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

パネルデータにおける個人内変動・個人間変動・生成メカニズム：モンテカルロシミュレーションによる検討

Within-Subject Variation, Between-Subject Variation, and Data Generation Mechanism: A Study Using Monte Carlo Methods

山本耕資（中央大学総合政策学部）

Koji YAMAMOTO

September 2011

No.51

東京大学社会科学研究所
INSTITUTE OF SOCIAL SCIENCE　UNIVERSITY OF TOKYO

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.51
2011 年 9 月

パネルデータにおける個人内変動・個人間変動・生成メカニズム：モンテカルロシミュレーションによる検討

山本耕資*　（中央大学総合政策学部）

要約　本稿では、パネルデータ分析の手法に関する基礎的な検討を行なう。これにより、以下の点を示す。(1)Allison モデルの性質を検討し、Allison モデルは、脱落があるデータに対しても、固定効果モデルとほぼ同等に偏りのない推定を行なうことができることを示す。(2)固定効果モデルそのものを用いながら、Allison モデルと同等に、個人間変動の効果を推定する、FE+OLS 方式を試用する。(3)Allison モデルの形の構造式を仮定すると「時間のパラドックス」が生じ、Allison 型の構造がありえないとすれば、Allison モデルの推定式は必ず特定化に失敗することを指摘し、その上で、個人内変動と個人間変動の効果を明確に構造レベルで区別し、かつ、時間の前後関係を明示的に考慮する、IEDE モデルを提案する。

付記　本研究の内容は、2010 年 10 月 27 日に東北大学で報告された。有益なコメントやアドバイスをくださった方々、とりわけ、石田浩氏と三輪哲氏に、筆者は深い謝意を抱いている。本研究は、筆者が受けた科学研究費補助金（若手研究（スタートアップ）、特別研究員奨励費）による研究活動の成果の一部である。

* 日本学術振興会特別研究員(PD)

1．はじめに

本稿はパネルデータの分析手法についての探求を行なう。パネルデータは、複数の同一主体（個人）を複数時点で観察して得られるデータであり、調査などが十全に行なわれれば、（人数）×（時点数）だけのデータ点が得られる。

パネルデータの利点の1つは、同じ個人の時間的な変化を見て、因果関係を考えることができる点にある。例えば、クロスセクションデータで、「金銭的に裕福であるほど幸福である」という結果が出て、幸福度から裕福度への因果的影響は考えられないとしても、果たして「裕福になると幸福になる」のかは不明である。パネルデータを用いれば、「裕福になった人々は幸福になった」かどうかを調べることができる。

これは、個人内の変動に着目した分析を行なうことによる。この例では、裕福度の個人内変動と、幸福度の個人内変動との関連性を調べることになる。2 変数が時間によって変化する変数であるからこそ、このような関連性を調べることが可能となる。

他方で、幸福度に個人内の変化以外の要素が影響するとも考えられる。例えば「南国生まれの人は（生まれつき）幸福度が高い」という仮説がありえるかもしれない。これは個人内変動ではなく個人間変動（個人間の相違）を問題にしている。この例で、出身地は時間とともに変化しない。

個人内変動と個人間変動の両方について、説明変数と被説明変数の関係がわかると、より豊富な知見が得られると思われる。パネルデータの分析でしばしば用いられている固定効果モデル(fixed-effect model)を用いれば、個人内変動のみに着目して、X の変化が Y の変化に影響するのかについて、より緩い仮定のもとで偏りなく推定できる。しかし、直接には個人間変動の効果を見ることができない、とされている。他方で、Allison (2009: 23-25)は、hybrid method と呼ばれる方法を提唱し、これによって、個人内変動の効果について、固定効果モデルと同等に偏りなく推定でき、かつ、個人間変動の効果を測れる、とした。この方法によるモデルを以下では Allison モデルと呼ぶ。

本稿の第 1 の目的は、この Allison モデルの性質を検討することにある。Allison によれば、Allison モデルでは、パネルデータでの脱落が発生するときに、固定効果モデルと推定結果（係数ではなく、標準誤差）が異なるという。本稿では、この差異がどのようなものであるのかを、モンテカルロシミュレーションによって示す。

本稿の第 2 の目的は、Allison モデルとは別に、固定効果モデルを用いながら、個人間変動の効果をも推定する方法（以下では、FE+OLS 方式と呼ぶ）について、検討することである。

この第 1 と第 2 の目的について、表に示したものが、表 1 である。固定効果モデルと Allison モデルを比較した場合、個人内変動の効果について、固定効果モデルにおいては、

脱落がある場合にも不偏性があると考えられるが、Allison モデルによる推定結果の性質については何が言えるのかを、本稿で明らかにしようとする。また、固定効果モデルでは、個人間変動の効果を分析できないとされているが、本稿では FE+OLS 方式を試用して検討する。

［表 1］

本稿の第 3 の目的は、データの生成メカニズムの「真の構造」との関係で、Allison モデルの定式化に検討を加えることにある。Allison モデルの発想は、データに中心化(centering)を加えるというものであり、これはより一般のマルチレベルモデルの発想に近いと言える。他方で、Allison モデルは固定効果モデルとは異なり、どのようなデータ生成の構造を「真の構造」と見なして推定しようとしているのかが明確ではなく、注意するべき点を残していると思われる。この点について検討するのが第 3 の目的である。

2．分析の枠組み

(1) 統計的な推定

古典的な統計学の考え方によれば、推定とは、「真の構造」を仮定して、その構造におけるパラメータを推し量る、ということを意味する。パネルデータの分析においても、古典的な統計的推定を行なう限り、考え方は同様である。

この推定の考え方を採る限りにおいて、計量モデルを推定する際、「真の構造」を意識することが望ましいと思われる。「真の構造」についての意識が曖昧である場合には、推定結果を正しく解釈できないことがありうるであろう。

さて、ここで言う「真の構造」は、自然に考えれば、データの生成メカニズムを反映するであろう。データの生成メカニズムを比喩的に表現すれば、「何が何にどう影響してそうなるのか」を示すものと言える。この点は、のちに 7 節で問題となる。

(2) 「真の構造」と「推定」のシミュレーション

本稿の分析では、「真の構造」として特定のものを仮定してパネルデータを計算機内につくり、推定モデルが「真の構造」をいかにうまくとらえられるか、あるいは、「真の構造」の何をとらえているのか、という観点から、複数のモデルの定式化を比較し、モデルの性質を調べる。

ここで、「真の構造」を示す構造モデル（数式で示せば構造式）と、推定モデル（推定式）とは概念的に異なるものである。分析において、構造式と推定式の形がズレる可能性も、本稿では意識される。このようなズレは、一般に、特定化の誤り(specification error)など

と呼ばれる。推定モデルの定式化にはバリエーションがありうるが、この定式化によって、同じデータに適用したときの分析結果が異なるように見えることがある。そのような状況を「真の構造」との関連でどのように解釈すればよいのかの例も、以下で考えることになる。

