

Discussion Paper Series

University of Tokyo
Institute of Social Science
Panel Survey

東京大学社会科学研究所 パネル調査プロジェクト
ディスカッションペーパーシリーズ

独立変数の変化の方向性を区別するモデルの
比較検討：
就業状況の変化と家事頻度を例に

Comparing Models that Differentiate the Directions of Changes in Variables:
An Exemplary Analysis of the Effects of Changes in Employment Status
on Housework

柳下実 (首都大学東京大学院・日本学術振興会)
不破麻紀子 (首都大学東京)

Minoru Yagishita

Makiko Fuwa

March 2020

No.116

独立変数の変化の方向性を区別するモデルの比較検討 ——就業状況の変化と家事頻度を例に——

柳下実（首都大学東京大学院・日本学術振興会）
不破麻紀子（首都大学東京）

要約

本稿では、前半部分で有田（2013）と Allison（2019）のコーディングを用いた固定効果モデルの推定結果を比較し、後半部分でモンテカルロシミュレーションを用いて、変化の方向性を区別する固定効果モデルの推定値の挙動を検討する。具体的には、前半部分では東京大学社会科学研究所が実施する働き方とライフスタイルの変化に関する全国調査（Japanese Life Course Panel Surveys, 以下 JLPS）を用いて、独立変数の変化の方向性を区別する固定効果モデルの例を示す。変化の方向性に着目したパネル分析アプローチは有田（2013）が一階差分モデルを利用し、推定できることを示した。しかし、一階差分モデルでは調査年が毎年でない、あるいは等間隔でない項目の分析が難しい。近年、固定効果モデルでもコーディングによって独立変数の変化の方向性を区別して推定できることが示された（Allison 2019）。本稿では JLPS で毎年調査される就業形態の変化が、wave 1, 3, 5, 6, 7, 8 で調査された家事頻度に与える影響を例に、有田と Allison のコーディングを用いた固定効果モデルで推定した。結果から、有田と Allison のコーディングを用いた固定効果モデルはほぼ同様の結果を示した。後半部分ではサンプルの限定やクラスタ頑健標準誤差の使用などによって、変化の方向性を区別する固定効果モデルの推定値に差が生じるのかを、モンテカルロシミュレーションを用いて検討した。結果から、独立変数はすべての wave で観測されるが、従属変数は一部の wave でのみ観測されるデータでも適切に推定できることが示された。また推定の際、調査への復帰を含まないサンプルに限定し、クラスタ頑健標準誤差を用いることが推奨される結果が得られた。JLPS には調査年が毎年でない、あるいは等間隔でない項目が含まれるため、JLPS を分析する際には変化の方向性を区別する固定効果モデルも有用であると考えられる。

謝辞

本研究は、日本学術振興会（JSPS）科学研究費補助金・特別推進研究（JP25000001, JP18H05204）、基盤研究（S）（JP18103003, JP22223005）、特別研究員奨励費（JP18J11821）、科研費基盤 C（JP19K02044）の助成を受けたものである。東京大学社会科学研究所（東大社研）パネル調査の実施にあたっては、社会科学研究所研究資金、株式会社アウトソーシングからの奨学寄付金を受けた。パネル調査データの使用にあたっては東大社研パネル運営委員会の許可を受けた。

1. 問題関心

本稿の前半部分では、独立変数の変化の方向や経路、状態の影響を検討する2つのモデルを比較検討し、後半部分では、これらの手法を用いてパネルデータを分析する際の実際の課題について、モンテカルロシミュレーションを用いて検討した。

具体的には、前半部分では固定効果モデルを用いて有田（2010, 2013）のコーディングと Allison（2019）の提唱したコーディングによる分析をおこなう。有田（2010, 2013）による通常の固定効果モデルの想定への批判により、固定効果モデルの想定が社会学者の関心とは必ずしも一致しないことが示された。有田（2013）によって批判されたのは、通常の固定効果モデルが持つ「変化の向きが異なってもその効果の程度（効果の絶対値）は等しい」、「変化とその効果は任意に分割／合併が可能である（経路に関わらず効果は一定である）」、「非変化時の状態が異なってもその効果は等しい」（有田 2013: 72）という想定である。しかし、就業や婚姻状況など社会経済的属性の変化等に着目する社会学的な分析では、これらの想定が当てはまらないことも予想される。例えば、従来の固定効果モデルでは正規雇用から非正規雇用へ移行することが従属変数に及ぼす影響は、非正規雇用から正規雇用への移行がもたらす影響と同じ（効果の絶対値は同じ）と想定される。しかし二重労働市場構造の日本的雇用システムのもとでは、正規雇用から非正規雇用への移行が内部市場から外部市場への移行と重なることが多いのに対し、外部市場の非正規雇用から正規雇用への移行では、正規職であっても職業の条件や機会が限定的であることも考えられる。このため、移行の向きによって従属変数に及ぼす影響も異なることが予想される。また、例えば同一の就業状態を継続する場合でも正規雇用を維持した場合と非正規雇用を維持した場合では、就業がもたらす影響が異なることも考えられる。有田（2010, 2013）は変化についてのコーディングと一階差分モデルとを用いて、これらの想定を緩める方法を示した。後述するように、有田のコーディングは、変化の向き、経路、非変化時の状態を区別し、それぞれに対応するパラメータを割り当てるものである。

しかし残念なことに、有田（2013）によって示された方法は広く利用されているとはいいがたい¹。現実のパネル調査では一階差分モデルの利用が難しいケースが多いことが要因の一つとして考えられる。第一に、一階差分モデルでは $t-1$ 時と t 時の差分をとるため、独立変数はすべてのウェーブで観測されるが、従属変数は一部のウェーブでのみ観測される場合に独立変数が従属変数に与える影響の分析が難しい。たとえば全部で3ウェーブの調査データがあり、独立変数の就業形態は全ウェーブで、従属変数の家事頻度は wave 1 と wave 3 のみで観測されているとする。就業形態が wave 1 で非就業、wave 2 で非正規、wave 3 で

¹ 竹内（2016）は、一階差分モデルを用いて検討をおこなっている。

非就業となるケースでは、一階差分モデルは従属変数が観測されている wave 1 と wave 3 の差分を取るため、wave 2 の情報をなかったものとして扱うこととなる²。第二に現実のパネル調査では、たとえば JLPS の家事労働に関する設問のように wave 1, 3, 5, 6, 7, 8 のように不均等間隔で観測される項目がある。このような場合に得られたデータのすべてを用いて分析しようとする一階差分モデルでは一年の差分と二年の差分とを含めて分析することになり、解釈が難しい。

そうした背景の中、近年 Allison (2019) はコーディングを工夫することで、固定効果モデルを用いても、変化の方向を区別した推定が可能であることを示した (Allison 2019)。具体的には、就業形態を例にとると、正規から非正規などの移行を捉えるダミー変数を作成し、さらにその変数の観測期間中の蓄積を捉える変数を作成し、モデルに含めるという方法である。Allison (2019) のコーディングによる固定効果モデルの利点としては、従属変数が観測されていないウェーブであってもそのウェーブの独立変数の情報を分析に利用できることがあり、またこのようなケースでも比較的シンプルなモデルで検討できることがあげられる。本稿では、JLPS データを用いて、Allison (2019) のコーディングによる固定効果モデルを検討する。さらに、有田 (2010, 2013) の変化についてのコーディングを用いて固定効果モデルで分析し、Allison (2019) のコーディングから得られた結果と比較する。

