4つの世帯識別符号（市町村番号・単位区符号・世帯番号・一連世帯番号）を用いて異なる月のレコードを照合すると、異なる世帯を同一世帯とみなす可能性がある。単位区符号が事後的に割り振られた番号であることが、その原因である。

単位区符号は調査開始月に応じて決定されるため、地理的には異なる単位区であっても同一の単位区符号を持つ。例えば、ある市町村で2月に調査開始した単位区を「021」とすると、その単位区では2月と8月に6世帯ずつ世帯を抽出され、翌年の2月に別の単位区と交代する。しかし、新たに調査開始された単位区のうちの1つにも「021」の番号が振られてしまう。

その単位区では新たに名簿が作成されるため、最終的に抽出された世帯の世帯番号・一連世帯番号が、交代前の単位区で抽出されていた世帯の世帯番号・一連世帯番号と一致する可能性がある。すなわち、異なる単位区に住む違う世帯であるにもかかわらず、4つの世帯識別の符号が完全に一致した世帯が存在することになる。これが、家計調査の世帯識別符号がパネルデータ構築に利用できない第一の理由である。

この可能性を考慮すれば、年間収入を照合キーに加えることは、論理チェックではなく実質的な意味を持っている。すなわち、4つの世帯識別の符号が完全に一致した世帯が存在したとしても、年間収入も一致しない限り同一世帯としてみなさないとすれば、誤って異なる世帯を同一世帯とみなす可能性は低くなる。しかし、年間収入は排他的に付与される符号ではなく、異なる世帯が同一の値をとりうるため「異なる世帯を同一世帯とみなす可能性」を完全には排除できない。つまり、年間収入を照合キーに加えることは、実際上の有効性はあるが、問題を完全に解決することはできない。

一方で、年間収入は、年間収入調査票の提出が遅れた場合に、その後提出された時点で修

正されるまで暫定的な推計値が代入される。つまり、同一世帯であっても年間収入が変化する可能性がある（美添・荒木，2006）。この手続きのため、年間収入を照合キーとすることで、新たに「同一世帯を照合できない誤り」を発生させる。この両方の意味で、年間収入を用いることは、照合の改善にはつながらない。

ここでは、年間収入を照合キーに加えずに世帯の交替を把握するために、異なる単位区は異なる符号となるように2つの情報を付加する。まず、単位区の調査開始年を単位区符号に付加する。さらに、同一単位区で1度目と2度目の抽出を区別できるように単位区符号に記号を付加する。同一の単位区であっても、2回の抽出で同一の世帯が選定されることはないため、別の単位区符号を付与することで、2度の抽出を識別できる。具体的には、各単位区の単位区符号の最初の2桁は1回目の最初の調査月になっているため、調査月との関係で最初の調査か2度目の調査かを決定することができる。例えば、1990年2月に調査開始のある単位区「021」については、1990年の2、3、4、5、6、7月は最初の調査で、8、9、10、11、12月および1991年の1月⁹⁾は2回目の調査となる。そこで、当該単位区の符号を、2月から7月までは「19900211」、それ以降翌年の1月までを「19900212」とする。この記号を付加することで、単位区の2回の調査で、同一の一連世帯番号・世帯番号を持つ別の世帯が存在しても、識別することが可能になり、異なる世帯を同一の世帯とみなす誤りを完全に排除できる。

3.3 同一世帯のレコードを照合できない誤り

単位区符号を修正した上で4つの世帯識別符号（市町村番号・単位区符号・世帯番号・一連世帯番号）を用いることで、照合された世帯は同一世帯とみなすことができる。しかし、必ずしも同一世帯のレコードが完全に照合できるとは限らない。その理由は、同一世帯の照合キーが月によって変化する可能性があるためである。以下では、世帯識別符号が変化する可能性について検討する。

具体的には、世帯識別符号のうち、市町村番号と世帯番号については同一世帯であっても調査期間中に変更される可能性がある。以下で、各符号が変更される理由を説明し、それに対する対応を述べる。

