

パネルデータの分析手法の展望[†]

北村 行伸

(一橋大学経済研究所 教授)

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Mátyás and Sevestre (1996)、Hsiao (2003)、Baltagi (2008)、Lee (2002)、Wooldridge (2010)、Arellano (2003)などを挙げることができる。また学界活動も活発で、最近に限っても、2013年6月8～9日には中国厦門大学でパネルデータの世界的権威であるCheng Hsiao教授のパネルデータ分析への貢献を記念した国際会議が開催され、2013年7月4～5日にはロンドン・シティー大学カス・ビジネススクールで第19回国際パネルデータ会議が開催されている。さらにJournal of Econometrics、Econometricaなどの計量経済学の主要研究雑誌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。

図表-1はアメリカの6大経済学雑誌に掲載された労働経済学分野の論文の使用データ類型別構成比を示したものであるが、表より明らかなようにパネルデータを用いた実証研究が、近年急激に拡大していることがわかる¹⁾。

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、日本語で書かれた研究書や展望論文、概説書はまだ少ないが、樋口、岩田 (1999)、樋口、太田、家計経済研究所 (2004)、北村 (2005)、樋口 (2005)、樋口・太田・新保 (2006)、シャオ (2007)、

千木良・早川・山本 (2011) など徐々に出版されるようになってきた。

本稿ではパネルデータ分析手法の考え方とその意義について論じる。その中で強調しておきたいのは、パネルデータは同一主体の時系列方向のデータが複数のクロスセクション・データとして入っているものであり、データとしてはクロスセクション・データの分析手法と時系列データの分析手法を組み合わせを使っていうことである。従って、パネルデータ分析で用いられる統計手法として全く新しい統計手法があるわけではなく、既存の手法をパネルデータの特徴に合わせて改良したものということである。このことは、既存の統計手法の問題点や論争点がそのままパネルデータ分析の方法にも持ち込まれてきていることも意味している。

日本におけるパネルデータ調査の概要やそれを用いた研究については、本特集の他の論文で扱われていると思うので、ここでは繰り返さないが、パネルデータ分析の後発国として、パネルデータを用いた実証研究やパネルデータ分析手法に関する理論的研究が、数多く発表されている。そこでは後発国の利点を生かして、先進国のパネルデータ収集上の問題点や限界を克服しつつ、研究がすすめられている。

2. パネルデータとは何か

パネルデータとは同一の対象を継続的に観察し記録したデータのことを指す²⁾。これは例えば、

図表-1 アメリカの6大経済学雑誌に掲載された労働経済学分野の論文の使用データ類型別構成比(%)

論文種別 (使用データ)	1965～1969	1970～1974	1975～1979	1980～1983	1994～1997	1998～2000	2001～2003	2004～2007
理論分析	14	19	23	29	21	28	20	21
実証分析								
マイクロデータ	11	27	45	46	66	58	68	70
内パネルデータ	1	6	21	18	31	30	36	43
実験データ	0	0	2	2	2	4	3	2
横断面データ	10	21	21	26	25	25	31	26
集計時系列データ	42	27	18	16	6	10	9	6
集計横断面データ	24	24	15	10	8	4	4	4
2次加工データ	14	3	3	4	2	1	1	0
論文総数	106	191	257	205	197	302	290	405

出所：慶應義塾大学 パネルデータ設計・解析センター作成

複数の個人に家計簿を継続して記録してもらい、それを集計したデータであるとか、上場企業が企業業務内容を有価証券報告書として毎年、財務省に提出するデータを同一企業ごとにまとめたデータであるとか、あるいは、多数の同一の労働者の勤務情報や賃金情報を毎月記録したデータなどのことをパネルデータという³⁾。