しかし、これらの方法を用いたパネル分析に進む前に、事前に検討されるべき点が 2 点ある。第一に、独立変数はすべてのウェーブで観測されるが、従属変数が一部のウェーブでしか観測されない場合に、推定値にはどのようなバイアスが生じるのか。第二に、パネル調査でしばしばみられる調査からの脱落や復帰は推定値にどのような影響をあたえるのか。そこで、本稿の後半部分では変化の方向性を区別する固定効果モデルを用いる際の実際の課題を、モンテカルロシミュレーションを用いて検討する。

本稿は、第一に働き方とライフスタイルの変化に関する全国調査 (Japanese Life Course Panel Surveys, 以下 JLPS) を用い、就業形態の変化が家事頻度にあたえる影響を例に、Allison (2019) のコーディングを利用することで、変化の向き、経路、非変化時の状態を区別するパラメータを設定する有田 (2013) のモデリングと同様の分析が固定効果モデルでも可能であることを示す。第二に、一部の wave でのみ従属変数が観測されるパネルデータやサンプルに脱落や復帰が生じるパネルデータなど、条件を操作したモンテカルロシミュレーションから得た推定値を比較する。一部のウェーブで従属変数が観測されたり、観測年が規則的でない、また独立変数に欠損が生じるといった現実のパネル調査データに近い条件下で、変化の方向性を区別する固定効果モデルを用いる際にどのような分析戦略が適切かを探り、変化

² 複雑なコーディングによる対処は考えうるが、変化の影響を捉える方法をより容易に利用できる方法を探るといって本稿の目的にはそぐわない。

の方向性が異なることを許容するモデリングの、従来よりも幅広い変数・データに対する利用可能性を探る。

2. 有田(2013)の方法と Allison(2019)の方法——就業形態を例に

本節ではまず、有田と Allison (2019) の方法を説明する。日本では変化のモデリングについて有田 (2013) の先駆的な業績がある。両者の議論で興味深いのは、有田が「ある時点の従属変数の水準を説明するためには、必ず別の変数の履歴を考慮しなければならない」(有田 2013: 80) と履歴の影響に言及するのと同様、Allison も“ Y_{it} depends on the entire previous history of changes in X ” (Allison 2019, 8) と指摘する点である。本稿では、JLPS の就業形態 (5 カテゴリー) を例に両者の方法を説明する³。本稿の例では非就業→非就業を基準カテゴリとしてコーディングした。表 1 に通常の固定効果モデル・有田・Allison のパラメータ設定を示した⁴。表 1 の (1) の表は通常の固定効果モデル (区別しないモデル) であり、 β は通常のダミー変数の係数を指す。表 1 (2) 有田のモデルの (a) は変化の向きを区別したモデル・(b) は向きに加え変化の経路を区別したモデル・(c) はさらに非変化時の状態を区別したモデルであり、表 1 (3) は Allison のモデルである。

有田 (2010, 2013) のモデル。有田の方法では変化の向き、経路、非変化時の状態を区別するためにそれぞれに対応するパラメータを割り当てる。表 1 (2) (a) 変化の向きを区別したモデルの表には、(1) 区別しないモデル (通常の固定効果モデル) で示した β に加え、変化の向きを区別するパラメータ (γ) が示される。(2) (b) さらに変化の経路を区別したモデルの表は変化の向きを区別するのに加え、変化の経路を区別するパラメータ (ζ) を示す。(2) (c) さらに非変化時の状態を区別したモデルの表は、これらに加え、非変化時の状態を区別するパラメータ (η) を示す⁵。

有田 (2013) による議論のポイントの一つは、「モデルはすべて互いに入れ籠 (nested) の関係にある」(有田 2013: 75) ことである。すなわち β からの乖離を捉える γ , ζ , η についてそれぞれに含まれる係数が同時に 0 であるかどうかを検定することで、通常の固定効果モデルのどの想定を緩和する必要がある、どの想定は緩和する必要がないのかを示せ

³ Allison (2019) は二値・連続変数について検討したが、複数のカテゴリを持つ変数については検討していない。有田の方法では複数カテゴリの変数を念頭に置くため本稿では 5 カテゴリの就業形態を例に用いる。

⁴ 表 1 は有田 (2013) の表 2 を参考にした。有田 (2013) と異なる点は、カテゴリが 3 カテゴリ (正規・非正規・自営) ではなく 5 カテゴリ (非就業・正規・非正規・自営・学生) であること、また今回の分析に合わせ基準カテゴリを非就業とした点である。それ以外は有田 (2013) のパラメータ設定を踏襲した。

⁵ Allison (2019) では非変化時の状態を区別するパラメータは議論されていない。

る。本稿の分析では有田のモデルは β が4個、 γ が4個、 ζ が12個、 η が4個、計24個のパラメータが必要である。次に述べるAllisonのモデルも α が24個必要であり、パラメータ数は同じである。有田(2010, 2013)は、こうしたパラメータ設定を用いて変化の向き、経路、非変化時の状態を区別した一階差分モデルによる推定をおこなった。有田も指摘するように、変化の方向性に着目する分析では「水準で水準を説明する」のではなく、「変化で変化を説明する」という異なる図式が想定される(有田 2013: 82)。

Allison (2019) のモデル。表1の(3)がAllison (2019) のパラメータ設定である。 β と区別するため α を用いる。Allison (2019) の方法では、まず例えば、正規から非正規などの移行を捉える変数を作成し、それをもとに移行の蓄積を捉える変数を作成し固定効果モデルに投入することで、一階差分モデルと同じように変化の向きによって係数が異なる推定を可能とする。有田(2013)も述べているように、変化の影響を分析するためには「対象時点間で該当する移動パターンを経験したケースを1、それ以外を0とするダミー変数」(有田 2013: 75)が必要であり、これはAllisonの方法でも同様である。そのため5カテゴリを持つ変数では、基準となる変数(非就業→非就業)を除き移動パターンを捉えるダミー変数は24個となる。使用するJLPS wave5-8では学生から自営への移動と自営から学生への移動は観測されなかったため、これらの変数はモデルに含められない。そのため今回の分析では作成するダミー変数は22個となる。Allisonの方法ではこれら22の移行を捉える変数を作成した後で、移行の蓄積を捉える変数を作成し、移行の蓄積を捉えた変数を固定効果モデルに投入する。移行の蓄積を捉える変数を作成するとは、たとえば正規→正規→非就業→非正規→非就業→非正規という移行が観測されたとすると、移行自体は隣接するペアでカウントされるが(正規→正規, 正規→非就業, ……), 移行の蓄積変数においては正規から正規への移行は1回、正規から非就業への移行が1回、非就業から非正規への移行が2回、非正規から非就業への移行が1回とカウントされるということである。

有田(2013)とAllison(2019)の方法は異なるが、目的は同じであり、互換的であると考えられる。上述したように、Allisonの方法は各セルにパラメータを割り振っており、有田のように固定効果モデルの3つの想定を緩めるためのパラメータ設定はしていない。変化の蓄積を捉える変数を用いることで、固定効果モデルでも有田の方法のパラメータ設定(γ , ζ , η)が可能であると考えられ、実際の分析において互換性を確認する。

表1. 通常の固定効果モデル・有田・Allisonのパラメータ設定

(1) 通常の固定効果モデル (区別しないモデル)