市町村番号の変更 市町村番号は、家計調査固有のものではなく総務省（2001年までは自治省）が導入している「地方公共団体コード」であり、町から市への移行・政令指定都市への移行などによって変更される。家計調査の調査期間中にこのコードが変更されると、同一世帯であっても市町村番号が変化することになり、レコードの照合に失敗する¹⁰⁾。

この変化に対応して同一世帯を照合するためには、月によらず市町村番号が統一されてい

なければならない。そこで、変更される前の市町村番号を変更後の番号に統一する。家計調査の調査市町村数は168だけであり、県庁所在市以外の市町村で調査期間中に市町村番号が変化するのは極めてまれであった¹¹⁾。しかし、2005年の「平成の大合併」により市町村合併が相次いだ影響で、2006年1月以降は市町村番号の変更が多く観察された。その意味では、近年の家計調査をパネルデータとして活用する場合には、十分な注意が必要である。

世帯番号の変更 世帯番号は、調査世帯の抽出後に割り当てた数字であり、1桁目は農林漁家であれば1、非農林漁家であれば2、2桁目は勤労者世帯は0、それ以外は1、とされる。すなわち、世帯番号によって世帯の格付を知ることができる。一方、3桁目は、単位区・世帯格付ごとに1から割り当てたもので、単位区あたりの調査世帯数が6であることから、最大で6までの数字となっている。また、世帯の格付が変更された場合に、1桁目は農林漁家から非農林漁家であれば6に変更され、逆のケースでは7、2桁目は勤労者世帯からそれ以外の世帯になると5、その逆は6、となる。

世帯番号は世帯の格付によって決まっているため、1桁目および2桁目の符号は世帯の格付が調査期間中に変更されるとそれに応じて変更される。言い換えれば、調査期間を通じて一定とは限らない。世帯番号が変化する可能性を考慮すると、世帯番号を照合キーに加えることで世帯の格付が変化した世帯のレコードの照合に失敗する。

この失敗は、調査客体の経済環境の変化によって生じるため、市町村番号の変更と比較して深刻なサンプルセクションバイアスを発生させる。例えば、勤労者世帯であった世帯が無職世帯になると世帯番号が変更され、その後のレコードとの照合はできないため、失業の所得・消費への影響を分析することは出来ない。これは、サンプル数の減少という情報の損失だけでなく、パネルデータとして本質的な性質を失わせることになる。家計調査をパネルデータとして最大限に活用するためには、特にこの点を改善することが必須である。

また、この世帯の格付の変更は世帯番号の3桁目も変更させる可能性がある。世帯の格付が変更された場合、当該世帯区分で世帯番号の3桁目にギャップが生まれることになる。この場合でも、原則として1桁目もしくは2桁目だけが変更されるはずであるが、当該単位区の世帯番号を再度割り当てられたと考えられるケースが観察された。これは、世帯の格付に変更がないにもかかわらず、同一単位区の他の世帯の格付が変更されたために世帯番号が変更されるケースである¹²⁾。

これらの問題に対応するために、世帯番号ではなく、世帯一連番号を用いることを考える。『家計調査年報』の「家計調査の概要」によれば、単位区が選定されると「調査員が単位区内を实地踏査して、単位区内に居住するすべての世帯をリストした「単位区世帯名簿」を作成する」とある。一連世帯番号は、この単位区世帯名簿に記載される番号であり、単位区内

の全世帯が世帯区分ごとにユニークに割り振られた番号である。この番号は世帯の格付の変更があっても変化せず、他の世帯の変化の影響も受けないため、照合に適した性質を持っている。しかし、単位区世帯名簿での一連世帯番号は、勤労者世帯・一般世帯・単身世帯の区分ごとに1から番号をつけているのに対し、個票データ上の一連世帯番号には世帯名簿作成時の世帯格付の情報は含まれていない。そのため、同一単位区内に世帯の区分は異なるが、同じ一連世帯番号を持った世帯が存在することになり、そのままでは世帯番号を代替することは出来ない。

