

## 第4講 ダイナミックパネル分析

### 4.1 はじめに

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、Balestra and Nerlove (1966) など 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 Arellano and Bond (1991) によって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。

本章ではダイナミック・パネルデータ分析の主要な結果をサーベイしているが、限られた紙幅では限定的なものにならざるを得ない。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、本書では全体としてクロスセクション方向の  $N$  が大きく、時系列方向の  $T$  が短いミクロ・パネルデータを扱っており、 $N$  も  $T$  も大きいマクロ・パネルデータについては扱ってこなかった。この分野も急速に研究が進んでおり、実証上の応用も増えている。それらの研究をフォローするのは不可能に近いが、Smith and Fuerter (2004) が現在のところ最も包括的なサーベイになっている。

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。これについては、日本語で読める教科書も良質のサーベイ論文もあるので、本章では明示的には扱わない<sup>1</sup>。ランダム係数モデル (the random coefficient model) もダイナミック・パネルデータ分析に含まれるべき問題であるが、本章では取り扱うことができない。関心

<sup>1</sup>日本語で読める教科書として大橋・浜田 (1995)、中村 (2001) を挙げておきたい。英語の最新の教科書として Singer and Willett (2003、第 9-11 章) がある。また、山口 (2001-2) はイベントヒストリー分析のサーベイとして有益である。サバイバル分析の手法を用いた最近の興味深い実証研究として阿倍・小黒 (2004) を挙げておく。

のある方は、Hsiao (2003, 第 6 章)、Maddala, Li, Trost and Joutz (1997)、Hsiao and Pesaran (2004) 等を参照されたい。

## 4.2 ダイナミック・パネルデータの考え方

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、

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \varepsilon_{it} \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (1)$$

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

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

ここで、 $\mu_i \sim iidN(0, \sigma_\mu^2)$  は固定効果を表しており、 $u_{it} \sim iidN(0, \sigma_u^2)$  は誤差項を表し、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  $\varepsilon_{it}$  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<sup>2</sup>。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。

より一般的には、Maddala (2001) はダイナミック・パネルデータ分析は次の二つのモデルに分類することが重要であることを示した。

### 系列相関モデル

$$y_{it} = \mathbf{x}'_{it}\beta + \mu_i + w_{it} \quad (3)$$

$$w_{it} = \rho w_{it-1} + u_{it} \quad |\rho| < 1 \quad (4)$$

### 状態依存モデル

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \mu_i + u_{it} \quad (5)$$

はダイナミックな要素が誤差項にあるモデルであり、はダイナミックな要素が被説明変数自体にあるモデルである。パネルデータ分析では後者を扱うことが多いが、前者についても研究されているので、以下では について解説した後、の問題に入りたい。

<sup>2</sup>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。

#### 4.2.1 系列相関モデル

系列相関が問題になるのは、それがモデルの特定化において重要な説明変数を落としている可能性を示唆しているからである。しかし、実際のデータではある程度の系列相関が見出されるのはむしろ当たり前であって、その問題をいかに軽減するかというのがここでの問題である。

パネルデータ分析において系列相関の問題を最初に取り上げたのは Lillard and Willis (1978) である。彼らは誤差項が AR(1) に従うケースをライフサイクル所得に当てはめ、マルコフ連鎖に従うと仮定した所得階層移動と対比する形で検討している。Lillard and Weiss (1979) では、アメリカにおける科学者の 1960-70 年における所得のダイナミックな変動と同時点内での所得変動に関して分析している。ここでも所得変動の誤差に系列相関がある場合を検討している。

Baltagi and Li (1991)、Wansbeek (1992) は (4) 式のような系列相関問題に対して、the Paris-Winsten (PW) transformation を用いて、系列相関を取り除いた後で最小二乗法で推定することを提案している。

代替的な推定方法として Maddala(2001)、Nerlove((2002、第 7 章) は一般化最小二乗法を提案している。(3) 式を LSDV 推定し、誤差項の推定値  $\hat{w}_{it}$  を得る。さらにそれを用いて (4) 式を OLS 推定し  $\hat{\rho}$  を得る。これらのパラメータを用いて (3) 式を次のように変換する。

$$\begin{aligned} y_{it}^* &= \mathbf{x}_{it}^* \beta + \mu_i^* + u_{it} & (6) \\ y_{it}^* &= y_{it} - \hat{\rho} y_{it-1} \\ \mathbf{x}_{it}^* &= \mathbf{x}_{it} - \hat{\rho} \mathbf{x}_{it-1} \\ \mu_i^* &= \mu_i (1 - \hat{\rho}) \end{aligned}$$

上式を再び LSDV 推定してやれば、 $\beta$  および  $\mu_i$  の一致推定を得ることができるというものである。

Bhargava, Franzini and Narendranathan (1982) は系列相関検定の定番である Durbin-Watson Statistic をパネルデータ用に一般化した。誤差項の系列相関を次のように定義すると、 $w_{it} = \rho w_{it-1} + u_{it}$ 、帰無仮説は  $H_0 : \rho = 0$  で対立仮説が  $H_1 : |\rho| < 1$  として、次の検定量を求めて判断することを提案している。

$$d = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{w}_{wit} - \hat{w}_{wit-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{w}_{wit}^2} \quad (7)$$

ここで  $\hat{w}_{wit}$  はウイズイン誤差である。誤差項が AR(1) でなく AR(n) になっても、この統計量に対応させることは容易である。

Baltagi and Li (1995) は系列相関と固定効果を同時検定するためのラグランジェ乗数検定を 3 つ提案している。AR(1) の系列相関がゼロでランダム効果を検定するための統計量、AR(1) の系列相関がゼロであり固定効果であることを検定するための統計量、ランダム効果の下で AR(1) の系列相関

がゼロかどうかを鑑定するための統計量、である。これらの統計量は誤差項が AR(1) であれ MA(1) であれ同じになることが示されている<sup>3</sup>。

#### 4.2.2 状態依存モデルの初期の研究

ダイナミック・パネル推定の研究は Balestra and Nerlove(1966) に始まる。彼らは パネルデータの特徴を生かさないプーリング OLS 推定は被説明変数のラグ項の係数は高くなるバイアスをもつ、固定効果推定は被説明変数のラグ項の係数を低くするバイアスをもつ、ダイナミック・モデルでは一般化最小二乗法は一致推定でも有効推定でもなくなる、という点をはじめて指摘し、その後のダイナミック・パネルデータ分析の方向性を決定づけた。についてはデータをプールするのかパネルデータとして使うのかという議論に発展した。Maddala (1971a, 1971b) では尤度比検定を用いることを提案している。については Nickell (1981) によって厳密に証明された。については代替的な推定方法として最尤法、操作変数法、一般化積率法 (GMM) など様々な推定方法の開発へと進展していった。この点については次節以後で詳しく解説する。

Maddala (2001) は最尤法推定を初期値に依存した条件付尤度関数と初期値には依存しない条件無し尤度関数に分けることが重要であると指摘している。条件無し尤度関数を仮定するためには、初期条件が定常状態あるいは均衡にあるような長期データであることを仮定していることになる。また条件付尤度関数では初期値が決定的に重要であり、かつ計算が大変であると論じている<sup>4</sup>。