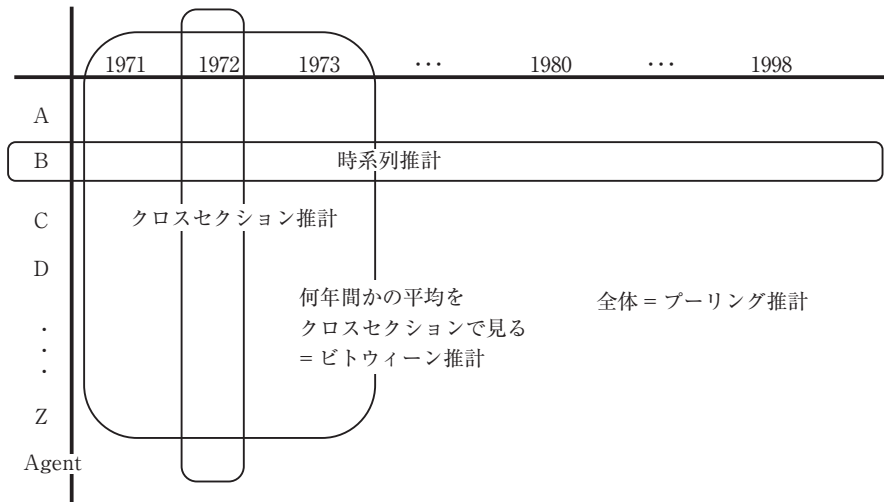
パネルデータを用いることの利点は第1に、これまでのクロスセクション・データや時系列データと比べた場合、観察点が格段に増加するので推定精度が上がるのが期待できる。もちろん、経済主体の多様性も増加するので、一概に推定精度が上がるとは言えない。むしろ、多様性を反映した分散不均一性を考慮した統計量を用いて推定精度を見る必要がある。しかし、膨大なクロスセクションの観察点（例えば、4,000人）を10年間にわたって継続的に調査すれば、パネルデータはクロスセクション・データの10倍の情報量があり、経験的に考えて、たとえ個人間の多様性を認めても、まだ統計的な情報量の多さによってもたらされる推定量の効率性、不偏性の上昇が期待できるのである。

第2に、パネルデータを用いることによって、観察不可能な経済主体間の違いを固定効果として抽出することが可能になる。経済モデルには理論的には想定できるが、現実には観察不可能な変数

がたくさんある。例えば、個人の能力の違いが所得の差として現れていると考えても間違いではないが、それをどのように測定するかについては合意はない。会社経営の質の違いが、企業業績の違いに反映されているとしても、それを観察することは不可能に近い。また、肥沃な農地とそうでない農地では、同じ労働量を投入して、同じ肥料を与えても、生産量に違いが出る。この肥沃度や水利のよさの違いを数量的に把握することは容易ではない。一般に、計量経済学では、被説明変数を説明するのに能力や質や肥沃度といった変数が利用可能でなければ、欠落変数として他の説明変数の推定量にバイアスを与えることが知られている。しかし、パネルデータ分析では、他の観察可能な変数による変動要因は全てコントロールした上で、観察不可能な変数を固定効果として捉えることで、観察不可能な変数を逆に抽出することができるようになる。観察不可能な個体差を固定効果推定として捉えるのではなく、独立確率分布に従う変数であると捉える場合には、これをランダム効果推定と呼ぶ。後で述べる通り、固定効果推定かランダム効果推定かは統計検定によって決定するのが望ましい。

第3に、パネルデータは時系列データの性質も持っており、経済主体がある時点の経済変動や政策に応じて、どのような反応を見せるかがわかる。

図表-2 パネルデータの構造



今日の経済理論は異時点間の最適化問題に関心を移しており、このような理論的想定が現実のデータとどの程度、整合的であるかを実証したいという動機は強い。これまで集計されたマクロ時系列データを用いて異時点間の最適化行動を検証することが多かったが、本来、最適化を行っているのは個別の経済主体であり、その行動を直接検証してはじめて理論の正当性を明らかにすることができるのである。

第4に、パネルデータやクロスセクション・データでは個票に記入されている数値を利用するので、記入ミス以外の集計誤差やバイアスは含まれていない。また、研究者が全ての個票の数値を観察できるので推計上の問題に関してもさまざまな解決方法を考えることができる。それに対してマクロデータでは、その集計過程が明らかではないので、研究者がマクロデータから集計誤差やバイアスを取り除くことはできない。

