パネル・データの分析

田中 勝人（一橋大学）

1. パネル・データ

N 個の主体（人，企業，団体，県，国など）の各々に関して T 期間にわたって観測されたデータ y_{it} $(i=1,\cdots,N;\, t=1,\cdots,T)$ をパネル・データという．

パネル・データ分析の利点：

(a) データ数が増えることにより，自由度が大きくなるので，推定精度が向上する．また，変数間の変動も大きくなるので，回帰モデルにおける多重共線の問題が回避できる．

(b) 主体間の異質性をモデルに取り込むことは，単一の時系列，あるいはクロスセクションのみでは不可能であるが，それが可能になる．例えば，A 銀行の住宅ローン（A 企業の研究開発投資）に関する時系列データだけでなく，他の銀行のローン（他の企業の研究開発投資）に関する時系列データも加えることにより，それぞれの銀行（企業）の特性をモデル化することができる．

(c) 主体間の異質性は，一般に観測不可能な主体固有の要因であり，そのような要因以外の全体の関係を分析することが主目的ならば，固有の要因を除去した分析が可能である．

2. パネル・データ・モデル

重回帰モデルの拡張として，次のモデルを考える．

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it}, \quad (i=1,\cdots,N;\ t=1,\cdots,T) \tag{1}$$

ここで，誤差項 $\{u_{it}\}$ は古典的な仮定をみたすとする．すなわち，

$$\mathrm{E}(u_{it}) = 0, \quad \mathrm{E}(u_{it}u_{js}) = \begin{cases} \sigma_u^2 & (i=j \text{ かつ } t=s \text{ のとき}) \\ 0 & (\text{その他のとき}) \end{cases}$$

をみたす．また，$\boldsymbol{\beta}$ は通常の回帰モデルの場合と同様に，未知の係数ベクトルである．

モデル (1) には，上記の他に，$\{\alpha_i\}$ と $\{\boldsymbol{x}_{it}\}$ が含まれる．$\{\boldsymbol{x}_{it}\}$ は説明変数の $p \times 1$ 確率ベクトルであり，誤差項 $\{u_{it}\}$ との独立性（**狭義外生性：strict exogeneity**）が仮定される．他方，α_i は，主体 i に特有の**個別効果**（individual effect）と呼ばれる量であり，次の 2 通りが仮定される．

(a) 通常の回帰モデルにおけるように，$\{\alpha_i\}$ は，定数のパラメータであると仮定する．この場合のモデルを**固定効果モデル**（Fixed effect model）という．

(b) $\{\alpha_i\}$ は，主体ごとに独立な確率変数であると仮定して，

$$\mathrm{E}(\alpha_i) = 0, \qquad \mathrm{E}(\alpha_i^2) = \sigma_\alpha^2, \qquad \mathrm{E}(\alpha_i u_{it}) = 0$$

をみたすものとする．この場合のモデルを**変量効果モデル**（Random effect model）と呼ぶ．

上記 2 つのモデルを総称して，パネル・データ・モデルと呼ぶ．説明変数のベクトル $\boldsymbol{x}_{it}$ と個別効果 α_i を確率的と考える理由は，これらの相関を考慮するためである．例えば，企業の生産活動の分析において，$\boldsymbol{x}_{it}$ は，資本や労働などの観測可能な投入量を表し，α_i は，当該企業の経営資源を表すと考えられる．経営資源は観測不可能な個別的属性であり，経営資源が豊富ならば，生産に必要な投入量は少なくてすむ．別の例として，家計の消費行動の分析においては，説明変数としては，当該家計の収入や世帯主の年齢等が考えられるが，個別効果としては当該家計固有の状況を想定することができる．

3. パネル・データ・モデルの推定

パネル・データ・モデル (1) の推定方法は，固定効果モデルの場合と変量効果モデルの場合とで異なるが，問題は，個別効果 $\{\alpha_i\}$ をどのように取り扱うかということである．それぞれのモデルを推定する際の問題点として，次の点を挙げることができる．

(a) 固定効果モデルにおいては，N が大きくなれば，パラメータ $\{\alpha_i\}$ も増えていく．また，T が小さいと，α_i を一致推定することは不可能である．