Sevestre and Trognon (1996, pp.130-33) は、Maddala (1971a) が系列相関を取り除くために変換する時に用いたウェイト  $\lambda$  にちなんで  $\lambda$ -class と名付けられたウイズイン推定、一般化最小二乗法、プーリング最小二乗法、ピトウィーン推定に関して、 $\beta = 0$  の場合のラグ項の係数  $\gamma(\lambda)$  のバイアスを推定した。バイアスの大きさの順序は次のようになった。

$$p \lim_{within} \hat{\gamma}(0) < \gamma < p \lim_{GLS} \hat{\gamma}(\lambda) < p \lim_{poolingOLS} \hat{\gamma}(1) < p \lim_{between} \hat{\gamma}(\infty) \quad (8)$$

一般化最小二乗法 (GLS) では系列相関の問題を取り除いたはずであるが、

<sup>3</sup>詳しくは Baltagi and Li (1995) か Baltagi (2001, pp.90-95) を参照されたい。

<sup>4</sup>ダイナミック・パネル分析において初期値をどこに取るかは実証研究上、極めて重要な問題である。例えば、経済成長を扱うときに、19世紀には世界経済の中心にあったイギリスが、英国病と呼ばれる低成長に陥っていた1970年と、戦後高度成長期を経て奇跡の復興を成し遂げていた1970年の日本と、植民地から独立してようやく国造りを始めた1970年のアフリカ諸国を並べて考えた時に、それらの国を集めたパネルデータが1970年から始まっているからといって1970年の経済をそれぞれの国の初期値として扱っていいとは決して言えないだろう。医学や生物学では、このような初期値の問題を考えるために、それぞれの主体の初期値が等しいと考えられる状態でデータを揃え、それからどのような時間経過で変化が起こるかを捉えようとするアプローチである生存時間解析 (survival analysis) が用いられている。経済問題にもこのアプローチが使われるようになってきたが、医学のようにある病原体に感染した時間から死亡するまでの経過のように、きれいに初期状態が確定できないことが多く、現在のところ応用は限定されているし、実際に使われているケースでも初期値の扱いは恣意的なものが多い。

ラグ項がモデル変換後の誤差項と相関しているためにバイアスが残っている (Ridder and Wansbeek (1990, pp.566-571) )

Trognon(1978) は外生変数が次のような AR(1) 過程に従い、 $x_{it} = \delta x_{it-1} + w_{it}$   $w_{it} \sim N(0, \sigma_w^2)$ 、初期値  $y_{i0}$  を固定と仮定すると、最尤法推定は端点解になり、一致推定にはならないことを示した。また時系列 T が短く、被説明変数のラグ項の係数  $\gamma$  が負であったり、かなり小さい場合に最尤法推定は一致推定とならない可能性が高いことも示した。この研究は次節で説明する Anderson and Hsiao (1981,1982) に引き継がれた。

### 4.3 最尤法推定と操作変数法推定

Anderson and Hsiao (1981,1982) はダイナミックなパネルデータの推定問題を検討している。とりわけ、初期値に関して様々な仮定を置いて、推定結果の違いを比較している。彼らは次のようなモデルを考えている。

$$w_{it} = \gamma w_{it-1} + \rho' \mathbf{z}_i + \beta' \mathbf{x}_{it} + \mu_i + u_{it} \quad i = 1, 2, \dots, N. \quad t = 1, 2, \dots, T \quad (9)$$

$$y_{it} = w_{it} + \eta_i \quad (10)$$

$$\mu_i = (1 - \gamma)\eta_i \quad E(\eta_i) = 0 \quad Var(\eta_i) = \sigma_\eta^2 = \sigma_\mu^2 / (1 - \gamma)^2 \quad (11)$$

ここで  $\mathbf{z}_i$  は時間とともに変動しない外生変数である。

最尤法と一般化最小二乗法は初期値に依存しているため、初期値に関しては次の4つの仮定をおく。

(仮定1)  $y_{i0}$  は固定。この場合は任意の初期値から始まり、 $(\mu_i + \rho' \mathbf{z}_i) / (1 - \gamma) + \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j} \gamma^j$  に収束すると想定されている。また固定効果  $\mu_i$  は平均ゼロで分散一定の変数がランダムに選択されたものであると想定されている。データをいつから開始するかということが  $y_{i0}$  とは無関係に任意に決定されている場合には、 $y_{i0}$  を固定と仮定することには問題がある。すなわち  $\mu_i$  と  $y_{i0}$  が無相関であり、かつ  $y_{i1}$ 、 $y_{i1} \dots$  に影響を与えるような固定効果  $\mu_i$  を想定することは難しいのである。

(仮定2)  $y_{i0}$  はランダムであり、 $\mu_i$  と  $u_{it}$  から独立であり、次のような関係に従っている。 $y_{i0} = \bar{y}_0 + \varepsilon_i$ 、ここで  $\bar{y}_0$  は0期における全体の平均であり、 $\varepsilon_i$  は *iid* に従う。これはさらに2つのサブグループに分かれる。

(仮定2a)  $y_{i0}$  と  $\mu_i$  は独立しており、初期値の違いは時間とともに消滅する。

(仮定2b)  $y_{i0}$  と  $\mu_i$  は独立しておらず、 $cov(y_{i0}, \mu_i) = \varphi \sigma_{y_0}^2$  である。この場合、初期値の違いが将来の  $y_{it}$  に影響を与え、長期的には  $[\varphi \varepsilon_i / (1 - \gamma)] = \lim_{t \rightarrow \infty} E[y_{it} - \rho' \mathbf{z}_i / (1 - \gamma) - \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j} \gamma^j | \varepsilon_i]$  に収束する。

(仮定3)  $w_{i0}$  は固定であり、 $y_{it} = w_{it} + \eta_i$  かつ  $\mu_i = (1 - \gamma)\eta_i$  なので、 $y_{it}$  と  $\mu_i$  は相関している。 $y_{i0}$  は任意の値から始っても  $\eta_i + \rho' \mathbf{z}_i / (1 - \gamma) + \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j} \gamma^j$  に収束する。データの開始時期と確率過程の開始時期は一致している必然性はない。

(仮定 4)  $w_{i0}$  はランダムであり、(3)と同様の関係に従っている。ここでは  $w_{it}$  の性質によって、4つのサブグループに分かれる。

(仮定 4a)  $w_{i0}$  はランダムで、平均は全体共通  $\theta_w$  で、分散は  $\sigma_u^2/(1-\gamma^2)$  であり、均一である。

(仮定 4b)  $w_{i0}$  はランダムで、平均は全体共通  $\theta_w$  で、分散は任意の  $\sigma_{w_0}^2$  となる。

(仮定 4c)  $w_{i0}$  はランダムで、平均は  $\theta_{i0}$  で、分散は  $\sigma_u^2/(1-\gamma^2)$  である。

(仮定 4d)  $w_{i0}$  はランダムで、平均は  $\theta_{i0}$  で、分散は任意の  $\sigma_{w_0}^2$  となる。

Anderson and Hsiao (1981,1982) はこのような仮定の下で、合計 8 通りの推定を行っている。初期値の仮定に応じて尤度関数が違ってくるが、ここでは全てについて厳密な展開をすることはしないで、一般型で表現しておく<sup>5</sup>。

$$L(\gamma, \rho, \beta, \gamma, \eta, \sigma_u^2, \sigma_w^2, \sigma_\mu^2) = (2\pi)^{-NT/2} |v|^{-N/2} \\ \times \exp\left\{-\frac{1}{2\sigma^2} \sum_i \sum_t (y_{it} - \gamma y_{it-1} - \rho' z_i - \beta' x_{it})^2\right\} \quad (12)$$