本稿の分析では、乱数を用いた実験的な推定を行なう。すなわち、モンテカルロシミュレーションを実施する。本稿の分析は、二重の意味でモンテカルロシミュレーションとなる。第1に、「真の構造」を仮定し、一般に撹乱項と呼ばれる部分に起因する数値を乱数で生成して、1 つの標本（サンプル）を仮想的に創出する。この標本のデータに対してモデルの推定を行なう。このとき、標本は1つであるが、ケース数は4000とする。第2に、6節では、理論母集団を想定して、そこからの複数の無作為標本を生成して（無作為に生成する過程で乱数を用いる)、複数回の推定を行ない、擬似的に統計量の標本分布を得て、推定モデルによる異同を示す。このとき、複数の標本（サンプル）を用いている（データセットが多数生成される）ことになる。

3.「真の構造」の設定

(1) 構造A

本節では、「真の構造」として仮定される構造式を示す。まず、構造 A を次のように設定する。この構造は、伝統的に、固定効果モデルが想定してきたものに近い。以下では、i は個人のインデックス、t は時点のインデックスである。

[1] $$Y_{it} = \beta X_{it} + \delta Z_i + u_i + e_{it}$$

ただし、

$$\beta = 100 \quad , \quad \delta = 50$$

$$Z_i \sim N(0,1) \quad , \quad u_i \sim N(0,10^2) \quad , \quad e_{it} \sim N(0,10^2)$$

$$X_{it} = \rho Z_i + \rho \frac{u_i}{10} + \sqrt{1-2\rho^2} R_{it} \quad , \quad R_{it} \sim N(0,1) \quad , \quad \rho = 0.5$$

ここで、Z_i、u_i、e_i、R_{it}は、与えられた分布に従うように乱数で具体的に生成する。理論的には、X_{it}の平均は0、分散は1となる（分散の理論値を1にするために、X_{it}の算出式に $\sqrt{1-2\rho^2}$ という係数が出現する)。X_{it}と Z_iとの相関係数は ρ、X_{it} と u_iとの相関係数も ρ となる。

言葉で説明すると、X_{it}は観察される time-varying な変数(時間とともに変化する変数)、Z_iは観察される time-invariant な変数(時間によって変化しない変数)、u_iは time-invariant な撹乱項（観察されない)、e_iは time-varying は撹乱項（観察されない）である。

前記のとおり、この構造の形式は、伝統的に固定効果モデルが想定してきたものに近い。ただし、X_{it} と u_i とが相関するなどの、相関の構造に注意されたい。

仮想データの生成に際しては、4000 名の個人からなる標本を考える。脱落のないデータでは、4 時点での観測結果が得られるとする。

なお、以下では時点固有の効果（時代効果または加齢効果）は考えない。通常の計量分析では、時点の変化を先行変数とする擬似効果を除去するため、時点ダミー変数などの使用が推奨される点を追記する。

(2) 構造 B

次に、構造 B を次のように設定する。この構造の形式は、Allison モデルの推定式をそのまま当てはめたものに相当する。

[2] $$Y_{it} = \beta(X_{it} - \overline{X}_i) + \gamma\overline{X}_i + \delta Z_i + u_i + e_{it}$$

ただし、

$$\beta = 100 \ , \ \gamma = 75 \ , \ \delta = 50$$

$$Z_i \sim N(0,1) \ , \ u_i \sim N(0,10^2) \ , \ e_{it} \sim N(0,10^2)$$

$$X_{it} = \rho\frac{u_i}{10} + \sqrt{1-\rho^2}R_{it} \ , \ R_{it} \sim N(0,1) \ , \ \rho = 0.5$$

ここで、Z_i、u_i、e_i、R_{it} は、与えられた分布に従うように乱数で具体的に生成する。また、X_{it} の平均は 0、分散は 1 となり、X_{it} と u_i との相関係数は ρ となる。

上述のように、この構造の形式は、Allison モデルの推定式を構造式として使用したものである。ただし、すべての変数が独立ではない点に注意されたい。

構造 A と同様に、構造 B においても、仮想データを生成する際には、4000 名の個人からなる標本を考え、脱落のないデータでは、4 時点での観測結果が得られるとする。

4．脱落がない場合の推定結果

(1) 構造 A

脱落がないと仮定して、構造 A のデータセットを 1 つ生成して、各種の推定モデルを適用した結果を、表 2 に示した。

［表 2］

ここで、構造式は式[1]であるので、X の係数は β であり、その真値は 100 である。Z の

係数は δ であり、その真値は 50 である。推定結果を検討する際には、真値が推定値の 95%信頼区間に入るか否かが、うまく推定できているかの 1 つの目安と言える。

本稿の主眼は、固定効果モデルと Allison モデルについての検討にあるが、参考までに、・プールされた OLS と単純な変量効果モデルの推定も実施した。表 2 に示された、これらのモデルの推定値には、バイアスが含まれている。例えば、プールされた OLS と単純な変量効果モデルのいずれも、β の推定値として、真値の 100 より大きな値を推定している。これは、X と u との間に相関があるからである。

さて、表 2 の固定効果モデルの結果に着目する。固定効果モデルにおける β の推定値はほぼ 100 であり、正しく推定していると言える。

さらに、固定効果モデルの結果を利用して、個人内変動の効果を推定することを試みる。先述のとおり、その方法をここでは「FE+OLS 方式」と呼んでいる。より具体的には、まず固定効果モデルの係数推定値を用いて、

$$[3] \qquad Y_{it} = \hat{\beta}X_{it} + \hat{v}_i$$

となるような $\hat{v}_i$ を各個人について推定する。これは、いわゆる個人効果（固定効果）を具体的に推定していることになる。次に、この推定された $\hat{v}_i$ を被説明変数とする OLS を行ない、Z の効果を推定する。表 2 によれば、この「FE+OLS 方式」では、δ が正しく推定される。

Allison モデルについては、2 種類の推定式を用いた。まず、Allison モデル(1)は、X の個人内差分と Z を投入して、変量効果モデル（ここでは GLS による）の推定手続きで推定するものである。推定式は次のようになる。

$$[4] \qquad Y_{it} = \hat{\alpha} + \hat{\beta}(X_{it} - \overline{X}_i) + \hat{\delta}Z_i$$

他方で、Allison モデル(2)は、Allison モデル(1)に加えて、X の個人内平均を投入するものである。推定式は次のようになる。

$$[5] \qquad Y_{it} = \hat{\alpha} + \hat{\beta}(X_{it} - \overline{X}_i) + \hat{\gamma}\overline{X}_i + \hat{\delta}Z_i$$

表 2 において、Allison モデル(1)の結果に着目しよう。ここで、X の個人内差分の係数は、固定効果モデルの X の係数と一致している。また、それらの間で、標準偏差も一致する。しかし、Z の係数は真値ではなく、100 程度の値となる。この理由は以下のように説明できる。構造 A の構造式[1]を変形すると、

[6]
$$\begin{aligned} Y_{it} &= \beta X_{it} + \delta Z_i + u_i + e_{it} \\ &= \beta(X_{it} - \overline{X}_i) + \beta\overline{X}_i + \delta Z_i + u_i + e_{it} \end{aligned}$$

となるが、X と Z は相関しているので、独立な変数 θ を仮定すると、

[7]
$$\overline{X}_i = \rho Z_i + \theta \quad , \quad Z_i \perp \theta$$

と表せる。よって、式[6]は、

[8]
$$\begin{aligned} Y_{it} &= \beta(X_{it} - \overline{X}_i) + \beta(\rho Z_i + \theta) + \delta Z_i + u_i + e_{it} \\ &= \beta(X_{it} - \overline{X}_i) + (\beta\rho + \delta) Z_i + \beta\theta + u_i + e_{it} \end{aligned}$$

