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[*2]。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。

6.2 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、

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \varepsilon_{it} \quad i = 1, 2, ..., N;\ t = 1, 2, ...T \tag{6.1}$$

ここで、γ はスカラー、x'_{it} は $1 \times K$ 行列、β は $K \times 1$ 行列。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。

$$\varepsilon_{it} = \mu_i + u_{it} \tag{6.2}$$

ここで、$\mu_i \sim iidN(0, \sigma^2_\mu)$ は固定効果を表しており, $u_{it} \sim iidN(0, \sigma^2_u)$ 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 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）には大きいが、時系列方向（T）には小さいということである[*3]。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。

[1] 系列相関モデル

$$y_{it} = \mathbf{x}'_{it}\beta + \mu_i + w_{it} \tag{6.3}$$
$$w_{it} = \rho w_{it-1} + u_{it} \quad |\rho| < 1 \tag{6.4}$$

[2] 状態依存モデル

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \mu_i + u_{it} \tag{6.5}$$

---

*2 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。

*3 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。

[1] はダイナミックな要素が誤差項にあるモデルであり、[2] 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 [1] について解説した後、[2] の問題に入って行きたい。

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 (1991a)、Wansbeek (1992) は (6.4) 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。(6.3) 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 (6.4) 式を OLS 推定し $\hat{\rho}$ を得る。これらのパラメータを用いて (6.3) 式を次のように変換する。

$$
\begin{aligned}
y_{it}^{*} &= \mathbf{x}_{it}^{*\prime}\beta + \mu_i^{*} + u_{it} \qquad (6.6)\\
y_{it}^{*} &= y_{it} - \hat{\rho} y_{it-1}\\
\mathbf{x}_{it}^{*} &= \mathbf{x}_{it} - \hat{\rho}\mathbf{x}_{it-1}\\
\mu_i^{*} &= \mu_i(1-\hat{\rho})
\end{aligned}
$$

上式を再び LSDV 推定すれば、β および μ_i の一致推定を得ることができるというものである。

Bhargava, Franizini 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。

$$
d = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{w}_{wit} - \hat{w}_{wit-1})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{w}_{wit}^2} \qquad (6.7)
$$

ここで $\hat{u}_{wit}$ 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、この統計量を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[*4]。

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 [1] パネルデータの特色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、[2] 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、[3] 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。[1] 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。[2] については Nickell (1981) によって厳密に証明された。[3] 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[*5]。

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 λ 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、ビ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。

$$p \lim_{within} \hat{\gamma}(0) < \gamma < p \lim_{GLS} \hat{\gamma}(\lambda) < p \lim_{poolingOLS} \hat{\gamma}(1) < p \lim_{between} \hat{\gamma}(\infty) \tag{6.8}$$

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、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)）。

[*4] 詳しくは Baltagi and Li (1995) か Baltagi (2001,pp.90-95) を参照されたい。

[*5] 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はなってきたが、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。

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 γ 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。

6.3 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おいて、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。

$$w_{it} = \gamma w_{it-1} + \rho' \mathbf{z}_i + \beta' \mathbf{x}_{it} + \mu_i + u_{it} \quad i = 1, 2, ..N. \ \ t = 1, 2, ...T \tag{6.9}$$

$$y_{it} = w_{it} + \eta_i \tag{6.10}$$

$$\mu_i = (1 - \gamma)\eta_i \quad E(\eta_i) = 0 \quad Var(\eta_i) = \sigma_\eta^2 = \sigma_\mu^2 / (1 - \gamma)^2 \tag{6.11}$$

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ので、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、$(\mu_i + \rho' \mathbf{z}_i)/(1-\gamma) + \beta' \sum_{j=0} \mathbf{x}_{it-j} \gamma^j$ に収束すると想定されている。また固定効果 μ_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 μ_i と y_{i0} が無相関であり、かつ y_{i1}、y_{i1}... に影響を与えるような固定効果 μ_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、μ_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、ε_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 μ_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 μ_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 $[\varphi \varepsilon_i / (1-\gamma)] = \lim_{t \to \infty} E[y_{it} - \rho' \mathbf{z}_i / (1-\gamma) - \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j} \gamma^j | \varepsilon_i]$ 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 μ_i は相関している。y_{i0} は任意の値から始まっても $\eta_i + \rho' \mathbf{z}_i / (1 - \gamma) + \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j} \gamma^j$ 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 θ_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 θ_w で、分散は任意の $\sigma^2_{w_0}$ となる。