これらの世帯番号・一連世帯番号の問題に対応するために、ここでは、世帯番号の1桁目・2桁目の情報を用いて調査世帯がもとの単位区世帯名簿の時点で勤労者世帯に区分されていたか一般世帯であったかを識別し、その区分を一連世帯番号に付加した。具体的には、世帯番号1桁目が6, 7の世帯をそれぞれ1, 2に修正して、2桁目が5, 6の世帯はそれぞれ0, 1に修正し、それを一連世帯番号の1・2桁目とするのである。また、3桁目は一連世帯番号を使う場合には追加的な情報を持たないため削除する。この修正された一連世帯番号は、単位区内でユニークであり、なおかつ世帯固有で時間と共に変化しない識別符号となる。

4 照合のアルゴリズム

4.1 照合手順

本節では、以上の論点をまとめ、家計調査の識別符号の特徴を利用した照合方法の具体的なアルゴリズムを提示する。さらに、実際の照合における技術的な問題点を示す。上記の問題点を考慮した、具体的な照合手順は、以下の通りである。

1. 全ての市町村番号を最新のものに変更する
2. 単位区符号に調査開始年を付加する
3. 調査月と調査開始月に応じて、修正した単位区符号に1もしくは2を付加する
4. 1999年6月以前のレコードの世帯番号の最初に2を付加する
5. 世帯番号の1桁目を切り出し、6は1に、7は2に変更する
6. 世帯番号の2桁目を切り出し、5は0に、6は1に変更する
7. 1999年6月以前のレコードの一連世帯番号の最初に0を付加する
8. 修正した世帯番号1桁目と2桁目を一連世帯番号に付加する
9. 修正した市町村番号・単位区符号・一連世帯番号の3変数を照合キーとする

4.2 実際の適用における注意点

手順1については、『家計調査年報』の「付録：家計調査の経緯」に記載された調査市町

村の交代を反映させ、各調査月における調査対象市町村をリスト化する。その調査対象市町村が、調査期間中に市町村番号の変更があったかを調べることで変更履歴を作成することができる。¹³⁾ 1981年1月から2010年3月までで、調査期間中に市町村番号が変更されたものを表1にまとめている。

表1 市町村番号の変更の一覧

旧市町村名	旧コード	新市町村名	新コード	変更時点
福岡県太宰府町	40302	大宰府市	40221	1983年1月
広島県廿日市町	34322	廿日市市	34213	1989年1月
宮城県仙台市	04201	政令指定都市への昇格	04100	1990年1月
千葉県千葉市	12201	政令指定都市への昇格	12100	1993年1月
茨城県潮来町	08423	潮来市	08223	2002年1月
埼玉県浦和市等	11204	さいたま市	11244	2002年1月
千葉県白井町	12326	白井市	12232	2002年1月
埼玉県さいたま市	11244	政令指定都市への昇格	11100	2004年1月
島根県三刀屋町	32381	雲南市	32209	2005年1月
群馬県水上町	10446	みなかみ町	10449	2006年1月
埼玉県妻沼町	11403	熊谷市	11202	2006年1月
富山県新湊市	16203	射水市	16211	2006年1月
石川県能都町	17423	能登町	17463	2006年1月
静岡県静岡市	22201	政令指定都市への昇格	22100	2006年1月
静岡県相良町	22422	牧之原市	22226	2006年1月
愛知県西枇杷島町	23341	清須市	23233	2006年1月
三重県二見町	24462	伊勢市	24203	2006年1月
滋賀県近江町	25464	米原市	25214	2006年1月
兵庫県龍野市	28211	たつの市	28229	2006年1月
兵庫県神崎町	28441	神河町	28446	2006年1月
和歌山県粉河町	30322	紀の川市	30208	2006年1月
岡山県吉永町	33344	備前市	33211	2006年1月
広島県音戸町	34311	呉市	34202	2006年1月
山口県小野田市	35209	山陽小野田市	35216	2006年1月
長崎県多良見町	42306	諫早市	42204	2006年1月
熊本県鏡町	43463	八代市	43202	2006年1月
鹿児島県吹上町	46367	日置市	46216	2006年1月
沖縄県平良市	47206	宮古島市	47214	2006年1月
宮城県田尻町	04502	大崎市	04215	2007年1月
茨城県伊奈町	08482	つくばみらい市	08235	2007年1月
長野県丸子町	20341	上田市	20203	2007年1月
大分県安岐町	44325	国東市	44214	2007年1月
新潟県新潟市	15201	新潟市	15100	2008年1月
静岡県浜松市	22202	浜松市	22130	2008年1月
京都府山城町	26361	木津川市	26214	2008年1月