ここで  $v$  は誤差分散共分散行列を表わしており、誤差構成要素によって表現が違ってくる。

一般にはこの式を未知パラメータに関して最大化すれば、最尤法推定が得られる。最尤法推定の一致性は初期値と標本数  $N$  と時系列数  $T$  に依存している。ただし、この最適化は  $\sigma_\mu^2 = 0$  の場合には端点解となる。この点は Anderson and Hsiao (1981) で検討されている。

最尤法推定では初期値の仮定によって尤度関数の形状が変わり、また推定結果も変動することがわかった。ここでは代替的に初期値に依存しない推定法である操作変数法について論じたい。

固定効果推定であれランダム効果推定であれ、上の (9) (10) 式から一階の階差をとれば時間とともに変動しない  $z_i$  と  $\mu_i$  は消去されてしまう。すなわち、

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})' \beta + \gamma(y_{it-1} - y_{it-2}) + (u_{it} - u_{it-1}) \quad (13)$$

このモデルはラグ被説明変数の階差が誤差項  $u_{it}$  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<sup>6</sup>、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。

操作変数に応じて  $\beta$  と  $\gamma$  は次のように推計される<sup>7</sup>。

( $y_{it-2} - y_{it-3}$ ) を操作変数として使った場合は次のようになる。

<sup>5</sup>詳細については Anderson and Hsiao (1982) および Hisao (2003, 第 4 章) を参照されたい。

<sup>6</sup>具体的には  $y_{it-1}$  と  $u_{it-1}$  は (1) 式より明らかに相関している。

<sup>7</sup>以下の式の展開は Hsiao (2003, pp.85-86) を参照。

$$\begin{pmatrix} \gamma_{iv} \\ \beta_{iv} \end{pmatrix} = \left[ \sum_{i=1}^N \sum_{t=3}^T \begin{pmatrix} (y_{i,t-1} - y_{i,t-2})(y_{it-2} - y_{it-3}) & (y_{it-2} - y_{it-3})(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \\ (\mathbf{x}_{it} - \mathbf{x}_{it-1})(y_{it-2} - y_{it-3}) & (\mathbf{x}_{it} - \mathbf{x}_{it-1})(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \end{pmatrix} \right]^{-1} \quad (14)$$

$$\times \left[ \sum_{i=1}^N \sum_{t=3}^T \begin{pmatrix} y_{it-2} - y_{it-3} \\ \mathbf{x}_{it} - \mathbf{x}_{it-1} \end{pmatrix} (y_{it} - y_{it-1}) \right]$$

$y_{it-2}$  を操作変数として使った場合には次のように書ける。

$$\begin{pmatrix} \gamma_{iv} \\ \tilde{\beta}_{iv} \end{pmatrix} = \left[ \sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} y_{it-2}(y_{i,t-1} - y_{i,t-2}) & y_{it-2}(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \\ (\mathbf{x}_{it} - \mathbf{x}_{it-1})y_{it-2} & (\mathbf{x}_{it} - \mathbf{x}_{it-1})(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \end{pmatrix} \right]^{-1} \quad (15)$$

$$\times \left[ \sum_{i=1}^N \sum_{t=2}^T \begin{pmatrix} y_{it-2} \\ \mathbf{x}_{it} - \mathbf{x}_{it-1} \end{pmatrix} (y_{it} - y_{it-1}) \right]$$

実証上、ラグ変数の水準であれば最小期間は 2 期間ですむが、階差であれば最低 3 期間は必要になる。また 3 期間以上のデータであれば、最適操作変数としては  $(y_{i,t-1} - y_{i,t-2})$  と  $(y_{it-2} - y_{it-3})$  あるいは  $y_{it-2}$  の相関がどれぐらいあるかで判断すれば良い<sup>8</sup>。

第 2 段階として推計した  $\hat{\beta}$  と  $\hat{\gamma}$  を (9) 式に代入し、 $\rho$  を最小二乗法で推定する。

$$\bar{y}_{it} - \hat{\gamma}\bar{y}_{it-1} - \hat{\beta}'\bar{\mathbf{x}}_{it} = \rho'z_i + \mu_i + \bar{u}_{it} \quad i = 1, \dots, N \quad (16)$$

ここで、 $\bar{y}_i = \sum_{t=1}^T y_{it}/T$ ,  $\bar{\mathbf{x}}_i = \sum_{t=1}^T \mathbf{x}_{it}/T$ ,  $\bar{u}_i = \sum_{t=1}^T u_{it}/T$ 。

第 3 段階として分散共分散行列  $\sigma_u^2$ ,  $\sigma_\mu^2$  を推定する。

$$\sigma_u^2 = \frac{\sum_{i=1}^N \sum_{t=2}^T [(y_{it} - y_{it-1}) - \hat{\gamma}(y_{it-1} - y_{it-2}) - \hat{\beta}'(\mathbf{x}_{it} - \mathbf{x}_{it-1})]^2}{2N(T-1)} \quad (17)$$

$$\sigma_\mu^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \hat{\gamma}\bar{y}_{i-1} - \hat{\rho}'z_i - \hat{\beta}'\bar{\mathbf{x}}_i)}{N} - \frac{1}{T}\hat{\sigma}_u^2 \quad (18)$$

これらの推定が一致性を持つことは、初期値には依存していない。N が T あるいは両方が無限に近づくと、操作変数法パラメータ  $\gamma$ ,  $\beta$ ,  $\sigma_u^2$  は一致推定となる。 $\rho$ ,  $\sigma_\mu^2$  が一致推定になるのは N が無限になる場合のみであり、N が固定されている場合には、T が無限になろうと一致推定にはならない。

Anderson and Hsiao (1982) は N 固定で T が無限、T 固定で N が無限の場合に分けた上で、上述の 8 通りについて最尤法推定法、固定効果推定法、操作変数法推定を行った。推計結果をまとめると次のようになる。最尤法でラ

<sup>8</sup>Arellano (1989) は操作変数としてはラグ変数の水準  $y_{it-2}$  や  $y_{it-3}$  を用いた方がラグの階差  $(y_{it-2} - y_{it-3})$  を用いるより望ましいとしている。

グ項のパラメータ  $\gamma$  が一致推定にならないのは (仮定 3) で  $T$  が固定、 $N$  が無限の場合のみである。時間とともに変動しない変数の係数  $\rho$  は  $N$  固定、 $T$  無限の場合には常に一致推定にはならない。また (仮定 3) で  $T$  が固定、 $N$  が無限の場合にも一致推定にはならない。操作変数法では時間とともに変動しない変数の係数  $\rho$  については最尤法と同じで  $N$  固定、 $T$  無限の場合には一致推定にはならない。また、ラグ項のパラメータ  $\gamma$  は全ての場合で一推定になる。固定効果推定法では  $\rho$  は推定できないが、ラグ項のパラメータ  $\gamma$  が一致推定にならないのは  $N$  固定、 $T$  無限の全ての場合と (仮定 3) で  $T$  が固定、 $N$  が無限の場合である。ここでの結果はダイナミック・モデル推定は誤差項の条件や初期値の仮定というより、 $T$  や  $N$  の仮定により強く依存していることを物語っている。