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

(仮定 4d)w_{i0} はランダムで、平均は θ_{i0} で、分散は任意の σ^2_{w0} 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[*6]。

$$L(\gamma,\rho,\beta,\gamma,\eta,\sigma^2_u,\sigma^2_w,\sigma^2_\mu) = (2\pi)^{-NT/2}|v|^{-N/2} \times \exp\{-\frac{1}{2\sigma^2}\sum_i\sum_t(y_{it}-\gamma y_{it-1}-\rho'\mathbf{z}_i-\beta'\mathbf{x}_{it})^2\} \tag{6.12}$$

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 $\sigma^2_\mu = 0$ 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 (6.9)(6.10) 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 μ_i は消去されてしまう。すなわち、

$$y_{it}-y_{it-1} = (x_{it}-x_{it-1})'\beta+\gamma(y_{it-1}-y_{it-2})+(u_{it}-u_{it-1}) \tag{6.13}$$

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[*7]、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 β と γ は次のように推計される[*8]。

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

$$\begin{pmatrix}\gamma_{iv}\\ \beta_{iv}\end{pmatrix} = \left[\sum_{i=1}^N\sum_{t=3}^T\begin{pmatrix}(y_{i,t-1}-y_{i,t-2})(y_{it-2}-y_{it-3}) & (y_{it-2}-y_{it-3})(\mathbf{x}_{it}-\mathbf{x}_{it-1})') \\ (\mathbf{x}_{it}-\mathbf{x}_{it-1})(y_{it-2}-y_{it-3}) & (\mathbf{x}_{it}-\mathbf{x}_{it-1})(\mathbf{x}_{it}-\mathbf{x}_{it-1})'\end{pmatrix}\right]^{-1} \tag{6.14}$$

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

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。

[*6] 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。

[*7] 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。

[*8] 以下の式の展開は Hsiao (2003, pp.85-86) を参照。

$$\begin{pmatrix} \gamma_{iv} \\ \tilde{\beta}_{iv} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} y_{it-2}(y_{i,t-1} - y_{i,t-2}) & y_{it-2}(\mathbf{x}_{it} - \mathbf{x}_{it-1})') \\ (\mathbf{x}_{it} - \mathbf{x}_{it-1})y_{it-2} & (\mathbf{x}_{it} - \mathbf{x}_{it-1})(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \end{pmatrix} \right]^{-1} \times \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} y_{it-2} \\ \mathbf{x}_{it} - \mathbf{x}_{it-1} \end{pmatrix} (y_{it} - y_{it-1}) \right] \tag{6.15}$$

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[*9]。

第 2 段階として推計した $\hat{\beta}$ と $\hat{\gamma}$ を (6.9) 式に代入し，ρ を最小二乗法で推定する。

$$\bar{y}_{it} - \hat{\gamma}\bar{y}_{it-1} - \hat{\beta}'\bar{\mathbf{x}}_{it} = \rho'\mathbf{z}_i + \mu_i + \bar{u}_{it} \quad i = 1, ..N \tag{6.16}$$

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 σ_u^2、σ_μ^2 を推定する。

$$\sigma_u^2 = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} [(y_{it} - y_{it-1}) - \hat{\gamma}(y_{it-1} - y_{it-2}) - \hat{\beta}'(\mathbf{x}_{it} - \mathbf{x}_{it-1})]^2}{2N(T-1)} \tag{6.17}$$

$$\sigma_\mu^2 = \frac{\sum_{i=1}^{N} (\bar{y}_i - \hat{\gamma}\bar{y}_{i,-1} - \hat{\rho}'\mathbf{z}_i - \hat{\beta}'\bar{\mathbf{x}}_i)}{N} - \frac{1}{T}\hat{\sigma}_u^2 \tag{6.18}$$

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 か 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 γ、β、σ_u^2 は一致推定となる。ρ、σ_μ^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。最尤法でラグ項のパラメータ γ 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 ρ 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 ρ 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 γ は全ての場合で一推定になる。固定効果推定法では ρ は推定できないが、ラグ項のパラメータ γ 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

[*9] 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。

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[*10]。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, γ) を決定するということである。すなわち、

$$\min[\sum_{i=1}^{N} \triangle \mathbf{u}_i^{*\prime} \Omega^{-1} \triangle \mathbf{u}_i^*] \tag{6.19}$$

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

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。

6.4　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[*11]。

$$E[y_{is}, (u_{it} - u_{i,t-1})] = 0, \quad s = 0, 1, ...t-2, \quad t = 2, ...T \tag{6.20}$$

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。

$$\frac{1}{n} \sum_{i=1}^{n} y_{is}[(y_{it} - y_{i,t-1}) - (y_{i,t-1} - y_{i,t-2})'\gamma - (x_{it} - x_{i,t-1})'\beta] = 0 \tag{6.21}$$