出所)『家計調査年報』各年版より筆者作成

手順2・3の単位区符号については、1回目と2回目の調査を区別する符号をつけるために調査開始月である単位区符号の上2桁を切り出した。その月が、そのレコードが記録され

ている調査月よりも6か月以内であれば1を、6ヶ月以上前であれば2を付与して3ケタの数字とした。

手順4・7の世帯番号・一連世帯番号の修正は、1999年6月まで農林漁家世帯が家計調査の対象となっていなかったことと、一連世帯番号が1999年7月から3桁に変更されたことに対応したものである。これは単純に、1999年6月以前の全てのレコードの世帯番号に非農林漁家の符号である2を付加し、一連世帯番号には0を付加することで対応できた。修正した世帯番号の1桁目・2桁目は、論理的にそれぞれ1, 2, 6, 7および0, 1, 5, 6の4つの値しかとらないが、それは実際のデータでも確認できた。

手順5・6は、もとの単位区世帯名簿で各世帯がどのように格付されていたかを復元するために、世帯番号1・2桁目の変更履歴を用いてもとの格付に戻したものである。世帯一連番号は調査開始前の状況に応じて、勤労者世帯と非勤労者世帯で別に割り当てられており、名簿での格付に戻すことで一連世帯番号が単位区内の世帯毎に固有の符号となる。

最後の修正一連世帯番号の作成には注意が必要である。原理的には、修正された世帯番号の上2桁を一連世帯番号に付加することで、各月の全てのレコードがユニークに識別できるはずである。しかし、実際のデータでは「同一月内に、同一格付で、同一の一連世帯番号を持つ」2つの世帯が存在するケースが観察された。個票の秘匿の観点から、個別のデータについては示さないが、他の項目などから推定する限り一連世帯番号の記録ミスの可能性が高い。そこで、ここでは重複する一連世帯番号のレコードの一方を、他と重複しない番号に置き換えることで対応した。¹⁴⁾

4.3 照合結果と先行研究との比較

ここでは、1981年1月から2010年3月までの期間（351ヶ月分）を対象として、実際に提示したアルゴリズム（以下新方式）と先行研究で最も一般的である年間収入も含めた5つの照合キーを用いた林（1986）の方法（以下林方式）によって、世帯のパネルデータを構築する。また、2つの結果を比較することで、照合方法の改善のインパクトを評価する。

家計調査のサンプル数は毎月約8,000世帯であり、期間全体で2,784,657レコードが利用可能である。これらのレコードは、各世帯は原則として6ヶ月調査されるため、約45万世帯程度になるよう照合されるはずである。また世帯ごとの複数レコードについて、それぞれ何回目の調査であるかも特定できる。

表2は、照合した全レコードが、新方式・林方式で何回目の調査としてカウントされているかをまとめた表である。この表の第1行および第1列の「合計」を見ると、新方式・林方式それぞれで1回目の調査を受けたとみなされている世帯の数が分かる。各世帯が少なくとも1回は調査されていることから、この合計が2,784,657レコードがそれぞれ何世帯のデー

タと認識されたか相当する。この表 2 によれば、新方式は504,613世帯であるのに対し林方式では548,697世帯であり、「調査世帯数」は新方式の方が少ない。これは、林方式には「別世帯を同一世帯とみなす」可能性もあったが、実際には「同一世帯を別世帯とみなす」ケースが多いことを示している。