Hsiao, Pesaran and Tahmiscioglu (2002) や Fujiki, Hsiao and Shen (2002) では代替的に Chamberlain (1982,1984) を嚆矢とする推定方法である最小距離推定法 (Minimum Distance Estimation: MDE) を用いることを主張している<sup>9</sup>。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  $(\beta, \gamma)$  を決定するということである。すなわち、

$$\min \left[ \sum_{i=1}^N \Delta \mathbf{u}_i^* \Omega^{-1} \Delta \mathbf{u}_i^* \right] \quad (19)$$

ここで  $\Omega$  は  $\Delta \mathbf{u}_i^*$  の共分散行列、 $\Delta \mathbf{u}_i^* = [\Delta y_{i1} - \beta \Delta x_{i1} - \gamma \Delta y_{i0}, \Delta y_{i2} - \beta \Delta x_{i2} - \gamma \Delta y_{i1}, \dots]$

この方法は誤差分布が正規分布に従うことを要求していないので、有効推定ではないが、 $N$  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。

#### 4.4 一般化積率法推定

操作変数法で推計するアプローチに対して、Arellano and Bond (1991)、Ahn and Schmidt (1995) ら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  $y$  の水準は誤差項の階差とは無相関であることを示すことができる<sup>10</sup>。

$$E[y_{is}, (u_{it} - u_{i,t-1})] = 0, \quad s = 0, 1, \dots, t-2, \quad t = 2, \dots, T \quad (20)$$

Arellano and Bond (1991) はこのような情報も利用した一般化積率法 (GMM) を用いることを提唱している。

$$\frac{1}{n} \sum_{i=1}^n y_{is} [(y_{it} - y_{i,t-1}) - (y_{i,t-1} - y_{i,t-2})' \gamma - (x_{it} - x_{i,t-1})' \beta] = 0 \quad (21)$$

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

<sup>9</sup>MDE の詳細については Chamberlain (1982,1984)、Lee(2002, 第 3 章) を参照。

<sup>10</sup>すなわち直交条件が成立する。これは Holtz-Eakin(1988)、Holtz-Eakin, Newey and Rosen(1988) によって指摘された。

被説明変数のラグ項の階差を次々に取っていくと、それに対応して使える操作変数は  $(y_{i1}, y_{i2}, y_{i3}, \dots, y_{iT-2})$  と増えていく。これを行列で表現すると次のようになる<sup>11</sup>。

$$W_i = \begin{bmatrix} [y_{i1}] & 0 & \dots & 0 \\ 0 & [y_{i1}, y_{i2}] & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & [y_{i1}, \dots, y_{iT-2}] \end{bmatrix} \quad (22)$$

操作変数が誤差ラグ項と無相関である条件 (20) は次のようにベクトル形式で書き直せる。

$$E(W_i' \Delta u_i) = 0 \quad (23)$$

ここで (1) 式のようなダイナミック・モデルを考えてみよう。

$$(y_{it} - y_{it-1}) = (y_{it-1} - y_{it-2})\gamma + (x_{it} - x_{it-1})'\beta + (u_{it} - u_{it-1}) \quad (24)$$

$$\Delta y_{it} = \Delta y'_{it-1}\gamma + \Delta x'_{it}\beta + \Delta u_{it} \quad i = 1, 2, \dots, N$$

この式に対して操作変数行列  $W_i$  を掛けてパラメータを推定するのが Arellano and Bond (1991) の一般化積率法 (GMM) である。(24) 式を次のように書き換えて、未知のパラメータについて積率法を当てはめるのである。

$$W' \Delta y_{it} = W' \Delta y'_{it-1}\gamma + W' \Delta x'_{it}\beta + W' \Delta u_{it} \quad (25)$$

$\gamma$ 、 $\beta$  について解くと次のように表せる。

$$\hat{\gamma}_{GMM} = [(\Delta y_{it-1})' W \hat{V}_N^{-1} W' (\Delta y_{it-1})]^{-1} [(\Delta y_{it-1})' W \hat{V}_N^{-1} W' (\Delta y_{it})] \quad (26)$$

$$\hat{\beta}_{GMM} = [(\Delta x_{it-1})' W \hat{V}_N^{-1} W' (\Delta x_{it-1})]^{-1} [(\Delta x_{it-1})' W \hat{V}_N^{-1} W' (\Delta y_{it})] \quad (27)$$

ここで  $V_N = \sum_{i=1}^N W_i' (\Delta u_i) (\Delta u_i)' W_i$  である<sup>12</sup>。

説明変数  $x_{it}$  が厳密に外生変数であれば、 $E(x_{it}u_{is}) = 0, \forall t, s = 1, 2, \dots, T$  であり、 $x_{it}$  が  $\mu_i$  と相関している場合には、(24) 式のようなダイナミック・モデルに関して全ての  $x_{it}$  が操作変数になり得る。もし  $x_{it}$  が外生変数ではなく先決変数 (predetermined) であり、 $E(x_{it}u_{is}) \neq 0$  for  $s < t$ 、 $E(x_{it}u_{is}) = 0$  for  $s \geq t$  の場合は、操作変数になり得るのは  $(x_{i1}, x_{i2}, \dots, x_{is-1})$  までである。先決変数と外生変数が混在しているような場合でも、操作変数行列  $W_i$  を適切に書き換えればよい。

<sup>11</sup>以下は Baltagi (2001, p.132) を参照。

<sup>12</sup>実際、Arellano and Bond (1991, p.279) では (25) 式を均一分散を仮定した一般化最小二乗法で推計した one-step 推定と初期値や誤差項の分布に制約を課さずに GMM で推計した two-step 推定を用いている。誤差項が *iid* に従う場合には漸近的に等しくなる。

Arellano and Bond (1991) の GMM 推定で重要な仮定は誤差項に系列相関がないということである。この点が確保されていなければ、操作変数として使えないものが出てくるし、その結果として GMM 推定は一致推定でなくなりバイアスをもつことになる。そこで次のような系列相関検定が Arellano and Bond (1991) で提案されている。

$j$  次の誤差ラグ項の自己相関係数の平均を次のように定義する<sup>13</sup>。

$$r_j = \frac{1}{T-3-j} \sum_{t=4+j}^T r_{tj} \quad (28)$$

ここで  $r_{tj} = E(\Delta u_{it} \Delta u_{it-j})$  である。帰無仮説  $H_0 : r_j = 0$  として検定量は次のように定義する。

$$m_j = \frac{\hat{r}_j}{SE(\hat{r}_j)} \quad (29)$$

ここで  $\hat{r}_j$  は標本から得られた  $\Delta \hat{u}_{it}$  と  $\hat{r}_{tj} = N^{-1} \sum_{i=1}^N \Delta \hat{u}_{it} \Delta \hat{u}_{it-j}$  に基づいて計算されたものを使う。

Arellano and Bond (1991) では操作変数に関する Sargan (1958) の過剰識別制約テストも導入している。

$$s = \Delta \hat{u}' W [\sum_{i=1}^N W_i' (\Delta \hat{u}_i) (\Delta \hat{u}_i)' W_i]^{-1} W' (\Delta \hat{u}) \sim \chi_{p-k-1}^2 \quad (30)$$