3. パネルデータの構造

パネルデータの基本構造は図表-2で表せば理解しやすいだろう。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ごとの平均を取り、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を集めて平均をとれば、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ごと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できない。

このような関係を数式で表すと次のようになる。

$$y_{it} = \alpha + \beta x_{it} + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

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モデルを想定する⁴⁾。

$$u_{it} = \mu_i + \lambda_t + v_{it} \quad (2)$$

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、 v_{it} は攪乱項を表す。

(1) 式のようなモデルに対して、まず、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。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2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(LSDV)と呼ばれる推定方法を用いる。

第3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、一元配置ランダム効果推定法を用いる。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 $y_{it} - \theta \bar{y}_i$ を $X_{it} - \theta \bar{X}_i$ 上で回帰する一般化最小二乗法(GLS)を用いる。ここで θ は個別ランダム効果と攪乱項の加重比を表す。

第4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、年ごとに生じた共通のショックの効果を排除するために時間(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。

本稿で強調したいことは、第1に、パネルデータ分析手法は相互に関連しているということである。考えればわかるように、パネルデータは時系列方向の情報とクロスセクション方向の情

報を含んでおり、パネルデータ分析で集約する情報は、これらの情報を加重平均したものとなっている。

第2に、仮説検定の哲学に従って、誤ったモデルを残さない、あるいは誤ったモデルに基づいて結論を導くことのないように細心の注意を払うことが必要だということである。計量経済学の検定では(1)理論が実際の数値と矛盾しないことを検討すること、(2)与えられたデータに対して適切な分析方法を用いているかを検討すること、を主たる目的としている。パネルデータを用いることの利点は、そのデータサイズの大きさのおかげで、さまざまな分析手法の中から適切な手法を選び、より適切な分析ができるということにある。同時に、さまざまな誤差が複合的に入り込んでおり、それを解きほぐすことによって、理論の問題が明らかになるという側面もある。これらの作業の重要性を強調しておきたい。

第3に、パネルデータではデータが不完備になることは常に起こることであって、むしろパネルデータの常態であると考えべきであるということである。本節で論じるように、多くの場合は推計上あまり大きな問題にはならないが、不完備の程度について事前にチェックしておくことが大切である⁵⁾。

4. ダイナミック・パネルデータ分析

経済現象は基本的には経済主体がダイナミックな枠組みの中で、最適化行動を行った結果であるという認識から、最近の経済学は、異時点間の資源分配の最適化を分析の中心にして、投資、消費、雇用、金融政策、財政政策などの議論が組み立てられている。パネルデータを用いる最大のメリットの一つに、同一経済主体の異時点間の変動、すなわち動学的最適化をデータとして捉え、それを実証的に検証できるということがある。個別経済主体の初期値を知りダイナミックな変動過程(運動方程式)を知ることができれば、将来の変動や政策反応を予測できることになる。これがパネルデータを経済学者が利用したがる大きな理由に

なっている。

パネルデータの動学的側面については、1960年代より意識されてきたことではあるが、1980年代の時系列分析の発展を受けて、本格的に進展してきた。とりわけ動学的最適化にマッチした形で誕生してきた一般化積率法（GMM）がパネルデータ分析に導入されて以来、急速な発展を遂げている⁶⁾。

さらに、生存時間解析（サバイバル分析）あるいはデュレーション・モデルとして知られている動学的分析は医学、生物学を中心とした自然科学の分野で広く応用されているし、政治学、社会学の分野を中心に社会科学の分野でも最近利用されるようになってきた。

一般にパネルデータでダイナミックな関係とは、被説明変数のラグ項が説明変数に入っていることをさす。すなわち以下のような構造をしている。

$$y_{it} = \gamma y_{it-1} + \beta x_{it} + \varepsilon_{it} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T$$