表 2 新方式と林方式による調査回数の比較

		新方式による調査回数						合計
		1	2	3	4	5	6	
林方式による調査回数	1	503,714	33,373	4,773	2,794	2,244	1,799	548,697
	2	234	454,565	31,771	4,622	2,668	2,080	495,940
	3	187	227	438,642	30,532	4,434	2,458	476,480
	4	19	179	27	419,570	29,164	4,082	453,041
	5	17	20	20	27	403,867	26,817	430,768
	6	30	21	18	20	26	377,020	377,135
	7	409	45	21	18	18	23	534
	8	0	389	41	21	16	17	484
	9	0	0	369	41	21	15	446
	10	0	0	0	354	39	20	413
	11	0	0	0	0	341	35	376
	12	0	0	0	0	0	321	321
	13	3	1	0	0	0	0	4
	14	0	3	1	0	0	0	4
	15	0	0	3	1	0	0	4
	16	0	0	0	3	1	0	4
	17	0	0	0	0	3	0	3
	18	0	0	0	0	0	3	3
	合計	504,613	488,823	475,686	458,003	442,842	414,690	2,784,657

出所) 家計調査・個票データより筆者作成

また、調査回数は各世帯で最大 6 回となるはずである。新方式では全ての世帯の調査月数は 6 以下となっているが、林方式では 6 回目以降の調査とされているレコードが5,800（世帯数にして534世帯）存在している。これは、異なる世帯を同一世帯とみなす誤りが発生していることを示唆している。年間収入も照合キーとして利用しているため、異なる世帯を照合する確率は十分に低い（全世帯の0.1%程度）が、照合キーに加えるだけでは誤って異なる世帯のレコードを照合する可能性を完全には排除できていない。さらに、調査の途中で世帯が交代すると、誤った世帯を照合したとしても調査月数が 6 ヶ月を超えない場合もあり、ここで見ている「誤った照合をしている可能性」は実際の誤りの下限に過ぎない。

さらに、本来照合すべきレコードが照合できなかったケースがどの程度存在しているかを見るために、新方式で6ヶ月間調査されたと判断された世帯が林方式で何回調査されたかをまとめたものが表3である。これによると、林方式の照合では9%にあたる世帯が6回未満しか調査できていない世帯とみなされており、本来は接続すべきレコードの照合に失敗している。¹⁵⁾

表3 林方式での調査回数の分布

	世帯数	構成比率 (%)	累積比率 (%)
1	27,799	6.7	6.7
2	3,338	0.8	7.51
3	2,291	0.55	8.06
4	2,007	0.48	8.54
5	1,781	0.43	8.97
6	376,592	90.81	99.79
6回以上	882	0.21	100
合 計	414,690		

出所) 家計調査・個票データより筆者作成

注) 新方式で6回調査された世帯のみを用いた分布。

林方式で照合に失敗した世帯の多くが、勤労者世帯から非勤労者世帯へ格付が変わった世帯、もしくはその逆である。言い換えれば、世帯主が引退・失業・就職・転職をした世帯がパネルデータから除外されてしまっていたのである。しかも、そうした世帯の割合は10%程度あり、無視できないレベルである。

5 ま と め

本稿では、先行研究での家計調査のレコードの照合手順の問題点を指摘し、その改善方法を論じた。先行研究の照合方法では、異なる世帯のレコードを同一世帯とみなす、本来接続すべきレコードの照合に失敗するといったケースが多くあった。それに対し、本研究で提示した照合方法を用いることでこれらの問題を回避することができる。

照合が失敗する原因のうち、世帯番号の変更については、経済活動と照合の失敗には関連があった。これは、照合の失敗によって実質的にサンプルの脱落とみなすことが、ある種のAttrition Biasを発生させることを意味する。しかも、そうしたケースが10%程度存在しており、無視できないインパクトを持っていた。