(b) 変量効果モデルにおいては，$\{\alpha_i\}$ も誤差項の中に組み入れることになり，$\{\alpha_i\}$ に関連するパラメータは，分散の σ_α^2 のみとなるので取り扱いが容易になる．ただし，説明変数 $\{x_{it}\}$ と $\{\alpha_i\}$ に相関があれば，β の最小 2 乗推定は一致性を失うことになる．

4. 固定効果モデルの推定

代表的な推定方法として，(F-1) LSDV 推定（最小 2 乗ダミー変数推定: Least Squares Dummy Variables），(F-2) 階差推定，(F-3) グループ内推定（within estimation），(F-4) グループ間推定（between estimation）がある．

(F-1) LSDV 推定

主体ごとのダミー変数として，$D_{1i}, \cdots, D_{Ni}$ を導入する．ここで，

$$D_{ji} = \begin{cases} 1 & (j = i \text{ のとき}) \\ 0 & (\text{その他のとき}) \end{cases}$$

である．このとき，LSDV 推定は，モデル

$$y_{it} = \alpha_1 D_{1i} + \cdots + \alpha_N D_{Ni} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}, \quad (i = 1, \cdots, N;\ t = 1, \cdots, T) \tag{2}$$

を考えて，最小 2 乗推定する方法である．ただし，このモデルにおいては，$\boldsymbol{x}_{it}$ の中に定数項を含めてはならない．

モデル (2) は，パラメータ $\alpha_1, \cdots, \alpha_N, \boldsymbol{\beta}$（全部で $N + p$ 個．この他に $\sigma_u^2 = \mathrm{E}(u_{it}^2)$）を含む．これらの OLS 推定量を $\hat{\alpha}_1, \cdots, \hat{\alpha}_N, \hat{\boldsymbol{\beta}}, \hat{\sigma_u}^2$ とすると，これらは次の正規方程式をみたす．

$$\begin{aligned} \bar{y}_i &= \hat{\alpha}_i + \bar{\boldsymbol{x}}'_i\hat{\boldsymbol{\beta}}, \quad (i = 1, \cdots, N) \\ \sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it} y_{it} &= \hat{\alpha_1} T \bar{x}_1 + \cdots + \hat{\alpha}_N T \bar{x}_N + \sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}'_{it} \end{aligned}$$

ここで，

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$$

これより，

$$\begin{aligned}\hat{\alpha}_i &= \bar{y}_i - \bar{\boldsymbol{x}}_i'\hat{\boldsymbol{\beta}}, \quad (i = 1, \cdots, N)\\ \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\hat{\boldsymbol{\beta}} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(y_{it} - \bar{y}_i)'\end{aligned}$$

が得られる．

なお，モデル (2) は，時間効果のダミー DT_{Tt} $(t = 1, \cdots, T)$ を入れて，

$$y_{it} = \alpha_1 D_{1i} + \cdots + \alpha_N D_{Ni} + \gamma_1 DT_{1t} + \cdots + \gamma_T DT_{Tt} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it} \tag{3}$$

のように拡張することもできる．推定は，個別効果のみが入っている場合と同様である．

個別効果の有無に関する検定は，

$$H_0:\ \alpha_1 = \cdots = \alpha_N \qquad \text{vs} \qquad H_1: \neq H_0$$

を F 検定すればよい．実際，

$$F = \frac{(RSS(H_0) - RSS(H_1))/(N-1)}{RSS(H_1)/(NT - N - p)}$$

は，H_0 のもとで自由度 $(N-1, NT-N-p)$ の F 分布に従う．ただし，$RSS(H_0)$ と $RSS(H_1)$ は，それぞれ，H_0 と H_1 のもとでの回帰の残差 2 乗和である．

(F-2) 階差推定

個別効果 $\{\alpha_i\}$ を消去するするために，階差変換

$$\Delta y_{it} = y_{it} - y_{i,t-1}, \quad \Delta\boldsymbol{x}_{it} = \boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1}, \quad \Delta u_{it} = u_{it} - u_{i,t-1}$$

を定義して，モデル (1) に適用すると，個別効果が除去され，

$$\Delta y_{it} = \Delta\boldsymbol{x}_{it}'\boldsymbol{\beta} + \Delta u_{it} \tag{4}$$