ここで、 ε_{it} は一元配置誤差構成要素モデルに従っているとする。

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

ダイナミック・パネル推定をめぐる大きな問題はラグ被説明変数が誤差項 ε_{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、時系列方向（ T ）には小さいということである⁷⁾。これは誤差項 u_{it} とが系列相関していない場合にも当てはまる。

5. 非線形パネル分析

ミクロ経済学の多くは選択問題を扱っている。消費者の消費財の選択、学校や就労先の企業の選択、居住地・住宅の選択、結婚・離婚の選択、出産の決定、退職の決定、再就職先の選択、保険への加入、証券投資の決定、企業の市場への参加の決定、企業の投資決定、企業の雇用決定、年金や退職給付制度の決定など数えればきりが無い。

近年、ミクロ経済学がゲーム理論で説明されるようになってきたが、これが可能になったのはミクロ経済学の多くの問題がゲームのように次の動きを選択するという形で設定されているからであると言える。これらの選択問題あるいは意思決定問題を実証的に分析してみるということはミクロ経済理論を検証するという意味できわめて重要なことであり、かつ現実的にも興味深いものである。

実証研究において、選択問題を扱おうとすると、選択した場合を1、選択しなかった場合を0とおくことで、本質的には質的な情報を数量化して、選択行動を統計的にモデル化することが可能となる。

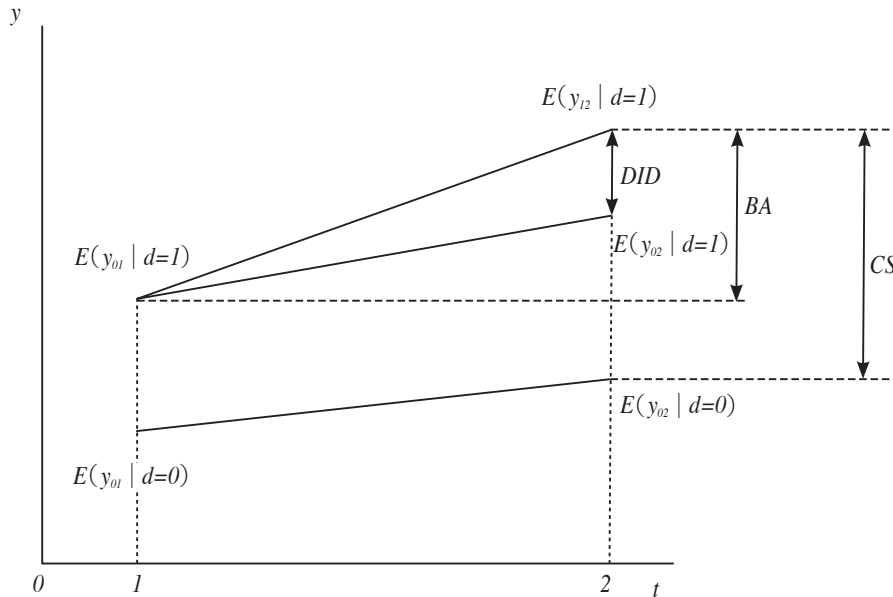
このアプローチの拡張として、選択すれば任意の正の数値をとるが、選択しなければ0であるという場合が考えられる。例えば、株式への投資は、投資を選択しなければ0にとどまるが、選択すればあとは投資家の投資額は個々人で違ってくる。選択した後の量の決定が任意の場合には、このようなアプローチが有益になってくるのである。

これまでこのような選択問題は主としてクロスセクション・データを用いて実証されてきたし、その統計手法もクロスセクション・データを中心に開発されてきた。当然予想されるようにパネルデータを用いれば、経済主体の意思決定の問題はより精緻に分析することができる。しかし、同時に、クロスセクション・データの選択モデルがすでに非線形モデルであり、その推計はかなり複雑になっているのだが、それをパネルデータに拡張するためには、かなり強い制約をおく必要が出てくる。

現在のところ、クロスセクション・データを用いた分析方法がある多項反応データを用いた順序プロビット、多項ロジット、ネステッド・ロジットなどのパネルデータ分析への応用はまだ時間がかかると思われる。これらのパネルデータ分析への拡張は将来の課題として残っている。