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

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 $(y_{i1}, y_{i2}, y_{i3}, ...y_{iT-2})$ と増えていく。これを行列で表現すると次のようになる[*12]。

$$W_i = \begin{bmatrix} [y_{i1}] & 0 & ... & 0 \\ 0 & [y_{i1}, y_{i2}] & ... & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & ... & [y_{i1}, ...y_{iT-2}] \end{bmatrix} \tag{6.22}$$

[*10] MDE の詳細については Chamberlain (1982,1984)、Lee(2002, 第3章) を参照。
[*11] 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。
[*12] 以下は Baltagi (2001, p.132) を参照。

操作変数が誤差ラグ項と無相関である条件 (6.20) は次のようにベクトル形式で書き直せる。

$$E(W_i' \vartriangle u_i) = 0 \tag{6.23}$$

ここで (6.1) 式のようなダイナミック・モデルを考えてみよう。

$$\begin{aligned}(y_{it} - y_{it-1}) &= (y_{it-1} - y_{it-2})'\gamma + (x_{it} - x_{it-1})'\beta + (u_{it} - u_{it-1}) \\ \vartriangle y_{it} &= \vartriangle y_{it-1}'\gamma + \vartriangle x_{it}'\beta + \vartriangle u_{it} \qquad i = 1, 2...N\end{aligned} \tag{6.24}$$

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。(6.24) 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。

$$W' \vartriangle y_{it} = W' \vartriangle y_{it-1}'\gamma + W' \vartriangle x_{it}'\beta + W' \vartriangle u_{it} \tag{6.25}$$

γ、β について解くと次のように表せる。

$$\hat{\gamma}_{GMM} = [(\vartriangle y_{it-1})'W\hat{V}_N^{-1}W'(\vartriangle y_{it-1})]^{-1}[(\vartriangle y_{it-1})'W\hat{V}_N^{-1}W'(\vartriangle y_{it})] \tag{6.26}$$

$$\hat{\beta}_{GMM} = [(\vartriangle x_{it-1})'W\hat{V}_N^{-1}W'(\vartriangle x_{it-1})]^{-1}[(\vartriangle x_{it-1})'W\hat{V}_N^{-1}W'(\vartriangle y_{it})] \tag{6.27}$$

ここで $V_N = \sum_{i=1}^{N} W_i'(\vartriangle u_i)(\vartriangle u_i)'W_i$ である[*13]。

説明変数 x_{it} が厳密に外生変数であれば、$E(x_{it}u_{is}) = 0, \forall t, s = 1, 2, ..., T$ であり、x_{it} が μ_i と相関している場合には、(6.24) 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 $(x_{i1}, x_{i2}, ..., x_{is-1})$ 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。

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[*14]。

$$r_j = \frac{1}{T-3-j} \sum_{t=4+j}^{T} r_{tj} \tag{6.28}$$

[*13] 実際、Arellano and Bond (1991, p.279) では (6.25) 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。

[*14] Arellano (2003, pp.121-23) を参照。

ここで $r_{tj} = E(\triangle\ u_{it}\ \triangle\ u_{it-j})$ 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。

$$m_j = \frac{\hat{r}_j}{SE(\hat{r}_j)} \tag{6.29}$$

ここで $\hat{r}_j$ は標本から得られた $\triangle\ \hat{u}_{it}$ と $\hat{r}_{tj} = N^{-1}\sum_{i=1}^{N} \triangle\ \hat{u}_{it}\ \triangle\ \hat{u}_{it-j}$ 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。

$$s = \triangle\ \hat{u}'W[\sum_{i=1}^{N} W_i'(\triangle\ \hat{u}_i)(\triangle\ \hat{u}_i)'W_i]^{-1}W'(\triangle\ \hat{u}) \sim \chi^2_{p-k-1} \tag{6.30}$$

ここで p は行列 W における行数、$\triangle\ \hat{u}$ は (6.25) 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。

$$E(y_{is}\ \triangle\ u_{it}) = 0 \quad t = 2, ...T,\ s = 0, 1, ...t-2 \tag{6.31}$$

さらに、次のような直交条件も利用できることを示した。

$$E(u_{iT}\ \triangle\ u_{it}) = 0 \quad t = 2, ...T-1 \tag{6.32}$$

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。(6.32) 式は γ が 1 に近いか、σ_μ^2/σ_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、(6.31)(6.32) 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。

(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 Estomator) 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 γ が 1 に近づくか、固定効果 μ_i の分散が大きい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 $E(y_{i1} \bigtriangleup u_{i3}) = 0$ に限定した。これによってパラメータ γ 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。

$$\bigtriangleup y_{i2} = \pi y_{i1} + \mu_i + u_{i2} \quad i = 1, 2, ...N \tag{6.33}$$