	非就業	正規	非正規	自営	学生
非就業	0	β_1	β_2	β_3	β_4
正規	$-\beta_1$	0	$-\beta_1+\beta_2$	$-\beta_1+\beta_3$	$-\beta_1+\beta_4$
非正規	$-\beta_2$	$-\beta_2+\beta_1$	0	$-\beta_2+\beta_3$	$-\beta_2+\beta_4$
自営	$-\beta_3$	$-\beta_3+\beta_1$	$-\beta_3+\beta_2$	0	$-\beta_3+\beta_4$
学生	$-\beta_4$	$-\beta_4+\beta_1$	$-\beta_4+\beta_2$	$-\beta_4+\beta_3$	0

(2) 有田のモデル

(a) 変化の向きを区別したモデル

	非就業	正規	非正規	自営	学生
非就業	0	$\beta_1+\gamma_1$	$\beta_2+\gamma_2$	$\beta_3+\gamma_3$	$\beta_4+\gamma_4$
正規	$-\beta_1$	0	$-\beta_1+\beta_2+\gamma_2$	$-\beta_1+\beta_3+\gamma_3$	$-\beta_1+\beta_4+\gamma_4$
非正規	$-\beta_2$	$-\beta_2+\beta_1+\gamma_1$	0	$-\beta_2+\beta_3+\gamma_3$	$-\beta_2+\beta_4+\gamma_4$
自営	$-\beta_3$	$-\beta_3+\beta_1+\gamma_1$	$-\beta_3+\beta_2+\gamma_2$	0	$-\beta_3+\beta_4+\gamma_4$
学生	$-\beta_4$	$-\beta_4+\beta_1+\gamma_1$	$-\beta_4+\beta_2+\gamma_2$	$-\beta_4+\beta_3+\gamma_3$	0

(b) さらに変化の経路を区別したモデル

	非就業	正規	非正規	自営	学生
非就業	0	$\beta_1+\gamma_1$	$\beta_2+\gamma_2$	$\beta_3+\gamma_3$	$\beta_4+\gamma_4$
正規	$-\beta_1$	0	$-\beta_1+\beta_2+\gamma_2+\zeta_1$	$-\beta_1+\beta_3+\gamma_3+\zeta_3$	$-\beta_1+\beta_4+\gamma_4+\zeta_5$
非正規	$-\beta_2$	$-\beta_2+\beta_1+\gamma_1+\zeta_2$	0	$-\beta_2+\beta_3+\gamma_3+\zeta_6$	$-\beta_2+\beta_4+\gamma_4+\zeta_9$
自営	$-\beta_3$	$-\beta_3+\beta_1+\gamma_1+\zeta_4$	$-\beta_3+\beta_2+\gamma_2+\zeta_7$	0	$-\beta_3+\beta_4+\gamma_4+\zeta_{11}$
学生	$-\beta_4$	$-\beta_4+\beta_1+\gamma_1+\zeta_8$	$-\beta_4+\beta_2+\gamma_2+\zeta_{10}$	$-\beta_4+\beta_3+\gamma_3+\zeta_{12}$	0

(c) さらに非変化時の状態を区別したモデル

	非就業	正規	非正規	自営	学生
非就業	0	$\beta_1+\gamma_1$	$\beta_2+\gamma_2$	$\beta_3+\gamma_3$	$\beta_4+\gamma_4$
正規	$-\beta_1$	η_1	$-\beta_1+\beta_2+\gamma_2+\zeta_1$	$-\beta_1+\beta_3+\gamma_3+\zeta_3$	$-\beta_1+\beta_4+\gamma_4+\zeta_5$
非正規	$-\beta_2$	$-\beta_2+\beta_1+\gamma_1+\zeta_2$	η_2	$-\beta_2+\beta_3+\gamma_3+\zeta_6$	$-\beta_2+\beta_4+\gamma_4+\zeta_9$
自営	$-\beta_3$	$-\beta_3+\beta_1+\gamma_1+\zeta_4$	$-\beta_3+\beta_2+\gamma_2+\zeta_7$	η_3	$-\beta_3+\beta_4+\gamma_4+\zeta_{11}$
学生	$-\beta_4$	$-\beta_4+\beta_1+\gamma_1+\zeta_8$	$-\beta_4+\beta_2+\gamma_2+\zeta_{10}$	$-\beta_4+\beta_3+\gamma_3+\zeta_{12}$	η_4

(3) Allisonのモデル

	非就業	正規	非正規	自営	学生
非就業	0	α_1	α_3	α_6	α_{10}
正規	α_2	α_4	α_7	α_{11}	α_{15}
非正規	α_5	α_8	α_{12}	α_{16}	α_{19}
自営	α_9	α_{13}	α_{17}	α_{20}	α_{22}
学生	α_{14}	α_{18}	α_{21}	α_{23}	α_{24}

注) 有田 (2013) の表2をもとに今回の分析に合わせ書き換えた.

一階差分モデルではなく固定効果モデルを用いることによって異なる点もあり、推定できる係数に差が生じる。一階差分モデルでは移行を捉えるダミー変数間で共線性が生じるのに対して、固定効果モデルでは移行の蓄積を捉える変数の一つと時間 (wave) との間に共線性が生じる。つまり、一階差分モデルでは共線性の問題からすべての移行について係数を計算することはできず、基準カテゴリを推定の際に除外する必要がある。しかし、固定効果モデルでは移行の蓄積を捉えた変数間では共線性の問題が生じないため、すべての移行について係数を計算できる。しかし移行の蓄積を捉える変数の少なくとも一つは時間と共線的な関係にあるため、移行の蓄積を捉える変数をすべて含めると時間を連続変数で投入した場合は時間の係数が推定できず、時間をダミー変数で投入した場合は時間について推定できる係数が1つ減る。

なお、Allison (2019) において Allison のコーディングによる固定効果モデルと一階差分モデルの同等性が示されているが、本稿では補完的な分析として、JPLS データを用いて Allison (2019) のコーディングを用いた固定効果モデルと一階差分モデル (unstructured, toeplitz1) の比較もおこなった (結果は図示していない)。分析では、従属変数・独立変数ともに等間隔で観測されているケースを想定した。結果によると、Allison のコーディングによる固定効果モデルと一階差分モデルから同様の結果が得られることが示された。

3. データと方法

有田と Allison のモデルの比較に使用するデータは東京大学社会科学研究所が 2007 年から実施する働き方とライフスタイルの変化に関する全国調査 (JLPS) w5-8 である。対象者は 2006 年 12 月末に 20 歳から 40 歳であった。本稿の従属変数は家事頻度である。家事頻度は食事の用意、洗濯、家の掃除、日用品・食料品の買い物の 4 項目について、毎日に 7、週に 5~6 日に 5.5、週に 3~4 日に 3.5、週に 1~2 日に 1.5、月に 1~3 日に 0.5、ほとんどしないに 0 を割り当て、合算した変数を用いる。家事頻度についてクロンバックのアルファ係数は家事頻度が含まれる wave で 0.8 以上であった。

独立変数は非就業を基準とする就業形態である。就業形態の履歴が家事頻度に与える影響を検討するため、独立変数を計算する際に使用するデータは就業形態がすべてのウェーブで観測されたケースに限定する⁶。今回は、分析手法の例のため統制変数は含めず分析する。Allison (2019) のコーディングでは、就業形態 (5 カテゴリ) × 就業形態 (5 カテゴリ)

⁶ 一度でも観測されないウェーブがあるケースは分析に含まれない。利用するウェーブが長くなればなるほど、一度観測されなくなったがその後観測されるようになったケース (復活したケース) が分析に含まれなくなるため、そのバイアスには注意が必要である。