最新の研究では、例えば、Honoré and de Paula (2008) が、ゲーム論的な設定の下で、意思決定が同時に行われるような状況を分析している。具体的には、退職時期を夫婦で決める場合の

図表-3 政策評価の構造



ナッシュ均衡の選択の仕方を、統計分析によって識別する方法を提示している⁸⁾。

6. パネルデータと政策評価

パネルデータを利用することのもう一つの利点は、政策評価あるいはプログラム評価がより精緻にできるということにある。パネルデータを用いた政策評価の手法としてはDifference-in-Differences (DID) 推定が知られている⁹⁾。以下では、この概念を説明したい。

簡単化のために2期間の場合を考える。ある政策あるいはプログラムが第2期 ($t=2$) に導入され、第1期 ($t=1$) にはそのような政策は導入されていないとする。その政策あるいはプログラムに参加した人（グループ）を処置グループと呼び、 $d=1$ と表示する。参加しなかった人（グループ）を比較グループと呼び、 $d=0$ と表示する。政策評価の対象となる結果は y_{jt} と表示され、 $j=1$ は政策あるいはプログラムに参加した場合、 $j=0$ は参加しなかった場合を表す。このような前提の下に、処置を施すことによる効果（処置効果: Treatment Effect=TE）を求めることを政策評価あるいはプ

ログラム評価という。

政策評価の手法の比較は図表-3を見ることでイメージできる。まず、第2期のクロスセクション・データしか利用できないのであれば、政策に参加した人と参加しなかった人の結果の差を見るしかその効果を測る方法はない。これはCross Section 推定と呼ばれ図表-3ではCSとして表されている。見ての通り、TEはかなり大きく出ることがわかる。

$$CS = E(y_{12} | d=1) - E(y_{02} | d=0)$$

次に、政策参加者のみのデータしかない場合¹⁰⁾、比較グループの情報がないので、参加しなかった場合の結果が不明である。これはBefore and After 推定 (BA) と呼ばれ、次のように定義できる。

$$BA = E(y_{12} | d=1) - E(y_{01} | d=1)$$

BA 推定も現実の処置効果としては過大評価をしている可能性が高い。というのは、図表-3からわかるように不参加者であっても、第1期から第2期への間に結果が変化することがある。これは

マクロ経済の状況によって所得が上昇したりするケースに相当する。もちろん逆に景気後退によって、結果が第1期よりさらに落ち込むこともあり得る。これも特定の政策やプログラムでは対応できないマクロ経済上の変化であり、その影響を取り除いて処置効果を評価しなければならない。そこで考えられた手法がDifference-in-Differences推定(DID)で、次のように定義することができる。

$$\begin{aligned} \text{DID} &= E(y_{12} - y_{02} \mid d = 1) \\ &= E(y_{12} - y_{02} \mid d = 1) + E(y_{01} - y_{02} \mid d = 1) \\ &= E(y_{12} - y_{02} \mid d = 1) - E(y_{01} - y_{02} \mid d = 0) \end{aligned}$$

式の展開から明らかなように、政策に参加した人が参加しなかった時に第2期に経験するであろう結果 y_{02} は実現した値ではなく、仮想現実(counterfactuals)である。問題はいかにこの仮想現実の結果をうまく導くかということにかかっているが、ここでは、参加しなかった人が第1期から第2期にかけて結果を変化させた分は、政策に参加しなくても上昇(下落)する分としてBA推定からさらに差し引くということをしている。

もちろん、このように処置効果が簡単に推計できれば問題はないが、実際には、政策に参加した人が参加しなかった時に経験するであろう結果が、最初から参加しなかった人の時系列変化と同じトレンドを持つことは常に成り立つ関係ではない。また、Ashenfelter's dipとして知られているように、政策参加者は、参加直前には一時的に結果を下落させる可能性があり、直前と直後の結果を比べるのではなく、過去の何期間かの平均と事後の結果を比べた方がいいという場合も起こりうる。