を得る．この階差モデルに対して OLS を適用するのが階差推定である．なお，説明変数 $\boldsymbol{x}_{it}$ には，定数項を含めてはならない．

(F-3) グループ内推定

最初のモデル (1) を考える．

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it}$$

このモデルにおいて，主体ごとの時間平均を取ると，

$$\bar{y}_i = \alpha_i + \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \bar{u}_i \tag{5}$$

が得られる．したがって，2 つの表現の差を取ると，α_i が除去されて，

$$y_{it} - \bar{y}_i = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + u_{it} - \bar{u}_i, \quad (i = 1, \cdots, N;\ t = 1, \cdots, T)$$

を得る．なお，説明変数 $\boldsymbol{x}_{it}$ には定数項を含めてはならない．この最後の式から $\boldsymbol{\beta}$ を OLS 推定する方法がグループ内推定である．名称の由来は，主体ごとの平均（=グループ内平均）からの偏差を使った推定であることによる．

このようにして得られる $\boldsymbol{\beta}$ のグループ内推定量は，個別効果をダミー変数で表したモデルを使った LSDV 推定量と一致することがわかる．なお，LSDV 推定は，時間効果も含むようなモデルにも適用されるので，グループ内推定よりも一般的である．

(F-4) グループ間推定

最初のモデル (1) に対して，主体ごとの時間平均を取ると，(5) のように，

$$\bar{y}_i = \alpha_i + \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \bar{u}_i, \quad (i = 1, \cdots, N) \tag{6}$$

が得られる．この式において，α_i を $\alpha_i = \alpha$（定数）とみなして，OLS 推定する方法がグループ間推定である．個別効果の違いが無視できない場合には，適用できない方法である．

5. 変量効果モデルの推定

変量効果モデルは，個別効果 α_i を

$$\mathrm{E}(\alpha_i) = 0, \qquad \mathrm{E}(\alpha_i^2) = \sigma_\alpha^2, \qquad \mathrm{E}(\alpha_i u_{it}) = 0$$

となるような確率変数であると仮定する．したがって，モデル (1) を

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}, \quad (i = 1, \cdots, N;\ t = 1, \cdots, T) \tag{7}$$

と書き換えて，$v_{it} = \alpha_i + u_{it}$ を誤差項とみなすことができる．新たな誤差項 $\{v_{it}\}$ は，主体間では独立であるが，時間ごとに独立ではない．実際，

$$\mathrm{E}(v_{it}) = 0, \qquad \mathrm{E}(v_{it}v_{js}) = \begin{cases} \sigma_\alpha^2 + \sigma_u^2 & (i = j \text{ かつ } t = s \text{ のとき}) \\ \sigma_\alpha^2 & (i = j \text{ かつ } t \neq s \text{ のとき}) \\ 0 & (i \neq j \text{ のとき}) \end{cases}$$

となる．さらに，第 2 節で述べたように，個別効果 α_i と説明変数 $\boldsymbol{x}_{it}$ の相関を考慮しなければならない場合がある．

以下では，まず，個別効果と説明変数は無相関であると仮定する．このとき，変量効果モデルの代表的な推定方法として，(R-1) pooled OLS 推定，(R-2) GLS 推定を考える．

(R-1) pooled OLS 推定

モデルとして，

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{v}_i, \qquad (i = 1,\ \cdots,\ N) \tag{8}$$

を考える．ここで，$\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})'$, $X_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT})'$, $\boldsymbol{v}_i = (v_{i1}, \cdots, v_{iT})'$.

モデル (8) に対して OLS 推定する方法が pooled OLS 推定であり，推定量は，

$$\hat{\boldsymbol{\beta}}_p = \left(\sum_{i=1}^N X_i'X_i\right)^{-1} \sum_{i=1}^N X_i'\boldsymbol{y}_i$$

により求められる．個別効果と説明変数が無相関の場合は，$\hat{\boldsymbol{\beta}}_p$ は一致性をもつが，効率的ではない．なぜならば，誤差項 $\boldsymbol{v}_i$ の共分散行列は，

$$\Omega = \mathrm{E}(\boldsymbol{v}_i\boldsymbol{v}_i') = \begin{pmatrix} \sigma_\alpha^2 + \sigma_u^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_u^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_u^2 \end{pmatrix} \tag{9}$$