ここで  $p$  は行列  $W$  における行数、 $\Delta \hat{u}$  は (25) 式における推定誤差を表わしている。Arellano and Bond (1991) では誤差ラグ構造が 1 階と 2 階の場合を考慮している。

Ahn and Schmidt(1995) は  $y$  の水準からだけではなく、 $y$  と誤差項の階差 ( $u_{it} - u_{it-1}$ ) 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。

$$E(y_{is} \Delta u_{it}) = 0 \quad t = 2, \dots, T, \quad s = 0, 1, \dots, t-2 \quad (31)$$

さらに、次のような直交条件も利用できることを示した。

$$E(u_{iT} \Delta u_{it}) = 0 \quad t = 2, \dots, T-1 \quad (32)$$

上の 2 種類の直交条件を合わせると  $T(T-1)/2 + (T-2)$  の制約が加わることになる。(32) 式は  $\gamma$  が 1 に近いが、 $\sigma_\mu^2/\sigma_u^2$  が大きい場合には、有効な情報 (パラメータの分散が小さい) を提供してくれることがモンテカルロ実験からわかった。

Ahn and Schmidt(1995) によれば、(31)(32) 式は通常の誤差項が満たすべき仮定より緩やかな仮定である。すなわち次のように書くことができる。

(仮定 1) 全ての  $i$  と  $t$  に対して、 $cov(u_{it}, y_{i0})$  は等しい。ちなみに、従来の仮定では  $cov(u_{it}, y_{i0}) = 0$  である。

<sup>13</sup>Arellano (2003, pp.121-23) を参照。

(仮定 2) 全ての  $i$  と  $t$  に対して、 $cov(u_{it}, \mu_i)$  は等しい。ちなみに、従来の仮定では  $cov(u_{it}, \mu_i) = 0$  である。

(仮定 3) 全ての  $i$  と  $t \neq s$  に対して、 $cov(u_{it}, u_{is})$  は等しい。ちなみに、従来の仮定では  $cov(u_{it}, u_{is}) = 0$  である。

この(仮定 1-3)に基づく GMM 推定は先に論じた Chamberlain (1982,1984) の最小距離推定 (Minimum Distance Estimator) と漸近的には等しくなり、漸近的に有効推定となることが示されている。

Blundell and Bond (1998) はこれまで GMM 推定の問題点と指摘されてきた操作変数の弱相関問題と、これは GMM 推定固有の問題ではないが Anderson and Hsiao (1981,1982) 以来、ダイナミック・パネル・モデルの本源的な課題とされてきた初期値問題とを取り上げて、それを解決する目的で従来の GMM をシステム GMM に拡張した。

まず、操作変数の弱相関問題であるが、Arellano and Bond の GMM 推定は  $\gamma$  が 1 に近づくか、固定効果  $\mu_i$  の分散が大ききときにはバイアスが大きくなることが知られている。Blundell and Bond (1998) はこの問題を明らかにするために  $T=3$  と仮定し、直交条件を  $E(y_{i1} \Delta u_{i3}) = 0$  に限定した。これによってパラメータ  $\gamma$  は適度識別される。この場合、GMM 推定は次のような操作変数法推定になる。

$$\Delta y_{i2} = \pi y_{i1} + \mu_i + u_{i2} \quad i = 1, 2, \dots, N \quad (33)$$

ここで  $\gamma$  が 1 に近づくか、固定効果  $\mu_i$  の分散が大きければ、パラメータ  $\pi$  は 0 に近づいても不思議はない。この場合、 $y_{i1}$  は  $\Delta y_{i2}$  とは弱相関になる。このことは次の式からも明らかである。すなわち  $E(y_{i1}\mu_i) > 0$  で、 $\sigma_\mu^2 = var(\mu_i)$ 、 $\sigma_u^2 = var(u_{it})$  の時、パラメータ  $\pi$  の確率極限は次のように表せる。

$$p \lim \hat{\pi} = (\gamma - 1) \frac{k}{(\sigma_\mu^2 / \sigma_u^2) + k} \quad k = \frac{(1 - \gamma)}{(1 + \gamma)} \quad (34)$$

Blundell and Bond (1998) は GMM 推定のバイアスはかなりの程度、Nelson and Startz (1990) や Staiger and Stock (1997) によって論じられた操作変数の弱相関に起因していることを示したのである。

初期値の問題に関しては次のように考えている。すなわち、Ahn and Schmidt (1995) らが主張した直交条件に加えて、 $T-3$  本の直交条件を導入する。

$$E(u_{it} \Delta y_{it-1}) = 0 \quad t = 4, 5, \dots, T \quad (35)$$

ところで、 $\Delta y_{i2}$  は観察可能であるから、次の直交条件も導入可能である。

$$E(u_{i3} \Delta y_{i2}) = 0 \quad (36)$$

この条件は初期値  $y_{i1}$  がどのように発生したかに依存している。すなわち初期値を次のように定義する<sup>14</sup>。

<sup>14</sup>これは  $y_{i0}$  がそれ以前の  $y_{i0-t}$  と全て等しくなる(定常状態)と仮定した場合に得られる

$$y_{i1} = \frac{\mu_i}{1-\gamma} + u_{i1} \quad (37)$$

$t = 2$  以後の関係式は与えられているので、 $y_{it}$  の全ての流れが確定する。ところで、(36) 式は次のように書き換えることができる。

$$E[(\mu_i + u_{i3})(u_{i2} + (\gamma - 1)u_{i1})] = 0 \quad (38)$$

であり、これはさらに次のような十分条件に書き換えることができる。

$$E(u_{i1}\mu_i) = E(u_{i1}u_{i3}) = 0 \quad i = 1, 2, \dots, N \quad (39)$$

この条件が意味しているのは、初期値  $y_{i0}$  からの乖離  $u_{i1}$  が初期値のレベル  $\mu_i/(1-\gamma)$  とは無相関であることである。

Blundell and Bond (1998) では  $\gamma$  が 1 に近いが、 $\sigma_\mu^2/\sigma_u^2$  が大きい場合でも、(35)(36) 式を用いることで GMM 推定のバイアスを劇的に改善できることを示した。そのために GMM 推定において操作変数行列  $\mathbf{Z}_i^+$  を次のように定義したものをを使ったシステム GMM を提案した。

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & 0 \\ 0 & 0 & 0 & \dots & \Delta y_{iT-1} \end{bmatrix} \quad (40)$$

ここで  $\mathbf{Z}_i$  は  $(T-2) \times m$  行列で従来の GMM 推定で用いられた操作変数行列である。

$$\mathbf{Z}_i = \begin{bmatrix} y_{i1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{iT-2} \end{bmatrix} \quad (41)$$

このように従来の GMM 推定を拡張することによって、 $\sigma_\mu^2/\sigma_u^2 = 1, T = 4$  の場合の従来の GMM 推定とシステム GMM 推定の分散比率をとると、 $\gamma = 0$  では 1.75、 $\gamma = 0.5$  では 3.26、 $\gamma = 0.9$  では 55.40 と  $\gamma$  の値が上昇するに従って、バイアスの差が拡大していくことが明らかになった。逆に言えば、システム GMM を用いることで、従来 GMM 推定が破綻すると言われてきた、 $\gamma$  が 1 に近いが、 $\sigma_\mu^2/\sigma_u^2$  が大きい場合でも、かなりバイアスを抑えることができることを示したのである。