このDID推定は、一定の条件の下では、すでに紹介したパネルデータ分析の固定効果推定と同値であり、パネルデータがなければ推計できない手法である。図表-3より明らかなように、DID推定はCS推定やBA推定よりもより厳密に処置効果を推定し、その結果としてTEは低く出ている。

このような手法は特定のミクロレベルの職業訓練や教育の効果の測定に用いられることが多いが、マクロ経済政策の評価にも使えるという研究

がPesaran and Smith (2012)で提示されている。興味深い研究なので紹介しておく、彼らの方法は、事前の評価と事後の評価を行うことで、結果に影響は与えるが、政策には影響を受けない変数をコントロールすることによって、純粋な政策効果を識別しようとするものである。具体的にはイングランド銀行の量的緩和策が長短金利スプレッドを100ベースポイント縮小した効果の測定を行っている。

7. パネルデータ分析の今後

第5節や第6節の最後に紹介した、Honoré and de Paula (2008)やPesaran and Smith (2012)らの最新のパネルデータ分析の手法について一般的に言えることは、現実直面する複雑な意思決定の過程や特定の政策の純粋な効果の抽出を、パネルデータを使って巧みに行っているということである。

パネルデータ分析の手法が今後どのような展開になるかは、現実の経済問題を理解する上で必要になる情報を最も効果的に抽出する方法を考えていく上で決まってくるものと思われる。現在日本で構築されているパネルデータはアメリカのPanel Study of Income Dynamics (PSID)やNational Longitudinal Surveys (NLS)などとの類似性はあるが、質問内容などに独自の側面も持っている。それは、日本の経済環境の中で、必然的に関心の高い問題への設問が多いことを反映しているからである¹¹⁾。例えば、少子高齢化の問題、未婚・晩婚化の問題、年金や介護の問題、高齢者医療の問題、住宅問題、子供の託児施設や育児休業制度の問題、いじめや学力などの義務教育の問題、大学教育の効果や国際競争力の問題、企業財務の問題、企業設備投資とグローバル化の問題、企業の生産性の問題、日本経済全体の成長戦略、貧困問題、財政健全化と税制改革の問題など挙げればきりが無いほど、重要かつ喫緊の問題が山積されている。

これらの問題に実証的な証拠を基に、頑強な政策を提言するためには、情報量の豊富なミクロ

データ、とりわけパネルデータを用いた実証研究が不可欠であろう。そのためにも、現在、日本で進行しているパネルデータ調査が途切れることなく継続され、貴重な情報が蓄積されていくことが、まず確保されなければならない。その上で、具体的な政策課題に対して、最もふさわしい分析方法を適用して、現状で最も信頼に足る政策提言や実証結果が蓄積されていくことが望まれる。

かつて、クロスセクション・データでの貢献ではあるが、アメリカで、イェール大学のJames Tobin教授が耐久消費財の需要関数を推定する際に考えた推定手法がトービット（Tobit）と呼ばれて、広範に使われるようになったこと、同じく、シカゴ大学のJames Heckman教授が女子労働供給行動を推定する際に考えた推定手法がヘキット（Heckit）あるいはヘックマンの2段階推定法と呼ばれ、これまた実証研究で広範に用いられているようになったことがある。日本のパネルデータを分析する中で、実証上きわめて有用で、汎用性の高い分析手法が開発されることを著者はひそかに期待している。そのためには、計量経済学者が理論研究に没頭するのではなく、実証研究に取り組んで、実際にパネルデータを使いながら、問題を設定し、それに解決を与えるための実用的な分析手法を開発することが必要になるだろう。実証研究者と計量経済学者との共同研究も有効かもしれない。

もう一つ、著者が指摘しておきたいのは、現在のような経済環境に不確実性が高まっており、大きな社会変動や経済危機が発生しやすい時には、歴史的事実から学ぶことが大切だということである。実際に、歴史統計のパネルデータ化も進行している。例えば、著者の所属する一橋大学経済研究所附属社会科学統計情報研究センターでは、日本の戦間期に農林省が毎年行っていた『農家経済調査』のサンプルの多くが継続的に同一家計となっていることに着目し、パネルデータとして、整理し、実証研究に使えるようなデータベースの形成を行っている。このデータベースを用いることによって、昭和恐慌以後の日本の農家経済が、社会変動に対して、どのような対応をしたの

