パネルデータ分析の新展開

北村行伸[§]
一橋大学経済研究所
2002年10月16日

要　旨

本論文は最近のパネルデータ分析の考え方、その意義、手法についてサーベイを行い、後半でパネルデータ分析の応用例を紹介する。パネルデータ分析の重要な点は、第一にデータの性質に応じて適切な分析手法を選択すべきであるということ、第二にパネルデータ分析手法選択に当たっては各種の検定テストを徹底的に行って、できるだけ頑強な推定を行うべきであるということ、第三にパネルデータは一般に不完備であるが、情報量を失うことのないようになるべくデータを利用する方法を考えるべきであること、第四にミクロ経済学の理論を直接検定することによってミクロ経済理論へのフィードバックを行うことができるということである。最後に、パネルデータを上手く利用すれば理論と整合的な政策分析が可能になるし、逆に政策分析にはパネルデータが不可欠になるということである。

キーワード： パネルデータ、固定効果、ランダム効果、ダイナミックパネル推定、不完備パネルデータ
JEL classification: C8, C23, C33

[§]本論文の作成にあたっては多くの人から恩恵を受けている。まずパネルデータ分析の計量経済学的側面に関しては Cheng Hsiao 教授、Richard Blundell 教授から常に刺激を受けている。共同研究者としては、藤木裕氏、中村恒氏、坂本和靖氏の協力に感謝したい。とりわけ藤木裕氏は私をパネルデータ分析の世界に導いてくれただけではなく、その後も共同研究を続けていただいていることに対して感謝の意を評したい。論文の作成にあたっては、いつもながら原美起氏、馬場路子氏の協力を得たことに対して感謝したい。

1 はじめに

近年、パネルデータが利用可能になり、実証研究でも盛んに使われるようになってきた。パネルデータ分析の手法についても日々新たなアプローチが提案されている。パネルデータ分析に関する主要な文献としてはMaddala(1993)、Mátyás and Sevestre (1996)、Hsiao(2002)、Baltagi(2001)、Lee(2002)、Wooldridge(2002) などを挙げることができるが、Journal of Econometrics、Econometrica などではパネルデータ推定に関する論文が頻繁に掲載されている。また、Greene(2000)、Maddala(2001)、Johnston and DiNardo(1997)、Wooldridge(2000) などの計量経済学の教科書でもパネルデータ分析の手法が紹介されるようになってきている。さらには、Handbook of Econometrics (North-Holland) には Chamberlain (1984) と Arellano and Honoré (2001) の 2 本のパネルデータに関するサーベイ論文が含まれている。

欧米と比べると、日本でのパネルデータ利用の歴史は浅く、日本語で書かれた展望論文は少ない[1]。そこで、本論文では最近のパネルデータ分析の考え方、その意義、手法についてサーベイを行ない、後半でパネルデータ分析の実際の応用について触れてみたい[2]。

潜在的にパネルデータ分析で重要となると思われる論点をあらかじめ列挙しておくと、(1) パネルデータの新規作成はもとより、既存のクロスセクション・データを連結してパネル化する作業は今後益々進むものと思われる。(2) 時系列分析の手法がパネルデータ分析に取り入れられることによってダイナミック・パネル分析が進む。(3) 検定テストが標準化される。(4) ミクロ経済理論の実証分析が進むことで、ミクロ経済学の理論へのフィードバックが可能になる。

論文の構成は以下の通りである。第 2 節でパネルデータ分析の基本的な考え方とその意義について議論したあと、第 3 節でパネルデータ分析手法のサーベイを行う。ここでは、固定効果推定、ランダム効果推定、モデル選択のための検定、ダイナミック・パネル推定、パネル単位根推定、質的従属変数パネル推定、不完備パネルデータについて扱う。第 4 節では、パネルデータ分析の応用について論じる。具体的には、筆者がこれまで行ってきたパネルデータ分析の中から、特徴的なアプローチを選んで説明する。最後に第 5 節では全体を取りまとめる。

[1]数少ない例外として北沢 (2001) がある。

[2]理解しやすいようにできるだけスカラー表現を用いてベクトル表現は控えた。実際の式の導出やコンピュータ・プログラムについては省略したが、機会を改めてそれらの結果も報告するつもりである。

2 パネルデータ分析の考え方とその意義

パネルデータ分析の基礎は近代統計学の生みの親である Ronald A. Fisher の一連の研究（1932, 1971, 1973a, 1973b)、とりわけ分散分析にある。分散分析とは、ある処理を与えた処理群（treatment group）と対照群（controll group）を比較して、処理の効果を明らかにするための手法であり、二つの群の分散を比較するという検定を行うことから分散分析と呼ばれている。Ronald A. Fisher はこのアプローチを広い意味で実験計画法の中で用いている。

実験計画法とは目的の結果を導出するために、実験をいかに効率的に計画するかという問題にかかわるものであり、実験結果を左右するような因子を適切に選び出しそれを組み合わせ、個々の因子の影響を抽出できるような実験を、なるべく効率的に行うように実験をデザインするということである[3]。Fisher はこのような実験計画に関して 3 原則を提唱している。それは、(1) 反復、(2) 無作為化、(3) 局所管理である。第一に、1 つの処理に対してもデータはばらつくので、そのばらつきを評価できるように何度か同じ処理に対して実験を行う必要がある、第二に系統的な誤差を確率的な誤差に転化するために実験を無作為化する必要がある、第三に、実験全体の無作為化や均一化が難しい場合には、局所的に均一化するような管理が必要になる。

ところで、経済は一般に管理実験ができない。とすれば実験計画法がどのように役立つのだろうか。我々の考えではパネルデータ分析とは、パネル経済データをあたかも実験データのように扱う点に特徴がある。すなわち、経済には様々なショックや確率的な変動が連続的に生じており、分析対象としたい経済関係に与える影響を様々な因子（要因）に分類してコントロールした上で、分析するという意味で実験計画法の手法が利用できるのである。具体的には、以下で論じる誤差構成要素モデル (erro components model) とは、誤差項を純粋な撹乱項とそれ以外の因子による誤差を分離しようとするものであり、こうすることによって、無作為化と局所管理ができることになる。現実の経済データをあたかも管理実験データであるかのように扱い、その結果、様々な属性をもった heterogeneous な主体それぞれの属性をコントロールしながら、最も関心の高い説明変数の被説明変数への効果を抽出することができるようになるということである。

パネルデータの基本構造は図 1 で表せる。プーリング・データとは時系列、クロスセクションのデータを全て合体して全ての変数が共通の母集団から発生していると考えて、データを一括して扱うケースである。ビトウィーン・データとは、プーリング・データに近い考え方だが、時系列方向に個別主体毎の平均を取り、それをクロスセクション・データとして分析するものである。このデータの扱い方は一回限りのクロスセクション・データでは個別主体が特定の時間効果を受けているために推定にバイアスがかかる恐れがあるが、個別主体について時系列方向で何回分かのデータ集めて平均をとれば、

[3] 以下の議論は広津（1992、第 1 章）を参照している。

そのような特定時点の効果を緩和することができるという考え方に基づいている。このデータでは時系列方向の変動ではなく、個別主体間の違いを見ることに主眼をおいたものである。それに対して、時系列データあるいはウィズイン・データとは個別主体毎の時系列方向のデータのみを扱うもので、データが時系列内で大きく変動する場合には、プーリング・データやビトウィーン・データとして扱うことは出来ない。

このような関係を数式で表すと次のようになる。

$$y_{it} = \alpha + X_{it}^{\prime}\beta + u_{it} \quad i = 1, ...N;\ t = 1, ...T \tag{1}$$

ここで i は個別経済主体（例えば、個人、家計、企業、国家）を表し、クロスセクション方向の情報であり、ｔ は時間を表し、時系列方向の情報を与える。誤差に関して一般的な二元配置誤差構成要素（two-way error component）モデルを想定する[4]。

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \tag{2}$$

ここで、μ_i は観察不可能な経済主体独自の個別効果を表し、λ_t は観察不可能な時間効果、ν_{it} は撹乱項を表す。$\lambda_t = 0$ の場合は、(2) 式は一元配置誤差構成要素 (one-way error component) モデルととなる。

$$u_{it} = \mu_i + \nu_{it} \tag{3}$$

(1) 式のようなモデルに対して、先ず、利用可能なデータをクロスセクション、時系列に関係なく無差別にプーリングした上で OLS 推定を行う (pooling estimation)。これは全ての経済主体が同じ定数項、同じ傾きを持つと仮定しているモデルであり、個別の異質性、ダイナミズムは存在しないことを意味する。第二に、経済主体の異質性を考慮して、モデルの傾きは同一だが、定数項がそれぞれの主体で異なっているという一元配置固定効果推定法（one-way fixed effect estimation）で推計してみる。この場合、固定効果としてダミー変数が入ってくるので、最小二乗ダミー変数モデル (Least Squares Dummy Variable Model ; LSDV) と呼ばれる推定方法を用いる。第三に、定数項が個別に固定的なものというよりランダムに決まっていると考えると一元配置ランダム効果推定法 (one-way random effect estimation) を用いる。ここでは個別ランダム効果が説明変数と無相関であることを仮定して、誤差項の分散共分散行列を勘案して、変換した $y_{it} - \theta\overline{y}_i$ を $X_{it} - \theta\overline{X}_i$ 上で回帰する一般化最小二乗法 (Generalized Least Squares; GLS) を用いる。ここで θ は個別ランダム効果と撹乱項の加重比を表す。第四に、一元配置固定効果推定法や一元配置ランダム効果推定法のそれぞれに、年毎に生じた共通のショックの効果を取り除くために時間（年）ダミーを導入することもある。これらはそれぞ

[4]原理的には n 次元配置誤差構成要素モデルを考えることは可能だが（例えば、個別主体、時間、地域、コーホート、産業などの誤差要素が考えられる）、計量経済学の標準的な説明としては二元配置モデルを扱うのが一般的であので、ここでもそれに従っている。

れ、二元配置固定効果推定法と二元配置ランダム効果推定法と呼ばれる。これはサンプル期間中に生じた経済全体に影響を与えた景気循環や構造変化などの影響をコントロールしようとするものである。このようにモデルを拡張していき、それぞれのモデルが与えられたパネルデータにどのように適合するかを検定して、適切にパネルデータを利用することが重要になってくる。

3　パネルデータ分析手法のサーベイ

パネルデータ分析の手法としては、大きく、静学的アプローチと動学的アプローチに分けることができる。静学的アプローチとしては固定効果推定法、ランダム効果推定法があり、動学的アプローチには操作変数法によるものと、一般化積率法 (GMM) によるものがある。また動学的アプローチの拡張としてパネル単位根推定がある。さらに進んだトピックとしてはクロスセクション・データ分析の手法を応用した質的従属変数パネル推定についても概観する[5]。また、パネルデータ分析の場合、モデル選択が重要になってくるので、モデル選択検定法についても説明する。

3.1　固定効果推定

先の (1) 式と (2) 式で表されるパネルモデルをさらに簡単化 ($\alpha = 0, \lambda_t = 0$ と仮定) して固定効果推定を説明する。基本的な考え方は次のようにまとめることが出来る。

$$y_{it} = \beta x_{it} + \mu_i + \nu_{it} \tag{4}$$

ここで μ_i は固定効果を表す。$\nu_{it} \sim IN(0, \sigma^2)$ とする。

まず、モデル全体（プーリング）の平均を $\tilde{x}$ と $\tilde{y}$ と定義すると

$$\tilde{x} = \frac{1}{TN}\sum_{t=1}\sum_{i=1} x_{it} \tag{5}$$

$$\tilde{y} = \frac{1}{TN}\sum_{t=1}\sum_{i=1} y_{it} \tag{6}$$

この変数から各個体の残差平方和と x と y それぞれに関する残差を相乗し

[5]本論文では、連立方程式パネル推定、可変パラメータ・パネル推定などは扱わない。

たものを次のように定義する。

$$S_{xx}^{pool} = \sum_{i=1}\sum_{t=1}(x_{it} - \tilde{x})^2 \tag{7}$$

$$S_{xy}^{pool} = \sum_{i=1}\sum_{t=1}(x_{it} - \tilde{x})(y_{it} - \tilde{y})^2 \tag{8}$$

$$S_{yy}^{pool} = \sum_{i=1}\sum_{t=1}(y_{it} - \tilde{y})^2 \tag{9}$$

同一個体内で被説明変数と説明変数の時系列 (ウィズイン) 平均を次のように定義する。

$$\bar{x}_i = \frac{1}{T}\sum_t x_{it} \tag{10}$$

$$\bar{y}_i = \frac{1}{T}\sum_t y_{it} \tag{11}$$

この変数からの個体内（ウィズイン）時系列平均の残差平方和と x と y それぞれに関する残差を相乗したものを次のように定義する。

$$S_{xx}^{with} = \sum_{i=1}\sum_{t=1}(x_{it} - \bar{x}_i)^2 \tag{12}$$

$$S_{xy}^{with} = \sum_{i=1}\sum_{t=1}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \tag{13}$$

$$S_{yy}^{with} = \sum_{i=1}\sum_{t=1}(y_{it} - \bar{y}_i)^2 \tag{14}$$

次に、個体間（ビトウィーン）の残差平方和を次のように定義する。

$$S_{xx}^{btw} = \sum_{i=1} T(\bar{x}_{it} - \tilde{x})^2 \tag{15}$$

$$S_{xy}^{btw} = \sum_{i=1} T(\bar{x}_{it} - \tilde{x})(\bar{y}_i - \tilde{y}) \tag{16}$$

$$S_{yy}^{btw} = \sum_{i=1} T(\bar{y}_i - \tilde{y})^2 \tag{17}$$

以上より、次のような関係が導かれる。

$$S_{xx}^{pool} = S_{xx}^{with} + S_{xx}^{btw} \tag{18}$$

$$S_{xy}^{pool} = S_{xy}^{with} + S_{xy}^{btw} \tag{19}$$

$$S_{yy}^{pool} = S_{yy}^{with} + S_{yy}^{btw} \tag{20}$$

推計される β に関しては、プーリング推定（OLS）（β^{pool}）、ウィズイン推定（β^{with}）、ビトウィーン推定（β^{btw}）に対しておのおのの次のような関係が導かれる。

$$\beta^{pool} = [S_{xx}^{pool}]^{-1} S_{xy}^{pool} = [S_{xx}^{with} + S_{xx}^{btw}]^{-1}[S_{xy}^{with} + S_{xy}^{btw}] \tag{21}$$

$$\beta^{with} = [S_{xx}^{with}]^{-1} S_{xy}^{with} \tag{22}$$

$$\beta^{btw} = [S_{xx}^{btw}]^{-1} S_{xy}^{btw} \tag{23}$$

これより次の式を得る。

$$S_{xy}^{with} = S_{xx}^{with}\beta^{with} \tag{24}$$

$$S_{xy}^{btw} = S_{xx}^{btw}\beta^{btw} \tag{25}$$

これを (21) に代入すると次のように整理できる。

$$\beta^{pool} = m^{with}\beta^{with} + m^{btw}\beta^{btw} \tag{26}$$

ここで

$$m^{with} = [S_{xx}^{with} + S_{xx}^{btw}]^{-1}S_{xx}^{btw} = I - m^{btw} \tag{27}$$