ここで γ が 1 に近づくか、固定効果 μ_i の分散が大きければ、パラメータ π は 0 に近づいても不思議はない。この場合、y_{i1} は $\bigtriangleup y_{i2}$ 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 π の確率極限は次のように表せる。

$$p\lim \hat{\pi} = (\gamma - 1)\frac{k}{(\sigma_\mu^2/\sigma_u^2) + k} \quad k = \frac{(1-\gamma)}{(1+\gamma)} \tag{6.34}$$

Blundell and Bond (1998) は GMM 推定のバイアスはかなりの程度、Nelson amd Startz (1990a,b) 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 Schumidt (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。

$$E(u_{it} \bigtriangleup y_{it-1}) = 0 \quad t = 4, 5, ...T \tag{6.35}$$

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

$$E(u_{i3} \bigtriangleup y_{i2}) = 0 \tag{6.36}$$

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[*15]。

[*15] 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。

$$y_{i1} = \frac{\mu_i}{1-\gamma} + u_{i1} \tag{6.37}$$

$t = 2$ 以後の関係式は与えられているので、y_{it} の全ての流列が確定する。ところで、(6.36) 式は次のように書き換えることができる。

$$E[(\mu_i + u_{i3})(u_{i2} + (\gamma - 1)u_{i1})] = 0 \tag{6.38}$$

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。

$$E(u_{i1}\mu_i) = E(u_{i1}u_{i3}) = 0 \quad i = 1, 2, ...N \tag{6.39}$$

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 γ が 1 に近いか、σ_μ^2/σ_u^2 が大きい場合でも、(6.35)(6.36) 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使ったシステム GMM を提案した。

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i & 0 & 0 & ... & 0 \\ 0 & \triangle\, y_{i2} & 0 & ... & 0 \\ 0 & 0 & \triangle\, y_{i3} & ... & 0 \\ . & . & . & ... & 0 \\ 0 & 0 & 0 & ... & \triangle\, y_{iT-1} \end{bmatrix} \tag{6.40}$$

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。

$$\mathbf{Z}_i = \begin{bmatrix} y_{i1} & 0 & 0 & ... & 0 & ... & 0 \\ 0 & y_{i1} & y_{i2} & ... & 0 & ... & 0 \\ . & . & . & ... & . & ... & . \\ 0 & 0 & 0 & ... & y_{i1} & ... & y_{iT-2} \end{bmatrix} \tag{6.41}$$

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 γ 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、γ が 1 に近いか、σ_μ^2/σ_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。

6.5 STATA コード

ここでは『21 世紀出生児縦断調査』を用いてダイナミックパネル推定を行ってみよう。これまで北村 (2007、2008, 2009) で行ってきた出生児の身体成長パターンの測定を行ってきた。これまでの研究ではダイナミックパネル推定ではなく、身長・体重の対数を出生からの経過日数とその二乗、および子育費用によって説明する固定効果モデルを採用してきた。

ここではこれまでのモデルに一次のラグ項を入れている。例えば体重については次のようなモデルを考えている。

$$\begin{aligned}\ln bdywht_{it} &= \alpha + \gamma \ln bdwht_{it-1} + \beta \ln survivalday_{it} + \delta \ln survaivalday_{it}^2 \\ &\quad + \zeta \ln \cos tofchildcare_{it} + u_{it}\end{aligned}$$

ここで $\ln bdywht =$ 体重の対数値、$\ln survivalday =$ 生存日数の対数値、$\ln \cos tofchildcare =$ 子育費用の対数値。身長の場合は $\ln bdywht$ に代えて $\ln bdyhgt$ を用いる。

推計に使った STATA コードは次のようになる。

```
/*dynamic panel analysis*/
tsset idno rirekino
/**weight**/
/*logarithm*/
reg lnbdywht lnbdywht_1 survivalday survival2 lncostofchildcare  /*OLS*/
xtreg lnbdywht lnbdywht_1 survivalday survival2 lncostofchildcare, mle /*Maximum
Likelihood*/

xtabond lnbdywht survivalday survival2 lncostofchildcare, lags(1) pre( lnbdyhgt_1)
artests(2) vce(gmm) /*Arellano and Bond GMM*/
estat sargan

xtabond lnbdywht survivalday survival2 lncostofchildcare, lags(1) pre( lnbdyhgt_1)
artests(2) vce(robust) /*Arellano and Bond GMM*/
estat abond
/**hight**/
/*logarithm*/
reg lnbdyhgt lnbdyhgt_1 survivalday survival2 lncostofchildcare /*OLS*/
xtreg lnbdyhgt lnbdyhgt_1 survivalday survival2 lncostofchildcare, mle /*Maximum
Likelihood*/
xtabond lnbdyhgt survivalday survival2 lncostofchildcare, lags(1) pre(lnbdywht_1)
artests(2) vce(gmm) /*Arellano and Bond GMM*/
```

```
estat sargan
xtabond lnbdyhgt survivalday survival2 lncostofchildcare, lags(1) pre(lnbdywht_1)
artests(2) vce(robust) /*Arellano and Bond GMM*/
estat abond
```