ここでの照合方法は、より適切なパネルデータとしての利用に不可欠である。今後の家計調査の個票データを利用した研究での活用に期待したい。

注

- 1) この論文を改訂し英語に訳したものが、Hayashi (1997) Ch. 5 である。
- 2) Kohara, Ohtake, and Saito (2002) では、年間収入を用いずに識別符号のみを用いて世帯照合をしている。
- 3) ここでは、統計法第33条による利用を想定している。そのため、ファイルの内容は、研究者の申請によって異なる。
- 4) 実際、1981年1月から2010年3月までの全ての月のデータにおいて、これらの符号が完全に重複したレコードは存在しなかったことは確認できた。
- 5) 市町村内のブロックは、基本的に国勢調査の調査区を組み合わせたものであり、世帯数が1,500から3,000未満になるように調整したもの。
- 6) 市町村コードは、「政府の統計窓口：e-stat」の「市町村名・コード」<http://www.e-stat.go.jp/SG1/hyoujun/initialize.do> で参照することができる。
- 7) 農林漁家は1999年7月から調査対象となったため、それ以前は世帯番号は最初の1桁目を除いた2桁である。
- 8) 実際に Kohara, Ohtake, Saito (2002) ではこの4変数だけで照合が行われている。
- 9) ただし、調査市町村の入れ替えがあった場合には、この限りではないため、実際の運用についての調査が必要である。
- 10) 家計調査の市町村番号は、通常地方公共団体コードが変更された翌暦年の1月から変更されており、必ずしも新たな市の誕生等と同時に変更されていない。
- 11) 家計調査では、県庁所在地および川崎市・北九州市以外の市町村は、原則5年調査される。町村部など十分な世帯を確保できない場合には、2年程度の場合もある。
- 12) ただし、この変更が行われても集計上の影響は無いため、パネルデータを構築しない場合には問題の無い変更である。
- 13) 市町村コードの変更履歴については、「政府の統計窓口：e-stat」の「市町村名・コード」<http://www.e-stat.go.jp/SG1/hyoujun/initialize.do> で参照することができる。
- 14) 一連世帯番号に記録ミスがあっても、家計調査公表データの結果には影響はない。またミスがあると考えられるレコードは総サンプル数約220万レコードのうち約15レコード（4世帯分）だけであり、無視できるレベルである。
- 15) 6ヶ月以上調査されたとみなされている世帯は、少なくとも照合されるべき6ヶ月に関しては照合ができていることから、照合の失敗はないと考えている。

参 考 文 献

- Hayashi, F. (1997) *Understanding Saving: Evidence from the United States and Japan*, MIT press: Cambridge.
- Hori, M. and S. Shimizutani (2005a) "Price Expectations and Consumption under Deflation: Evidence from Japanese Household Survey Data," *International Economics and Economic Policy*, vol. 2, pp. 127-151.

- Hori, M. and S. Shimizutani (2005b) "Consumer Response to the 1994 Tax Cut: Evaluating the Japanese First Tax Cut in the 1990s," *Hitotsubashi Journal of Economics*, vol. 46, pp. 85-97.
- Hori, M. and S. Shimizutani (2006) "Did Japanese Consumers Become More Prudent During 1998-1999? Evidence From Household Level Data," *International Economic Journal*, vol.20, pp.197-209.
- Kohara, M., F. Ohtake, and M. Saito (2002) "A Test of the Full Insurance Hypothesis: The Case of Japan," *Journal of the Japanese and International Economies*, vol.16, pp.335-352.
- Shimizutani S. (2006) "Consumer Response to the 1998 Tax Cut: Is a Temporary Tax Cut Effective?" *Journal of the Japanese and International Economies*, vol.20, pp.69-287.
- 清水谷諭・堀雅博 (2005) 「ボーナス制度と家計貯蓄率：サーベイデータによる再検証」『経済研究』第56巻, pp. 234-247.
- 林文夫 (1986) 「恒常所得仮説の拡張とその検証」『経済分析』vol. 101 pp. 1-23.
- 美添泰人・荒木万寿夫 (2006) 「1980年代以降における家計の消費と資産に関する実証研究」『総務省統計研修所リサーチペーパー』第5号.