(26) 式はプーリング推定（OLS）はウィズイン推定とビトウィーン推定の加重平均であることを示している。もしウィズイン推定の変動が小さく m^{with} が小さい場合には、ビトウィーン推定とプーリング推定は近似してくる。

3.2 ランダム効果推定

固定効果推定では各主体に対してダミーを割り当てるために、N が大きくなれば、推定すべきパラメータの数が膨大なものになり、その結果、推定における自由度は著しく低下する。固定効果 μ_i をランダム（確率変数、$\mu_i = IID$（0、σ_μ^2)）だと仮定すれば、この問題は回避できる。ランダム効果推定を用いるのに適しているのは母集団から N 個を無作為に抽出したような場合である[6]。

ランダム効果モデルでは、固定効果 μ_i を確率変数として扱う。μ_i は撹乱項 u_{it} から独立している。

$$\mu_i \sim IID(0, \sigma_\mu^2) \tag{28}$$

$$u_{it} \sim IID(0, \sigma_u^2) \tag{29}$$

単純化のために説明変数はひとつとし、μ_i を確率変数とする。誤差項は次のように表せる。

$$v_{it} = \mu_i + u_{it}$$

ここで

$$\begin{aligned} cov(v_{it}, v_{is}) &= \sigma_u^2 + \sigma_\mu^2 \quad for\ t = s \\ &= \sigma_\mu^2 \qquad for\ t \neq s \\ cov(v_{it}, v_{js}) &= 0 \qquad for\ \forall_{t,s} \quad if\ i \neq j \end{aligned}$$

[6]家計に関するサンプル調査等がこれに相当する。

これは同一主体内の誤差項 μ_i と u_{it} が相関していることを表し、効率的な推定値を得るためには一般化最小二乗法（generalized least square: GLS）を用いる必要がある。Maddala (1971a,b) に従って以下のように簡略化できる。

$$\beta^{rnd}_{GLS} = \frac{S^{with}_{xy} + \theta S^{btw}_{xy}}{S^{with}_{xx} + \theta S^{btw}_{xx}},\ \theta = \frac{\sigma^2_u}{\sigma^2_u + T\sigma^2_\alpha} \tag{30}$$

先の (21)(22)(23) より

$$\beta^{pool} = \frac{S^{pool}_{xy}}{S^{pool}_{xx}}$$

$$\beta^{with} = \frac{S^{with}_{xy}}{S^{with}_{xx}}$$

であり、プーリング推定（OLS）とウィズイン推定（LSDV：least squares with dummy variables）はランダム効果推定（GLS）においてそれぞれ $\theta = 1$ と $\theta = 0$ に相当する特殊ケースであることがわかる。

3.3　モデル選択のための検定

計量経済学の検定には一般に二つの意味がある。第一に、理論モデルが正しいかどうかを検定するということである。これは、理論的に演繹された関係式が統計的に支持されるかどうかを確認する作業であり、仮説検定の基礎にある考え方である。しかし、すでに第 2 節で述べたように、パネルデータ分析は多様な経済主体の多様な状況に対する反応を含んでいるという意味で、同質的な経済主体の行動関係を仮説検定するということだけでは不十分である。

第二には、与えられたデータに対して適切な推定方法が用いられているかどうかを検証するということである。これは、実験計画法あるいは分散分析を基礎にして、集められたデータから、様々なノイズやコントロールできる要因を取り除いていく作業の一環と考えることができる。一般にパネルデータは時系列データやクロスセクションデータよりも多くの観察点を含み、事前に全体的な傾向を把握することが難しいことが多い。このような場合、手元にあるパネルデータをどのような手法で分析するのが望ましいかは、データ自体に決めさせるというのが正当な考え方である。そのためには、システマティックな検定手続きが必要になる[7]。

以下では第 2 節で説明した様々な推定法に基づく静学モデルを対象にしたモデル選択の手順を説明する[8]。

[7] もちろん、データ自体に決めさせると言っても、理論モデル設定が悪く、用いられた変数に間違えがあれば、統計量は不十分となる。ここでは不均一分散や系列相関、誤差項の正規性について論じる余裕はないが、これらの問題は基本的にはモデルの特定化に問題があることによって生じていると考えられるので、統計的にこれらの問題が検出されれば、理論モデルに戻って、検討し直す必要があり、小手先の修正は勧められない。

[8] パネルデータは年次データであったり、あるいは 5 年に一回のデータであったりして、データ頻度が低いことが多く、実際には静学モデルを用いることが多い。むしろ、年次データに動学モデルを当てはめても、データ自体が動学的な調整過程を捉えていない可能性が高い。

第一に、固定効果推定法が正当化されるかどうかは、先ず、一元配置固定効果推定法（LSDV）とプーリング推定法（OLS) を比べる。これは一元配置固定効果推定法における経済主体別の定数項が全て等しいという制約が課される場合がプーリング推定法であるとして、その制約について F 検定でテストする。つまり、制約が無効であるとの帰無仮説が棄却されると、一元配置固定効果推定法が正当化される。次に、二元配置固定効果推定法がプーリング推定法や一元配置固定効果推定法に対して正当化されるかどうかも同様に係数制約問題に関する F 検定を通して判断できる。

第二に、ランダム効果推定法がプーリング推定法に対して正当化されるかどうかについてはラグランジュ乗数法 (Lagrange Multiplier test) を用いる。すなわちプーリング推定法（OLS）の誤差項が平均的にゼロであるとの帰無仮説についてラグランジュ乗数統計量をもとめカイ二乗検定を行うのである。この結果、帰無仮説が棄却されると、誤差項に経済主体毎の異質性が存在することを意味し、ランダム効果推定法（GLS）が正当化される。

第三に、固定効果推定法とランダム効果推定法の間のモデル選択は Hausman test を用いる。これは、個別主体要因が、説明変数と無相関であるとの帰無仮説を立て、それをカイ二乗検定するものである。仮説が棄却されると固定効果推定法が正当化されることとなる。

上述の手続きをトーナメント方式で順次行って行くことによって最適な推定方法が選択できる。モデル選択の構造は図 2 にまとめられている。具体的な検定テストの考え方は以下で順に説明していきたい。

3.3.1　F 検定

これまでの議論から明らかなように、まず時系列推定法、すなわち各経済主体別の定数項および傾きが全てばらばらの場合と、それらが全て等しいという制約が課されるプーリング推定法の場合について F 検定でテストする。

因みに、ここで用いる F 検定は分散分析で用いる F 検定に準じるものである。すなわち、ある帰無仮説の下で推定される残差平方和（RSS_0）を自由度 (ν_0) で割ったものは、その自由度のカイ二乗分布に従い、対立仮説の下でも同様に残差平方和 (RSS_1) を自由度 (ν_1) で割ったものを求め、帰無仮説と対立仮説の比をとったものは F（ν_0, ν_1）分布に従うという関係を用いる。

$$F(\nu_0, \nu_1) = \frac{RSS_0/\nu_0}{RSS_1/\nu_1} \tag{31}$$

具体的には、全ての定数項と傾きが共通であるとの帰無仮説の下に、第一自由度 $(N-1)(k+1)$、第二自由度 $(NT - N(K+1))$ の F 分布に従う。こ

の検定量は次のように表せる。

$$F(pool\ vs\ time\ series) = \frac{(RSS_{Pool} - RSS_{TimeSeries})/(N-1)(k+1)}{(1 - RSS_{TimeSeries})/(NT - N(k+1))} \tag{32}$$

ここで、RSS_{pool} はプーリング推定法の残差平方和、$RSS_{TimeSeries}$ は時系列推定法の残差平方和を表している。他の変数はすでに定義した通りである。

次に、単純なプーリング推定法が棄却された場合、各主体の定数項はばらばらだが、傾きは等しいという一元配置固定効果推定法を帰無仮説として、時系列推定法を F 検定でテストする。具体的な F 分布は次のようになる。

$$F(oneway\ fixed\ vs\ time\ series) = \frac{(RSS_{of} - RSS_{TimeSeries})/(N-1)}{(1 - RSS_{TimeSeries})/(N(T-1)k)} \tag{33}$$

ここで、RSS_{of} は一元配置固定効果推定法の残差平方和である。ここで帰無仮説が棄却されなければ、一元配置固定効果推定法が採択される。

さらに、一元配置固定効果推定法が時系列推定法に対して採択されたとして、これがプーリング推定法に対して正当化されるかどうかを検定する。すなわち、一元配置固定効果推定法において定数項が全て等しい場合がプーリング推定法である。このとき第一自由度 $(N-1)$、第二自由度 $(NT-(N+1)$) の F 分布に従う。

$$F(pool\ vs\ oneway\ fixed) = \frac{(RSS_{pool} - RSS_{of})/(N-1)}{(1 - RSS_{of})/(NT - (N+1))} \tag{34}$$

最後に、一元配置固定効果推定法がプーリング推定法に対して採択されたとして、これが二元配置固定効果推定法に対して正当化されるかどうかを検定する。すなわち、二元配置固定効果推定法における時間ダミーのパラメータがすべてゼロである場合が一元配置固定効果推定法となるとすると、F 分布は第一自由度 $(T-1)$、第二自由度 $(NT-(N+1))-(T-1))$ に従う。

$$F(oneway\ fixed\ vs\ twoway\ fixed) = \frac{(RSS_{of} - RSS_{tf})/(T-1)}{(1 - RSS_{tf})/(NT - (N+1) - (T-1))} \tag{35}$$

ここで、RSS_{tf} は二元配置固定効果推定法の残差平方和である。

このように、F 検定を順次行うことによって時系列 (個別) 検定とプーリング検定、一元配置固定効果推定、二元配置固定効果推定の間に序列をつけることが出来る。

3.3.2　Hausman 検定

Hausman 検定はモデル特定化を検証するために用いられている。帰無仮説 H_0 はモデルの特定化が正しいというものであり、対立仮説 H_1 はモデルの特

定化に誤りがあるというものである。これは、簡単に言えば、二つの仮説に基づいて推定されたパラメータが等しいかどうかを検定して、等しくなければ、モデルの特定化に問題があるということになる。

次のようなモデルを考えよう[9]。

$$y = \beta x + u \tag{36}$$

OLS 推定するためには、x は u から独立していなければならない。仮説検定の形で表すと次のようになる。

$$H_0 : x \text{ と } u \text{ は互いに独立}$$
$$H_1 : x \text{ と } u \text{ は互いに依存}$$

ここで、Hausman 検定のための設定として、β に関して次のような二つの推定が得られたとしよう。

$\hat{\beta}_0$は H_0の下で一致かつ有効推定であるが、H_1の下では一致推定ではない。
$\hat{\beta}_1$は H_0でも H_1の下でも一致推定であるが、H_0の下では有効推定ではない。

そこで $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$ と定義する。ここから、Hausman は次の関係を導いた[10]。

$$var(\hat{q}) = var(\hat{\beta}_1) - var(\hat{\beta}_0) \quad H_0\text{の下でそれぞれの分散を推定}$$

$\hat{V}(\hat{q})$ を $var(\hat{q})$ の一致推定とすると、次の統計量は自由度 1 のカイ二乗分布に従うかどうかで帰無仮説 H_0 を検定できる。

$$m = \frac{\hat{q}^2}{\hat{V}(\hat{q})} \sim \chi^2(1) \tag{37}$$

具体的にパネルデータ分析に関して、固定効果かランダム効果のどちらが望ましいかというテストをするためには次のような仮説を検定することになる。

$$H_0 : \text{ランダム効果は説明変数 } x_{it} \text{と相関していない}$$
$$H_1 : \text{ランダム効果は説明変数 } x_{it} \text{と相関している}$$

H_0 の下ではランダム効果推定（$\hat{\beta}_r$）が有効一致推定である。固定効果推定 $\hat{\beta}_f$ は帰無仮説に関係なく一致推定となる。ここで $q = \hat{\beta}_f - \hat{\beta}_r$ と定義し、