Dependent Variable: weight	OLS		MLE		GMM	
	Coef.	t-statistics	Coef.	z-statistics	Coef.	z-statistics
llnbdywht L1.					-0.033	-7.02
lnbdywht_1	0.246	231.58	0.088	133.71	0.261	20.06
survivalday	0.000	-56.75	0.000	114.60	0.000	100.14
survival2	1.24E-07	122.16	7.21E-09	11.78	-6.44E-09	-9.52
lncostofchildcare	0.013	35.14	0.003	12.73	0.000	-0.66
_cons	2.122	1,582.70	2.093	2,377.90	1.200	27.04
/sigma_u			0.098			
/sigma_e			0.057			
rho			0.750			
Diagnostic Test						
Number of observation	179,479		179,479		116,725	
Number of groups	43,169		43,169		36,854	
R-squared	0.841					
Adj R-squared	0.841					
Root MSE	0.107					
LR chi2(3)			450,209.24			
Sargan test					chi2(28) = 1584.489 Prob > chi2 = 0.0000	
Wald test					Chi2(4) = 314446.08 Prob > chi2 = 0.0000	
Arellano-Bond test for residual AR(1)					z = -24.467 Prob > z = 0.0000	
Arellano-Bond test for residual AR(2)					z = 8.675 Prob > z = 0.0000	
Instruments for differenced equation						
GMM-type:					L(2/.).lnbdywht L(1/.).lnbdyhght_1	
Standard:					D.survivalday D. survival2	

図表 6.1 ダイナミックパネル推定（体重）

図表 6.1 はプーリング OLS 推定、最尤法（M L E）推定, Arellano-Bond GMM 推定の結果を報告してある。GMM 推定については Sargan の過剰識別制約検定を行ったが、これが有効であるという帰無仮説は棄却された。すなわち、このモデルでは操作変数が直行条件を満たしていないか、外生変数が本当は内生変数であるか、モデルの重要な説明変数が落ちているかなどの問題があることがわかる。また Arellano-Bond の系列相関検定でも、系列相関が AR(1) でも AR(2) でも残っていることから、操作変数の直行条件が満たされていない可能性が示唆されている。

ここでは体重のラグ項のパラメータ γ 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。

$$MLE(0.088) < OLS(0.246) < GMM(0.261)$$

GMM 推定は先に述べたように問題が多いが、推定値は OLS とほぼ等しい。

それに対して図表 6.2 では身長を推定した。GMM 推定の統計的問題点は全く同じであるが、γ の値は次のような順序になる。

Dependent Variable: hight	OLS		MLE		GMM	
	Coef.	t-statistics	Coef.	z-statistics	Coef.	z-statistics
llnbdywht L1.					0.085	13.35
lnbdywht_1	0.213	196.5	0.079	103.98	−0.008	−3.50
survivalday	5.09E−05	32.81	0.000	206.8	0.000	155.71
survival2	2.19E−08	55.72	−1.75E−08	−65.38	−2.13E−08	−64.56
lncostofchildcare	0.003	22.37	0.000	−4.49	−0.001	−5.62
_cons	3.510	959.77	3.960	1553.86	3.937	180.96
/sigma_u			0.033			
/sigma_e			0.023			
rho			0.674			
Diagnostic Test						
Number of observation	170,124		170124		113742	
Number of groups			42528		35748	
R-squared	0.925					
Adj R-squared	0.925					
Root MSE	0.038					
LR chi2(3)			524,682.810			
Sargan test					chi2(28) = 993.153 Prob > chi2 = 0.0000	
Wald test					Chi2(4) = 442673.89 Prob > chi2 = 0.0000	
Arellano-Bond test for residual AR(1)					z = −37.335 Prob > z = 0.0000	
Arellano-Bond test for residual AR(2)					z = 4.3317 Prob > z = 0.0000	
Instruments for differenced equation						
GMM-type:					L(2/.).lnbdywht L(1/.).lnbdyhght_1	
Standard:					D.survivalday D. survival2	

図表 6.2　ダイナミックパネル推定（身長）

$$GMM(-0.008) < MLE(0.079) < OLS(0.213)$$

この場合は GMM 推定の値は最低になり、OLS 推定が最大値を取っている。この二つの推定では MLE 推定が 0.08 程度、OLS 推定が 0.2 程度と安定している。どちらかを選ぶとすれば、経験的に考えて、OLS 推定がここでは選ばれていいのではないだろうか。

しかし、我々の判断としてはここで用いた $\ln\cos tofchildcare$ 変数が内生変数である可能性が高いと思う、そこで次章ではこの問題について取り組みたい。

参考文献

阿部修人・小黒曜子（2004）「社長交代と外部出身取締役－－ Semi Parametric 推計による分析」、『経済研究』、55(1)、pp.72-84.