となる。これにより、Y を X の個人内差分と Z のみで説明する推定式を用いると、Z の係数は($\beta\rho+\delta$)=100 程度となる。要するに、Z の係数が真値からズレるのは、$\beta\overline{X}_i$ の項が存在しないことによる一種の omitted variable bias と考えられる。別の言い方をすれば、構造式と推定式のズレによる、特定化の誤りによると言える[1]。

次に、表 2 の Allison モデル(2)の結果に着目する。ここでも、X の個人内差分の係数は、固定効果モデルの X の係数と一致する。それらの標準偏差も一致する。

他方で、この結果において、X の個人内平均の係数推定値は、式[6]から考えると、直感的には 100 となることが期待されるが、113 程度の値が推定された。このうち、8 程度のバイアスの部分については、次のように説明できる。

構造式[1]より、R_{it} の分散は 1 であるが、4 時点での平均をとった $\overline{R}_i$ の分散は 1/4 である。ここで、

[9]
$$\overline{X}_i = \rho Z_i + \rho\frac{u_i}{10} + \sqrt{1-2\rho^2}\,\overline{R}_i$$

である。また、u_i を $\overline{X}_i$ に回帰させると、その係数は、

[1] このような誤りを「危険視する」べきか否かは分析の目的などに依存すると思われる。例えば、クロスセクションデータでの OLS による分析において、先行変数や媒介変数が含まれる場合と含まれない場合とで、ある説明変数の係数が異なってくるとき、この現象を「危険視する」ことも「有用な情報だと見る」ことも可能であることを本稿では指摘しておきたい。

[10] $$\frac{Cov(\overline{X}_i, u_i)}{Var(\overline{X}_i)} = \frac{10\rho}{\rho^2 + \rho^2 + \frac{1}{4}(1-2\rho^2)} = \frac{40\rho}{1+6\rho^2} = 8$$

となる。すなわち、ある独立な変数 τ を用いて、

[11] $$u_i = 8\overline{X}_i + \tau \quad , \quad \overline{X}_i \perp \tau$$

と表せる。式[11]を式[6]に代入すると、

[12] $$\begin{aligned} Y_{it} &= \beta(X_{it} - \overline{X}_i) + \beta\overline{X}_i + \delta Z_i + u_i + e_{it} \\ &= \beta(X_{it} - \overline{X}_i) + \beta\overline{X}_i + \delta Z_i + 8\overline{X}_i + \tau + e_{it} \\ &= \beta(X_{it} - \overline{X}_i) + (\beta + 8)\overline{X}_i + \delta Z_i + \tau + e_{it} \end{aligned}$$

となるので、Allison モデル(2)で推定すると、$\overline{X}_i$ の係数は β よりも大きく推定されることになる。構造 A では、X は Z とも相関しているために、X の係数のバイアスによって、Z の係数も真値ではなくなり（43 程度になり）、また、X の係数のバイアスが式[12]で予想される 8 程度よりも大きくなるものと思われる。要するに、これらは、撹乱項と説明変数が相関していることによるバイアスと言える。

本項の結果をまとめよう。構造 A は、固定効果モデルが伝統的に想定してきた構造であった。この構造を仮定した場合、Allison モデルが推定する個人内変動の効果は確かに固定効果モデルと一致する。他方で、Allison モデルの time-invariant な変数の係数の解釈には注意が必要であることが明らかとなった。

(2) 構造 B

本項では、構造 B に各種の推定モデルを適用した結果について述べる。推定結果は表 3 に示される。

［表 3］

ここでの構造式は式[2]であるので、X（ないし X の個人内差分）の係数は β であり、その真値は 100 である。X の個人内平均の係数は γ であり、その真値は 75 である。Z の係数は δ であり、その真値は 50 である。以下の推定結果においては、プールされた OLS と単

純な変量効果モデルの結果については説明を割愛する。

表 3 に示された固定効果モデルの結果（「差分使用」と記されていないもの）では、X の係数はうまく推定されている。ここでさらに FE+OLS 方式を適用した場合、X の個人内平均の係数は-14 程度と推定される。これは直感的には 75 となるように考えられるが、このズレは次のように説明される。構造式[2]より、

[13]
$$\begin{aligned} Y_{it} &= \beta(X_{it} - \overline{X}_i) + \gamma\overline{X}_i + \delta Z_i + u_i + e_{it} \\ &= \beta X_{it} + (\gamma - \beta)\overline{X}_i + \delta Z_i + u_i + e_{it} \end{aligned}$$

となり、他方で、構造式[2]より、R_{it} の分散は 1 であるが、4 時点での平均をとった $\overline{R}_i$ の分散は 1/4 であるため、

[14]
$$\overline{X}_i = \rho\frac{u_i}{10} + \sqrt{1-\rho^2}\,\overline{R}_i$$

である。ここで、u_i を $\overline{X}_i$ に回帰させると、その係数は、

[15]
$$\frac{Cov(\overline{X}_i, u_i)}{Var(\overline{X}_i)} = \frac{10\rho}{\rho^2 + \frac{1}{4}(1-\rho^2)} = \frac{40\rho}{1+3\rho^2} = 11.43$$

となる。すなわち、ある独立な変数 λ を仮定して、

[16]
$$u_i = 11.43\overline{X}_i + \lambda \quad , \quad \overline{X}_i \perp \lambda$$

と表せる。式[16]を式[13]に代入すると、

[17]
$$\begin{aligned} Y_{it} &= \beta X_{it} + (\gamma - \beta)\overline{X}_i + \delta Z_i + u_i + e_{it} \\ &= \beta X_{it} + (\gamma - \beta)\overline{X}_i + \delta Z_i + (11.43\overline{X}_i + \lambda) + e_{it} \\ &= \beta X_{it} + (\gamma - \beta + 11.43)\overline{X}_i + \delta Z_i + \lambda + e_{it} \end{aligned}$$

となる。よって、$\overline{X}_i$ の係数は$(\gamma$-β+11.43)=-13.57 程度に推定されることになる。

次に、固定効果モデル（差分使用）の結果に着目しよう。この推定モデルは、X の情報を、個人内差分と個人内平均とに分解した上で、固定効果モデルの推定手続きに投入した

ものである。この結果によれば、X の個人内差分の係数は、直感的に期待されるように、100 程度に推定されている。他方で、この推定に FE+OLS 方式を適用した場合の、X の個人内平均の係数推定値は、86 程度となっている。これは、γ の真値 75 に、式[15]で示されるバイアスである 11 程度が加算された結果であると考えられる（この場合に式[17]に相当する式を導くと、$\overline{X}_i$ の係数に-β が存在しない形となる）。

Allison モデルの結果の検討に移る。表 3 に示された、Allison モデル(1)の結果においては、X の個人内差分の係数は、固定効果モデルにおけるものと一致する。それらの標準誤差も一致する。また、Z の効果も期待されるとおりに正しく推定されている。

Allison モデル(2)の結果でも、X の個人内差分の係数は、固定効果モデルのそれと一致する。それらの標準誤差も一致する。また、Z の効果も正しく推定している。他方で、X の個人内平均の係数は、直感的に期待される γ=75 ではないが、固定効果モデル（差分使用）の FE+OLS 方式におけるものと等しい数値として推定されている。すなわち、Allison モデル(2)の個人内平均の係数は、固定効果モデル（差分使用）と同様にバイアスされていると考えることができる。