[9] 以下の説明は Maddala(2001、pp.494-495）に従っている。

[10] まず、$var(\hat{q}) = var(\hat{\beta}_1) - var(\hat{\beta}_0)$ が成り立つためには、$cov(\hat{\beta}_0, \hat{q}) = 0$ を証明しなければならない。この証明は Maddala(2001, pp.495-496) で与えられている。

$V(q) = V(\hat{\beta}_f) - V(\hat{\beta}_r)$ を Hausman 検定、$m = \hat{q}'[\hat{V}(\hat{q})]^{-1}\hat{q} \sim \chi^2(k)$ を計算することになる。β が $k \times 1$ ベクトルであり、V_1 と V_0 が行列式で表される場合には、上の Hausman 検定量は次のように書き換えられる。

$$m = \hat{\mathbf{q}}'[\hat{\mathbf{V}}(\hat{\mathbf{q}})]^{-1}\hat{\mathbf{q}} \sim \chi^2(k) \tag{38}$$

結果はスカラーの場合と全く同じである[11]。

3.3.3 Breusch-Pagan 検定

もう一つのモデル特定化テストとしては固定効果の分散がゼロかどうかを検定する Breusch-Pagan 検定がある。

次のようなモデルを考えよう。すなわち、定数項 μ は個別固定効果 μ_i ではなく、すべての i に対する定数である。

$$y_{it} = \mu + \mathbf{x}_{it}\beta + u_{it} \tag{39}$$

これは個別固定効果が存在しないので最小二乗法推定ができる。その残差項を $\hat{u}_{it}$ とすると、つぎのような統計量を定義できる。

$$S_1 = \sum_{i=1} \left(\sum_{t=1} \hat{u}_{it} \right)^2 \tag{40}$$

$$S_2 = \sum_{i=1}\sum_{t=1} \hat{u}_{it}^2 \tag{41}$$

これに対して Lagrange Multiplier (LM) 統計を次のように定義すると、この統計量は自由度 1 のカイ二乗分布に従うはずである。

$$LM = \frac{NT}{2(T-1)} \left(\frac{S_1}{S_2} - 1 \right)^2 \sim \chi^2(1) \tag{42}$$

この検定が有意であれば、個別固定効果 (プーリング推定) を棄却し、ランダム効果推定を採択することが望ましいことを意味する。

Hausman 検定と Breusch-Pagan 検定をあわせるとプーリング推定、ランダム効果推定、固定効果推定の間に序列をつけることが出来る。

[11] ここで用いる k は推計すべきパラメータ β から定数項と時間ダミーを除いたパラメータの数を表している。

3.4 ダイナミック・パネル推定

近年、経済主体は動学的最適化に基づいて行動しているという経済理論が主流であり、実証研究でも経済主体のダイナミクな調整に関心が集まっている。

一般にパネル・データでダイナミックな関係とは、被説明変数のラグが説明変数に入っていることをさす。すなわち、

$$y_{it} = \delta y_{it-1} + x'_{it}\beta + u_{it} \quad i = 1, 2, ..., N; t = 1, 2, ...T \tag{43}$$

ここで、δ はスカラー、x'_{it} は $1 \times K$ 行列、β は $K \times 1$ 行列。u_{it} は一元配置誤差構成要素モデルに従っているとする。

$$u_{it} = \mu_i + \nu_{it} \tag{44}$$

ここで、$\mu_i \sim IID(0, \sigma^2_\mu)$ は個別固定効果を表しており, $\nu_{it} \sim IID(0, \sigma^2_\nu)$ は撹乱項を表し、相互に独立である。

ダイナミック・パネル推定を巡る大きな問題はラグ被説明変数が誤差項 u_{it} と相関していること、そしてデータがクロスセクション方向（N）には大きいが、時系列方向（T）には小さいということである[12]。これは撹乱項 v_{it} が系列相関していない場合にも当てはまる。この問題に対しては二つの解決方法が提案されている。一つは Anderson and Hsiao (1981)、Arellano(1989)、Hahn,Hausman and Kuerteiner (2002) らによる操作変数法であり、いま一つは、Arellano and Bond (1991)、Ahn and Schmidt (1995) らによる一般化積率法（GMM）である[13]。

固定効果推定であれランダム効果推定であれ、上の (19) 式から一階の階差をとれば μ_i は消去されてしまう。すなわち、

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})'\beta + \delta(y_{it-1} - y_{it-2}) + (\nu_{it} - \nu_{it-1}) \tag{45}$$

このモデルはラグ被説明変数の階差が撹乱項 v_{it} の階差と相関しているという意味では問題が残っているが[14]、操作変数法を用いて推定することで内生性バイアスを取り除くことができる。すなわち、有効ではないが一致推定を得ることができる。具体的には $(y_{it-2} - y_{it-3})$ かそれぞれのラグ変数、y_{it-2} 、y_{it-3} が $(y_{it-1} - y_{it-2})$ に対する操作変数として使われる[15]。

[12]時系列が短いという問題に対しては一般に時間軸は長くなくてもよいと考えることができる。むしろ経済主体のダイナミックな調整パラメータは時間と共に変化する可能性が高いので、それが一定とみなされる期間（例えば５年）ぐらいに限定したほうがいいとも言える。調整スピードが速い場合には１年以内に調整が終わり、前年の実績（ラグ変数）はほとんど説明力をもたないといケースもある（4.3 節参照）。

[13]このアプローチはさらに Arellano and Bover (1995), Blundell and Bond (1998) らによって拡張されている。

[14]具体的には y_{it-1} と ν_{it-1} は（19）式より明らかに相関している。

[15]Arellano(1989) はラグ変数の水準 y_{it-2} や y_{it-3} を操作変数として用いる方が望ましいとしている。

このアプローチに対して、Arellano and Bond (1991)、Ahn and Schmidt (1995) は操作変数法は重要な情報を用いていないので、有効でないと論じている。例えば、一階の階差モデルを想定すると、2 期ラグをとった y の水準は撹乱項の階差とは無相関であることを示すことができる[16]。

$$E[y_{is},(\nu_{it}-\nu_{it-1})]=0,\quad s=0,1,...t-2,\quad t=2,....T \tag{46}$$

これに対応した一般化積率法（GMM) は次のように表せる。

$$\frac{1}{n}\sum_{i=1}^{n} y_{is}[(y_{it}-y_{it-1})-(x_{it}-x_{it-1})'\beta-\delta(y_{it-1}-y_{it-2})]=0 \tag{47}$$

$$s=0,....,t-2,\quad t=2,....,T$$

Ahn and Schmidt(1995) は y の水準からだけではなく、y と撹乱項の階差 $(\nu_{it}-\nu_{it-1})$ との間からも重要な情報 (ここでは直交条件) が得られることを示している。これは次のように表せる。

$$E[y_{is}(\nu_{is+1}-\nu_{is})-y_{is+1}(\nu_{is+2}-\nu_{is+1})]=0 \tag{48}$$

$$E[(y_{it}-x_{it}{'}\beta)y_{it}-(y_{it-1}-x_{it-1}{'}\beta)y_{it-1}]=0 \tag{49}$$

$$t=2,....T$$

ダイナミック・パネル推定に関する操作変数法と一般化積率法を巡る論争は、現在最も活発に行われており、いまだに決着はついていない。例えば、Binder, Hsiao, and Pesaran (2000)、Hsiao, Pesaran and Tahmiscioglu (2002)、Hsiao (2002) などでは、理論的に直交条件を加えることで GMM 推定の有効性を増すことはありうるが、有限サンプルの下ではあまりに多くの直交条件を課すことには問題があり、実証的には下方バイアスが増すと論じている。また、操作変数法と GMM 推定に関するモンテカルロ実験（T=5、N=50、2500 回反復）の結果、最尤法は 1 %下方バイアスがあり、GMM は場合によっては 15 − 20 %の下方バイアスが見られる。操作変数法にもバイアスは見られるが GMM と比べると小さいことが示されている。

Hsiao, Pesaran and Tahmiscioglu (2002) や Fujiki, Hsiao and Shen (2002) で代替的に彼らが提示している推定方法は最小距離推定法 (Minimum Distance Estimation: MDE) と呼ばれるものである[17]。基本的な考え方は、誤差項の階差 2 次式を最小化するようにパラメータ (β,δ) を決定するということである。すなわち、

$$\min[\sum_{i=1}\triangle\nu_i^*\Omega^{-1}\triangle\nu_i^*] \tag{50}$$

[16] すなわち直交条件 (orthogonality conditions) が成立する。これは Holtz-Eakin(1988)、Holtz-Eakin, Newey and Rosen(1988) によって指摘された。

[17] MDE の詳細については Lee(2002,Chap3) を参照。

ここで Ω は $\triangle\nu_i^*$ の共分散行列、$\triangle\nu_i^* = [\triangle y_{i1} - \beta\triangle x_{i1} - \delta\triangle y_{i0}, \triangle y_{i2} - \beta\triangle x_{i2} - \delta\triangle y_{i1}, ...]$

この方法は有効ではないが、N が大きければ漸近的に一致推定となる。しかも計算ははるかに簡単になる。モンテカルロ実験の結果でも、MDE 推定のバイアスは少なくとも GMM 推定より小さい。推定値の平均平方誤差で比べても、MDE 推定は最尤法よりは大きいが、GMM 推定より小さい。Hahn, Hausman and Kuersteiner (2002) は階差の取り方を 1 階ではなく例えば 3 階（すなわち、$y_n - y_{n-3}$）のような長階差 (long differences;LD) をとる事で操作変数の説明力を高め、バイアスを引き下げ、結果として最小距離 (MDE) を縮めることができると論じている。

同時に、Arellano and Bond (1991)、Kiviet (1995)、Ziliak（1997)、Blundell and Bond (1998)、Alonso-Borrego and Arellano (1999) は GMM 推定に関してモンテカルロ実験を行い、一回階差の誤差に対するモーメント条件を用いた GMM 推定は、識別制約が弱く、しかもクロスセクション方向のサンプルサイズ N に対して相対的にモーメント条件が多すぎる場合には、バイアスが大きいことを確認している。とはいえ、Blundell, Bond and Windmeijer(2000) は「データの厳密な検証と GMM 推定の制約条件の適切な選択によってダイナミック・パネル・データに関する GMM 推定の問題は回避できる」だろうと主張している。

3.5 パネル単位根推定

近年の時系列分析の中では変数の定常性が重要な問題となっており、それを検定するための様々なテストが考案されている。パネルデータが非定常な変数である場合にも spurious 推定の問題が出てくる。

パネルデータは一般にクロスセクション方向に膨大なサンプルがあるために、時系列だけではサンプル数が不足して検定テストの精度が落ちるといった問題を回避できると考えられている。しかし、時系列の帰無仮説、対立仮説とパネル単位根検定とでは異なっている。以下ではいくつかの検定を紹介したい。

最もよく知られた検定は Levin-Lin (LL) test(1992,1993) であるが、他にも Im-Pesaran-Shin (IPS) test(1997)、Maddala-Wu (MW) test (1999) などが提案されている。

次のようなモデルを考えよう。

$$y_{it} = \rho y_{it-1} + e_{it} \quad i = 1, 2, ...N \tag{51}$$

一般に第 1 主体の単位根を検定する場合、t 値による単位根検定は次のよ

うに定義される。

$$H_0 : \rho_1 = 1 \quad vs \quad H_1 : \rho_1 < 1 \tag{52}$$

このようなテストの検定力は低いので、Levin-Lin (LL) test では次のような検定を提示した。

$$H_0 : \rho_1 = \rho_2 = ... = \rho_N = \rho = 1 \quad vs \quad H_1 : \rho_1 = \rho_2 = ... = \rho < 1 \tag{53}$$

これら 2 つの検定は帰無仮説も対立仮説も異なっており、代替的な検定とは言えない。O'Connell(1998) は、Levin-Lin test は同時点のクロスセクション内での誤差相関が推定上、大きな歪みを与えていることを指摘し、そのような誤差因子をコントロールする必要性を主張している。Im-Pesaran-Shin (IPS) test は次のような検定を提案し、Levin-Lin test の一般化であると主張している。

$$H_0 : \rho_i = 1 \; for \; all \; i \quad vs \quad H_1 : \rho_i < 1 \; at \; least \; one \; i$$

しかし Maddala (2001, p.554) で指摘されているように、これは N 個の単位根検定を個別に行っていることと同値であり、Levin-Lin test はすべての主体に対して単位根があるという複合仮説を検定していることになる。

それぞれの単位根検定が Augmented Dickey-Fuller test によって同じラグ構造の下で検定されているとすれば、N 主体それぞれの t 統計は平均 M で分散 σ^2 の分布に従い、t 統計全体の平均 $\bar{t}$ は平均 M、分散 σ^2/N の分布に従う。Maddala-Wu test は N 主体の独立した単位根検定を集計して検定するというもので、個別検定を集計して検定するという Ronald A. Fisher (1973a) のアイディアを応用したものである。すなわち、P_i を i 主体の単位根検定の有意水準に関する P 値とすると、$\lambda = -2 \sum_{i=1}^{N} \log_e P_i$ は自由度 $2N$ の χ^2 分布に従うことから、N 主体単位根検定の全体的な検定はカイ二乗検定（P_λ test）により行うというものである。Maddala and Wu (1999) のブートストラップ実験によれば、Fisher 流のカイ二乗検定が定常性テストとしても共和分テストとしても最もパフォーマンスが良いとしている。Choi(1999a) は Fisher 検定をさらに拡張して、他の検定に対して Fisher 検定が優位にあることを、より厳密に示した。