知久馬　（1995）『生存時間解析　SAS による生物統計』、東大出版会

北村行伸 (2005) 『パネルデータ分析』、岩波書店

北村行伸 (2009) 『ミクロ計量経済学入門』、日本評論社

北村行伸（2007）「21 世紀成年者縦断調査に基づく子供の成長パターンの測定」、厚生労働科学研究費補助金・統計情報高度利用総合研究事業平成 18 年度報告書『パネル調査（縦断調査）に関する総合的分析システムの開発研究』、2007 年 3 月、pp.101-123.

北村行伸 (2008) 「21 世紀出生児縦断調査に基づく子供の成長パターンの測定（II）」、厚生労働科学研究費補助金・統計情報総合研究事業　平成 19 年度　総括研究報告書『パネル調査（縦断調査）に関する総合的分析システムの開発研究』、2008 年 3 月、pp.71-89.

北村行伸 (2009) 「21 世紀出生児縦断調査に基づく子供の成長パターンの測定（III）」、厚生労働科学研究費補助金・統計情報総合研究事業　平成 20 年度　総括研究報告書『パネル調査（縦断調査）に関する統合的高度統計分析システムの開発研究』、2009 年 3 月、pp.127-151.

中村　剛（2001）『Cox 比例ハザードモデル』、医学統計学シリーズ、朝倉書店

山口一男（2001-2)「イベントヒストリー分析 (1)-(15)」、『統計』2001 年 9 月号－ 2002 年 11 月号

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

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

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

Arellano, M. (2003) *Panel Data Econometrics*, Oxford University Press.

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.

Balestra, P. and Nerlove, M. (1966) "Pooling Cross-Section and time-Series Data in the Estimation of a Dynamic Model: the Demand for Natural Gas," *Econometrica*, 34, pp.585-612.

Baltagi, B.H. (2001) *Econometric Analysis of Panel Data*, 2nd ed, John Wiley & Sons.

Baltagi, B.H. and Li, Q. (1991a) "A Transformation that Will Circumvent the Problem of Autocorrelation in an Error Component Model," *Journal of Econometrics*, 48, pp.385-393.

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

Chamberlain, G. (1982) "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, pp.5-46.

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

Fisher, R.A.(1973a) *Statistical Methods for Research Workers*, 14th ed, Hafner Publishing.

Fujiki, H., Hsiao, C. and Shen, Y. (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.

Holtz-Eakin, D.(1988) "Testing for Individual Effects in Autoregressive Models", *Journal of Econometrics*, 39, pp.297-307.

Holtz-Eakin, D., Newey, W. and Rosen, H.S. (1988) "Estimating Vector Autoregressions with Panel Data, " *Econometrica*, 56, pp.1371-1395.

Hsiao, C. (2003) *Analysis of Panel Data 2nd ed.*, Cambridge University Press.

Hsiao, C. and Pesaran, M.H. (2004) "Random Coe cient Panel Data Models," mimeo.

Lillard, L.A. and Weiss, Y. (1979) "Components of Variation in Panel Earnings Data: American Scientists 1960-70," *Econometrica*, 47(2), pp.437-454.

Lillard, L.A. and Willis, R.J.(1978) "Dynamic Aspects of Earning Mobility," *Econometrica*, 46(5), pp.985-1012.

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

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

Maddala, G.S. (2001) *Introduction to Econometrics*, 3rd ed, John Weily & Sons.

Maddala, G.S., Li, H., Trost, R.P. and Joutz, F. (1997) "Estimation of Short-Run and Long-Run Elasticities of Energy Demand from Panel Data Using Shrinkage Estimators," *Journal of Business and Econmic Statistics*, 15, pp.90-100.

Nelson, C.R and Startz, R. (1990a) "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One," *Journal of Business*, 63(1), pp.X125-S140.

Nelson, C.R and Startz, R. (1990b) "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica* 58 (4), pp.967-976.

Nerlove, M. (2000) "Growth Rate Convergence, Fact or Artifact? An Essay onPanel Data

Econometrics, in Krishnakumar, J. and Ronchetti, E. (eds) *Panel Data Econometrics: Future Directions*, North Holland, Chapter 1, pp.3-33.

Nerlove, M. (2002) *Essays in Panel Data Econometrics*, The Cambridge University Press.

Nickell, S. (1981) "Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, pp.1417-1426.

Ridder, G. and Wansbeek, T. (1990) "Dynamic Models for Panel Data," in Von der Plog, R (ed.) *Advanced Lectures in Quantitative Economics*, Academic Press, pp.557-582.