本項の結果をまとめると以下のようになる。ここで仮定された構造 B は、Allison モデルの推定式を使用したものである。ここでも、Allison モデルが推定する個人内変動の効果は固定効果モデルと一致する。他方で、固定効果モデルでも、Allison モデルでも、time-invariant な変数の係数の解釈には注意が必要であることが明らかとなった。

(3) この節のまとめ

固定効果モデルを推定してさらに FE+OLS 方式を用いる場合と、Allison モデルとは、「互角」と言ってよい。いずれも、個人内変動の効果については同等の結果を示す。個人間の変動の効果（time-invariant な変数の効果）については、いずれでも、撹乱項と説明変数の相関や、特定化の誤り（としての omitted variable bias）に注意するべきである、という点で、変わりがない。強いて言えば、「真の構造」がどちらに適しているかのみが両者の「使いやすさ」を分ける、と言える。

5．脱落の設定

(1) MAR 的な脱落

前節までは、（人数）×（時点数）のデータ点がすべて得られている状況でのデータの推定であった。すなわち、脱落（あるいは欠落）のないパネルデータの存在を仮定していた。これに対して、以下では、データ生成の際に、脱落を発生させる。この際、脱落はその後の観察の不可能性を意味するものとし、「脱落後の復活」は考慮しない。

まず、「MAR 的な脱落」を考える。MAR とは missing at random を意味しており、観

察される変数のみに依存して脱落が生じることを指す。ここでは、MAR 的な脱落を、確率的に、を次のように発生させる。時点 1 では 4000 名すべてについて情報が観察されるとする。

[18]　個人 i の情報が時点(t-1)で観察されていて、かつ、

$X_{i1}+Z_i+\sqrt{2}\omega>0$ のとき、個人 i の情報は時点 t でも観察される。

ただし、ω は N(0,1)に従う乱数である。

これは、X_{i1} と Z_i の係数がそれぞれ $1/\sqrt{2}$ であるプロビットモデルによって、脱落確率が表現されることを意味している。X と Z は観察される変数であるため、この脱落メカニズムは MAR 的であると言える。

(2) NI 的な脱落

他方で、「NI 的な脱落」の場合も考慮する。NI とは non-ignorable を意味しており、観察されない変数にも依存して脱落が生じることを指す。ここでは、NI 的な脱落を、確率的に、次のように発生させる。時点 1 では 4000 名すべてについて情報が観察されるとする。

[19]　個人 i の情報が時点(t-1)で観察されていて、かつ、

$\frac{u_i}{10}+\frac{e_{it}}{10}+\sqrt{2}\omega>0$ のとき、個人 i の情報は時点 t でも観察される。

ただし、ω は N(0,1)に従う乱数である。

これは、u_i/10 と e_{it}/10 の係数がそれぞれ $1/\sqrt{2}$ であるプロビットモデルによって、脱落確率が表現されることを意味している。u_i と e_{it} は（推定モデルにとって）観察されない変数とされているから、この脱落のメカニズムは NI 的であると言える[2]。

(3) 構造 A で脱落があるデータでの推定結果

本項では、構造 A を仮定しながら、脱落のあるデータセットを 1 つ生成して、各種の推定モデルの結果を示す。

まず、仮想データにおける脱落の状況を表で示す。構造 A で MAR 的な脱落がある場合の脱落状況を表 4 に示した。他方で、構造 A で NI 的な脱落がある場合の脱落状況を表 5 に示した。MAR 的な脱落の場合と比べて、NI 的な脱落の場合には、時点が進んでからも

[2] ただし、このメカニズムは、Y そのものの値に依存した脱落を仮定していない。

新たに脱落していく割合が大きい。これは、以下の理由による。MAR 的な脱落の場合は、脱落しない個人は脱落しにくい（不変の）属性を持っているために、初期に脱落しなかった個人はその後には脱落しにくいのに対し、NI 的な脱落の場合には、e_{it} が時点によって異なる値をとるために、MAR 的な場合と比べて、時点が進んでからも新たに脱落していく可能性が高いのである。

［表 4］

［表 5］

さて、構造 A で MAR 的な脱落があるデータに対する推定結果は、表 6 に示される。個人内変動の効果に関しては、固定効果モデルと Allison モデル(1)(2)の係数は一致している。しかし、標準誤差はわずかに異なっている。

［表 6］

次に、構造 A で NI 的な脱落があるデータに対する推定結果を表 7 に示した。ここでも、個人内変動の効果に関しては、固定効果モデルと Allison モデル(1)(2)の係数は一致しているが、標準誤差がわずかに異なっていることがわかる。

［表 7］

(4) 構造 B で脱落があるデータでの推定結果

本項では構造 B で脱落のあるデータを生成させて分析を行なう。まず、構造 B で MAR 的な脱落があるデータについて、脱落状況を表 8 に示した。また、構造 B で NI 的な脱落を発生させたデータについて、その脱落状況を表 9 に示した。

［表 8］

［表 9］

構造 B で MAR 的な脱落があるデータでの推定結果を、表 10 に示した。ここでも、個人内変動の効果に関しては、固定効果モデルと Allison モデル(1)(2)の係数は一致しているものの、標準誤差がわずかに異なる。

［表 10］

構造 B で NI 的な脱落がある場合の推定結果を、表 11 に示した。ここでも、個人内変動の効果に関しては、固定効果モデルと Allison モデル(1)(2)の係数は一致し、他方で、標準誤差はわずかに異なっている。

［表 11］

6．1000 回の抽出シミュレーションによる標本分布の確認

前節で確認したように、脱落が発生する場合には、固定効果モデルと Allison モデルとで、個人内変動の効果を示す係数の、標準誤差が相違する。そこで本節では、固定効果モデル、Allison モデル(1)、Allison モデル(2)における、脱落がある場合の挙動をより詳しく調べる。標準誤差（誤差分散）に違いは何を意味しているのか、どれがより「正しい」標準誤差（誤差分散）を推定しているのかが、ここでの焦点である。

以下では、理論母集団から乱数を用いて標本を生成し、モデルを推定し、X または X の個人内差分の係数と誤差分散を記録する、という作業を 1000 回繰り返す。このとき、標本の生成は仮想的な抽出であると見なせる。

まず、構造 A で MAR 的な脱落がある場合のシミュレーション結果を表 12 に示した。表 12 で用いられている統計量の記号は以下のような意味を持つ。

［表 12］

E(Beta-hat)： $E(\hat{\beta})$ を示す。推定された係数を、1000 回の推定について平均したものである。

E(Var-hat(Beta-hat))： $E[\hat{Var}(\hat{\beta})]$ を示す。推定された誤差分散（標準偏差の 2 乗）を、1000 回の推定について平均したものである。

Var(Beta-hat)： $Var(\hat{\beta})$ を示す。推定された係数の、1000 回の推定における分散である。1000 回が十分な試行回数である限りにおいて、これは「実際の」誤差分散であると見なせる。

Prob[L95%C.I.<=Beta<=U95%C.I.]： Prob$[L_{95\%C.I.} \leq \beta \leq U_{95\%C.I.}]$ を示す。推定された係数の 95%信頼区間に真値が入る確率を表現している。理論上は、この数値は 0.95 になることが予定されている。

