第 12 回：パネル・データ
【教科書第 9 章第 1 節，第 3 節～第 4 節】

北村　友宏

2025 年 12 月 16 日

本日の内容

1. パネル・データとは

2. 固定効果モデルの Within 推定

3. 変量効果モデルの FGLS 推定

パネル・データ

- 複数の個体を複数の時点にわたり，一定の時間間隔で観測したデータをパネル・データ（panel data）という．
  - e.g., 47 都道府県，1999 年〜2014 年，5 年間隔

パネル・データのモデル

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{it} + u_{it}, \\
i &= 1, 2, \cdots, n, \\
t &= 1, 2, \cdots, T,
\end{aligned}
$$

を推定することを考える.

- i ：個体識別番号
  - e.g., 個人番号
- n ：個体数
- t ：時点識別番号
  - e.g., 年
- T ：時点数

ここで，誤差項 u_{it} が 2 つの部分からなるとして，

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

のように書く．

- μ_i ：個別効果（individual effect）
  - 個体に特有で時間を通じて一定の効果
  - e.g., 生まれもっての個人の能力
- ε_{it} ：その他要因
  - $E(\varepsilon_{it} \mid x_{it}) = 0.$
- 説明変数と相関する個別効果を固定効果（fixed effect）という．
- 説明変数と相関しない個別効果を変量効果（random effect）という．

固定効果モデル

- 個別効果が説明変数と相関することを仮定したモデルを**固定効果モデル（fixed effect model）**という.
- 固定効果モデルの推定方法
  - 1. Least Squares Dummy Variable (LSDV)（この授業では説明は省略）
  - 2. Within 推定

Within 推定

$$y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \varepsilon_{it} \tag{1}$$

の各個体 $i = 1, 2, \cdots, n$ それぞれについて，$t = 1$ から $t = T$ までの（時点間）平均をとると，

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} y_{it} &= \frac{1}{T}\sum_{t=1}^{T}(\beta_0 + \beta_1 x_{it} + \mu_i + \varepsilon_{it}) \\
&= \beta_0 + \beta_1 \cdot \frac{1}{T}\sum_{t=1}^{T} x_{it} + \mu_i + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} \\
\Leftrightarrow \bar{y}_i &= \beta_0 + \beta_1 \bar{x}_i + \mu_i + \bar{\varepsilon}_i \qquad (2)
\end{aligned}$$

- $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$.

(1) と (2) の左辺同士・右辺同士を引き算すると，

$$\begin{aligned} y_{it} - \bar{y}_i &= \beta_0 - \beta_0 + \beta_1(x_{it} - \bar{x}_i) + \mu_i - \mu_i + \varepsilon_{it} - \bar{\varepsilon}_i \\ \Leftrightarrow \tilde{y}_{it} &= \beta_1 \tilde{x}_{it} + \tilde{\varepsilon}_{it}. \end{aligned} \tag{3}$$

- $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{x}_{it} = x_{it} - \bar{x}_i$, $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

$$\Downarrow$$

個別効果 μ_i が消去された！

- $E(\varepsilon_{it} \mid x_{it}) = 0$ の仮定より，$\mathrm{Cov}(x_{it}, \varepsilon_{it}) = 0$.
- $\bar{x}_i$ と $\bar{\varepsilon}_i$ はそれぞれ x_{it} と ε_{it} の平均.

$\Rightarrow \mathrm{Cov}(\tilde{x}_{it}, \tilde{\varepsilon}_{it}) = 0$ となり，(3) は OLS で一致推定できる.

- 時点間平均との差を級内変動（within group variation）という.
  - e.g., ここでいう $\tilde{y}_{it}, \tilde{x}_{it}, \tilde{\varepsilon}_{it}$.

(3) からは定数項も消去されてしまった！

$$\Downarrow$$

全個体・全時点で共通の項を (3) の両辺に足せば，定数項が復活する.

$$\Downarrow$$

足す項が観測可能な（データから計算できる）項でなければ，左辺が観測不能となり推定できなくなる.

$$\Downarrow$$

y_{it} の全個体・全時点での（個体間・時点間）平均を (3) の両辺に足す.