の 25 セルある遷移行列について、基準カテゴリである非就業継続（非就業→非就業）を除いた 24 セルすべてに別々のパラメータを割り振る。Allison (2019) では online appendix として R, SAS, Stata でのコーディングを公開している。また本稿の巻末の Appendix にも一部、分析に利用した Stata を用いたコーディングを示した。正規から非就業、非就業から正規への移行を例に説明する。まず、正規から非就業の移行を捉える変数 (regnlf) を作成し、就業形態が観測されたケースに 0 を代入する。次に、一つ前の wave と現在の wave の値を比較し、一つ前の wave で正規で、現在の wave で非就業のケースには 1 を代入する。非就業から正規への移行を捉える変数 (nlfreg) も同様に作成する。そのうえで、これらの変数を最初から最後の wave まで足しあげ、移行の蓄積を捉える変数 (regnlfcum, nlfregcum) を作成する。これらの変数を 24 セルすべてについて作成する。

有田 (2010, 2013) の方法では、無職を基準とした通常の正規ダミー、非正規ダミー、自営ダミー、学生ダミー (β) をモデルに投入したうえで、表 1 に示したそれぞれの移行について、向き・経路・非変化の状態によって係数の大きさが異なるのかどうかを示すダミー変数群 (γ , ζ , η) を投入する。詳述すると以下のようなになる。基準となる就業形態ダミー⁷ (β_1 : 正規の際に 1・それ以外で 0 を取る, β_2 : 非正規の際に 1・それ以外で 0 を取る, β_3 : 自営の際に 1・それ以外で 0 を取る, β_4 : 学生の際に 1・それ以外で 0 を取る), 非就業から非正規など変化の向きの影響を示すダミー変数 (γ_1 : 非就業→正規, 非正規→正規, 自営→正規, 学生→正規の移行が観測された際に 1 を取る, γ_2 : 非就業→非正規, 正規→非正規, 自営→非正規, 学生→非正規の移行が観測された際に 1 を取る, γ_3 : 非就業→自営, 正規→自営, 非正規→自営, 学生→自営の移行が観測された際に 1 を取る, γ_4 : 非就業→学生, 正規→学生, 非正規→学生, 自営→学生の移行が観測された際に 1 を取る), 変化の経路の効果を示すダミー変数 (ζ_1 : 正規→非正規の移行が観測された際に 1 を取る, ζ_2 : 非正規→正規の移行が観測された際に 1 を取る, ζ_3 : 正規→自営の移行が観測された際に 1 を取る, ζ_4 : 自営→正規の移行が観測された際に 1 を取る, ζ_5 : 正規→学生の移行が観測された際に 1 を取る, ζ_6 : 非正規→自営の移行が観測された際に 1 を取る, ζ_7 : 自営→非正規の移行が観測された際に 1 を取る, ζ_8 : 学生→正規の移行が観測された際に 1 を取る, ζ_9 : 非正規→学生の移行が観測された際に 1 を取る, ζ_{10} : 学生→非正規の移行が観測された際に 1 を取る, ζ_{11} : 自営→学生の移行が観測された際に 1 を取る, ζ_{12} : 学生→自営の移行が観測された際に 1 を取る), 非変化時の状態の効果を示すダミー変数 (η_1 : 正規→正規の移行が観測された際に 1 を取る, η_2 : 非正規→非正規の移行が観測された際に 1 を取る, η_3 : 自営→自営の移行が観測された際に 1 を取る, η_4 : 学生→学生の移行が観測された際に 1 を取る) である。有田の方法を固定効果モデルで再現するためには、 γ ,

⁷ 有田 (2013) でも述べられているように、 β は通常のダミー変数である。

と、 η について蓄積変数を作成する。

これらのコーディングは交互作用項と同様であり，そのように考えると理解しやすい (York & Light 2017)．有田 (2013) が「それぞれの仮定緩和によって許容される効果の相違分を表すパラメータ」(有田 2013: 74) と述べたように，有田 (2013) の方法は通常の固定効果モデルが置く想定を緩めるため，それとの乖離を許すパラメータ，いわゆる交互作用項を導入する (γ , ζ , η)．例えば，一般に教育年数が男女の賃金に与える影響を分析する際は，教育年数および教育年数と女性ダミーの交互作用項を導入して推定する．この際，交互作用項は，男性が教育年数から受けるリターンと女性が教育年数から受けるリターンとの乖離度合いを示す．これらと同様に， γ , ζ , η は通常の固定効果モデルの想定を緩めた際の効果の乖離分を表すという点で，交互作用項と同様である。

本稿では有田 (2013) と Allison (2019) のコーディングに従った固定効果モデルを比較するため，JLPS の wave 5, 6, 7, 8 を用いて推定する．分析については有田 (2013) の表 3 や Appendix の Stata コード，および Cox (2002) も参照のこと。

4. 有田(2013)の方法と Allison(2019)の方法の比較結果

本節では通常の固定効果モデル，および有田と Allison のコーディングによる結果を示す．表 2 に通常の固定効果モデル，有田のコーディング，Allison のコーディングによる推定結果を示した．表 2 (1) の区別しないモデルは，表 1 の (1) に相当し，有田のコーディングによる結果は (2) 区別するモデルの (c) に相当する変化の向き・経路・非変化時の状態を区別したモデル (有田のモデル) を示した⁸．右側には Allison のコーディングによる，移行をすべて区別して推定した結果を示した。

まず，固定効果モデルのどの想定を緩和する必要があるのかを検定した．結果から， γ に含まれる変数の係数がすべて 0 であるという仮説は 1%水準で棄却される ($F(4, 3368) = 4.25$, $Prob > F = 0.0020$)．また ζ に含まれる変数の係数がすべて 0 であるという仮説は 0.1%水準で棄却される ($F(10, 3368) = 4.59$, $Prob > F = 0.00$)．さらに， η に含まれる変数の係数がすべて 0 であるという仮説は 0.1%水準で棄却される ($F(4, 3368) = 6.81$, $Prob > F = 0.00$)．これらの結果から，今回の分析においては通常の固定効果モデルの 3 つの仮定のすべてを緩和するモデルが支持される。

結果から，区別なしのモデルで非就業と正規を比べると，正規では家事頻度が 3.8 日分低い．これは通常の固定効果モデルから得られた推定値であるため，水準が水準を規定す

⁸ (2) - (c) では学生から自営や自営から学生といった移動が観測されないため，パラメータを示していない。

るモデルである。このモデルでは変化の向きを区別しないので、非就業→正規の移行では家事頻度が3.8日分減り、正規→非就業の移行では家事頻度が3.8日分増えるということになる。

一方で有田のコーディングによる(c)向き・経路・非変化区別で推定された係数によると、正規→非就業の係数は $-\beta_1$ から、6.375であるのに対し、非就業→正規の係数は $\beta_1 + \gamma_1$ 、つまり $-6.375 + 2.626 = -3.749$ である。すなわち、非就業から正規への移行による家事頻度への効果の大きさは、正規から非就業への移行による効果と異なり、通常の固定効果モデルの想定は支持されていないことが示唆される。

次に、有田のモデルの向き・経路・非変化区別で推定された係数とAllisonのコーディングによって推定された係数が等しいことを確認する。正規→非就業はAllisonのコーディングでは6.375である。上述のように有田のコーディングでは6.375であり、一致する。次に、 γ を含む係数として非就業→正規をみる。Allisonのコーディングでは係数が-3.750である。有田のコーディングでは非就業→正規の係数は-3.749($-6.375 + 2.626$)であり、丸め誤差があるものの一致する⁹。次に ζ を含む係数をみると、たとえば正規から非正規の移動は $-\beta_1 + \beta_2 + \gamma_2 + \zeta_1$ 、つまり $-(-6.375) + -2.131 + -.047 + -2.952 = 1.245$ であるが、表2のAllisonのコーディングによる推定値は1.246であり、丸め誤差によって0.001の誤差が生じたが一致する。また、自営から非正規の移動は $-\beta_3 + \beta_2 + \gamma_2 + \zeta_7$ 、つまり $-(-.355) - 2.131 - .047 + .639 = -1.184$ であり、表2のAllisonのコーディングによる推定値と一致する。他の推定値についてもほぼ一致する結果が得られた。