3.6　質的従属変数パネル推定

これまでクロスセクションデータでよく用いられてきた質的（離散的）従属変数を用いた推定はパネルデータでも有効である[18]。具体的に例を挙げれば、車を買うかどうか、あるいは車を所有しているかどうか、住宅を買うかどうか、労働組合に参加するかどうか、結婚するかどうかなどの意思決定問題に計量経済学的に答えることができる。このような場合、従属変数 y_{it} は一般に選択しなければ 0、選択すれば 1 の 2 項選択の形をとることが多いが、経済主体 i が時間 t に意思決定をする（例えば、結婚する）確率を p_{it} と表せば、従属変数の期待値は $E(y_{it}) = 1 \cdot p_{it} + 0 \cdot (1 - p_{it}) = p_{it}$ 　となり、これは通常、なんらかの変数 (x_{it}) で説明される。

$$p_{it} = \Pr[y_{it} = 1] = E(y_{it}|x_{it}) = F(x'_{it}\beta) \tag{54}$$

クロスセクションデータを用いた実証研究では $F(x'_{it}\beta)$ の定式化としてプロビット・モデルとロジット・モデルがそれぞれ次のように定義されている。

プロビット・モデル

$$F(x'_{it}\beta) = \Phi(x'_{it}\beta) = \int_{-\infty}^{x'_{it}\beta} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \tag{55}$$

ロジット・モデル

$$F(x'_{it}\beta) = \frac{e^{x'_{it}\beta}}{1 + e^{x'_{it}\beta}} \tag{56}$$

これらのモデルでは、実際に何らかの意思決定がなされたとすると、従属変数が直接は観察できないある水準を超えたことを意味している。すなわち、

$$\begin{aligned} y_{it} &= 1 \quad if\ y^*_{it} > 0 \\ y_{it} &= 0 \quad if\ y^*_{it} \leq 0 \end{aligned} \tag{57}$$

ここで、$y^*_{it} = x'_{it}\beta + u_{it}$. つまり

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[u_{it} > -x'_{it}\beta] = F(x'_{it}\beta) \tag{58}$$

となる。

[18] この分野における基本文献は Maddala(1983, 1987) である。また、最近の文献には Gourieroux (2000)、Lee (2002) がある。残念ながら、ここでは Maltinominal logit, ordered probit, sequential Tobit、Count data などについては扱わない。

3.6.1 パネル・プロビット・モデルとパネル・ロジット・モデル

パネルデータの場合、誤差項に固定効果 μ_i が入ることで従来のプロビット分析、ロジット分析とは異なってくる。次のような固定効果モデルを考えよう。

$$y^*_{it} = x'_{it}\beta + \mu_i + \nu_{it} \tag{59}$$

$$\Pr[y_{it} = 1] = \Pr[y^*_{it} > 0] = \Pr[\nu_{it} > -x'_{it}\beta - \mu_i] = F(x'_{it}\beta + \mu_i) \tag{60}$$

ここで、T を固定すると、固定効果 μ_i のパラメータは N に応じて増加する。これは、パラメータ μ_i は固定された T に対して一致推定を得ることが出来ないことを意味している[19]。線形パネルデータ回帰モデルでは、パラメータ μ_i はウィズイン推定によって除去して、β だけに対しては一致推定を得ることができる[20]。

付随パラメータ問題は未解決であるが、Chamberlain (1980) は μ_i の最小十分統計量は $\sum_{t=1}^{T} y_{it}$ であることを示し、次のような条件付尤度関数を最大化して β のロジット推定を得ることを提唱した。

$$L_c = \prod_{i=1}^{N} \Pr(y_{i1}, \ldots y_{iT} \mid \sum_{t=1}^{T} y_{it}) \tag{61}$$

この方法では十分統計の定義により、推定されたパラメータ β は μ_i に依存しない。

このモデルは Chamberlin の提示した条件付最尤法と固定効果を考慮しない通常のロジット推定の差を Hausman の χ^2 検定の要領で検定できる。通常のロジット推定が有効一致推定であるのは固定効果がない場合であり、固定効果がある場合には一致推定にはならない。Chamberlin の推定は固定効果の有無にかかわらず一致しているが、固定効果がない場合には、有効ではなくなる[21]。

代替的なモデルとして Liang and Zeger (1986) が提案した Generalized Estimating Equations (GEE) Population-averaged Model がある。これは、ランダム効果線形推定法であり、プロビット推定法を線形近似した簡便法である。

---

[19] この問題は Neyman and Scott (1948) によって古典的付随パラメータ問題 (the classical incidental parameter problem) と呼ばれているものに相当する。Lancaster (2000) はこの問題は現在も解決されていないことを指摘した上で、固定効果の直交条件を見つけることが重要であると指摘している。

[20] Hsiao(2002) でも示されているように、β と μ_i が漸近的に独立であれば、線形モデルの最尤法で β の一致推計を得ることが出来る。これが非線形モデルの場合やプロビット・モデルの場合には一致推計を得ることが出来ない。

[21] これに対して、固定効果プロビット・モデルでは計算はロジット・モデルのように簡単ではない。一般に固定効果を含んだ最尤法は、N が大きく、T が固定されている場合には、一致推計量が得られない。Heckman (1981b) を参照。

3.6.2　パネル・トービット・モデル

これまで、従属変数が０か１の二項選択のモデルを考えてきたが、０と連続変数の選択はトービット・モデルで扱うことができる。パネルデータでは固定効果トービット・モデルは次のように定義できる[22]。

$$y^*_{it} = x'_{it}\beta + \mu_i + \nu_{it} \qquad \nu_{it} \sim IIN(0, \sigma^2_\nu) \tag{62}$$

$$\begin{aligned} y_{it} &= y^*_{it} \quad if\ y^*_{it} > 0 \\ y_{it} &= 0 \quad otherwise \end{aligned} \tag{63}$$

ここで $d_{it} = 1 \ \ if\ y^*_{it} > 0,\ d_{it} = 0 \ \ otherwise$ とすると、対数尤度関数は次のように定義できる。

$$LogL = \sum_{i,t}(1-d_{it})Log\Phi(\frac{-x_{it}\beta - \mu_i}{\sigma}) + \sum_{i,t} d_{it}\{-\frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_{it} - x'_{it}\beta - \mu_i)^2\} \tag{64}$$

線形モデルとは違い β と σ は μ_i に依存する。これまで何度も論じてきたように、パラメータ μ_i は固定された T に対して一致推定を得ることが出来ない。この不一致はパラメータ β と σ を通して発生する。

Heckman and MaCurdy (1980) は反復法 (iterative methods) によって推定することを提唱した[23]。すなわち、β と σ に対して初期値を与え、それを所与として、上述の対数尤度関数を μ_i に関して最大化する。その値を再び尤度関数に代入し、今度は β と σ に関して最大化し、新たな β と σ を得る。この作業を β と σ が収束するまで繰り返すのである。

Honoré(1992) は誤差項の分布を特定化しないセミパラメトリック推定を提唱している。具体的には（62）式より次のように定義する。

$$u_{ist}(b) = \max\{y_{is}, (x_{is} - x_{it})b\} - \max\{0, (x_{is} - x_{it})b\} \tag{65}$$

$b = \beta$ の場合、

$$\begin{aligned} u_{ist}(\beta) &= \max\{y_{is}, (x_{is} - x_{it})\beta\} - \max\{0, (x_{is} - x_{it})\beta\} \\ &= \max\{\mu_i + \nu_{it}, -x_{is}\beta, -x_{it}\beta\} - \max\{-x_{is}\beta, -x_{it}\beta\} \end{aligned} \tag{66}$$

[22] ランダム効果トービット・モデルはランダム効果プロビット・モデルを拡張することによって推計できる。しかし、これまでのところ Hausman and Wise (1979) などを例外として、あまり実証研究には用いられていない。

[23] 彼らは、従属変数のラグが説明変数に入っていないのならば、μ_i の一致推計を得ることが出来ないということはそれほど大きな問題ではないと述べている（1980, P.59)。

ここで $u_{ist}(\beta)$ は s と t に関して対称である。ν_{it} が $i.i.d.$ に従っているとすれば、$u_{ist}(\beta)$ と $u_{its}(\beta)$ も $i.i.d.$ に従う。このことから、次のモーメント条件が導かれる。

$$E[(\xi(\psi(u_{its}(\beta) - \psi(u_{ist}(\beta))))|x_{it}, \mu_i] = 0 \tag{67}$$

この条件を満たすようにGMM推定すれば β は一致推定となる。モンテカルロ実験によれば、N が小さければ β 推定は歪みを持つことが示されている。

3.7 不完備パネルデータ

これまで、パネルデータはすべて揃っていて欠損がない完備パネルデータを想定していた。しかし、実際のパネルデータは個人や企業が回答拒否して観察値が欠落していることがある（これを attrition 問題と呼ぶ）。また、さらには、企業であれば倒産したり、新規参入してくることもあるし、個人であれば、死亡したり、移転して追跡不可能になることもある。むしろ、パネルデータは不完備な状態の方が当たり前とさえ言える。では、不完備パネルデータを利用するために注意すべき問題点は何だろうか。

データの問題として、無作為（ランダム）にデータが欠測する場合と、有為に欠測する場合（例えば、企業倒産や個人のサンプルからの脱落）とでは意味が違ってくる。無作為（ランダム）欠測の場合、一般に不完備パネルデータであっても、その平均、分散の計算をデータサイズを適切に考慮して計算し、データサイズに応じたウェイト付けした加重最小二乗法 (weighted least square=WLS) を用いて推定すれば問題はない。問題はデータの欠測に何らかの法則性 (self-selection resones) があり、残ったサンプルが元のサンプルの性格と違ってくる場合である。この場合にはいわゆるサンプル・セレクション・バイアス問題に直面する[24]。

誤差項の分散に関する推定は ANOVA(分散分析) 法[25]や最尤 (ML) 法[26]が用いられている。ANOVA 法は、完備データに対しては最良不偏推定が得られることが知られている、不完備データに関しては、推定は誤差項の分散の関数として表されているが（Townsend and Searle (1971)）、不偏推定を得る

[24] よく知られている事例は、ニュージャージー (New Jersey) およびインディアナ州ギャリー（Gary）における所得維持政策実験である。ここでは、家計簿をつけることに便益を感じない参加者が脱落し、軍隊に召集された人も脱落し、さらにこの実験から何の恩恵も受けない高額所得者も脱落した。このように、一定の傾向を持った人々が脱落することで実験計画の無作為化と局所管理の原則が破られていった。Hausman and Wise (1979) はこの脱落問題の引き起こすバイアスを推計している。彼らによれば脱落のバイアスは小さいが有意であることが明らかにされている。

[25] ANOVA 法については (Searle (1971)、Townsend and Searle (1971)、Wallace and Hussain (1969)、Swamy and Arora (1972)、Fuller and Battese (1974)、Henderson (1953) などを参照。

[26] 最尤法については Jennrich and Sampson (1976)、Harville (1977)、Das (1979)、Corbeil and Searle (1976a,b)、Hocking (1985) などを参照。

ことは可能である。しかし、等分散性、無相関性は保障されていないので、最良不偏推定とはならない。最尤法は十分統計量の関数となり、一致推定であり、漸近的に有効推定となることが示されているが、誤差項の分散を推定するために多くの自由度が失われている。

Baltagi and Chang (1994) は不完備パネルデータを用いて一元配置誤差項モデルのモンテカルロ実験を行った。その結果、次のようなことが明らかにされた。（1）推定されたパラメータに関しては ANOVA 法による一般化最小二乗法の推定も、最尤法の推定もほとんどかわらないこと。（2）誤差項の個別分散推定においては ANOVA 法による推定は最尤法に比べて精度が低い。とりわけ、データの不完備度が高かったり、分散構成比 (variance component ratio) が 1 より大きい場合には、それが顕著となる。（3）不完備データから完備データ部分だけを摘出して推定することは、有効性を大幅に失う。

これらの結果より、不完備データだからといって、一概にそのサブセットである完備データにまで情報量を落とすことは薦められないし、現在では一般に用いられているパネル・データ推定プログラムでも不完備データに応じて自動的に推定を調整してくれるようになり、推定量が完備データと比べれば最良ではないとしても、不完備データの問題は大幅に縮小されるようになっている[27]。

4　パネルデータ分析の応用

パネルデータを用いると、これまで時系列データやクロスセクションデータでは不十分にしか扱えなかった経済問題が解明できるようになってきた。以下ではいくつかの応用例を示しながら、具体的に何が解明されてきたのかを説明したい。

4.1　国際資本移動に関するフェルドスタイン・ホリオカ・パラドックスにおける固定効果推定[28]

国際資本移動の自由度の計測は国際経済学の分野では主要な研究テーマであり、Feldstein and Horioka (1980) および Feldstein(1983) は、この問題に対する簡便なアプローチとして多くの研究者の関心を引いてきた。彼らのアプローチは OECD 加盟 23ヶ国の国別の名目国内投資の名目 GDP に対する比率（国内投資率）を名目国内貯蓄の名目 GDP に対する比率（国内貯蓄率）に対して回帰するというものであり、回帰係数が 1 に極めて近いとの結果を

[27] もちろん、不完備データにも程度があり、あまりにデータの欠落が多いようだと利用上問題がでてくることもあることには注意を要する。
[28] 本節は北村・藤木（1995）に依拠している。

得ている。数式で表すと以下のようになる。

$$(\frac{I}{GDP})_{it} = \alpha_i + \beta(\frac{S}{GDP})_{it} + u_{it,} \quad t = 1, ...T \tag{68}$$

ここで I は国内名目投資、S は国内名目貯蓄、GDP は名目 GDP を表す。

彼らの結果は、国内貯蓄の外生的な増加分はほとんど当該国の国内投資に振り向けられていることを意味しており、「限界的な新規の貯蓄は最も有利な利回りを保証する国に投資されるはずであり、国内貯蓄と国内投資には相関はない」という国際資本移動自由化の下での予測に反することになり、フェルドスタイン・ホリオカ・パラドックスと呼ばれるようになった。

