

「パネルデータ分析」*

北村行伸

一橋大学経済研究所

1. 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¹。これは例えば、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毎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いう²。

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ことにある。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、たとえ主体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沢山ある。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

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

¹ 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。

² 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。

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。後述する通り、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。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出来るのである。

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様々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出来ない。

2. パネルデータの構造

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、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。

図1 パネルデータの構造

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³。

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

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を用いる。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

³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。

ることが重要になってくる。

本稿で強調したいことは、第一に、パネルデータ分析手法は時系列とクロスセクション分析を組み合わせたものだということである。別の言い方をすれば、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。

第二に、仮説検定の哲学に従って、誤ったモデルを選択しない、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、様々な分析手法の中から適切な手法を選び、より適切な分析が出来るということにある。同時に、様々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あ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。

第三に、パネルデータではデータが途中で脱落してしまうという意味で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られるべきであるということである。多くの場合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⁴。

3. ダイナミックパネル分析

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

⁴具体的な推定方法の解説についてはHisiao (2003)、Baltagi (2001)、Wooldridge (2003)、北村 (2005)、樋口・太田・新保(2006)等を参照されたい。

に導入されて以来、急速な発展を遂げている⁵。

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。

$$y_{it} = \gamma y_{it-1} + \beta x_{it} + \varepsilon_{it} \quad i=1,2,\dots,N; \quad t=1,2,\dots,T \quad (3)$$

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。

$$\varepsilon_{it} = \mu_i + u_{it} \quad (4)$$

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 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⁶、時系列方向(T)には小さいということである⁶。これは誤差項 u_{it} が系列相関していない場合にも当てはまる。

4. 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

⁵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。

⁶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よいとも言える。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。

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志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極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でてくる。

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一般化されていない。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。

5. 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。このような様々な構造的な問題にきめ細かく対応し、かつそれを事後的に評価するためには、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。

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始めて、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。

参考文献

- [1] 北村行伸(2005)『パネルデータ分析』、岩波書店
- [2] 北村行伸(2006)「パネルデータの意義とその活用」『日本労働研究雑誌』No.551, pp.6-16
- [3] 樋口美雄、岩田正美(編)(1999)『パネルデータからみた現代女性 結婚・出産・就業・消費・貯蓄』、東洋経済新報社
- [4] 樋口美雄(編)(2005)『日本の家計行動のダイナミズム(1)慶応家計パネル調査の特性と居住・就業・賃金分析』、慶応義塾大学出版会
- [5] 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
- [6] Arellano, Manuel (2003) *Panel Data Econometrics*, Oxford University Press.
- [7] Baltagi, Badi H.(2001) *Econometric Analysis of Panel Data*, 2nd ed, Wiley.
- [8] Cameron, A.C. and Trivedi, P.K.(2005) *Microeconometrics: Methods and Applications*, Cambridge University Press.
- [9] Hisao, Cheng (2003) *Analysis of Panel Data*, 2nd ed, Cambridge University Press.
- [10] Wooldridge, Jeffrey. M.(2003) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press

図1 パネルデータの構造