⁹ Stataの内部でlincomコマンドを用いて計算すると、少なくとも小数点第三位までは結果が一致する。

表2. 有田コーディングとAllisonコーディングの結果 (wave 5, 6, 7, 8)

	(1) 区別なし	(2) 区別あり	Allisonのコーディング	
	(c) 向き・経路 非変化区別			
β 1: 正規	-3.768 *** (.339)	-6.375 *** (.669)	正規→非就業	6.375 *** (.669)
2: 非正規	-2.483 *** (.211)	-2.131 *** (.334)	非正規→非就業	2.131 *** (.334)
3: 自営	-1.866 *** (.346)	-.355 (.793)	自営→非就業	.355 (.793)
4: 学生	-2.001 ** (.714)	-1.017 (1.380)	学生→非就業	1.017 (1.380)
γ 1: 非就業→正規		2.626 ** (.899)	非就業→正規	-3.750 *** (.773)
2: 非就業→非正規		-.047 (.400)	非就業→非正規	-2.178 *** (.262)
3: 非就業→自営		.009 (1.010)	非就業→自営	-.346 (.584)
4: 非就業→学生		1.272 (1.880)	非就業→学生	.255 (1.730)
ζ 1: 正規→非正規		-2.952 ** (.943)	正規→非正規	1.246 † (.636)
2: 非正規→正規		1.051 (.930)	非正規→正規	-.568 (.422)
3: 正規→自営		-5.971 *** (.986)	正規→自営	.059 (.441)
4: 自営→正規		2.591 * (1.170)	自営→正規	-.804 † (.466)
5: 正規→学生		-2.352 (3.530)	正規→学生	4.279 (3.000)
6: 非正規→自営		-.922 (.964)	非正規→自営	.864 (.714)
7: 自営→非正規		.639 (1.030)	自営→非正規	-1.184 † (.629)
8: 学生→正規		2.784 (1.960)	学生→正規	.052 (1.010)
9: 非正規→学生		-1.116 (2.250)	非正規→学生	1.271 (1.510)
10: 学生→非正規		-.222 (1.750)	学生→非正規	-1.383 (1.110)
η 1: 正規→正規		-.067 (.078)	正規→正規	-.067 (.078)
2: 非正規→非正規		-.335 *** (.089)	非正規→非正規	-.335 *** (.078)
3: 自営→自営		-.425 *** (.129)	自営→自営	-.425 *** (.129)
4: 学生→学生		-.815 † (.492)	学生→学生	-.815 † (.492)

注) 観測数13,380, 個人3,369. 括弧内はクラスター頑健標準誤差である.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .1$

5. 固定効果モデルを用いた変化の推定において実用上解決すべき点の検討

本稿の後半部分では、パネル調査データを用いて独立変数のカテゴリ間の移行などの影響を、変化の方向性を区別する固定効果モデルで推定する際に、しばしば直面する2つの実際上の課題について、モンテカルロシミュレーションを用いて検討する。モンテカルロシミュレーションとは計量経済学で利用される方法であり、データ生成過程と乱数に基づいて多数のデータを発生させ、分析し、得られたパラメータの平均値と真のパラメータとの乖離を検討することで推定値の性質を検討する方法である (Cameron & Trivedi 2010)。

第一の問題は、独立変数はすべてのウェーブで観察されるが、従属変数は一部のウェーブでしか観測されないというケースがあるが、その条件下でも推定に問題が生じないのか。第二に、パネル調査では脱落や脱落後の調査への復帰がみられるが、復帰ケースを推定に含めることが推定へどのような影響をもたらすのか、である。

第二の脱落と復帰の問題は有田 (2013) が指摘したように、これらのモデルでは移行を捉えるために分析に利用できるケースが制約されることと関連する。たとえば有田 (2013) は、すべてのウェーブで独立変数が観測されたケースに分析を限定した。一般にパネル調査ではあるウェーブでは回答が得られなかったものの、その後回答が得られるケースが存在する。これは調査からの脱落、その後の復帰であるが、復帰したケースは脱落していた時点の独立変数の状態が不明なため、復帰したウェーブでは移行の分類ができず、基準カテゴリに分類される。一般には脱落からの復帰を推定に含めると、本来ならば何らかの移行が確認されていたケースが、誤って基準カテゴリに分類される。そのため、バイアスが生じると考えられる。

モンテカルロシミュレーションでは、第一の問題が推定に与える影響を評価するため、独立変数はすべてのウェーブで観測されたが、一部のウェーブでのみ従属変数が観測される条件と、すべてのウェーブで従属変数が観測される条件で得られた推定値を比較する。第二の問題が推定に与える影響を評価するため、脱落が生じないケースと脱落が生じるケース (復帰ケースを含める・含めない) などの条件によって、変化の向き・経路や非変化時の状態を区別する固定効果モデルの推定値に差が生じるのかを検討する。

5.1 モンテカルロシミュレーションの設定

モンテカルロシミュレーションでは、以下で説明するデータ生成過程に基づくデータを一万個生成し、各データを用いて固定効果モデルで分析し、それらの結果すべてをプールし得られた推定値の平均と真の値との乖離を検討する。データ生成過程は以下の通りである。簡便のため、独立変数は正規ダミー (基準: 非正規) とする。従属変数は家事頻度を想定して作成した。まず時不変の観察されない異質性 (u_i) を、平均 0、標準偏差 1 の標準正規分

布に従う乱数から作成する。次に、非正規ダミーの潜在変数を 0 から 1 を取る一様な乱数と時不変の観察されない異質性から作成する。この潜在変数について 60 パーセント以下を正規とする。使用するデータは非正規が 4 割、正規が 6 割となる。乱数を用いるため生成されたデータによって差があるが、非正規と正規の遷移行列では非正規→非正規が 84.4%、非正規→正規が 15.6%、正規→非正規が 10.4%、正規→正規が 89.6%である。

次に、この非正規ダミーから非正規→正規、正規→非正規、正規→正規、非正規→非正規の移行を識別する変数を作成し、さらに移行の蓄積を捉える変数を作成する。そのうえで、時不変の観察されない異質性と wave 1 時点の誤差として標準正規分布に従う誤差項 (e_{i1})、および wave 1 時点での正規・非正規間 (reg_{i1}) で 4.6 の差があると想定し、従属変数 (y_{i1}) を生成する (式 1)。なお y_{it} の i は個体を、 t は時点を示す。またパラメータは JLPS の事前分析から得た。

$$y_{i1} = -4.6 * reg_{i1} + e_{i1} + u_i \quad (1)$$

次に、wave 1 から wave 2 など各 wave の正規ダミーの移行から、以下のように wave 2 以降のデータを書き換える。正規→非正規では 1.6 の増加 (式 2)、非正規→正規では 1 の減少 (式 3)、正規継続では変化なし (式 4)、非正規継続で 0.13 の減少 (式 5) を想定する。さらにこれらの増減や変化なしには標準正規分布に従う誤差 (e_{it}) が含まれると想定する (Allison 2019: 8 も参照)。データ生成過程において、独立変数の移行が従属変数に与える影響は一定であると想定する¹⁰。

$$y_{it} = y_{it-1} + 1.6 \quad + e_{it} \quad (2)$$

$$y_{it} = y_{it-1} - 1 \quad + e_{it} \quad (3)$$

$$y_{it} = y_{it-1} + 0 \quad + e_{it} \quad (4)$$

$$y_{it} = y_{it-1} - 0.13 \quad + e_{it} \quad (5)$$

また、欠損値を含むデータを用いたモンテカルロシミュレーションでは正規ダミーと従属変数を作成した後に、正規ダミーに欠損を生じさせる。正規ダミーに欠損を生じさせるために 0 から 1 を取る一様な乱数を発生させ、一時点前が非正規の場合は 10%、一時点前が正規の場合は 5%のサンプルに欠損が生じるように設定した。これによって、全ウェーブをプールしたデータでは正規ダミーには 5%程度の欠損が生じ、5 波を通して正規ダミーが観