この結果を受けて、多くの研究者がこのパラドックスを解こうとしてきた。第一の方向は回帰式の説明変数に財政変数、国内投資、国内貯蓄両方に影響を与えるショック、非貿易財の効果などを加えて、国内貯蓄率にかかる回帰係数を低下させようとするものである。第二の方向は、本来のパネルデータではなく、時系列データやクロスセクション・データを用いて彼らの結果を再検証しようというものである[29]。

北村・藤木（1995) では、あえて Feldstein and Horioka(1980) と同じデータを用いて、パネルデータ分析の方法を適切に用いれば、彼らの結果は統計的に否定され、より現実的な回帰係数が得られることを示している。

すなわち、Feldstein and Horioka (1980) の手法はパネルデータ分析では、何年分かのデータの平均をとり、それをクロスセクション・データとして推定するビトウィーン推定であることが知られているが、この手法は全サンプルをプールして同一の係数、切片を仮定するプーリング推定に近いものである。ビトウィーン推定とプーリング推定の結果は共に回帰係数 β が 1 に近い。問題はこの推定方法が第 3 節で論じたモデル選択の結果選ばれるものであるかどうかということである。

北村・藤木 (1995) の結果では、(1) プーリング推定法と固定効果推定法の比較では、プーリング推定法が選ばれることはなく、国別の異質性を定数項の差異によりコントロールした固定効果推定法が常に選ばれる。(2) 固定効果推定法による回帰係数 β は有意に 1 より小さい、(3) さらに様々なパネルデータ分析の方法を時系列的にローリング推定を通して比較検討していくと、固定効果推定法が常に選択されるわけではなく、各国別に時系列データして扱う方が望ましいとの結果も得られている。しかし、経時的に回帰係数 β の変化を表 1 でたどると、その値は確実に 1 より低下してきていることが明らかである。

このことは、適切な推定方法をパネルデータに応じて選択すれば、フェルドスタイン・ホリオカ・パラドックスはそもそも存在しなかったということ

[29]最近の研究は国際資本移動の歴史的な発展を検定する際に Feldstein and Horioka(1980) のアプローチを用いるようになっている。この点に関しては Jones and Obstfeld (2001) などを参照されたい。

を示唆しているのである。なぜ、このような結果になったのかということを直感的に理解するためには図 3 を参照されたい。ここでは 3 カ国の 5 年分のデータが図 3 のように散らばっている場合、各国毎に見ると国内投資率と国内貯蓄率は無相関あるいは 1 よりもはるかに低い回帰係数に従っているいるにもかかわらず、各国データを 5 年間について平均してクロスセクション・データとして回帰すればその係数は 1 に近いものになるということである。このようなデータの適切な分析方法は、回帰係数は共通だが、切片が各国で違う固定効果推定を用いることであり、それは F 検定によりなんらの恣意性もなく検証できる。

4.2 企業収益性の固定効果推定[30]

パネルデータの応用で最も広範に用いられているのは財務諸表に基づいた企業データである。というのは、上場企業であれば有価証券報告書を大蔵省（財務省）に提出する義務があり、それは一般に公開されており、データベースとして整備されたものが市販されているからである。また、同一企業が営業を続けている限り、毎年報告する義務があり、それを接続すれば、もともとは毎年行われているのはクロスセクション調査であっても、パネル化することができるからである。本節で用いた『企業活動基本調査』は我が国の企業活動の実態を明らかにし、企業に関する施策の基礎資料を作ることを目的として 1992 年（平成 4 年）より統計法に基づく（旧）通商産業省企業活動基本調査規則によって実施されている指定統計調査である。この調査は日本標準産業分類に掲げる大分類 D（鉱業）、F（製造業）および I（卸売・小売業、飲食店）に属する事業所を有する企業のうち、従業員 50 人以上かつ資本金または出資金 3 千万円以上の会社全てを調査対象としている。サンプルサイズは各年おおよそ 26000 社である。調査項目としては（1）企業の設立形態、設立時期、（2）事業組織・従業員数、（3）資産・負債および資本ならび取引状況、（4）企業間取引および海外取引、（5）研究開発、（6）技術の所有および取引状況、（7）親会社・子会社・関連会社の状況などを調べている。本調査でも同一企業を接続したパネル化が可能となる。

企業活動の根本は収益を挙げることであり、企業の収益性には、企業活動の全ての要素が集約されているといっても過言ではない。収益性指標のうち代表的なものであり、かつ『企業活動基本調査』にある情報から作成できるものに総資産利益率（ROA=経常利益/総資産）がある。この指標を用いて企業収益性に関するパネルデータ分析を行ったのが北村 (2002b) である。ここでの考え方は、企業は財サービス等の生産物において寡占競争にさらされていると同時に、労働や資本などの投入財の市場でも競争にさらされている、さらには、企業財務の構造を通して様々なコーポレート・ガバナンス主体か

[30] 本節は北村（2002b）を参照している。

らの制約を受けると考えて、それらを総合的に把握しようという試みである。具体的な推定式は以下に示す通りであるが、26000 を超える企業に個別固定効果を割り当てても推定すべきパラメータが膨大な数にのぼり、しかもそこから有意な議論を導き出すのは難しいということ、そして産業政策の評価という観点から、個別企業毎ではなく、産業別に固定効果を推定した。

$$\begin{aligned} ROA_{it} &= \alpha_i + \phi S_{it} + \beta d\ln K_{it} + \mu S_{K_i} + \gamma d\ln L_{it} + \zeta S_{L_i} + \kappa d\ln Debt_{it} \\ &\quad + \vartheta S_{D_i} + \delta Debt/Asset_{it} + \eta (Debt/Asset_{it})^2 + \theta OwnCapRatio_{it} \\ &\quad + \nu LiquidityRatio_{it} + \xi Debt/Equity_{it} + \epsilon_{it} \end{aligned} \tag{69}$$

ここで、K ＝実質有形固定資産額、L ＝常時従業者数、D ＝実質負債、$Debt/Asset$ ＝負債資産比率、$(Debt/Asset)^2$ ＝負債資産比率の 2 乗、$OwnCapRatio$ ＝自己資本比率、$LiquidityRatio$ ＝流動性比率、$Debt/Equity$ ＝負債比率である。S_{it} ＝商品売上シェア、S_{K_i} ＝資本シェア、S_{L_i} ＝労働シェア、S_{D_i} ＝負債シェアである。

市場競争の企業業績への効果であるが、商品売上シェアは有意に正の効果を持ち、また、労働シェア、資本シェア、負債シェアもそれぞれ有意な説明力を持つことが確認された。労働シェアは正、資本シェアは負、負債シェアは負となっている。ここで用いたシェア変数であるが、労働、資本、負債に関しての市場競争の範囲が事前にはわからなかったので、全国、業種別、地域別、地域 × 業種別の 4 種類の競争市場を考えて、最も当てはまりのよい変数を選択した。その結果、労働は地域内で競争しており、資本は業種内で競争、負債は地域内で競争をしていると想定することが尤らしいということになった[31]。推定式には、選択された変数のみが報告されている。

企業財務もほとんどの変数が有意に効いており、ROA に対しては企業財務の健全性が重要であることが確認された。パネル・データ分析の結果としては、固定効果モデルがランダム効果モデルに対して、統計的に選択されることがわかる。産業毎の平均をクロス・セクションで回帰したビトウィーン推定の係数と固定効果推定の係数はかなり違うし、また t 値もビトウィーン推定ではあまり有意でなくなる。このことは、これまで、ミクロ統計の利用に当たっては、クロス・セクション推定が主流であったが、その推定結果の頑健性に疑問を生じさせる結果となっている。さらに言えば、クロス・スセクションでは有意でなかった変数も、パネル・データとして使うと有意になることもわかった。

上式における α_i はパネル推定における固定効果を表すものである。これは、説明変数では説明されない産業固有の効果を捉えたものである。図 4 は

[31]負債に関しては、地域 × 業種別の分類も同様に説明力を持ったが、業種によってはサンプル数が極めて少なくなるケースがありここでは地域とした。実際には、ここに集められたサンプルのほとんどは上場企業ではなく、地域の金融機関から資金調達をしていると考えられるので、この結果は納得のいくものである。

ROA の産業別平均値とパネル推定より得られた固定効果を描いたものである。この図から明らかなように、ROA の高い産業は固定効果も高いことがわかる。ここで発見された面白い事実は、ROA の分布が正規分布より大きく外れている産業、農林水産業、不動産業、金融・保険業などの固定効果が突出しているということである。これは、これらの業種では、モデルでは説明されなかった要因によって ROA が引き上げられていることを意味する。すなわち、少なくとも農林水産業、不動産業、金融・保険業[32]に関する限りは、政府から多大な補助金ならびに規制による保護を受けてきたことが明らかであり、これが固定効果の突出となって表れているのかもしれない[33]。逆に、ROA が正規分布していない場合のパネル推定の結果には大きな誤差が含まれている可能性もある。

もうひとつ固定効果が突出している運輸・通信業はどのように考えれば良いだろうか。この業種も規制が強く、行政介入が大きかったと見ることもできるが、ROA の分布に関する限りは、他の 3 業種とはちがい、かなりきれいに分布している。この業種では自由競争が行なわれているという解釈に立てば、固定効果の突出は、通信業のネットワーク外部性などの要因が働いていると考えることもできる。しかし、その場合は、事後的に推定された平均 ROA がそれほど大きくなく、何らかの要因が突出分を相殺するような働きをしていることになる。この問題はさらに追求する必要がある。

4.3　購買力平価に関するダイナミックパネル推定

国際金融の教科書である Obstfeld and Rogoff (1996, p.203) でも取り上げられているように、購買力平価を推計するための格好の素材としてマクドナルド・ハンバーガーの主力商品であるビッグ・マックの価格を世界中で比べて、購買力平価の指標とし、それが現行の為替レートとどのように異なっているかということをみることで、購買力平価説が成り立つかどうかが検証されている。一般には Obstfeld and Rogoff (1996, p.203) が論じているように、ハンバーガー製造には牛肉、食用油、ソースなど貿易財も含まれるが、レストランの地代、労働賃金など各国固有の非貿易財も含まれているため、購買力平価は成り立たない可能性が高い。しかし、Click（1996）が論じているように、各国の発展段階に応じた購買力格差（Balassa-Samuelson 効果）をコントロールすれば、購買力平価は成り立つようにも見える。実際にビッグマックの価格比と為替レートをプロットしてみると (図 5)、ほぼ 45% 線上に並んでいることがわかる。先の図 3 で見た国内貯蓄率と国内投資率のプロットと

[32] ここで取り上げられている金融・保険業は金融庁管轄下にある都市銀行、地方銀行、保険会社ではなく、経済産業省管轄下の製造業などから派生した比較的小規模なクレジット会社、保険会社であることに注意されたい。

[33] もうひとつの説明はこれらの産業のサンプル数が極めて小さいというサンプル・バイアスのせいだというものである。

は違い、各国別の傾きもほぼ 1 に近いことがわかる。Fujiki and Kitamura (2002) では、Click の枠組みに従いながら、より厳密な検定を行った。その結果、Balassa-Samuelson 効果を入れても厳密には購買力平価は成り立たないことがわかった。次に 3.4 節で論じたダイナミック・パネル分析の枠組みを用いて、マクドナルドの価格調整スピードを検定してみた。推計に用いたモデルは次に示すとおりである。

$$\ln(\frac{p_{it}}{p_{ust}}) = \alpha + \beta_1 \ln(e_{it}) + \beta_2 \ln(\frac{RGDP_{it}}{RGDP_{ust}}) + \gamma \ln(\frac{p_{it-1}}{p_{ust-1}}) + \varepsilon_{it} \qquad (70)$$

ここで p_{it} は i 国における t 年のビッグマック価格、p_{ust} はアメリカにおける t 年のビッグマック価格、e_{it} は i 国における t 年の名目為替レート (対米ドル)、$RGDP_{it}$ は i 国における t 年の一人あたり GDP、$RGDP_{ust}$ はアメリカにおける t 年の一人当たりの GDP、ε_{it} は誤差項。

推計結果は表 3 のとおりである。ここでのダイナミック・パネル推定はランダム効果推定を行う一般化最小二乗法推定（GLS）と Hsiao, Pesaran and Tahmiscioglu (2002) で提案された最小距離推定（Minimum Distance Estimator=MDE）を用いて行った。

ダイナミックな価格調整メカニズムを表す γ はほとんどゼロであり、価格調整スピードは極めて早く、一年内に為替変動や生産性ショックは吸収していることがわかった。また、問題の係数 β_1 はいずれの推計でもほぼ 1 になっており、為替レート変動に関するマクドナルドの価格調整の長期弾力性を $\beta_1/(1-\gamma)$ とすると、これもほぼ 1 になることがわかり、為替レートの変動は長期的には必ずビッグマック価格に反映されることを意味している。従って、厳密には購買力平価は成り立たなくとも、ビッグマック価格は為替レート変動を吸収しており、しかもかなり購買力平価に近い値をとっていること、そして、GLS 推定でも MDE 推定でも、結果にほとんど違いがないことから、価格調整メカニズムの推計に年次データを用いることはほとんど意味がないことがわかった。

4.4 結婚の意思決定におけるパネル・プロビット推定

日本の家計に関するパネルデータとして広く公開されている調査は財団法人家計経済研究所の『消費生活に関するパネル調査』であろう[34]。この調査は調査開始時 (1993 年) に 24 − 34 歳であった女性 1500 人を対象として行ってきたもので、1993 年以来毎年、同一個人に対して調査が実施されている。このデータのカバーしている調査内容は多岐にわたっているが、家計貯蓄や労働供給と並んで、結婚の意思決定という点についても調査している。そこで、先に見た質的従属変数パネル推定の応用として結婚問題について考えてみたい。