Sevestre, P. and Trognon, A. (1996) "Dynamic Linear Models," in Mátyás, L. and Sevestre, P. (eds.), *The Econometrics of Panel Data*, Chapter 7, pp.120-144, Kluwer Academic Publishers.

Smith, R.P. and Fuertes, A-M. (2004) "Panel Time-Series," mimeo.

Staiger, D. and Stock, J. (1997) "Instrumental Variables Estimation with Weak Instruments," *Econometrica*, 65, pp.557-586.

Trognon, A. (1978) "Miscellaneous Asymptotic Properties of Ordinary Least Square and Maximum Likelihood Methods in Dynamic Error Components Models," *Annales de L'INSEE*, 30-31, pp.631-357.

Vella, F. and Verbeek, M. (1998) "Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men," *Journal of Applied Econometrics*, 13, pp.163-183.

Wooldridge, J.M. (2003) *Introductory Econometrics: A Modern Approach*, 2nd edition, South-Western, Thomson Learning.

第 7 章

同時方程式パネルデータ分析

7.1　はじめに

多くの経済変数は相互依存関係にあり、完全に外生的に決まっている変数はむしろ珍しいといってもいいかもしれない。もちろん、これは人間の経済活動自体が相互依存関係の中で決まっているという事実を勘案すれば当然のことであろう[*1]。

しかし、現実問題として、我々は実証研究で、同じ変数をある時は被説明（内生）変数に使い、ある時は説明（外生）変数に使ってきた。例えば、企業収益は企業活動の目的であると考えれば、これは被説明（内生）変数として扱うべきであるが、投資関数の推計においては企業収益を説明（外生）変数に使うことがある。同様に、賃金も被説明（内生）変数として使ったり、労働供給関数では説明（外生）変数として使うことがある。

変数が内生的であるという議論をする場合には、暗黙の内に同時方程式あるいは経済システムを考えていることが多い。経済活動の相互依存関係を捉えるという意味で一般均衡モデルを用いることが重要であるということは、多くの経済学者に認められてきたことであるし、経済政策分析においては益々その重要性が増してきているとも言える。実証研究においても、内生変数の取り扱いに注意して、それがもたらすバイアスを適切に取り除く必要があるということは広く認識されてきている。本章でもその手法について解説することを主眼としている。

同時に 1970 年代後半以後、大型計量経済モデルに対する批判が相次ぎ、同時方程式の研究自体も少なくなってきた。同時方程式パネルデータ分析も Baltagi (1981b)、Hausman and Taylor (1981) などの 1980 年代の研究から大きな進展がなく、Hsiao(2003) においても、旧版である Hisao(1986) から唯一変更がなかったのが同時方程式の第 5 章である。

本章でも同時方程式の相互依存関係を本格的に取り込んだ完全情報 3 段階最小二乗法（3SLS）や完全情報最小距離推定法などについては、簡単な解説にとどめ、より実用的な 2 段階最小二乗法 (2SLS) による単一方程式に議論を絞っている。また、同時方程式といっても全ての変数が相互依存関係にあるのではなく、一部の変数が影響を与えあっており、他の変数は外生変数と扱っても問題がない場合が往々にしてある。この特殊ケースとして三角配列システムがあるが、これについて

[*1] 本章は北村（2005）『パネルデータ分析』（岩波書店）の第 6 章に基づいている。詳細については北村（2005）第 6 章を参照されたい。実証分析は新たに『21 世紀出生児縦断調査』を用いて行っている。

はここでは論じない[*2]。

7.2 同時方程式パネルデータ分析の考え方

同時方程式パネルデータとは、次のような構造をもっている。

$$
\text{主体 } i \begin{cases} \boxed{1} & \boxed{Y_{1i} = \beta_{1i} X_{1i} + v_{1it}} \\ 2 & \vdots \\ \vdots & \vdots \\ g & \vdots \\ \vdots & \vdots \\ G & Y_{Gi} = \beta_{Gi} X_{Gi} + v_{Git} \end{cases}
$$

$$
\text{主体 } j \begin{cases} \boxed{1} & \boxed{Y_{1j} = \beta_{1j} X_{1j} + v_{1jt}} \\ 2 & \vdots \\ \vdots & \vdots \\ g & \vdots \\ \vdots & \vdots \\ G & Y_{Gj} = \beta_{Gj} X_{Gj} + v_{Gjt} \end{cases}
$$