¹⁰ 有田の方法であれ、Allison の方法であれ、推定される係数は移行が従属変数に与える影響の平均である。

察されるケースが 75%, 4 波観察されるケースが 24%, 3 波観察されるケースが 1%程度となる。設定するサンプルサイズは 3,000, 観測数は 15,000 (T=5), 30,000 (T=10) とした。正規ダミーに欠損を許すデータセットでは 5%程度が欠損となるため, 有効なサンプルサイズは 3,000 より小さく, サンプルサイズは 2,250 (T=5), 1,589 (T=10), 観測数は 11,253 (T=5), 15,899 (T=10) 程度である。

モンテカルロシミュレーションで検討する条件群は観測の長さを 5 波と 10 波の二つを検討する。全国家族調査パネルスタディ (NFRJ-08Panel) のように 5 波が観察されたパネル調査や, JLPS のように観測が続いているパネル調査に当てはまるように 5 波と 10 波を選択した。次に, 脱落がないデータと脱落があるデータ (復帰ケースを含む推定と含まない推定), 通常の標準誤差とクラスタ頑健標準誤差の条件を操作する。実際に分析するパネル調査は脱落が生じるため, 脱落がない理想的なデータから得られた推定値と比較することで, どの条件下では適切な推定値が得られるのかを確認する。また, 有田 (2013) も推定では脱落が生じ, 独立変数の履歴がデータから復元できないケースを分析から除いた。復帰ケースを含めることで, サンプルサイズの増加につながるのか, それとも推定値に与えるバイアスが大きいのかは, 手法の利用者にとって有用な情報だと考える。また, 固定効果モデルであってもクラスタ頑健標準誤差の利用は計量経済学では推奨されるが (Cameron & Trivedi 2010), 社会学では言及が少ない (筒井ほか編 2016)。作成したデータの級内相関係数は 0.9 (T=5) と 0.85 (T=10) 程度であり, 個体間の分散が大きく, 個体内の分散は小さいデータである。分析には Stata 16.0 を使用した。

5.2 モンテカルロシミュレーションの結果

モンテカルロシミュレーションの結果を示す。表 3 が 5 波のデータと通常の標準誤差を用いたモンテカルロシミュレーションの結果である。脱落なしの条件では点推定のバイアスは小さい。標準誤差については, 一部のウェーブを用いた結果では標準誤差が大きい (正規→非正規では, 1.5 倍 ($0.065 \div 0.042$))。95%信頼区間のカバー率を見ると, 正規→非正規や非正規→正規ではカバー率が 9 割を超えたものの, 正規→正規や非正規→非正規では全ウェーブを用いても 8 割程度であり, カバー率が低い。これらの値は理想的な条件下で得られており, 点推定にはバイアスが小さいことから, 標準誤差に問題があることが示唆される。

脱落あり・復帰ケースを含む条件の結果では, 正規→非正規や非正規→正規といった移行の点推定にバイアスが生じる。真のパラメータは正規→非正規は正 (1.6), 非正規→正規は負 (-1) であるが, 生じたバイアスは正規→非正規で負, 非正規→正規で正であるため, 過小評価するバイアスが生じる。状態の維持を捉える変数に生じたバイアスは異なっており, 非正規→非正規の真のパラメータは負 (-0.13) であるが係数が負の方向に過大に推定される。バイアスの大きさは最大でそれぞれ, 正規→非正規で 7.1% ($0.113 \div 1.6$), 非正規

→正規で 1.4% ($0.014 \div 1$), 非正規→非正規で 7.7% ($0.010 \div 0.13$) である。点推定の標準誤差は脱落がない条件と比べ少し大きい。95%信頼区間のカバー率は, 正規→非正規や非正規→正規などで 6 割, 正規→正規や非正規→非正規で 8 割と低い。

最後に, 脱落あり・復帰ケースを含まない (一度でも脱落したケースは分析に含まない) 条件の結果では, 点推定のバイアスは復帰ケースを含む場合と比べ, 改善した。標準誤差は今まで検討した条件の中で一番大きい。信頼区間のカバー率も改善したが, 依然として 95%を下回る。

表3. 固定効果モデルを用いたモンテカルロシミュレーションの結果 (T = 5, 通常の標準誤差)

	正規→非正規	非正規→正規	正規→正規	非正規→非正規
脱落なし				
点推定のバイアス				
全ウェーブ	-.001	-.000	-.000	-.000
一部のウェーブ (wave 1, 3, 5)	-.001	-.000	-.000	-.000
点推定の標準誤差				
全ウェーブ	.042	.042	.008	.010
一部のウェーブ (wave 1, 3, 5)	.065	.065	.010	.013
95%信頼区間のカバー率				
全ウェーブ	.913	.912	.780	.777
一部のウェーブ (wave 1, 3, 5)	.941	.943	.889	.893
脱落あり・復帰ケースを含む				
点推定のバイアス				
全ウェーブ	-.075	.075	.000	-.009
一部のウェーブ (wave 1, 3, 5)	-.113	.114	.000	-.010
点推定の標準誤差				
全ウェーブ	.043	.045	.009	.012
一部のウェーブ (wave 1, 3, 5)	.067	.069	.011	.015
95%信頼区間のカバー率				
全ウェーブ	.572	.591	.779	.736
一部のウェーブ (wave 1, 3, 5)	.601	.611	.889	.849
脱落あり・復帰ケースを含まない				
点推定のバイアス				
全ウェーブ	.000	.000	-.000	.000
一部のウェーブ (wave 1, 3, 5)	.000	.000	-.000	.000
点推定の標準誤差				
全ウェーブ	.048	.049	.009	.013
一部のウェーブ (wave 1, 3, 5)	.076	.076	.012	.017
95%信頼区間のカバー率				
全ウェーブ	.914	.912	.771	.789
一部のウェーブ (wave 1, 3, 5)	.939	.941	.889	.896

注: 生成したデータは10,000である。

次に表4が10波のデータと通常の標準誤差を用いた結果である。脱落なしの条件では点推定のバイアスは小さい。標準誤差については、一部のウェーブを用いた結果では標準誤差が大きい（正規→非正規では、1.1倍（ $0.039 \div 0.035$ ））。95%信頼区間のカバー率を見ると、正規→非正規や非正規→正規ではカバー率が9割を超えたものの、正規→正規や非正規→非正規では7割程度であり、5波から得られた結果よりもカバー率が低い。

表4. 固定効果モデルを用いたモンテカルロシミュレーションの結果 (T = 10, 通常の標準誤差)