[34] この調査の詳しい解説については北村 (2002a) を参照されたい。

これまで結婚の意思決定に関する研究では、クロスセクションデータが使用されていたために、調査時点での結婚状態と同時期の社会経済変数を用いたプロビット分析が行われていた。ここで用いる『消費生活に関するパネル調査』では、パネルデータという特質を活かして、有配偶サンプルの結婚前の状態を捕捉することができる。これにより、結婚を選択したものとそれを行っていないものとの「無配偶状態」における比較という視点を分析にもちこむことができる。また、結婚によって、彼女たちの就業状況・所得などの社会経済的要素がどのように変化することもみることができる。要するに、「結婚をする」という意思決定をおこなったものと、それをしないで踏みとどまっているものとの差異はどこにあるのかということを明らかにしようということである。

調査時点ごとの未婚率（1-結婚率）を見ると、調査年度が後になるほど同じ年齢でも未婚率は高まることがわかる。これは晩婚化現象を反映したものであるが、このような晩婚化現象を時間（time or life-cycle）効果、コーホート効果、個別固定効果のうちそれぞれがどの程度説明できるのかも実証的に解明しなければならない。同じデータを使った先行研究には樋口美雄、阿部正浩 (1999) があるが、そこでは、結婚に関しては（1）既婚＝1、未婚＝0のダミーを被説明変数としたプーリング・プロビット分析と（2）未婚から結婚への状態の変化をプロビット分析と結婚年齢を変数にしたサバイバル分析を行っている。

北村・坂本 (2002) では、彼らの研究とは補完的に、パネルデータの特徴をより明示的に生かした、計量経済学的手法であるパネル・プロビット推定を用いた。

ここで、ランダム効果 ν_i が次のような正規分布 $N(0,\sigma_\nu^2)$ に従うと仮定する。

$$\Pr(y_i|x_i) = \int_{-\infty}^{\infty} \frac{e^{-\nu_i^2/2\sigma_\nu^2}}{\sqrt{2\pi}\sigma_\nu} \left\{ \prod_{t=1}^{T_i} F\left(x_{it}\beta+\nu_i\right) \right\} d\nu_i \tag{71}$$

ここで、$F(x_{it}\beta+\nu_i) = \begin{cases} \Phi\left(x_{it}\beta+\nu_i\right) & if \quad y_{it} \neq 0 \\ 1-\Phi\left(x_{it}\beta+\nu_i\right) & otherwise \end{cases}$、$\Phi$ は累積正規分布関数である。さらに、対数尤度関数 L を次のように定義する。

$$L = \sum_{i=1} w_i \log \Pr\left\{(y_i|x_i)\right\} \approx \sum_{i=1} w_i \log \left\{ \frac{1}{\sqrt{\pi}} \sum_{m=1} F\left(x_{it}\beta + \sqrt{2\frac{\rho}{1-\rho}}a_m^* \right) \right\} \tag{72}$$

w_i はパネル i に対するウエイトを表す。上述のような対数尤度関数を最大化することによって、非線型確率関数 F(.) のパラメータを推定するというのがパネル・プロビットの手法である。

実証手法の正当性は最終的には統計検定によって判断されなければならない。パネル・プロビット推定がいいのか、プーリング・プロビット推定がい

いのかは、テストをして決めるべき問題である。ここで用いる検定統計量は次のように導かれる。まず次の統計量を定義する。

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + 1} \tag{73}$$

ここで、σ_v^2 はパネル・レベルでの分散をあらわし、ρ は全分散に対するパネル・レベルでの分散の比率を表わす。ここで帰無仮説を $\rho = 0$ とした対数尤度比検定を行い、$\rho = 0$ が棄却できなければ、パネル・レベルでの分散は重要ではなく、パネル・プロビット推定はプーリング・プロビット推定と大きく変わりないことになる。逆に $\rho = 0$ が棄却された場合には、パネル・プロビット推定を用いるべきであることになる。

1990 年代中頃までは、日本の社会制度の中では、女性の社会進出が結婚の意思選択に負の影響を与えているという論調が強かった。しかし近年、晩婚化問題はむしろ「キャリアアップ」を行い、社会的自立することを目指す女性に起因するのではなく、親と同居することによって基礎的経済的経費を負担してもらい、自分たちは高い経済的環境にいる「パラサイトシングル」が晩婚化の大きな要因として注目されるようになってきた。北村・坂本 (2002) では、こうした議論を受け、家族との関係、特に前年に親と「同居」[35]していたかどうかということを説明変数に加えることで、家族関係と結婚の関係を分析を行った。

使用した変数には、前年の年収 (inc)、年収の 2 次項 (inc^2)、前年の貯蓄 (dep)、年齢 (age)、年齢の 2 次項 (age^2) 、前年の非正規就業 (パート) ダミー (Occupdu1)、前年の無業ダミー (Occupdu2)[36]を、そして、同居ダミー (Cohabpa、親と同居=1、親と別居=0)、被説明変数には、有配偶か無配偶かを示す配偶状態ダミーを用いる。

実証結果は表 4 に示してある。対数尤度比検定の結果、パネル・プロビット推定が支持され、プーリングプロビット推定は棄却された[37]。樋口・阿部（1999）で有意とされた前年の本人収入はここでも負に有意となることが確認された。これは所得を外生とした場合、高収入のほうが結婚確率が低くなっていることを意味している。年齢に関しては、一次と二次の項を入れて非線形性を検定し、一次の項が正 (結婚確率を高める)、二次の項が負 (結婚確率を低める) となっている。両方とも有意な結果が得られ、年齢に関する強い非線形性がこのサンプルから検出された。Wald カイ 2 乗テストによって最終的に選ばれたモデル 1-1 と 1-6 では、正規就業との比較でいうと結婚前に非正規就業（パート）あるいは無業を選択した方が結婚確率を高める結果と

[35]「同居」とは、家計経済研究所の定義に従い、「親と同一建物で、生計を共にしている (同居世帯)」、親と同一建物で、生計が別 (準同居世帯)、親と同一敷地内の別建物に居住 (準同居世帯) の 3 つのものをとする。

[36]正規就業を基準値とする。

[37]この結果はいわば自明である。すなわち、パネル・プロビット分析を行なうために固定的要因を除外し、期間内に結婚したサンプルをその間に変動した変数で説明するモデルが、結婚の意思決定と変数間の時間的因果関係を明示的に扱えないプーリング・プロビット推計より、情報量が多く、説明力が高くなるのは明らかだからである。

なっている。また同居ダミーも負に有意な結果となり、予想通り同居が結婚確率を低める結果となっていることがわかった。

もちろんこの結果には注意が必要である。未婚でいることを同居ダミーで説明するのは因果関係が逆ではないかという議論もあり得る。すなわち、とりあえず結婚しないのだから親と同居しておいて、結婚が決まってから独立するというロジックを考えれば、同居の意思決定が結婚（婚約）ダミーによって説明されるモデルもあり得るということである。しかし、現実的に判断すると、未婚と同居は 1 対 1 対応している訳ではなく、未婚であっても親と別居しているケースと結婚していても親と同居しているケースがあり、両者を比べると前者の方が多いことがわかる。このことは同居していれば、かなり高い確率で未婚であると言えるが、未婚だからといって親と同居している確率はそれよりも低いということである。すなわち、同居から未婚への因果関係の方が、未婚から同居への因果関係より、強い因果性が見出されると考えられる。これが、ここで用いたアプローチを正当化するための直感的な説明である。しかし、この点に関するさらなる検証が必要である。

5　おわりに

これまでの議論から明らかなように、パネルデータ分析はクロスセクション分析の手法と時系列分析の手法を取り入れた、究極的なデータ分析手法であり、これまでマクロ集計データや、クロスセクションデータ、時系列データだけでは検定できなかった個別経済主体の行動が詳細に分析できるようになった。

本論文の主要な論点としては、第一にパネルデータの性質に応じて適切な分析手法を選択すべきであるということである。パネルデータはその性質上、かなり多様な経済主体を含んだ heterogeneous なデータであり、単純な代表的個人モデルや代表的企業モデルでは捉えられない側面をもっている。そのようなデータは実験計画法の手法に習って、データの最適な使い方を熟慮すべきであり、アプリオリに推定方法が決まっているというものではないことを認識しておくべきである。

第二に、パネルデータ分析の手法の選択に当たっては、各種の検定テストを徹底的に行って、できるだけ頑強な推定を行うべきである。

第三に、パネルデータは膨大なサンプルを有することが多いが、その全てを利用しようとして、闇雲にサンプルを増加させることは計算効率上問題であるが、完備パネルデータだけを用いるために、データに欠測値のあるサンプルを落としてしまうことも問題となる。現在では、コンピュータ・プログラムは欠測値の問題は考慮して推定してくれるので、ことさらに欠測値を問題とすべきではない。

第四に、パネルデータの利点はミクロ経済学の理論を直接検定できるということであり、理論モデルへのフィードバックを行うことが可能になった。また逆にパネルデータを上手く利用すれば理論と整合的な政策分析が可能になる。今後の政策分析ではミクロ経済データ、とりわけ、パネルデータを用いることが当たり前になるだろう。またそのような方向に進むべきであると思われる。

参考文献

[1] 北沢良継（2001）「パネルデータ計量経済学の最近の動向」、『エコノミックス』（九州産業大学経済学会）、6(1)、pp.89-99.

[2] 北村行伸 (2002a)『家計行動のパネル統計』、一橋大学経済研究所附属日本経済統計情報センター統計資料シリーズ no.54.

[3] 北村行伸（2002b）「『企業活動基本調査』に基づく日本企業行動のパネル分析」、財団法人日本統計協会統計研究助成報告書（平成 13 年度）

[4] 北村行伸、藤木裕（1995）「国際比較研究へのパネルデータ分析の応用ー Feldstain-Horioka パラドックスの再検討」、『金融研究』、第 14 巻 1 号、pp.145-160.

[5] 北村行伸、中村恒（1998）「価格・数量調整過程の再検討ーマクロ・産業別パネル・データ分析ー」、日本銀行金融研究所 Discussion Paper No.98-J-5.

[6] 北村行伸、坂本和靖（2002）「結婚の意思決定に関するパネル分析」、一橋大学経済研究所 PIE Discussion Paper, no.109, 2002 年 8 月

[7] 樋口美雄、阿部正浩 (1999)「経済変動と女性の結婚・出産・就業のタイミング」、樋口美雄、岩田正美（編）『パネルデータからみた現代女性』、東洋経済新報社

[8] 広津千尋（1992）『実験データの解析ー分散分析を超えてー』、共立出版

[9] Ahn, S.C. and Schmidt, P. (1995) "Efficient Estimation of Models for Dynamic Panel Data", Journal of Econometrics, 68, pp.5-28.

[10] Alonso-Borrego,C. and Arellano,M.(1999) "Symmetrically Normalized Instrumental-variable Estiantion Using Panel Data", Journal of Business and Economic Statistics, 17, pp.36-49.

[11] Amemiya, T. (1985) Advanced Theory of Econometrics, Cambridge: Harvard University Press.

[12] Anderson, T.W. and Hsiao, C. (1981) "Estimation of Dynamic Models with Error Components," Journal of the American Statistical Association, 76, pp.598-606.

[13] Anderson, T.W. and Hsiao, C. (1982) "Formulation and Estimation of Dynamic Models Using Panel Data," Journal of Econometrics, 18, pp.47-82.

[14] Arellano, M. (1989) "A Note on the Anderson-Hsiao Estimator for Panel Data," Economics Letters, 31, pp.337-341.

[15] Arellano, M. and Bond, S. (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, pp.277-297.

[16] Arellano, M. and Bover, O. (1995) "Another Look at the Instrumental Variable Estimation of Error-components Models," Journal of Econometrics, 68, pp.29-52.

[17] Arellano, M. and Honoré, B. (2001) "Panel Data Models: Some Recent Developments," Handbook of Econometrics, Vol.5., Chapter 53, pp.3229-3296.

[18] Arellano, M.,O. Bover, and J. Labeaga (1999) "Autoregressive Models with sample Selectivity for Panel Data," in Analysis of Panels and Limited Dependent Variable Models, eds. by C. Hsiao, K. Lahiri, L.F. Lee and M.H. Pesaran, Cambridge: Cambridge University Press, pp.23-48.

[19] Baltagi, B.H.(2001) Econometric Analysis of Panel Data, 2nd ed, New York: John Wiley & Sons.

[20] Baltagi, B.H. and Y.J. Chang (1994) "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-Way Error Components Regression Model," Journal of Econometrics, 62, pp.67-89.

[21] Binder, M., C. Hsiao and M.H. Pesaran (2000) "Estimation and Inference in Short Panel Vector Autoregression with Unit Roots and Cointegration," mimeo.

[22] Blundell, R. and Bond, S. (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," Journal of Econometrics, 87, pp.115-143.

[23] Blundell, R., S. Bond and F. Windmeijer (2000) "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," Advances in Econometrics 15, pp.53-91.

[24] Chamberlain, G. (1980) "Analysis of Covariance with Qualitative Data," Review of Economic Studies, 47, pp.225-238.

[25] Chamberlain, G. (1984) "Panel Data, in Z. Griliches and M. Intrilligator (eds.), Handbook of Econometrics, North-Holland, Amsterdam., Vol.2., Chapter 22, pp.1247-1318.

[26] Chamberlain, G. and Z. Griliches (1975) "Unobservables with a Variance-Components Structure: Ability, Schooling and the Economic Success of Brothers," International Economic Review, 16, p.422-50.