表 12 のシミュレーション結果によれば、まず、推定される係数の期待値が、Allison モデルと固定効果モデルとで一致している。これは、各回のデータ生成（抽出）において、係数推定値が一致するからである。

次に、Allison モデルは固定効果モデルよりも、誤差分散を小さく見積もっているが、Allison モデルも固定効果モデルもともに、実際の分散よりは過大に推定している。95%信頼区間に真値が入る確率も、いずれにおいても 95%を越えている。すなわち両者とも、保守的な（実際より誤差を大きく見積もる）推定を行なっていることになる。

同様に、構造 A で NI 的な脱落がある場合のシミュレーション結果を表 13 に示した。ここでも、推定される係数の期待値は、Allison モデルと固定効果モデルとで一致している。他方で、表 13 は、Allison モデルと固定効果モデルのいずれも、誤差分散を実際のものより過小に推定していることを示している。ただし、95%信頼区間に真値が入る確率は、ほぼ 95%である。これは、外れ値的に性質の悪い標本が稀に抽出されてしまう（係数の標本分布が想定どおりの t 分布／正規分布ではない）ことを意味していると思われる。いずれにしても、区間推定についてはほぼ妥当な結果が得られると思われる。また、一定の仮定のもとで[3]、5%水準での有意性検定でもほぼ妥当な結果が得られると思われる。

［表 13］

さらに、構造 B についても、脱落があるデータについてのシミュレーションを実施した。MAR 的な脱落がある場合の結果が表 14、NI 的な脱落がある場合の結果が表 15 に示されている。いずれにおいても、構造 A の場合と同様のことが言える。

［表 14］
［表 15］

本節をまとめると、以下のようになろう[4]。脱落が MAR 的であれば、固定効果モデルも Allison モデルも誤差分散を保守的に見積もっている。脱落が NI 的であれば、固定効果モ

---

[3] より具体的には、他の条件が一定なら、係数の真値が異なる場合にも標本分布の形状が同一である（平行移動のみが生じる）という仮定を指している。

[4] ここでのシミュレーション結果は、ありうるすべての条件を網羅したものではないことに注意が必要である。異なる設定の場合には異なる結果が得られる可能性がある。例えば、X が複数であり、さらにそれらが相関している場合、ρ の値が異なる値である場合、u_i や e_{it} の分散や分布が異なる場合、脱落のメカニズムが異なる場合、u_i と X との相関が異なる場合などの、すべてにおいて同等の結果が得られるという保証はない。

デルも Allison モデルも誤差分散を過小に見積もっているが、95%信頼区間はほぼ正確であり、区間推定にはほぼ問題がないと思われる。

7. パネルデータの生成メカニズム

(1) Allison モデルと「時間のパラドックス」

以上の検討は、固定効果モデルと比較した場合に、個人内変動の効果について、Allison モデルがほぼ同等の有用性を有する、ということを意味するように思われる。これに対して、本節では、データの生成メカニズムという観点から、Allison モデルに付随しうる問題点を指摘し、その問題点を克服するようなシンプルな提案を行なう。

Allison モデルの推定式をそのまま構造式にしたものが、上記の構造 B であった。この、いわば Allison 型の構造式である式[2]を再掲する。

[2] $$Y_{it} = \beta(X_{it} - \overline{X}_i) + \gamma\overline{X}_i + \delta Z_i + u_i + e_{it}$$

ここで、4 時点のデータを考えると、$\overline{X}_i = (X_{i1} + X_{i2} + X_{i3} + X_{i4})/4$ である。

式[2]が実際にデータを生成させるメカニズムを表現する構造式であると仮定すると、「時間のパラドックス」とでも呼べるような現象が生じる。式[2]のもとでは、時点 1 の Y の値は、次式[20]で決まるはずである。これは、時点 1 の Y が $\overline{X}_i$ に影響されることを意味する。

[20] $$Y_{i1} = \beta(X_{i1} - \overline{X}_i) + \gamma\overline{X}_i + \delta Z_i + u_i + e_{it}$$

もし、$\overline{X}_i$ が決まるためにはあくまで時点 4 での X_{i4} の値が必要である、とすると、X_{i4} から $\overline{X}_i$ が決まり、$\overline{X}_i$ が Y_{i1} に影響するはずであるから、「未来が過去を規定している」という、直感に反するような構造が存在することになる。

他方で、もし、「$\overline{X}_i$ は時点 1 で決まるが、これはあくまで個人内の平均的な水準を表現しているだけである」としても、$\overline{X}_i = (X_{i1} + X_{i2} + X_{i3} + X_{i4})/4$ から $X_{i4} = 4\overline{X}_i - X_{i1} - X_{i2} - X_{i3}$ であるので、時点 3 ですでに X_{i4} は決定することになる。すなわち、「未来は 1 時点前に決まる」という現象が生じうることになる。ここで、X が内生変数であれば、過去にすでに決まっていたということはありうるかもしれない。しかし、Allison モデルは特に X を内生変数として扱うことを求めていない。また、X_{i1}、X_{i2}、X_{i3} は「自由に」値が決まりうるのに対して、X_{i4} のみが全く異なる（「内生的」な）決まり方をする点は自然だとは言えないであろう。

このような考察からすれば、Allison モデル型の構造は、パネルデータの生成メカニズムとして適切なのか、疑問が湧く。もし Allison モデル型の構造がありえないとすれば、Allison モデルの推定式は必ず特定化に失敗することになる。

これは、構造式と推定式との区別の問題に起因しているとも言える。固定効果モデルは、構造式と推定式を明確に分け、通常は推定式でのみ変数の差分を用いる。他方で、Allison モデルは、どのような構造を仮定しているのかが不明確である。

また、この問題は、「一般の」マルチレベルモデルと、パネルデータ分析のモデルとの関係を慎重に検討するべきであることをも示しているであろう。Allison モデルは、マルチレベルモデルにおける変数の中心化(centering)の発想によるものであると考えられる。パネルデータは、マルチレベルデータの一種であると考えることもできるが、マルチレベルモデルの考え方・分析手法のうちのそれぞれがパネルデータに適用できるかどうかは、何を分析したいのかに照らして吟味する必要があろう。

さらに述べれば、次の点を指摘できる。この問題は、本稿でシミュレーションを実施した際に、データを生成するにあたって、Allison 型の構造を用いると時間の経過を自然に表現できなかったことによって、気付かれたものである。これは、大げさに言えば、「世界を再現してみる」というシミュレーションの発想が、データ生成メカニズムについての考察の機会を与える、という有用性を持つことを意味している。

(2) 初期値効果-差分効果モデル(IEDE Model)

前項の考察をもとにして、本項ではシンプルなパネルデータ分析モデルを提示する。ここでは、以下の 2 点が意識される。第 1 に、Allison モデルや FE+OLS 方式が実現するように、個人内変動の効果と個人間変動の効果の両方を区別して分析することである。ここで、Allison モデルとは異なり、これら 2 つの効果を、構造レベルで明確に区別する。第 2 に、Allison 型の構造が内包するような、時間的なパラドックスを回避することである。これは、時間的な前後関係を明確に考慮することを意味する。