値であるが、すでに何度も論じてきたように、このような初期値を経済変数で実際に求めることは極めて難しい。また、初期値が個人の個体属性である固定効果と関連していると考えるのはモデルの構造上自然であるが、現実の固定効果というのは初期値の帰結として出てきたものも多く（例えば親の経済状態を反映した学歴）、初期値と固定効果の関係についてはさらに深い議論が必要である。

#### 4.5 推定方法の比較と評価

これまでの議論からも明らかなように、ダイナミック・パネルデータ分析の手法は多岐にわたっており、一つの方法に収斂しているとは言い難い。しかし、大きく分ければ第 4.3 節で紹介した最尤法と第 4.4 節で紹介した一般化積率法 (GMM) が 2 つの対立軸として残っている。実際、GMM は最小二乗法 (OLS)、一般化最小二乗法 (GLS)、操作変数法 (IV)、最小距離法 (MDE) などを特殊型として含む極めて一般的な推定方法ではあるが、最尤法とは相容れないものがある<sup>15</sup>。

これまでの議論に基づいて両者の利点と問題点を整理したのが図表 4.1 である。

図表 4.1 ダイナミック・パネル推定方法の比較

一般に GMM は関数型を仮定せずに直交条件に基づいてパラメータを推計するので、線形、非線型にかかわらず簡便なパラメータ推定が出来ることが強みであるが、サンプルが小さければ母集団の直交条件に一致する保証はなくなる。また、系列相関や単位根を持つようなデータに対しては有効性が急激に低下するなどの問題点が指摘されている。

最尤法は、はじめからデータのあるべき姿がわかっているのではなく、データから尤もらしい関数パラメータを推定するという考え方に基づいている。これも線形、非線型にかかわらず使えるし、サンプルサイズにはそれほど影響されない。しかし、初期値を与えたり、通常は関数型を事前に特定化する必要がある。

これらの利点と問題点を考慮しながら、どちらの推定方法が望ましいかを決めるとするのが現状であり、どちらか一方が正しく、他方が間違っているということではない。また、これらの推定上の問題は固定的なものではなく、様々な方向から改善される余地があり、実際、GMM でも最尤法でも膨大な研究がなされている。

計量経済学では推定バイアスを判断する方法としてモンテカルロ実験が用いられる。この実験は次のように行う<sup>16</sup>。事前に関心のあるパラメータの値を設定し、説明変数を乱数を用いて作り、さらに誤差項にも分散や系列相関に関して仮定をおいて乱数データを発生させる。これらを合わせて被説明変数を作り上げる。こうして人工的に作り上げたデータセットに対して各種の

<sup>15</sup>最尤法と積率法の対立は第 1 章で解説したように、フィッシャーとピアソンの対立以来延々と続いている。近年でも GMM 推定と最尤法の論争は続いているが、すでに見たように Blundell and Bond (1998) に見られるように GMM の推定方法に最尤法の問題意識を取り込むようになったり、最尤法の方でも empirical likelihood といって関数型を特定せずにノンパラメトリックに尤度を求めるという方法が精力的に開発されている。ちなみに、empirical likelihood に関しては Owen (2001)、Mittelhammer, Judge and Miller (2000) などを参照されたい。

<sup>16</sup>STATA を用いたモンテカルロ実験法の簡単な解説は Johnston and DiNardo (1997、第 11 章) を参照されたい。

推定方法を適用して、そのパラメータ推定値と事前に設定した真のパラメータ値を比較することで推定方法のバイアスを調べるのである<sup>17</sup>。

これまでの議論からも明らかなように、現実のデータは完全に独立した確率変数ではないし、誤差は様々な形で入ってくるので、不均一分散であったり強い系列相関があったりするので、モンテカルロ実験の結果ですべての決着がつく訳ではないことには注意を要する。しかし、管理実験できない経済学において、推定方法を比較検討する手法としてはモンテカルロ実験は極めて有効であり、広範に行われている。以下ではダイナミック・パネルデータ推定に関するモンテカルロ実験の結果を見てみよう。

まず GMM を中心とする実験としては次のような結果が得られている。

Arellano and Bond (1991) では  $N=100$ 、 $T=7$ 、1000 回反復、誤差均一分散で系列相関はないという設定の下で実験を行っている。そこでは (1) GMM 推定の小サンプルバイアスは小さい、(2) GMM 推定の分散は操作変数法 (IV) の分散より小さい、(3) two-step GMM 推定は下方バイアスがある、ことが報告されている。

Ziliak (1997) は直交条件を加えることで得られる有効性の改善とバイアスの増加との代替関係を論じている。Ziliak (1997) では  $N=532$ 、 $T=8$ 、誤差項は *iid* に従うモデルを考察しているが、被説明変数のラグ項が説明変数に入ると言う意味でのダイナミック・モデルではないことに注意されたい。ここで得られた結論は、直交条件が増加すると GMM の下方バイアスは拡大していくということである。また Keane and Runkle (1992) で論じられた forward filter 2SLS(FF)<sup>18</sup> が最もいいパフォーマンスを挙げていることを示唆している。

Ahn and Schmidt (1999) は推定の有効性を失うことなくどこまで直交条件を削減することができるかという問題を論じている。誤差項が不均一分散の場合には、いくつかの直交条件は推定を改善しない、すなわち不要であることが明らかにされた。

関連した論文として Crepon, Kramarz and Trognon (1997) がある。この論文では関心のあるパラメータと無関係なパラメータに分けて、無関係なパラメータを削除しても有効性を失うことはないことを示した。すなわちそれに応じた直交条件を減らすことができるということである。

Alonso-Borrego and Arellano (1999) は  $N=100$ 、 $T=4,7$ 、1000 回反復のモンテカルロ実験を行い、一回階差の誤差に対する直交条件を用いた GMM 推定では、操作変数の相関が弱い場合にはバイアスが大きくなることを確認している。

Alvarez and Arellano (2003) は固定効果推定、one-step GMM、制限情報最尤法 (LIML) に対して  $N$  と  $T$  を変化させた時にどのようなことが起こるかを考

<sup>17</sup> 様々な推定方法を様々なパラメータの組み合わせの下で比較するにあたっては同じ乱数を用いる必要があるので、乱数の発生を適切にコントロールしてやる必要がある。このためには乱数を発生させる時のコマンドである *seed* をきちんと管理しなければならない。STATA では *seed* は 123456789 に初期設定されている。

<sup>18</sup> この方法は Hayashi and Sims (1983) が時系列で行った forward filtering をパネルデータに適用したものである。この方法は先決変数の系列相関をフィルターをかけて取り除いてしまうもので、階差をとるだけの方法より有効であるとされている。

察したものである。T/N が正の定数に収束する場合には、下方バイアスはそれぞれ  $1/T$ 、 $1/N$ 、 $1/(2N-T)$  に比例する形で生じる。固定した T に対しては GMM と LIML は漸近的に等しいが、T が増加すると LIML のバイアスの方が GMM のバイアスより小さくなる。