	正規→非正規	非正規→正規	正規→正規	非正規→非正規
脱落なし				
点推定のバイアス				
全ウェーブ	.001	.000	.000	.000
一部のウェーブ (wave 1, 3, 5-10)	.000	.000	.000	.000
点推定の標準誤差				
全ウェーブ	.035	.035	.005	.006
一部のウェーブ (wave 1, 3, 5-10)	.039	.039	.004	.006
95%信頼区間のカバー率				
全ウェーブ	.907	.907	.711	.672
一部のウェーブ (wave 1, 3, 5-10)	.918	.920	.664	.666
脱落あり・復帰ケースを含む				
点推定のバイアス				
全ウェーブ	-.193	.204	.000	-.016
一部のウェーブ (wave 1, 3, 5-10)	-.192	.201	.000	-.015
点推定の標準誤差				
全ウェーブ	.034	.035	.004	.006
一部のウェーブ (wave 1, 3, 5-10)	.039	.040	.005	.007
95%信頼区間のカバー率				
全ウェーブ	.005	.005	.604	.362
一部のウェーブ (wave 1, 3, 5-10)	.013	.010	.669	.413
脱落あり・復帰ケースを含まない				
点推定のバイアス				
全ウェーブ	-.001	.001	.000	.000
一部のウェーブ (wave 1, 3, 5-10)	-.001	.000	.000	.000
点推定の標準誤差				
全ウェーブ	.049	.049	.005	.008
一部のウェーブ (wave 1, 3, 5-10)	.055	.056	.006	.009
95%信頼区間のカバー率				
全ウェーブ	.906	.903	.594	.605
一部のウェーブ (wave 1, 3, 5-10)	.922	.919	.658	.667

注: 生成したデータは10,000である。

次に、脱落あり・復帰ケースを含む条件の結果では、正規→非正規や非正規→正規や非正

規→非正規の点推定にバイアスが生じており、5波のデータよりも深刻である。バイアスの大きさは最大で正規→非正規で 12.1% ($0.193 \div 1.6$)、非正規→正規で 20.4% ($0.204 \div 1$)、非正規→非正規で 12.3% ($0.016 \div 0.13$) である。点推定の標準誤差については脱落がない条件とほぼ変わらない。95%信頼区間のカバー率は正規→非正規や非正規→正規などで 1%程度、正規→正規や非正規→非正規で 4~7割と、低い。

最後に、脱落あり・復帰ケースを含まない（一度でも脱落したケースは分析に含めない）条件の結果では、点推定のバイアスは復帰ケースを含む場合と比べ改善した。標準誤差は 5波のデータと同様、正規→非正規や非正規→非正規で一番大きい。正規→非正規や非正規→正規の 95%信頼区間のカバー率も改善したが、正規→正規や非正規→非正規のカバー率は 5波の結果より低く 6割程度である。

表 5 がクラスタ頑健標準誤差に基づいた、固定効果モデルを用いたモンテカルロシミュレーションの結果である。クラスタ頑健標準誤差を用いることで個体内での残差の相関に対処することができる。今回の条件では標準誤差のみを変更した。そのため点推定のバイアスは表 3・4 と表 5 で同様の傾向を示しており、点推定の結果は省略した。標準誤差の結果を見ると、表 3・4 よりも表 5 で標準誤差の推定値が大きい。例えば、正規→非正規（10波・脱落なし・全ウェーブ）の標準誤差の推定値は 0.035 であるが、クラスタ頑健標準誤差を用いると 0.041 となり、1.17 倍 ($0.041 \div 0.035$) である。表 5 で注目したい点は 95%信頼区間のカバー率が顕著に改善した点である。表 5 では 5波のデータでも、点推定の値にバイアスがある脱落あり・復帰ケースを含む条件の場合以外では、最もカバー率が低いのが非正規→正規の一部のウェーブ（wave 1, 3, 5/脱落あり・復帰ケースを含まない）の 0.947 である。また、10波のデータについても点推定の値にバイアスが生じる脱落あり・復帰ケースを含む条件の場合以外では、ほぼ 95%となった。

これらの結果から、今回のデータ生成過程の下では、第一に独立変数がパネルデータのすべてで観測され、従属変数がパネルデータの一部でしか観測されないケースでも、従属変数が観測されたデータのみから得られた係数に生じるバイアスは小さい。第二に、脱落プロセスにも影響されるが、脱落が生じるパネルでは脱落から復帰したケースを推定に含めると係数が過小に推定するバイアスや過大に推定するバイアスが生じる。そのため、推定は復帰ケースを含めないほうがよいことが示唆された。第三に、級内相関が高い本稿のようなデータではクラスタ頑健標準誤差を用いなければ、正規→正規や非正規→非正規といった状態の持続を捉える変数の標準誤差に生じるバイアスは大きく、95%信頼区間が不安定になることが示された。変化の方向性を区別する固定効果モデルを用いる際は、分析サンプルに復帰ケースを含めない、かつクラスタ頑健標準誤差を用いることが推奨される。

表5. 固定効果モデルを用いたモンテカルロシミュレーションの結果 (T = 5およびT = 10, クラスター頑健標準誤差)

	正規→非正規	非正規→正規	正規→正規	非正規→非正規
T = 5				
脱落なし				
点推定の標準誤差				
全ウェーブ	.047	.047	.013	.017
一部のウェーブ (wave 1, 3, 5)	.067	.067	.013	.016
95%信頼区間のカバー率				
全ウェーブ	.949	.949	.949	.948
一部のウェーブ (wave 1, 3, 5)	.949	.949	.950	.950
脱落あり・復帰ケースを含む				
点推定の標準誤差				
全ウェーブ	.052	.053	.014	.019
一部のウェーブ (wave 1, 3, 5)	.071	.073	.014	.018
95%信頼区間のカバー率				
全ウェーブ	.693	.708	.949	.925
一部のウェーブ (wave 1, 3, 5)	.643	.657	.950	.917
脱落あり・復帰ケースを含まない				
点推定の標準誤差				
全ウェーブ	.055	.056	.014	.020
一部のウェーブ (wave 1, 3, 5)	.078	.079	.014	.020
95%信頼区間のカバー率				
全ウェーブ	.949	.949	.947	.951
一部のウェーブ (wave 1, 3, 5)	.948	.947	.947	.951
T=10				
脱落なし				
点推定の標準誤差				
全ウェーブ	.041	.041	.009	.012
一部のウェーブ (wave 1, 3, 5-10)	.044	.044	.009	.012
95%信頼区間のカバー率				
全ウェーブ	.951	.952	.951	.951
一部のウェーブ (wave 1, 3, 5-10)	.947	.949	.953	.951
脱落あり・復帰ケースを含む				
点推定の標準誤差				
全ウェーブ	.050	.051	.010	.014
一部のウェーブ (wave 1, 3, 5-10)	.052	.054	.010	.013
95%信頼区間のカバー率				
全ウェーブ	.027	.022	.953	.788
一部のウェーブ (wave 1, 3, 5-10)	.041	.037	.954	.789
脱落あり・復帰ケースを含まない				
点推定の標準誤差				
全ウェーブ	.057	.057	.011	.019
一部のウェーブ (wave 1, 3, 5-10)	.062	.062	.011	.018
95%信頼区間のカバー率				
全ウェーブ	.950	.947	.951	.949
一部のウェーブ (wave 1, 3, 5-10)	.952	.949	.951	.949

注: T = 5とT = 10のいずれにおいても, 生成したデータは10,000である.