ここで提案されるモデルを、初期値効果-差分効果モデル(IEDE Model: Initial-value Effect and Difference Effect Model)と呼んでおく。これはその名のとおり、構造レベルで、初期値の効果と差分の効果を明確に区別する。

このモデルでは、構造レベルで、観察の初期時点の Y は初期時点で判明している情報のみから決定される、と仮定する。この決定の効果を初期値効果と呼ぶ。また、構造レベルで、Y の変化は X の変化によって決定される、と考える。この効果を差分効果と呼ぶ。通常の固定効果モデルも Allison モデルも、推定式で差分を用いるが、これを構造レベルで考え、かつ、初期値または前期からの差分を用いるのが、ここでのアイデアである。

一般的には、以下のように定式化される。構造式として、

[21] $$Y_{i1} = f(X_{i1}, Z_i; \boldsymbol{\theta}_{t=1})$$

[22] $$Y_{it} - Y_{i,t-1} = g(X_{it} - X_{i,t-1}; \boldsymbol{\theta}_{t-1,t}) \quad \text{for } t = 2, \ldots, T$$

となる。ただし、$f(\cdot)$、$g(\cdot)$はそれぞれ何らかの関数、T は扱われる時点の中での最終時点である。これらから、

[23] $$Y_{it} = f(X_{i1}, Z_i; \boldsymbol{\theta}_{t=1}) + \sum_{k=2}^{t} g(X_{ik} - X_{i,k-1}; \boldsymbol{\theta}_{k-1,k})$$

という形が導かれる。関数型を線形で特定化した場合、例えば次のように書ける。

[24] $$Y_{i1} = \alpha + \gamma X_{i1} + \delta Z_i + u_i + \varepsilon_{i1}$$

[25] $$Y_{it} - Y_{i,t-1} = \beta_{t-1,t}(X_{it} - X_{i,t-1}) + \varepsilon_{it} \quad \text{for } t = 2, \ldots, T$$

ここでさらに、次式[26]のような仮定を置くと、

[26] $$\beta_{t-1,t} = \beta \quad \text{for } t = 2, \ldots, T$$

構造は次式[27]で表せることになる。

[27] $$Y_{it} = \alpha + \beta(X_{it} - X_{i1}) + \gamma X_{i1} + \delta Z_i + u_i + e_{it}$$

ここで注意するべき点を 2 点指摘したい。第 1 は、観察の始期をどのように定めるのかが解釈の上で重要である、ということである。仮に、ある時点で出生した複数の個人を、出生時点から観察し続けることができ、例えば「体格」と「（泣き）声の大きさ」の関係を示すデータを用いるなら、上記のパラメータ $\boldsymbol{\theta}_{t=1}$ や γ は、「出生以前に生じた効果」と解釈することができ、他方で、それとは全く別に、パラメータ $\boldsymbol{\theta}_{t-1,t}$ や β は「出生後の経験による効果」などと解釈できよう。しかし、もし観察始期が個人の出生時点ではなく、例えば対象者が 15 歳である時点であれば、観察始期時点での個人間の差異の効果は、実際には「出生以前に生じた効果」と「出生後の経験による効果」とが混濁したものとなろう。さらに、例えば観察始期における対象者の年齢がばらついているならば、この混濁の度合いのばらつきが大きくなることが予想される。

第 2 に、式[27]においては、ε_{it} の分散がすべて等しくても、e_{it} の分散は一定だとは限ら

ない。仮に ε_{it} が時点間で無相関であれば、e_{it} の分散は t が大きくなると増大していくことになる。いわば、「誤差が蓄積していく」ような構造が存在しうる。

以上は、IEDE モデルの構造の表現である。このモデルの推定方法はいくつか考えられる。1 つの選択肢は、FE+OLS 方式による推定である。この場合、上記のような、構造レベルにおける非等分散性が、推定式のレベルにも反映されるため、これを考慮する推定が望まれる。

他方で、式[25]からは、一階差分モデルの適用も自然な選択肢となると思われる。一階差分モデルに付随する、ありうる弱みとして、隣接する時点間で両方有効なデータ点が存在する場合でないと適用できないという点が挙げられる。いわば、時点の「欠け」に弱いと言えるかもしれない。しかし、時点の「欠け」の情報を積極的に利用するという考え方もありうる。例えば、「何時点のデータ点が欠けると、差分の撹乱項の分散がどれだけ増えるか」という対応関係を推定モデルに組み入れられれば、ε_{it} の時点間の相関の程度を推定できる可能性があると思われる。

8．結論

本稿では、通常の固定効果モデルとの比較において、Allison モデルの性質を検討し、また、FE+OLS 方式について試行した。その結果、Allison モデルは、FE+OLS 方式を含む固定効果モデルと、「互角」と言ってよい性質を示すことが明らかとなった。すなわち、Allison モデルは、脱落があるデータに対しても、固定効果モデルとほぼ同等に偏りのない推定を行なうことができる一方、FE+OLS 方式は、固定効果モデルそのものを用いながら、Allison モデルと同等に、個人間変動の効果を推定することができることが示された。

本稿ではさらに、Allison モデルに付随しうる問題点を指摘した上でそれに関わる提案を行なった。すなわち、Allison 型の構造を仮定すると「時間のパラドックス」が生じ、Allison 型の構造がありえないとすれば、Allison モデルの推定式は必ず特定化に失敗することを指摘した。その上で、個人内変動と個人間変動の効果を明確に構造レベルで区別し、かつ、時間の前後関係を明示的に考慮する、IEDE モデルを提案した。

本稿での検討は、パネルデータ分析の手法に関する基礎的な貢献となるものである。

引用文献

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Sage Publications.

図表

表1. 固定効果モデルとAllisonモデル

	固定効果モデル	Allisonモデル
個人内変動の効果の推定		
脱落がない場合の不偏性	○	○
脱落がある場合の不偏性	○	？→検討
個人間変動の効果の推定が可能か	×？ →FE+OLS方式を検討	○

表2. 構造Aで脱落のないデータでの推定結果

変数	プールされたOLS				固定効果モデル				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限			下限	上限
X	106.6297	0.1168	106.4008	106.8586	100.0882	0.1273	99.8388	100.3377	103.9490	0.1169	103.7199	104.1782
Z	46.8092	0.1186	46.5767	47.0418	(dropped)				48.1791	0.1464	47.8922	48.4661
定数	0.0447	0.1025	-0.1562	0.2456	0.6625	0.0791	0.5074	0.8177	0.0506	0.1336	-0.2113	0.3124
					FE+OLS方式							
Z					50.1522	0.1775	49.8041	50.5002				
定数					-0.6035	0.1774	-0.9514	-0.2557				
個体数	-				4000				4000			
観測数	16000				16000				16000			

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.0882	0.1273	99.8388	100.3377	100.0882	0.1273	99.8388	100.3377
Xの個人内平均					113.2220	0.1973	112.8354	113.6086
Z	101.3013	1.1164	99.1132	103.4895	43.4403	0.1585	43.1297	43.7509
定数	0.2774	1.1158	-1.9096	2.4643	0.0303	0.1222	-0.2091	0.2698
個体数	4000				4000			
観測数	16000				16000			