となり，古典的仮定をみたさないからである．

(R-2) GLS 推定

変量効果モデル (7) の誤差項は独立ではなく，(9) のような共分散構造を持っている．そこで，(9) に現れるパラメータ σ_α^2 と σ_u^2 を一致推定量で置き換えて，これらを真のパラメータの代わりに代入して，$\hat{\Omega}$ を作る．その上で，GLS 推定量

$$\hat{\boldsymbol{\beta}}_G = \left(\sum_{i=1}^N X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^N X_i'\hat{\Omega}^{-1}\boldsymbol{y}_i\right)$$

を定義する．

- **$\hat{\Omega}^{-1}$ の計算について**

$$\Omega = \sigma_u^2 I_T + \sigma_\alpha^2 \boldsymbol{e}_T\boldsymbol{e}_T', \quad I_T:\ T\times T \text{ 単位行列}, \quad \boldsymbol{e}_T = (1,\cdots,1)':\ T\times 1$$

と表されるから，行列の逆転公式を使って，

$$\Omega^{-1} = \left(\sigma_u^2 I_T + \sigma_\alpha^2 \boldsymbol{e}_T\boldsymbol{e}_T'\right)^{-1} = \frac{1}{\sigma_u^2}(I_T - \frac{\sigma_\alpha^2\boldsymbol{e}_T\boldsymbol{e}_T'}{\sigma_u^2+T\sigma_\alpha^2}) = \left(\frac{1}{\sigma_u}(I_T-\gamma\boldsymbol{e}_T\boldsymbol{e}_T')\right)^2 = \frac{1}{\sigma_u^2}C_T^2$$

と表すことができる．ここで，

$$C_T = (I_T - \lambda\boldsymbol{e}_T\boldsymbol{e}_T'/T), \quad \lambda = T\gamma = 1-\sqrt{\frac{\sigma_u^2}{\sigma_u^2+T\sigma_\alpha^2}}$$

である．したがって，GLS 推定は，次の変換モデル

$$C_T\boldsymbol{y}_i = \hat{C}_T X_i\boldsymbol{\beta} + \hat{C}_T\boldsymbol{v}_i, \Leftrightarrow\ y_{it}-\lambda\bar{y}_i = (\boldsymbol{x}_{it}-\lambda\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + v_{it}-\lambda\bar{v}_i$$

を OLS 推定することと同等である（実際には，λ を推定量で置き換える）．λ の定義から，GLS 推定に関して，次の性質を導くことができる．

- GLS 推定量の性質

  (a) $T\to\infty$，あるいは σ_α^2 が大きくなれば，λ が 1 に近づくので，GLS 推定は，固定効果モデルにおける LSDV 推定，あるいはグループ内推定に等しくなる．

  (b) 個別効果 α_i の変動が小さくなれば，λ は 0 に近づくので，GLS 推定は，単に $\boldsymbol{y}_i$ を X_i に回帰した pooled OLS 推定に等しくなる．

GLS 推定では，分散の一致推定が必要となるが，これらは次のように行うのが普通である．

- **σ_u^2 と σ_α^2 の一致推定**

  (i) まず，式 (8) の pooled OLS 推定により，残差 $\hat{\boldsymbol{v}}_i$ に基づいて，$\sigma_v^2 = \sigma_\alpha^2+\sigma_u^2$ の一致推定量

$$\hat{\sigma_v}^2 = \frac{1}{NT-p}\sum_{i=1}^N \hat{\boldsymbol{v}}_i'\hat{\boldsymbol{v}}_i = \frac{1}{NT-p}\sum_{i=1}^N\sum_{t=1}^T \hat{v}_{it}^2$$

  を求める．

(ii) 次に，$\sigma_\alpha^2 = \mathrm{E}(v_{it}v_{is})$ $(t \neq s)$ であり，

$$\mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} v_{it}v_{is}\right) = \sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \sigma_\alpha^2 = \sigma_\alpha^2 \frac{T(T-1)}{2}$$

となることから，σ_α^2 の一致推定量として，

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - p}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \hat{v}_{it}\hat{v}_{is}$$