Blundell and Bond (1998) は既に見たように、操作変数を拡張したシステム GMM を提唱した。N=100,200,500、T=4 という条件の下でモンテカルロ実験を行い、システム GMM はバイアスを劇的に減少させることを示した。

次に、最尤法を中心とする実験からは次のような結論が得られている。

Binder, Hsiao, and Pesaran (2000)、Hsiao, Pesaran and Tahmiscioglu (2002)、Hsiao (2003) 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=50,500、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。また同じモンテカルロ実験では、MDE 推定のバイアスは少なくとも GMM 推定より小さく、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、GMM 推定より小さいことが示されている。

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ることができると論じている。

Wansbeek and Bekker (1996) では最適操作変数を選択し、それをを用いてダイナミック・モデルを推定しているが、最尤法の分散の方が操作変数法の分散よりはるかに小さいことを発見している。また操作変数法および GMM 推定と最尤法との分散の差は Ahn and Schmidt (1995) で提案されている追加的な直交条件を加えれば減少することが示されている。

このようにモンテカルロ実験の結果は多岐にわたっているし、その結果も必ずしも同じ方向に出ている訳ではない。その理由は第一にモンテカルロ実験のデザインにあると思われる。説明変数の自己相関や誤差項の系列相関などをどれくらい明示的に取り入れるかで結果が大きく変わっている。第二に小サンプル・バイアスについては、Alvarez and Arellano (2003) で始められていることであるが、T と N の関係、すなわち T/N あるいは N/T の関係がパラメータへ及ぼす影響を調べる必要がある。現状では N と T が恣意的に与えられており、両者の関係は考慮されていない。

とはいえモンテカルロ実験の手法に関しては対立があるわけではなく、包括的な実験手法が確立されれば、現状よりも確固とした結論に達し、よりバイアスの少ない有効な推定方法が共有されるようになるだろう。

#### 4.6 パネル単位根推定

経済学で単位根を持つデータとは一般にはランダムウォークに従っている時系列データを指すが、ランダムウォークに従うとは、取引が効率的に行われ全ての情報が資産価格に反映されているので、今日の取引価格は昨日の価格とほぼ同じであり、違いがあるとすれば全く予想されなかった新しい情報が入ってきたためであると考えられている。

このような性格を有するデータは主として金融市場で取引されている財産価格、例えば為替や金利などが中心になっているが、いずれもある程度長い時系列データを前提としている。

このようなアプローチがパネルデータ分析の枠組みで論じられるようになったのは、為替などの金融データが利用可能になり（例えば、Frankel and Rose (1996)、Pedroni(2001)）、また国際機関を中心に様々なマクロ、セミマクロのクロスカントリー・パネルデータが公表され、マクロ経済学の実証分析で使われるようになってきたという背景もある（例えば、Sala-i-Martin (1996)、Nerlove (2000)、Quah (1996)）。加えて、地域データをパネル化して時系列変動と地域間変動という2方向の情報を用いて、地域経済の問題を解明しようとする研究も出てきており、そこでも時系列分析から発生したパネルデータ分析の手法が用いられるようになってきた（例えば、Nagahata, Saita, Sekine and Tachibana (2004)）。

本節では、これらの発展のうち、近年の時系列分析の中で標準的な手法となっている変数の定常性検定をパネルデータ分析に拡張したものを紹介するととどめる。パネルデータが非定常な変数である場合にも時系列データと同様に 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は異なっている。

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(2003)、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。

$$y_{it} = \gamma y_{it-1} + u_{it} \quad i = 1, 2, \dots, N \quad (42)$$

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。

$$H_0 : \gamma_1 = 1 \quad vs \quad H_1 : \gamma_1 < 1 \quad (43)$$

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。

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_N = \gamma = 1 \quad vs \quad H_1 : \gamma_1 = \gamma_2 = \dots = \gamma_N = \gamma < 1 \quad (44)$$

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。

$$H_0 : \gamma_i = 1 \text{ for all } i \text{ vs } H_1 : \gamma_i < 1 \text{ at least one } i$$

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  $\sigma^2$  の分布に従い、 $t$  統計全体の平均  $\bar{t}$  は平均  $M$ 、分散  $\sigma^2/N$  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  $\chi^2$  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 ( $P_\lambda$  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(1999) 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。

#### 4.7 STATA コード

ここでは Wooldridge (2003) が教育目的で公開しているデータ (<http://www.msu.edu/~ec/wooldridge/book2.htm>) に入っている WAGEPAN.DTA というデータを使ってダイナミック・パネルデータ推定を行う。このデータはもともと Vella and Verbeek (1998) で用いられたものである。彼らの研究では労働組合加入率が低下している中で、労働組合に入ることによるプレミアムを測定している。しかし、組合プレミアムはその他の固定的要因に影響を受けており、その効果を確定することは難しいことを論証している。Wooldridge (2003) はこのデータを使って賃金関数を測定し、結婚することによるプレミアムを確認している。

ここで推計しているのは次のようなダイナミックな賃金関数である。

$$\begin{aligned} \ln wage_{it} = & \alpha + \gamma \ln wage_{it-1} + \beta \exp er_{it} + \delta \exp er_{it}^2 + \zeta hours_{it} + \eta hours_{it-1} \\ & + \theta union_i + \kappa educ_i + \lambda married_i + \nu poorhlth_i + \mu_i + \nu_t + u_{it} \end{aligned}$$

ここで  $\ln wage$  = 賃金の対数値、 $exper$  = 労働経験年数、 $hours$  = 年間労働時間、 $union$  = 労働組合に属していれば 1 をとるダミー変数、 $educ$  = 教育年数、 $married$  = 結婚していれば 1 をとるダミー変数、 $poorhlth$  = 健康状態が悪い場合に 1 をとるダミー変数、 $\nu_t$  = 年ダミーを表している。

推計に使った STATA コードは次のようになる。

```
/**Dynamic Panel **/  
/*Pooled OLS*/  
reg lwage lwage_1 d81 d82 d83 d84 d85 d86 d87 exper expersq  
hours hours_1 union educ married poorhlth  
  
/*LSDV*/  
xtreg lwage lwage_1 d81 d82 d83 d84 d85 d86 d87 exper expersq  
hours hours_1 union educ married poorhlth, fe  
est store fixed  
xtreg lwage lwage_1 d81 d82 d83 d84 d85 d86 d87 exper expersq  
hours hours_1 union educ married poorhlth, re  
xttest0  
est store random  
hausman fixed random  
  
/*Anderson-Hsiao IV Estimation*/  
xtivreg lwage lwage_1 d81 d82 d83 d84 d85 d86 exper expersq (  
hours hours_1 = union educ married poorhlth ), re ec2sls  
  
/*Anderson-Hsiao Maximum Likelihood Estimation*/  
xtreg lwage lwage_1 d81 d82 d83 d84 d85 d86 d87 exper expersq  
hours hours_1 union educ married poorhlth, mle  
  
/*Arellano-Bond GMM Estimation*/  
xtabond lwage hours hours_1 d81 d82 d83 d84 d85 d86 d87,  
lags(1) inst(exper expersq union educ married poorhlth) artests(2)  
xtabond lwage hours hours_1 d81 d82 d83 d84 d85 d86 d87,  
lags(1) inst(exper expersq union educ married poorhlth) artests(2)  
robust
```