かを具体的に分析することができる。また、それを補完するデータとして日本の農家以外の家計がつけてきた家計簿データをパネルデータとして整理しなおして、分析に使うという試みも現在進行中である¹²⁾。過去の歴史は変えることはできないが、当時には使えなかった分析手法を応用することで、歴史に新しい解釈を与えることは可能であるし、そこから新たな教訓を導き出すこともできるかもしれない。

8. おわりに

21世紀に入り、少子高齢化の時代、産業構造の大変革の時代と呼ばれているように、経済社会構造は大きく変動している。それに対応して経済政策の分野でも各種の規制緩和や構造改革が標榜されている。例えば、労働の分野を見ても、団塊世代の退職問題や若者の非就業化、転職の増加などに対応した政策が必要とされている。そのような構造的な問題にきめ細かく対応するためには、継続的に同一経済主体の経済行動を追跡したパネルデータを利用することが望ましい。また、超緩和的な金融政策はどこかで方向転換して、正常な金融市場取引の環境を復活させ、その中で伝統的な金融政策を行う時が来るはずである。そのための出口政策の実施のタイミングやミクロ経済主体への影響の違いを知るためにも、パネルデータを蓄積し、それを有効に活用して、間違いのない政策判断を行う必要がある。

繰り返すまでもないが、社会が変動している時に、スナップショット的なクロスセクション・データを用いただけでは厳密な見識は得られない。同一経済主体が状況の変化にどのように対応するかを深く観察することによってはじめて、冷静な政策評価ができるし、意味のある政策含意も得られる。この意味でもパネルデータに基づく研究の重要性は増すことはあっても、低下することはないだろう。

†本稿は北村（2005, 2006, 2007）から引用し、加筆・修正を加えたものである。

注

- 1) その背景には利用可能なパネルデータ調査の蓄積と統計ソフト（例えばStata, EViews, LIMDEP）で、パネルデータ分析がきわめて簡単にできるようになったことがあるだろう。
- 2) パネルデータをロンジチューディナルデータと呼ぶこともある。
- 3) より厳密にはパネルデータとはクロスセクション・データを各主体ごとに時系列方向に拡大したデータであると定義できる。従って、各国の同一時点でのマクロや金融変数のクロスカントリー・データを時系列データを用いて拡張してもパネルデータとして扱うことができる。すなわち、パネルデータとは必ずしもミクロの経済主体について調査したデータに限定されるものではないということである。
- 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るので、ここでもそれに従っている。
- 5) 具体的な推定方法の解説についてはHsiao (2003)、Baltagi (2008)、Wooldridge (2010)、北村 (2005) 等を参照されたい。
- 6) ダイナミック・パネルデータ分析の理論的側面について、さらに知りたい方はArellano (2003) が包括的な参考文献となっているので参照されたい。また、近年の日本人研究者のこの分野への貢献として千木良・早川・山本 (2011) がある。
- 7) 時系列が短いという問題に対しては逆に時間軸は長くなくてもよいと考えることもできる。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とも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ケースもある。
- 8) このアプローチと関連して、同時方程式の推定手法としてAnderson (2005) やRivers and Vuong (1988) が提案した制限情報最尤法 (LIML) をパネルデータ分析に応用している研究にAkashi and Kunitomo (2012) がある。
- 9) 政策評価の計量経済学的手法に関してはLee (2005)、Caliendo (2006)、Cameron and Trivedi (2005) を参照されたい。
- 10) これはプログラム参加者に対してアンケートを行ってプログラムの事後評価をする場合や、そもそもプログラムへの参加申込者に事前に記入してもらった情報のうち、最終的に参加した人の情報のみが残っているケースであり、比較的多く見られるデータの形態である。
- 11) 厚生労働省で調査している『21世紀縦断調査』とりわけ『21世紀出生児縦断調査』は出生時からの全国規模でのパネル調査であり、世界的に見ても、この規模で継続的に行われているパネル調査としてはきわめてまれで