6. 考察と結論

本稿ではJLPSのwave5-8データを用いて変化の方向性を区別する固定効果モデルにおいて、有田(2013)とAllison(2019)の方法が互換的であることを示し、またモンテカルロシミュレーションを用いてさまざまな条件を操作することで、固定効果モデルの推定値に差が生じるのかを検討した。変化の向き、経路、および非変化時の状態を区別するパラメータを設定するという有田の方法の優れた点は、Allisonの変化の蓄積を捉える変数を用いることで固定効果モデルでも利用できる。さらに、モンテカルロシミュレーションの結果から、独立変数がすべてのウェーブで観測され、従属変数が一部のウェーブでしか観測されない場合でも、一部のウェーブを用いた推定では標準誤差が大きい、係数の挙動は同様であった。そのため、独立変数は就業状態などですべてのウェーブで観測されるが、従属変数が一部のウェーブでのみ観測される場合に、今回示した変化の向き・経路と非変化時の状態を区別する方法を用いても、本稿のシミュレーションの想定が正しければ係数には大きな問題は生じないと考えられる。

また、移行の影響を捉える固定効果モデルを利用する際には、クラスタ頑健標準誤差の利用が強く推奨される結果が得られた。今回の結果では通常の標準誤差に基づく結果では標準誤差・95%信頼区間が適切に機能していない。もちろんクラスタ数が小さい場合やサンプルサイズが小さいなど、データによってクラスタ頑健標準誤差が適切かどうかは異なると考えられるが、今回のシミュレーションのサンプルサイズ(3000程度)と同等のデータを用いた分析においてはクラスタ頑健標準誤差の利用が推奨される。

次に脱落については、今回のデータ生成過程の想定が実際のパネル調査データにも当てはまるとすると、復帰ケースを分析から除外することで、点推定に生じるバイアスを避けることが可能であることが示された。復帰ケースを分析に含めることによって得られるサンプルサイズの増加によるメリットよりも、係数へのバイアスによるデメリットが大きいと考えられる。点推定へのバイアスには過小評価と過大評価が見られた。そのため復帰ケースを含めた分析と含めない分析の係数を比較しバイアスの方向が一致しなくとも、それをもって復帰ケースを含めた推定に問題がないと判断するのは妥当でないと推察される。

一方で、今回のモンテカルロシミュレーションの結果が実際のパネルデータ分析の際の指針となるかどうかは、分析されるデータと今回のシミュレーションの想定との乖離度合いによると考えられる。本稿ではできるだけ実際のパネル調査に近いデータ生成を試みたが、項目が異なれば脱落プロセスや級内相関なども異なる可能性が高い。特に脱落に関しては、独立変数の脱落のみを考慮するという非常に簡素な条件を設定しており、複雑な従属変数の欠損プロセスや多重代入による補正の影響などの検討は今後の課題である。

また、変化を捉える方法を用いて実際に分析する際の難点として先行研究でも挙げられ

るのは、カテゴリとその組み合わせの多さである。今回の JLPS を用いた分析では移動が確認されず推定が不可能な学生→自営と自営→学生の移行と非就業→非就業継続（基準カテゴリ）を除き、他の移動について推定したため $5 \times 5 - 3 = 22$ パラメータが推定される。変数の作成が複雑になり、また表が大きくなり非常に読みにくく、分析者も結果を十分には整理しきれないかもしれない。さらに通常の固定効果モデルと同じように、独立変数において移行はランダムに生じるわけではないという点にも注意が必要である。本稿の議論・知見によってパネルデータを用いた変化の分析の利用が広まることを期待する。

参考文献

- Allison, P., 2019, “Asymmetric Fixed-effects Models for Panel Data,” *Socius*, 5: 1–12.
- 有田伸, 2010, 「変化の方向とパターンを区別したパネルデータ分析の可能性——従業員上の地位の変化をもたらす所得変化を事例として」『東京大学社会科学研究所 パネル調査プロジェクトディスカッションペーパーシリーズ No. 35』.
- 有田伸, 2013, 「変化の向き・経路と非変化時の状態を区別したパネルデータ分析——従業員上の地位変化をもたらす所得変化を事例として」『理論と方法』28(1): 69–86.
- Cameron, A. and P. Trivedi, 2010, *Microeconometrics using Stata: Revised Edition*, Stata Press.
- Cox, N., 2002, “Speaking Stata: How to move step by: step,” *Stata Journal*, 2(1): 86–102.
- 竹内麻貴, 2016, 「定年による家事分担の変化」筒井淳也・水落正明・保田時男編『パネルデータの調査と分析・入門』ナカニシヤ出版, 95–104.
- 筒井淳也・水落正明・保田時男編, 2016, 『パネルデータの調査と分析・入門』ナカニシヤ出版.
- York, R. and R. Light, 2017, “Directional Asymmetry in Sociological Analyses,” *Socius*, 3:1–13.

Appendix

分析に使用した Stata コードの一部を示す。これらは著者らの分析に使用したものである。転用する際は分析者の責任で利用していただきたい。

```
//empty "types of employment"
//pattern 1-1 regular (1: seiki) to nlf (4: not in labor force)
bysort PanelID (wave): gen      regnlf=0 if !missing(emptyt)
bysort PanelID (wave): replace regnlf=1 if emptyt[_n-1] == 1 & emptyt[_n]
== 4

//pattern 1-1 nlf to regular
bysort PanelID (wave): gen      nlfreg=0 if !missing(emptyt)
bysort PanelID (wave): replace nlfreg=1 if emptyt[_n-1] == 4 & emptyt[_n]
== 1

bysort PanelID (wave): gen regnlfcum = sum(regnlf)
bysort PanelID (wave): gen nlfregcum = sum(nlfreg)
```

東京大学社会科学研究所パネル調査プロジェクトについて

労働市場の構造変動、急激な少子高齢化、グローバル化の進展などにもない、日本社会における就業、結婚、家族、教育、意識、ライフスタイルのあり方は大きく変化を遂げようとしている。これからの日本社会がどのような方向に進むのかを考える上で、現在生じている変化がどのような原因によるものなのか、あるいはどこが変化してどこが変化していないのかを明確にすることはきわめて重要である。

本プロジェクトは、こうした問題をパネル調査の手法を用いることによって、実証的に解明することを研究課題とするものである。このため社会科学研究所では、若年パネル調査、壮年パネル調査、高卒パネル調査、中学生親子パネル調査の4つのパネル調査を実施している。

本プロジェクトの推進にあたり、以下の資金提供を受けた。記して感謝したい。

文部科学省・独立行政法人日本学術振興会科学研究費補助金

基盤研究 S：2006 年度～2009 年度、2010 年度～2014 年度 基盤研究 C：2013 年度～2016 年度 特別推進研究：2015 年度～2017 年度 若手研究 A：2015 年度～2018 年度
基盤研究 B：2016 年度～2020 年度 特別推進研究：2018 年度～2024 年度

厚生労働科学研究費補助金

政策科学推進研究：2004 年度～2006 年度

奨学寄付金

株式会社アウトソーシング（代表取締役社長・土井春彦、本社・静岡市）：2006 年度～2008 年度

東京大学社会科学研究所パネル調査プロジェクト ディスカッションペーパーシリーズについて

東京大学社会科学研究所パネル調査プロジェクトディスカッションペーパーシリーズは、東京大学社会科学研究所におけるパネル調査プロジェクト関連の研究成果を、速報性を重視し暫定的にまとめたものである。



東京大学社会科学研究所 パネル調査プロジェクト
<https://csrda.iss.u-tokyo.ac.jp/panel/>