表3. 構造Bで脱落のないデータでの推定結果

変数	プールされたOLS				プールされたOLS（差分使用）				固定効果モデル			
	係数	標準誤差	95%信頼区間 下限	95%信頼区間 上限	係数	標準誤差	95%信頼区間 下限	95%信頼区間 上限	係数	標準誤差	95%信頼区間 下限	95%信頼区間 上限
X	94.0573	0.1082	93.8452	94.2694					100.0721	0.1039	99.8684	100.2757
Xの個人内差分					100.0721	0.1249	99.8273	100.3168				
Xの個人内平均					86.2885	0.1419	86.0103	86.5667				
Z	49.9676	0.1098	49.7523	50.1829	50.0612	0.0952	49.8747	50.2478	(dropped)			
定数	0.0196	0.1098	-0.1955	0.2348	0.0358	0.0951	-0.1506	0.2222	0.6076	0.0791	0.4525	0.7627
									FE+OLS方式			
Xの個人内平均									-13.7836	0.1968	-14.1694	-13.3977
Z									50.0612	0.1320	49.8025	50.3200
定数									-0.5718	0.1319	-0.8304	-0.3133
個体数	–				–				4000			
観測数	16000				16000				16000			

変数	固定効果モデル（差分使用）				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間 下限	95%信頼区間 上限	係数	標準誤差	95%信頼区間 下限	95%信頼区間 上限
X					97.0670	0.1023	96.8665	97.2676
Xの個人内差分	100.0721	0.1039	99.8684	100.2757				
Xの個人内平均	(dropped)							
Z	(dropped)				49.9314	0.1469	49.6434	50.2193
定数	0.8298	0.0791	0.6746	0.9849	0.0134	0.1468	-0.2744	0.3012
	FE+OLS方式							
Xの個人内平均	86.2885	0.1968	85.9026	86.6743				
Z	50.0612	0.1320	49.8025	50.3200				
定数	-0.7940	0.1319	-1.0526	-0.5354				
個体数	4000				4000			
観測数	16000				16000			

(to be continued)

表3. *(continued)*

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.0721	0.1039	99.8683	100.2758	100.0721	0.1039	99.8684	100.2757
Xの個人内平均					86.2885	0.1968	85.9027	86.6743
Z	51.1010	0.9240	49.2900	52.9120	50.0612	0.1320	49.8025	50.3199
定数	0.2148	0.9235	-1.5952	2.0249	0.0358	0.1319	-0.2227	0.2943
個体数	4000				4000			
観測数	16000				16000			

表4.　構造AにおけるMAR的な脱落の状況（表6で用いられるデータの例）

時点	回答率
1	100.0%
2	49.8%
3	35.1%
4	27.7%

Note:　回答率は当初の回答者の数である4000を分母とするものである。

表5.　構造AにおけるNI的な脱落の状況（表7で用いられるデータの例）

時点	回答率
1	100.0%
2	51.1%
3	28.6%
4	17.8%

Note:　回答率は当初の回答者の数である4000を分母とするものである。

表6. 構造Aで脱落がMAR的であるデータでの推定結果

変数	プールされたOLS				固定効果モデル				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限			下限	上限
X	106.4373	0.1598	106.1241	106.7504	100.2557	0.2059	99.8520	100.6594	104.2169	0.1604	103.9026	104.5313
Z	46.7900	0.1626	46.4713	47.1087	(dropped)				48.1078	0.1937	47.7281	48.4874
定数	0.9168	0.1504	0.6220	1.2116	20.3689	0.1211	20.1314	20.6064	0.6698	0.1794	0.3182	1.0215
					FE+OLS方式							
Z					50.2274	0.2034	49.8286	50.6261				
定数					-20.2162	0.2033	-20.6148	-19.8177				
個体数	-				4000				4000			
観測数	8502				8502				8502			

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.2557	0.2059	99.8522	100.6592	100.2557	0.2042	99.8555	100.6558
Xの個人内平均					109.2701	0.2306	108.8181	109.7221
Z	100.5305	1.3100	97.9630	103.0980	45.5543	0.2042	45.1540	45.9546
定数	-5.7765	1.3096	-8.3433	-3.2097	0.6671	0.1710	0.3319	1.0023
個体数	4000				4000			
観測数	8502				8502			

Note: Allisonモデルにおける個人内差分・個人内平均は、脱落していない部分（unbalanced data）から算出した。脱落した部分も含むbalanced dataから算出した場合とでは推定結果が異なる。

表7. 構造Aで脱落がNI的であるデータでの推定結果

変数	プールされたOLS				固定効果モデル				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限			下限	上限
X	106.5335	0.1636	106.2127	106.8542	100.3114	0.2139	99.8921	100.7307	104.2883	0.1638	103.9673	104.6093
Z	46.8585	0.1671	46.5310	47.1861	(dropped)				47.9837	0.2010	47.5898	48.3777
定数	3.7637	0.1459	3.4776	4.0497	5.5612	0.1141	5.3375	5.7848	2.4445	0.1825	2.0867	2.8022
					FE+OLS方式							
Z					50.0236	0.2069	49.6180	50.4293				
定数					-4.2711	0.2068	-4.6765	-3.8656				
個体数	–				4000				4000			
観測数	7897				7897				7897			

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.3114	0.2138	99.8923	100.7305	100.3114	0.2119	99.8961	100.7267
Xの個人内平均					109.0381	0.2316	108.5843	109.4920
Z	101.3850	1.3424	98.7539	104.0160	45.5593	0.2107	45.1463	45.9724
定数	2.4332	1.3416	-0.1964	5.0627	2.1279	0.1746	1.7858	2.4701
個体数	4000				4000			
観測数	7897				7897			

Note: Allisonモデルにおける個人内差分･個人内平均は、脱落していない部分（unbalanced data）から算出した。脱落した部分も含むbalanced dataから算出した場合とでは推定結果が異なる。

表8.　構造BにおけるMAR的な脱落の状況（表10で用いられるデータの例）

時点	回答率
1	100.0%
2	49.7%
3	33.7%
4	25.2%

Note:　回答率は当初の回答者の数である4000を分母とするものである。

表9.　構造BにおけるNI的な脱落の状況（表11で用いられるデータの例）

時点	回答率
1	100.0%
2	51.1%
3	28.6%
4	17.8%

Note:　回答率は当初の回答者の数である4000を分母とするものである。表5の数値と一致するのは乱数の種と脱落に影響する変数が同じであるからである。

表10. 構造Bで脱落がMAR的であるデータでの推定結果

変数	プールされたOLS				プールされたOLS（差分使用）				固定効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限			下限	上限
X	93.8953	0.1517	93.5980	94.1927					100.1437	0.1709	99.8087	100.4787
Xの個人内差分					100.1437	0.2224	99.7078	100.5796				
Xの個人内平均					89.6969	0.1823	89.3396	90.0542				
Z	50.1008	0.1532	49.8004	50.4012	50.0855	0.1424	49.8064	50.3646	(dropped)			
定数	-1.2049	0.1605	-1.5195	-0.8902	-0.8485	0.1495	-1.1415	-0.5555	13.2081	0.1104	12.9916	13.4246
									FE+OLS方式			
Xの個人内平均									-8.4177	0.2265	-8.8616	-7.9737
Z									50.0103	0.1933	49.6313	50.3893
定数									-13.8850	0.1945	-14.2663	-13.5037
個体数	–				–				4000			
観測数	8342				8342				8342			