ある。この調査はもともと少子化を経験している日本の実態を継続的に捉えることを目的として実施されたものであり、そのほか『21世紀成年者縦断調査』『21世紀中高年者縦断調査』も行われており、全体として日本における子供、成年、中高年をカバーした政策対応型パネル調査となっている。

- 12) 中村 (1993) はこのアプローチの嚆矢となるものであるが、当時はパネルデータ分析という意識がなく、分析手法も未開発であったが、現在の分析能力を用いて、再度分析を行うことは意味があると考えている。Atkinson and Cowell (1983) も所得プロセスに関する初期の貢献である。

文献

- 北村行伸, 2005, 『パネルデータ分析』岩波書店。
 ———, 2006, 「パネルデータの意義とその活用」『日本労働研究雑誌』551: 6-16。
 ———, 2007, 「パネルデータ分析」『ESP』505: 24-27。
 ———, 2009, 『ミクロ計量経済学入門』日本評論社。
 千木良弘朗・早川和彦・山本拓, 2011, 『動学的パネルデータ分析』知泉書館。
 中村隆英編, 1993, 『家計簿からみた近代日本生活史』東京大学出版会。
 樋口美雄・岩田正美編, 1999, 『パネルデータからみた現代女性』東洋経済新報社。
 樋口美雄・太田清・家計経済研究所編, 2004, 『女性たちの平成不況』日本経済新聞社。
 樋口美雄・太田清・新保一成, 2006, 『入門パネルデータによる経済分析』日本評論社。
 樋口美雄編, 2005, 『日本の家計行動のダイナミズム (1) 慶應家計パネル調査の特性と居住・就業・賃金分析』慶應義塾大学出版会。
 Akashi, Kentaro and Naoto Kunitomo, 2012, "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," *Journal of Econometrics*, 166: 167-183。
 Anderson, T. W., 2005, "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics*, 127: 1-16。
 Arellano, Manuel, 2003, *Panel Data Econometrics*, Oxford: Oxford University Press。
 Atkinson, A. B. and F. A. Cowell, 1983, *Panel Data on Incomes*, London: London School of Economics。
 Baltagi, Badi H., 2008, *Econometric Analysis of Panel Data*, 4th ed., Chichester: Wiley。
 Caliendo, Marco, 2006, *Microeconomic Evaluation of Labour Market Policies*, Berlin: Springer。
 Cameron, A. C. and P. K. Trivedi, 2005, *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press。
 Honoré, Bo and Áureo de Paula, 2008 "Interdependent Durations," mimeo。

- Hsiao, Cheng, 2003, *Analysis of Panel Data*, 2nd ed., Cambridge: Cambridge University Press. (= 2007, 国友直人訳『ミクロ計量経済学の方法——パネルデータ分析』東洋経済新報社。)
- Lee, Myoung-jae, 2002, *Panel Data Econometrics*, San Diego: Academic Press.
- Lee, Myoung-jae, 2005, *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford: Oxford University Press.
- Mátyás, László and Sevestre Patrick eds., 1996, *The Econometrics of Panel Data*, 2nd ed., Dordrecht: Kluwer Academic Publishers.
- Pesaran, M. Hashem and Ron P. Smith, 2012, "Counterfactual Analysis in Macroeconometrics: An Empirical Investigation into the Effects of Quantitative Easing," CESifo Working Papers, No. 3879.
- Rivers, Douglass and Quang H. Vuong, 1988, "Limited Information Estimators and Exogeneity Test for Simultaneous Probit Models," *Journal of Econometrics*, 39: 347-366.
- Wooldridge, Jeffrey. M., 2010, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge: The MIT Press.

きたむら・ゆきのぶ 一橋大学経済研究所 教授。
主な著書に『パネルデータ分析』（岩波書店, 2005）。
応用計量経済学、公共経済学専攻。