[27] Choi, I. (1999) "Unit Root Tests for Panel Data," Working Paper, Department of Economics, Kookmin University, Korea.

[28] Click, R.W. (1996) "Contrarian MacParty," Economics Letters 53, pp.209-212.

[29] Corbeil, R.R. and S.R. Searle (1976a) "A Comparison of Variance Component Estimators," Biometrics, 32, pp.779-791.

[30] Corbeil, R.R. and S.R. Searle (1976b) "Restricted Maximum Likelihood (REML), Estimation of Variance Components in the Mixed Model," Technometrics 18, pp.31-38.

[31] Das, K. (1979) Asymptotic Optimality of Restricted Maximum Likelihood Estimates for the Mixed Model," Calcutta Statistical Association Bulletin 28, pp.125-142.

[32] Deaton, A. (1985) "Panel Data from Time Series of Cross-Sections," Journal of Econometrics, 30, pp.109-126.

[33] Feldstein, M.(1983) "Domestic Saving and International Capital Movements in the Long Run and the Short Run", European Economic Review, 21, pp.129-151.

[34] Feldstein, M.and Hirioka, C.(1980) "Domestic Savings and International Capital Flows", Economic Journal, 90, pp.314-329.

[35] Fisher, R.A. (1932) Statistical Methods for Research Workers, 4th ed., Edinburg: Oliver and Boyd.

[36] Fisher, R.A.(1971) The Design of Experiments, New York: Hafner Publishing.

[37] Fisher, R.A.(1973a) Statistical Methods for Research Workers, 14th ed, New York: Hafner Publishing.

[38] Fisher, R.A.(1973b) Statistical Methods and Scientific Inference, New York: Hafner Publishing.

[39] Fujiki, H., C. Hsiao, and Y. Shen.(2002) "Is There a Stable Money Demand Function under the Low Interest Rate Policy? A Panel Data Analysis", Monetary and Economic Studies, 20(2), pp.1-23.

[40] Fujiki, H. and Y. Kitamura. (1995) "Feldstein-Horioka Paradox Rivisited", Monetary and Economic Studies, 13(1), pp.1-16.

[41] Fujiki, H. and Y. Kitamura. (2002) "The BigMac Standard: An Statistical Illustration", Hitotsubashi University, mimeo.

[42] Fuller, W.A. and G.E. Battese (1974) "Estimation of Linear Models with Cross-Error Structure," Journal of Econometrics, 2, pp.67-78.

[43] Gourieroux, C. (2000) Econometrics of Qualitative Dependent Variables, Cambridge: Cambridge University Press.

[44] Greene, W.H.(2000) Econometric Analysis, 4th ed., New Jersey: Prentice-Hall.

[45] Griliches, Z. and J.A. Hausman (1986) "Errors in Variables in Panel Data," Journal of Econometrics, 31, pp.93-118.

[46] Hahn,J., Hausman, J. and Kuersteinerm G.(2002) "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects", MIT, mimeo.

[47] Harville, D.A. (1977) "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," Journal of the American Statistical Association 72, pp.320-340.

[48] Hausman, J.A. (1978) "Specification Tests in Econometrics," Econometrica, 46, pp.1251-1271.

[49] Hausman, J.A. and D. Wise (1979) "Attrition Bias in Experimental and Panel Data: the Gary Income Maintenance Experiment," Econometrica, 47, pp.455-473.

[50] Hausman, J.A., B.H. Hall and Z. Griliches (1984) "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," Econometrica, 52, pp.909-938.

[51] Hayashi, F. (2000) Econometrics, Princeton: Princeton University Press.

[52] Heckman, J.J. (1981b) "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete time-Discrete Data Stochastic Process," in C.F. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge.

[53] Heckman, J.J. (1981c) "Heterogeneity and State Dependence," in S. Rosen (ed.), Studies in Labor Markets, Chicago University Press, Chicago.

[54] Heckman, J.J. and T.E. MaCurdy (1980) "A Life-Cycle Model of Female Labor Supply," Review of Economic Studies, 52, pp.681-690.

[55] Henderson, C.R., Jr. (1953) "Estiation of Variance Components," Biometrics, 9, pp.226-252.

[56] Hocking, R.R. (1985) The Analysis of Linear Models, Monterey: Brooks/Cole Company.

[57] Honoré, B.E. (1992) "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," Econometrica, 60, pp.533-565.

[58] Hsiao, C. (1992a) "Random Coefficients Models," in The Econometrics of Panel Data, eds. by L. Matyas and P. Sevestres, Kluwer, 1st ed., pp.223-241, 2nd ed. (1996), pp.410-428.

[59] Hsiao, C. (1992b) "Logit and Probit Models," in L. Mátyás and P. Sevestre (eds.), The Econometrics of Panel Data, Kluwer, Dordrecht, pp.223-241.

[60] Hsiao, C. (1992c) "Nonlinear Latent Variables Models," in L. Mátyás and P. Sevestre (eds.), The Econometrics of Panel Data, Kluwer, Dordrecht, pp.242-261.

[61] Hsiao, C. (2002) Analysis of Panel Data 2nd ed., Cambridge: Cambridge University Press.

[62] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (1999) "Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models," in Analysis of Panels and Limited Dependent Variables Models, eds. by C. Hsiao, L.F. Lee, K. Lahiri and M.H. Pesaran, Cambridge: Cambridge University Press. pp.268-296.

[63] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (2002) "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," Jounral of Econometrics, 109, pp.107-150.

[64] Im, K., M.H. Pesaran and Y. Shin (1998) "Testing for Unit Roots in Heterogeneous Panels," Econometrica, forthcoming.

[65] Jenrich, R.I. and P.F. Sampson (1976) "Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation," Technometrics 18, pp.11-17.

[66] Johnston, J. and DiNardo, J. (1997) Econometric Methods, 4th ed., Singapore: McGraw-Hill.

[67] Jones, M.T. and M. Obstfeld (2001) "Saving, Investment, and Gold: A Reassessment of Historical Current Account Data," in Money, Capital Mobility, and Trade: Essays in Honor of Robert Mundell, eds. by G.A. Carlvo, R. Dornbusch and M. Obstfeld, Cambridge: The MIT Press.

[68] Kiviet, H.H. (1995) "On Bias Inconsistency and Efficiency in Various Estimators of Dynamic Panel Data Models," Journal of Econometrics, 68, pp.53-78.

[69] Lancaster, T. (2000) "The Incidental Parameter Problem Since 1948, Journal of Econometrics 95, pp.391-413.

[70] Lee, M.J.(2002) Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables, San Diego: Academic Press.

[71] Levin, A. and C.F. Lin (1992) "Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties," Discussion Paper #92-93 (University of California at San Diego).

[72] Levin, A. and C.F. Lin (1993) "Unit Root Test in Panel Data: New Results," Discussion Paper #93-56 (University of California at San Diego).

[73] Levin, A. and C.F. Lin (1996) "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties," Journal of Econometrics, forthcoming.

[74] Maddala, G.S. (1971a) "the Use of Variance Components Models in Pooling Cross Section and Time Series Data," Econometrica, 39, pp.341-358."

[75] Maddala, G.S. (1971b) "The Likelihood Approach to Pooling Cross-Section and time Series Data," Econometrica, 39, pp.939-53.

[76] Maddala, G.S. (1983) Limited-Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press.

[77] Maddala, G.S. (1987) "Limited Dependent Variable Models Using Panel Data," The Journal of Human Resources, 22, pp.307-338.

[78] Maddala, G.S. (ed.) (1993) The Econometrics of Panel Data, Vols.I, II, Edward Elgar, Cheltenham.

[79] Maddala, G.S.(2001) Introduction to Econometrics, 3rd ed, New York: John Weily & Sons.

[80] Matyas, L. and P. Sevestre (1996) The Econometrics of Panel Data: A Handbook of the Theory with Applications," 2nd ed., Dorecht: Klumer Academic Publishers.

[81] Neyman, J. and E.L. Schott (1948) "Consistent Estimates Based on Partially Consistent Observations," Econometrica, 16, pp.1-32.

[82] Obstfeld, M. and K. Rogoff.(1996) Foundations of International Macroeconomics, Cambridge: The MIT Press.

[83] Searle, S.R. (1971) Linear Models, Wiley, New York.

[84] Searle, S.R. (1987) Linear Models for Unbalanced Data, Wiley, New York.

[85] Swamy, P.A.V.B. and S.S. Arora (1972) "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," Econometrica, 40, pp.261-275.

[86] Townsend, E.C. and S.R. Searle (1971) "Best Quadratic Unbiased Estimation of Variance Components from Unbalanced Data in the One-Way Classification," Biometrics 27, pp.643-657.

[87] Wallace, T.D. and A. Hussain (1969) "The Use of Error Components Models in Combining Cross-Section and time-Series Data, Econometrica, 37, pp.55-72.

[88] Wooldridge, J.M. (2000) Introductory Econometrics: A Modern Approach, South-Western College Publishing.

[89] Wooldridge, J.M. (2002) Econometirc Analysis of Cross Section and Panel Data, The MIT Press: Cambridge.

図１ パネルデータの構造

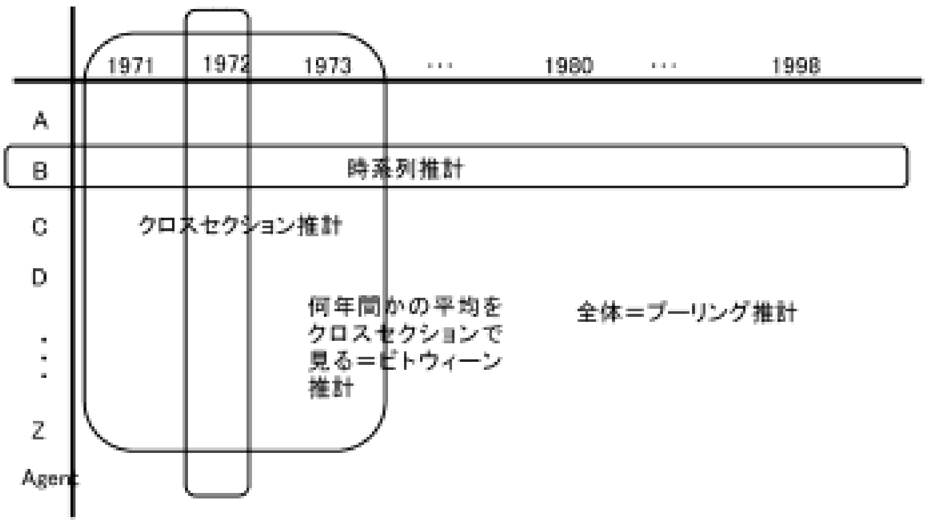
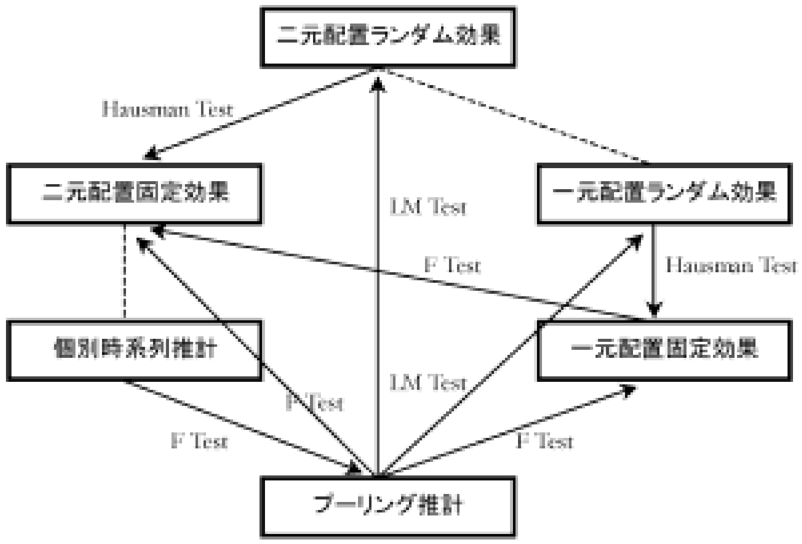