変数	固定効果モデル（差分使用）				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
X					96.8316	0.1438	96.5498	97.1135
Xの個人内差分	100.1437	0.1709	99.8087	100.4787				
Xの個人内平均	(dropped)							
Z	(dropped)				49.9925	0.1969	49.6066	50.3785
定数	21.6008	0.1095	21.3862	21.8155	-0.6262	0.1995	-1.0172	-0.2353
	FE+OLS方式							
Xの個人内平均	83.8312	1.3180	81.2471	86.4152				
Z	49.1210	0.8839	47.3882	50.8539				
定数	-31.5296	0.8832	-33.2612	-29.7979				
個体数	4000				4000			
観測数	8342				8342			

(to be continued)

表10. *(continued)*

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.1437	0.1708	99.8089	100.4785	100.1437	0.1696	99.8114	100.4760
Xの個人内平均					90.9821	0.2254	90.5404	91.4238
Z	50.1564	1.2526	47.7014	52.6114	50.0373	0.1856	49.6736	50.4011
定数	-9.6071	1.2522	-12.0614	-7.1528	-0.7854	0.1880	-1.1540	-0.4169
個体数	4000				4000			
観測数	8342				8342			

Note: 個人内差分･個人内平均は、脱落していない部分（unbalanced data）から算出した。脱落した部分も含むbalanced dataから算出した場合とでは推定結果が異なる。

表11. 構造Bで脱落がNI的であるデータでの推定結果

変数	プールされたOLS				プールされたOLS（差分使用）				固定効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限			下限	上限
X	93.9617	0.1528	93.6621	94.2613					100.2543	0.1746	99.9119	100.5966
Xの個人内差分					100.2543	0.2325	99.7985	100.7101				
Xの個人内平均					90.1672	0.1806	89.8133	90.5212				
Z	49.9911	0.1568	49.6838	50.2985	50.0177	0.1463	49.7310	50.3045	(dropped)			
定数	2.0254	0.1571	1.7175	2.3334	2.6018	0.1475	2.3126	2.8910	1.9267	0.1120	1.7070	2.1463
									FE+OLS方式			
Xの個人内平均									-8.2537	0.2198	-8.6847	-7.8227
Z									49.9880	0.1935	49.6087	50.3673
定数									-0.5897	0.1934	-0.9688	-0.2107
個体数	–				–				4000			
観測数	7897				7897				7897			

変数	固定効果モデル（差分使用）				単純な変量効果モデル			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
X					96.8706	0.1445	96.5873	97.1539
Xの個人内差分	100.2543	0.1746	99.9119	100.5966				
Xの個人内平均	(dropped)							
Z	(dropped)				49.9170	0.1998	49.5254	50.3086
定数	17.1659	0.1088	16.9525	17.3793	1.4263	0.1994	1.0355	1.8172
	FE+OLS方式							
Xの個人内平均	87.0335	1.3547	84.3775	89.6896				
Z	50.1605	0.9085	48.3794	51.9416				
定数	-14.9113	0.9078	-16.6912	-13.1314				
個体数	4000				4000			
観測数	7897				7897			

(to be continued)

表11. *(continued)*

変数	Allisonモデル(1)				Allisonモデル(2)			
	係数	標準誤差	95%信頼区間		係数	標準誤差	95%信頼区間	
			下限	上限			下限	上限
Xの個人内差分	100.2543	0.1745	99.9122	100.5964	100.2543	0.1732	99.9148	100.5938
Xの個人内平均					91.3966	0.2203	90.9648	91.8285
Z	51.2070	1.2944	48.6700	53.7441	49.9825	0.1883	49.6134	50.3515
定数	2.5339	1.2937	-0.0017	5.0695	1.7544	0.1882	1.3855	2.1233
個体数	4000				4000			
観測数	7897				7897			

Note: 個人内差分･個人内平均は、脱落していない部分（unbalanced data）から算出した。脱落した部分も含むbalanced dataから算出した場合とでは推定結果が異なる。

表12. 構造Aで脱落がMAR的である場合の1000回の抽出シミュレーションの結果

	固定効果モデル	Allisonモデル(1)	Allisonモデル(2)
E(Beta-hat)	100.0031	100.0031	100.0031
E(Var-hat(Beta-hat))	0.04404	0.04402	0.04331
Var(Beta-hat)	0.04135	0.04135	0.04135
Prob[$L_{95\%C.I.}$<=Beta<=$U_{95\%C.I.}$]	0.960	0.960	0.959

Note: 構造Aで個体数が4000、時点数が4のパネルデータを理論母集団から1000回抽出し、MAR的な脱落を発生させ、固定効果モデル･Allisonモデル(1)･Allisonモデル(2)を推定した際の、XないしXの個人内差分の係数と誤差分散を記録し、その標本分布を示した。

表13. 構造Aで脱落がNI的である場合の1000回の抽出シミュレーションの結果

	固定効果モデル	Allisonモデル(1)	Allisonモデル(2)
E(Beta-hat)	99.998051	99.998051	99.998051
E(Var-hat(Beta-hat))	0.04795	0.04794	0.04694
Var(Beta-hat)	0.04947	0.04947	0.04947
Prob[$L_{95\%C.I.}$<=Beta<=$U_{95\%C.I.}$]	0.952	0.952	0.948

Note: 構造Aで個体数が4000、時点数が4のパネルデータを理論母集団から1000回抽出し、NI的な脱落を発生させ、固定効果モデル･Allisonモデル(1)･Allisonモデル(2)を推定した際の、XないしXの個人内差分の係数と誤差分散を記録し、その標本分布を示した。

表14. 構造Bで脱落がMAR的である場合の1000回の抽出シミュレーションの結果

	固定効果モデル	Allisonモデル(1)	Allisonモデル(2)
E(Beta-hat)	100.00014	100.00014	100.00014
E(Var-hat(Beta-hat))	0.03049	0.03047	0.03003
Var(Beta-hat)	0.02900	0.02900	0.02900
Prob[$L_{95\%C.I.}$<=Beta<=$U_{95\%C.I.}$]	0.958	0.958	0.956

Note: 構造Bで個体数が4000、時点数が4のパネルデータを理論母集団から1000回抽出し、MAR的な脱落を発生させ、固定効果モデル･Allisonモデル(1)･Allisonモデル(2)を推定した際の、XないしXの個人内差分の係数と誤差分散を記録し、その標本分布を示した。

表15. 構造Bで脱落がNI的である場合の1000回の抽出シミュレーションの結果

	固定効果モデル	Allisonモデル(1)	Allisonモデル(2)
E(Beta-hat)	99.998409	99.998409	99.998409
E(Var-hat(Beta-hat))	0.03197	0.03195	0.03145
Var(Beta-hat)	0.03298	0.03298	0.03298
Prob[$L_{95\%C.I.}$<=Beta<=$U_{95\%C.I.}$]	0.952	0.952	0.950

Note: 構造Bで個体数が4000、時点数が4のパネルデータを理論母集団から1000回抽出し、NI的な脱落を発生させ、固定効果モデル･Allisonモデル(1)･Allisonモデル(2)を推定した際の、XないしXの個人内差分の係数と誤差分散を記録し、その標本分布を示した。

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ともない、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、高卒パネル調査の３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

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
http://ssjda.iss.u-tokyo.ac.jp/panel/