例えば $\hat{\beta}_1$ を推計したい場合、

1. 各主体はクロスセクションで Y_{1i} を選択すると同時に Y_{gi} $(g = 1, \ldots, G)$ を選択している。
2. これを t 時点毎に行っている。
3. 一般の同時方程式問題では、各 i について各時点共通の β を推計している。
4. この β_g が異なる主体 i, j で共通しているとすると、誤差項の構造に違いに配慮した推計をする必要が出てくる (固定効果 A_i を用いる)。
5. 一般のパネル推計では時系列方向の平均 $\bar{Y} = \beta \bar{X}$ を用いたが、同時方程式パネル推計では、個別主体の同時方程式制約下での β の推定量と主体間のパネル推定量の差を最小にするような推定が有効かつ一致するように求めることが問題となる。

一般に、同時方程式体系を考えたとしても、パネルデータ分析の枠組みでは、その中から、関心の高い一本の方程式を選んで、それについて内生性の問題に配慮した推計方法を考えることが多

[*2] 三角配列システムを推計する方法は Chamberlain(1977) によって操作変数法が、Chamberlain and Griliches (1975) によって最尤法が提案されている。最尤法は有効推定ではあるが計算が難しい。操作変数法は有効ではないが、計算が簡単であり、一致推定は得られる。詳細については Hsiao(2003, chapter 5), pp.129-136 を参照されたい。

い[*3]。

$$\underset{NT\times 1}{\mathbf{y}_g} = \mathbf{Y}_g\gamma_g + \mathbf{X}_g\mathbf{B}_g + \mathbf{v}_g \qquad (7.1)$$
$$= \mathbf{w}_g\theta_g + \mathbf{v}_g \quad g = 1,\ldots,G$$

しかし、この方法は基本的には制限情報推定法 (limited-information estimation method) に従っており、同時方程式体系全体から得られる情報を完全に反映した推計 (完全情報推定法：full-information estimation method) を行えば、より有効な推定量を得られる。

ここで

$$\mathbf{y} = (\mathbf{y}\prime, \ldots, \mathbf{y}'_G)', \quad \mathbf{v} = (\mathbf{v}'_1, \ldots, \mathbf{v}'_G)'$$
$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{o} & \cdots & \mathbf{o} \\ \mathbf{o} & \mathbf{w}_2 & & \mathbf{o} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{o} & \mathbf{o} & \cdots & \mathbf{w}_G \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_G \end{bmatrix}$$

G 本の構造方程式体系は次のように表わすことができる。

$$\mathbf{y} = \mathbf{w}\theta + \mathbf{v} \qquad (7.2)$$

(7.2) 式を 3 段階最小 2 乗法（3SLS）で推計して有効なのは誤差項（$v_{1_{it}}, v_{2_{it}}, \ldots v_{G_{it}}$）が i と t に関して iid の場合のみである。不均一分散や系列相関がある場合には、完全情報最小距離推定 (the full information minimum-distance estimator) か一般化 3 段階最小 2 乗法（G3SLS) を用いる必要がある。

完全情報最小距離推定は次のように定式化される。T は固定、N は無限であるとすると、T 期の推計式に対して個人 i の G 本の構造方程式を考える。

$$\begin{aligned} \mathbf{y}_{1_i} &= \mathbf{w}_{1_i}\theta_1 + \mathbf{v}_{1i} \\ \mathbf{y}_{2_i} &= \mathbf{w}_{2_i}\theta_2 + \mathbf{v}_{2i} \\ &\vdots \\ \mathbf{y}_{G_i} &= \mathbf{w}_{G_i}\theta_G + \mathbf{v}_{G_i} \quad i = 1,\ldots N \end{aligned} \qquad (7.3)$$

最小距離推定値 θ は次の目的 (距離) 関数を最小化するように $\hat{\theta}$ を選ぶことによって得られる。

$$\left[\hat{\tilde{\pi}} - \tilde{\mathbf{f}}(\theta)\right] \hat{\tilde{\Omega}}^{-1} \left[\hat{\tilde{\pi}} - \tilde{\mathbf{f}}(\theta)\right]$$

ここで $\hat{\tilde{\pi}}$ は $\mathbf{y}_i$ の無制約の最小 2 乗法推定量、$\hat{\tilde{\mathbf{\Omega}}}$ は $\hat{\tilde{\pi}}$ と個人 i における $\tilde{\pi}$ の差の分散共分散行列の一致推定量 ($\tilde{\mathbf{\Omega}} = \sqrt{N}(\hat{\tilde{\pi}} - \tilde{\pi})$)。もし $\mathbf{v}_i$ が iid であれば $\sqrt{N}(\hat{\theta} - \theta)$ は漸近的に正規分布となる。

G3SLS は次のように求められる。

$$\hat{\theta}_{G3SLS} = (S_{wx}\hat{\Psi}^{-1}S'_{wx})^{-1}(S_{wx}\hat{\Psi}^{-1}S_{xy}) \qquad (7.4)$$

[*3] 以下の議論は Hsiao (2003,pp.124-6) を参照している。