図表 4.2-4.4

図表 4.2 はプーリング OLS 推定、固定効果推定、ランダム効果推定、最尤法推定の結果を報告してある。ここでは賃金のラグ項のパラメータ  $\gamma$  がどのような値をとるかに関心がある。パラメータの順序は次のようになる。

$$FE(0.092) < MLE(0.17) < RE(0.49) = OLS(0.49)$$

通常のパネルデータ推定ではハウスマン検定により固定効果推定が選択されることになるが、推定されたパラメータは低すぎるように思われる。これは、賃金のラグ項の内生性を考慮していない可能性があり、それを改善する目的で操作変数法 (IV) 推定と one-step GMM 推定を行った結果が図表 4.3-4.4 に報告されている。ここではパラメータ  $\gamma$  は次のようになる。GMM と IV では操作変数の入り方が違うのでサンプル数も異なるので直接比較するには注意を要するが、内生性バイアスはいずれの場合も改善されている。

$$GMM(0.31) < IV(0.57)$$

図表4.1 ダイナミック・パネル推定の比較

	GMMおよび操作変数法	最尤法
利点	<ul style="list-style-type: none"> <li>1) 固定効果・ランダム効果を区別する必要がない</li> <li>2) 初期値や誤差項の分布に関して特定化する必要がない</li> </ul>	<ul style="list-style-type: none"> <li>1) 直交条件は時間Tに依存しない</li> <li>2) 有限(小)サンプルの下でも有効推定を得られる</li> <li>3) 被説明変数のラグ項のパラメータに依存しない</li> <li>4) 時間とともに変化しない変数を推定できる</li> </ul>
問題点	<ul style="list-style-type: none"> <li>1) 直交条件は時間Tの2乗の割合で増えていく</li> <li>2) 有限(小)サンプルの下は有効推定が得られない可能性がある(特にTが小さい場合)</li> <li>3) <math>\lambda</math>が1に近づくとGMMのバイアスが拡大する</li> <li>4) 時間とともに変化しない変数の推定ができない(固定効果が推定できない)</li> <li>5) 誤差系列相関がある場合には使えない操作変数が出てくる</li> </ul>	<ul style="list-style-type: none"> <li>1) 固定効果・ランダム効果を推定するための特定化が必要になる</li> <li>2) 初期値や誤差項の分布に関する特定化が必要である</li> </ul>

出典) Hsiaoの講義録(2003年春)より著者作成

图表4.2

Dependent Variable: lwage	Pool OLS		Fixed		Random		MLE	
	Estimated Coefficient	t-statistics	Estimated Coefficient	t-statistics	Estimated Coefficient	z-statistics	Estimated Coefficient	z-statistics
lwage_1	0.4900	44.31	0.0924	8.04	0.4900	44.31	0.1696	14
d81	0.3583	14.91	0.7329	4.14	0.3583	14.91	0.1518	7.04
d82	0.3121	12.18	0.0148	0.88	0.3121	12.18	0.1182	4.33
d83	0.3176	11.30	-0.0135	-0.81	0.3176	11.30	0.1188	3.41
d84	0.3453	11.37	-0.0143	-0.88	0.3453	11.37	0.1458	3.41
d85	0.3547	10.87	-0.0201	-1.25	0.3547	10.87	0.1659	3.26
d86	0.3813	10.95	-0.0115	-0.72	0.3813	10.95	0.2005	3.39
d87	0.3959	10.72	(dropped)		0.3959	10.72	0.2341	3.47
exper	0.0335	3.30	0.1406	17.73	0.0335	3.30	0.0949	7.5
expersq	-0.0016	-2.64	-0.0056	-9.88	-0.0016	-2.64	-0.0047	-8.44
hours	-0.0001	-11.72	-0.0001	-13.68	-0.0001	-11.72	-0.0001	-13.44
hours_1	0.0001	10.09	0.0001	6.18	0.0001	10.09	0.0001	7.62
union	0.0893	7.03	0.0614	4.06	0.0893	7.03	0.0764	5.39
educ	0.0489	12.45	(dropped)		0.0489	12.45	0.0765	8.68
married	0.0663	5.70	0.0320	2.22	0.0663	5.70	0.0493	3.69
poorh1th	-0.0480	-1.16	-0.0082	-0.22	-0.0480	-1.16	-0.0148	-0.41
_cons	-0.1671	-2.72	1.0394	24.87	-0.1671	-2.72	0.0701	0.56
<b>Diagnostic Test</b>								
Number of observation	4316		4316		4316		4316	
Number of groups (ari)	545		545		545		544	
R-sq: within	---		0.2764		0.1735		---	
between	---		0.0672		0.8268		---	
overall	0.4623		0.1484		0.4623		---	
Log Likelihood	---		---		---		-1149.4376	
F test that all u <sub>i</sub> =0:			F(544, 3757) = 6.19 Prob>F = 0.0000					
sigma_u			0.3633		0		0.273	
sigma_e			0.2740		0.2740		0.2753	
rho			0.6374		0		0.4958	
Breusch and Pagan Lagrangian multiplier test for random effects:			chi2(1) = 580.65 Prob > chi2 = 0.0000					
Hausman specification test			chi2(13) = -6324.07					
Likelihood-ratio test of sigma_u = 0 for MLE			chibar2(01) = 936.32 Prob>chibar2 = 0.000					

図表4.3

Dependent Variable: lwage	IV	
	Estimated Coefficient	z-statistics
hours	0.0000	0.25
hours_1	0.0006	6.66
lwage_1	0.5692	34.36
d81	0.4165	8.83
d82	0.2822	7.38
d83	0.2314	6.03
d84	0.1848	5.56
d85	0.1491	4.90
d86	0.1509	5.06
exper	0.0702	4.43
expersq	-0.0033	-3.48
_cons	-1.1540	-7.55

  

Diagnostic Test	
Number of observation	4316
Number of groups	545
R-sq: within	0.0532
between	0.3976
overall	0.1900
Wald test	chi2(11) = 1610.58 Prob>chi2 = 0.0000
sigma_u	0.0000
sigma_e	0.4483
rho	0.0000

注1) 被操作変数 = hours, hours\_1、操作変数 = lwage\_1, d81, d82, d83, d84, d85, d86, exper, expersq, union, educ, married, poorhlth,

注2) Baltagi(2001) の the error component two-stage least square (EC2SLS) に従っている。

图表4.4

Dependent Variable: lwage		GMM	
one-step results	Estimated Coefficient	Robust z-statistics	
lwage_1	0.3111	7.81	
hours	-0.0002	-8.95	
hours_1	0.0001	6.33	
d82	-0.0284	-1.76	
d83	-0.0187	-1.42	
d84	-0.0021	-0.17	
d85	-0.0100	-0.89	
d86	-0.0055	-0.37	
_cons	0.0382	9.59	
<b>Diagnostic Test</b>			
Number of observation		3189	
Number of groups		545	
Sargan test	chi2(26) = 84.13 Prob>chi2 = 0.0000		
Wald test	chi2(8) = 138.34		
Arellano-Bond test for residual AR(1)	z = -8.63 Prob>z = 0.0000		
Arellano-Bond test for residual AR(2)	z = 2.20 Prob>z = 0.0280		

注1) 操作变数 = exper、expersq、union、edu、married、poorhlth。