図 2　モデル選択の構造



注）　矢印は各検定・診断テストについて、矢の根元が帰無仮説、矢の先が対立仮設を表している。

図 3　国別の異質性を無視したことによる
フェルドスタイン・ホリオカ・パラドックスの発生

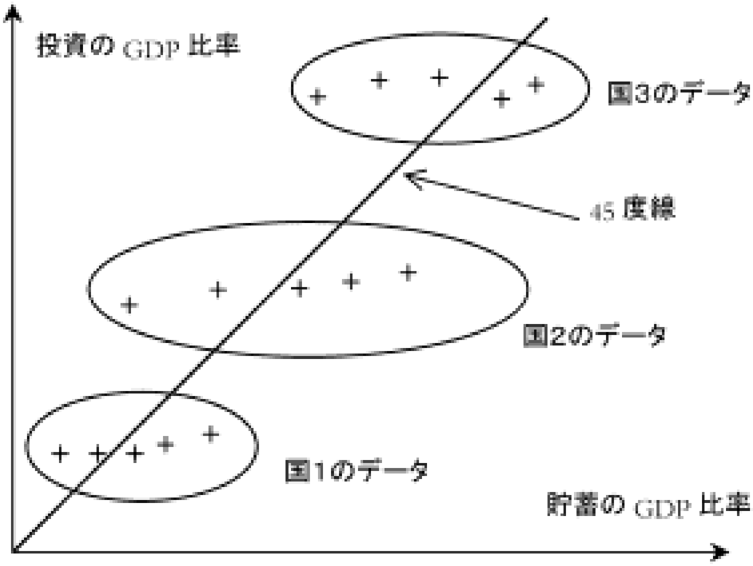
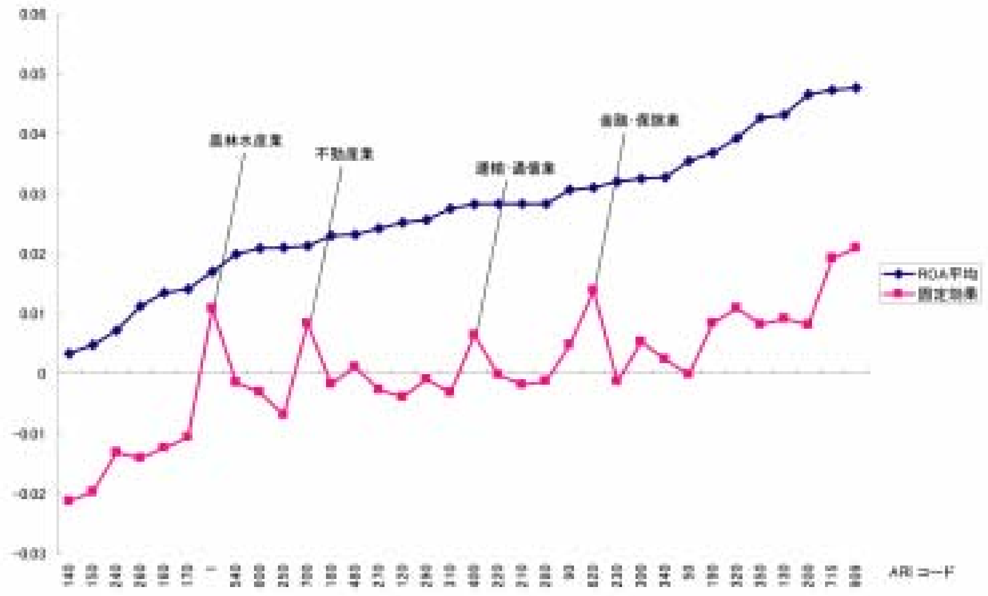


図4　ROA 平均と固定効果



ARIコード	業種名
140	繊維工業
150	衣服・その他の繊維製品製造業
240	なめし革・同製品・毛皮製造業
260	鉄鋼業
160	木材・木製品製造業（家具を除く）
170	家具・装備品製造業
1	農林水産業
540	小売業
600	飲食店
250	窯業・土石製品製造業
700	不動産業
180	パルプ・紙・紙加工品製造業
480	卸売業
270	非鉄金属製造業
120	食料品製造業
290	一般機械器具製造業
310	輸送用機械器具製造業
400	運輸・通信業
220	プラスチック製品製造業
210	石油製品・石炭製品製造業
280	金属製品製造業
90	建設業
620	金融・保険業
230	ゴム製品製造業
300	電気機械器具製造業
340	その他の製造業
50	鉱業
190	出版・印刷・同関連産業
320	精密機械器具製造業
350	電気・ガス・熱供給・水道業
130	飲料・たばこ・飼料製造業
200	化学工業
715	サービス業
908	持株会社

図 5　ビックマック平価と為替レート

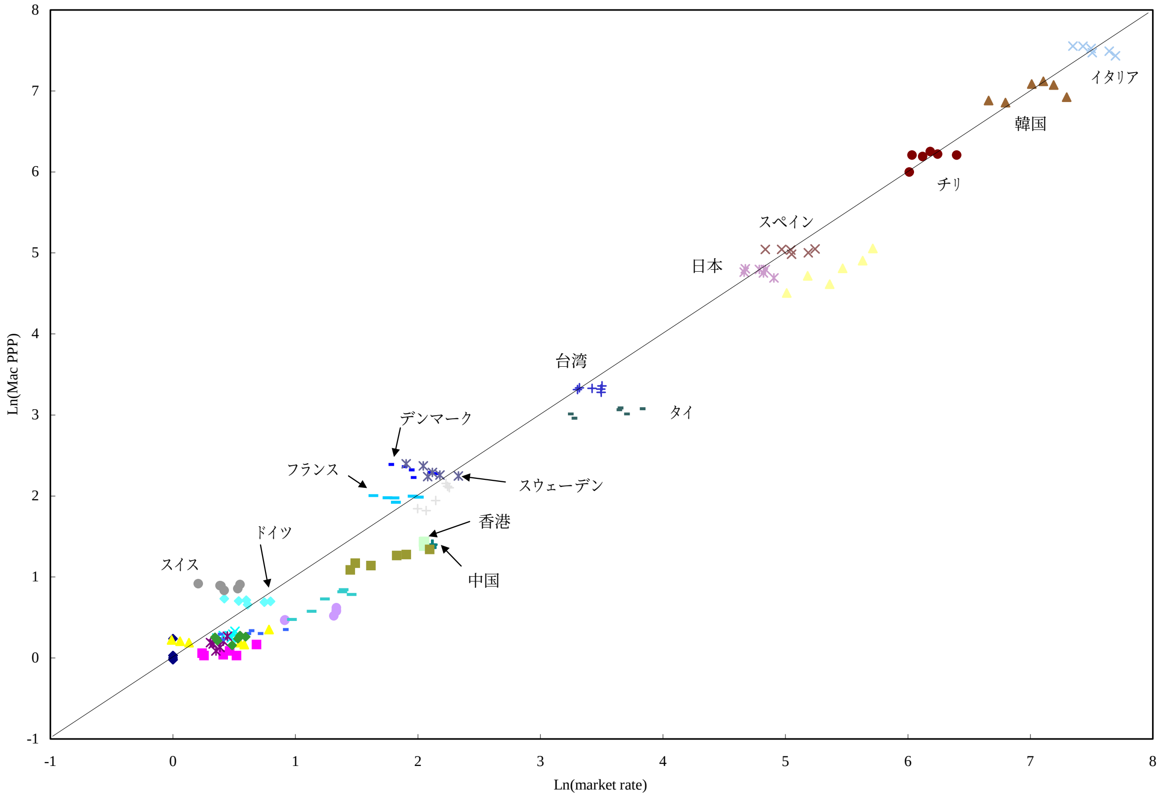


表1　フェルドスタイン･ホリオカ･パラドックスの推定

計測始期	計測終期	推 定 法	推 定 値	標準誤差
1960	1964	一元固定	0.532	0.091
1961	1965	国　別		
1962	1966	国　別		
1963	1967	国　別		
1964	1968	国　別		
1965	1969	一元固定	0.527	0.114
1966	1970	一元固定	0.587	0.116
1967	1971	一元固定	0.672	0.119
1968	1972	一元固定	0.639	0.095
1969	1973	国　別		
1970	1974	二元固定	0.493	0.115
1971	1975	国　別		
1972	1976	国　別		
1973	1977	国　別		
1974	1978	国　別		
1975	1979	国　別		
1976	1980	二元固定	0.351	0.102
1977	1981	国　別		
1978	1982	国　別		
1979	1983	二元固定	0.450	0.098
1980	1984	二元固定	0.534	0.099
1981	1985	二元固定	0.423	0.127
1982	1986	国　別		
1983	1987	一元固定	0.191	0.100
1984	1988	一元固定	0.303	0.098
1985	1989	国　別		

表 2　ROA 推定式

被説明変数：ROA	Fixed		Random		Between	
	Estimated Coefficient	t-statistics	Estimated Coefficient	t-statistics	Estimated Coefficient	t-statistics
商品売上シェア(全国)	0.2992	6.70	0.3014	6.83	0.7820	2.52
ln(常時従業員数)	0.0411	36.74	0.0414	36.93	0.4518	6.43
労働者シェア(地域)	0.2908	5.68	0.2845	5.57	0.7493	0.40
ln(実質総資産)	0.0019	3.86	0.0020	3.94	0.0646	1.12
有形固定資産シェア(業種)	-0.2945	-6.12	-0.2916	-6.41	-0.0778	-0.35
ln(実質負債)	-0.0028	-3.10	-0.0027	-2.97	-0.0137	-0.20
負債シェア(地域)	-0.1927	-4.83	-0.1883	-4.73	-1.0966	-1.59
負債資産比率	-0.0210	-5.30	-0.0207	-5.23	-0.0532	-0.31
負債資産比率の2乗	-0.0252	-15.92	-0.0253	-15.94	-0.1326	-1.39
自己資本比率	0.0340	11.54	0.0344	11.68	0.0084	0.10
流動性比率	-0.0006	-6.15	-0.0006	-6.21	-0.0119	-2.33
負債比率	0.0000	-2.13	0.0000	-2.09	0.0024	2.43
定数項	0.0490	14.91	0.0480	14.29	0.1543	1.33
Diagnostic Test						
Number of observation	90,838		90,838		90,838	
Number of groups (ari)	34		34		34	
R-sq: within	0.1729		0.1729		0.0800	
between	0.3204		0.3233		0.9051	
overall	0.1803		0.1803		0.0839	
F test that all u_i=0:	F(33, 90792) =	42.01				
Breusch and Pagan Lagrangian multiplier test for random effects:			chi2(1)=4924.98 Prob>chi2=0.0000			
Hausman specification test			chi2(12)=17.76 Prob>chi2=0.1233			

表 3 ビックマック購買力平価のダイナミックパネル推定

被説明変数 ： ln (Pit / Pt)、推計期間 ： 1996-2001

推定方法	効果	α	β1	β2	γ	長期弾力性
MDE	Fixed		1.0150	0.3052	0.0109	1.0262
	(S.E.)		(0.0128)	(0.0478)	(0.0120)	
MDE	Fixed		0.9995		0.0229	1.0230
	(S.E.)		(0.0151)		(0.0142)	
GLS	Random	-1.4641	1.0192	0.3196	-0.0089	1.0103
	(S.E.)	(6.5785)	(0.0155)	(0.0587)	(0.0141)	
GLS	Random	-0.2643	1.0009		0.0226	1.0241
	(S.E.)	(4.7374)	(0.0184)		(0.0174)	

注）サンプルは以下のとおり。
Argentine, Australia, Brazil, Britain, Canada, Chile, China, Denmark, France, Germany, Hong Kong, Hungary, Italy, Japan, Malaysia, Mexico, New Zealand, Poland, Russia, Singapore, South Africa, Korea, Spain, Sweden, Switzerland, Taiwan, Thailand.

表4　結婚確率のプロビット推定

Random-effect Probit Model

被説明変数：結婚ダミー (既婚=1,独身=0)	モデル 1-1			モデル 1-2			モデル 1-3			モデル1-4		
	推計値	漸近的t値		推計値	漸近的t値		推計値	漸近的t値		推計値	漸近的t値	
inc(年収)	-0.008	-12.18	***	-0.020	-10.79	***	-0.009	-8.73	***	-0.023	-8.89	***
inc^2(年収の2次項)	---	---		0.000	7.58	***	---	---		0.000	8.01	***
age(年齢)	0.408	12.44	***	0.343	9.86	***	0.898	5.20	***	1.179	5.80	***
age^2(年齢の2次項)	---	---		---	---		-0.029	-2.85	***	-0.044	-3.78	***
dep (貯蓄)	---	---		---	---		-0.002	-4.53	***	-0.003	-6.86	***
Cohabpa(同居)	-2.010	-10.65	***	-2.019	-9.91	***	-2.571	-8.07	***	-1.977	-7.19	***
Occupdu1(パート)	0.628	2.33	**	0.737	2.45	**	0.612	1.39		0.082	0.21	
Occupdu2(無業)	1.355	5.33	***	0.661	2.34	**	1.282	3.87	***	0.555	1.38	
constant	1.422	2.08		3.747	4.84	***	1.278	1.32		1.467	1.16	
診断テスト												
サンプル数	4,411			4,411			2,335			2,535		
グループ数	1,339			1,339			945			945		
Wald chi2	390.91			337.05			202.11			170.45		
rho	0.92			0.92			0.94			0.92		
Log likehood	-990.37			-961.08			-676.49			-656.47		
Likelihood ratio test of rho=0	1,281.62			1,215.18			781.39			722.54		
Prob>=chibar2	0.000			0.000			0.000			0.000		

GEE Population-averaged Model

被説明変数：結婚ダミー (既婚=1,独身=0)	モデル1-5			モデル 1-6			モデル 1-7			モデル 1-8		
	推計値	漸近的t値		推計値	漸近的t値		推計値	漸近的t値		推計値	漸近的t値	
inc(年収)	-0.002	-8.06	***	-0.005	-7.60	***	-0.002	-4.55	***	-0.005	-7.25	***
inc^2(年収の2次項)	---	---		0.000	3.98	***	---	---		0.000	4.50	***
age(年齢)	0.099	11.62	***	0.096	11.05	***	0.235	6.60	***	0.236	6.29	***
age^2(年齢の2次項)	---	---		---	---		-0.008	-3.74	***	-0.008	-3.67	***
dep (貯蓄)	---	---		---	---		-0.001	-4.16	***	-0.001	-4.93	***
Cohabpa(同居)	-0.616	-10.94	***	-0.577	-9.95	***	-0.528	-8.01	***	-0.513	-7.51	***
Occupdu1(パート)	0.223	3.39	***	0.167	2.60	***	0.199	2.45	**	0.124	1.54	
Occupdu2(無業)	0.335	5.08	***	0.174	2.67	***	0.402	4.51	***	0.188	2.24	**
constant	0.542	2.62	***	1.029	4.93	***	-0.379	-1.33		0.298	1.12	
診断テスト												
サンプル数	4,411			4,411			2,535			2,535		
グループ数	1,339			1,339			945			945		
Wald chi2	410.76			450.19			304.84			351.72		

注）***は1%水準、**は5%水準、*は10%水準で有意。