を得る．

(iii) $\sigma_u^2 = \sigma_v^2 - \sigma_\alpha^2$ であるから，σ_u^2 の一致推定量として，

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_\alpha^2$$

を得る．

6．個別効果と説明変数の相関

前節で議論した変量効果モデルの推定では，個別効果 α_i と説明変数 $\boldsymbol{x}_{it}$ は無相関であることを仮定した．しかし，第 2 節で述べたように，一般には相関を考慮する必要がある．

- **個別効果と説明変数に相関がある場合の推定量の性質**

(a) 個別効果と説明変数が無相関ならば，変量効果モデルに適用した pooled OLS，GLS 推定は一致性をもつ．特に，GLS 推定は，漸近的（$N \to \infty$）に有効である．しかし，相関がある場合は，pooled OLS，GLS 推定ともに一致性を失う．

(b) 固定効果モデルに適用した LSDV 推定，階差推定，グループ内推定は一致性をもつ．特に，LSDV 推定，グループ内推定は，漸近的（$N \to \infty$）に有効である．

実際の分析においては，個別効果と説明変数の相関の有無に関する検定が必要になる．そのために，回帰モデルの特定化に関する検定として広く使われている**ハウスマン検定**を使うことができる．

7．ハウスマン検定（Hausman test）

次の仮説検定を考える．

$$H_0: \ \mathrm{E}(\alpha_i \boldsymbol{x}_{it}) = \mathbf{0} \quad \text{vs} \quad \mathrm{E}(\alpha_i \boldsymbol{x}_{it}) \neq \mathbf{0}$$

帰無仮説のもとでは，変量効果モデルに適用した GLS 推定量 $\hat{\boldsymbol{\beta}}_{RE}$ が有効である．他方，対立仮説のもとでは，固定効果モデルに適用した LSDV 推定量 $\hat{\boldsymbol{\beta}}_{FE}$ が有効である．それぞれの推定量の漸近分散は，H_0 のもとで，

$$\begin{aligned}
\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{RE}) &= \sigma_u^2 \left[\mathrm{E}(X_i'(I_T - \lambda \boldsymbol{e}_T\boldsymbol{e}_T'/T)X_i)\right]^{-1}/N, \\
\mathrm{Avar}(\hat{\boldsymbol{\beta}}_{FE}) &= \sigma_u^2 \left[\mathrm{E}(X_i'(I_T - \boldsymbol{e}_T\boldsymbol{e}_T'/T)X_i)\right]^{-1}/N
\end{aligned}$$

となる．ここで，前者の方が後者よりも行列の意味で小さい．実際，

$$\mathrm{E}(X_i'(I_T - \lambda \boldsymbol{e}_T \boldsymbol{e}_T'/T)X_i) - \mathrm{E}(X_i'(I_T - \boldsymbol{e}_T \boldsymbol{e}_T'/T)X_i) = (1-\lambda)\mathrm{E}(X_i'(I_T - \boldsymbol{e}_T \boldsymbol{e}_T'/T)X_i) > 0$$

となる．このとき，上記の検定問題に対するハウスマン検定統計量は，

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})' \left[\hat{V}(\hat{\boldsymbol{\beta}}_{FE}) - \hat{V}(\hat{\boldsymbol{\beta}}_{RE})\right]^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})$$

で定義される．ただし，$\hat{V}(\hat{\boldsymbol{\beta}}_{FE})$ と $\hat{V}(\hat{\boldsymbol{\beta}}_{FE})$ は，共分散行列の推定量である．ハウスマン検定統計量は，H_0 のもとで，自由度 p の χ^2 分布に従い，値が大きいときに H_0 は棄却される．

8. 狭義外生性が成立しない場合

今までの議論では，説明変数と誤差項の独立性，すなわち，狭義外生性を仮定してきた．しかし，狭義外生性が成り立たない場合には，注意が必要である．例えば，説明変数が左辺の y_{it} のラグ付き変数を含むような変量効果モデル，さらに，ラグ付き内生変数を含み，AR 型の誤差項をもつような固定効果および変量効果モデルでは，狭義外生性が成立しない．このような場合の対処法について，詳しくは

J.W. Wooldridge(2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press

を参照されたい．