(1) の全個体・全時点での（個体間・時点間）平均をとると，

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}y_{it}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}(\beta_0+\beta_1 x_{it}+\mu_i+\varepsilon_{it})\right] \\
\Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i &= \frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1\bar{x}_i+\mu_i+\bar{\varepsilon}_i) \\
\Leftrightarrow \bar{\bar{y}} &= \beta_0+\beta_1\bar{\bar{x}}+\bar{\mu}+\bar{\bar{\varepsilon}}. \qquad (4)
\end{aligned}
$$

- $\bar{\bar{y}} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$, $\bar{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_i$, $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu_i$, $\bar{\bar{\varepsilon}} = \frac{1}{n}\sum_{i=1}^{n}\bar{\varepsilon}_i$.

(3) と (4) の左辺同士・右辺同士を足し合わせると，

$$
\begin{aligned}
(\tilde{y}_{it} + \bar{\bar{y}}) &= (0 + \beta_0) + \beta_1(\tilde{x}_{it} + \bar{\bar{x}}) + (0 + \bar{\mu}) + (\tilde{\varepsilon}_{it} + \bar{\bar{\varepsilon}}) \\
&= (\beta_0 + \bar{\mu}) + \beta_1(\tilde{x}_{it} + \bar{\bar{x}}) + (\tilde{\varepsilon}_{it} + \bar{\bar{\varepsilon}}) \\
\Leftrightarrow \tilde{\tilde{y}}_{it} &= \alpha + \beta_1 \tilde{\tilde{x}}_{it} + \tilde{\tilde{\varepsilon}}_{it}. \qquad (5)
\end{aligned}
$$

- $\tilde{\tilde{y}}_{it} = \tilde{y}_{it} + \bar{\bar{y}}, \alpha = \beta_0 + \bar{\mu}, \tilde{\tilde{x}}_{it} = \tilde{x}_{it} + \bar{\bar{x}}, \tilde{\tilde{\varepsilon}}_{it} = \tilde{\varepsilon}_{it} + \bar{\bar{\varepsilon}}$.

$\Rightarrow$ 定数項が α として復活した！
$\tilde{\tilde{x}}_{it}$ と $\tilde{\tilde{\varepsilon}}_{it}$ はそれぞれ $\tilde{x}_{it}$ と $\tilde{\varepsilon}_{it}$ に定数を足したもの．
$\Rightarrow \mathrm{Cov}(\tilde{\tilde{x}}_{it}, \tilde{\tilde{\varepsilon}}_{it}) = 0$ となり，(5) も OLS で一致推定できる．
$\Rightarrow$ (5) を OLS で推定する手法は，被説明変数と説明変数の級内変動を用いた級内推定（Within 推定）．

実証分析例：固定効果モデル

3020 人分の 2007 年と 2009 年のパネル・データを用い，病気やけがをすると生活の満足度が下がるかどうかを調べるモデルを Within 推定する．

$$
\begin{aligned}
life_{it} = \beta_0 + \beta_1 shock_i + \beta_2 y2_t + \beta_3 shock_i y2_t \\
+ \beta_4 income_{it} + \mu_i + \varepsilon_{it}
\end{aligned}
$$

- $life_{it}$: 生活満足度（不満を 0, 満足を 4 とする 5 段階のランク）
- $shock_i$: 病欠ダミー（2008 年に病気やけがで休んだことがある＝ 1, ない＝ 0）
- $y2_t$: 2009 年ダミー（2009 年＝ 1, 2007 年＝ 0）
- $income_{it}$: 年収（万円）
- i : 個人番号
- t : 年

簡単化のため，$income_{it}$ と個別効果 μ_i とその他効果 ε_{it} が全て 0 とすると，$shock_i$ と $y2_t$ それぞれのとりうる値の組み合わせの場合の被説明変数 $life_{it}$ の値は，

	$shock_i = 1$	$shock_i = 0$	差
$y2_t = 1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
$y2_t = 0$	$\beta_0 + \beta_1$	β_0	β_1
差	$\beta_2 + \beta_3$	β_2	β_3

$$\Downarrow$$

2009 年における，病欠ありの人となしの人との間の生活満足度の差 $\beta_1 + \beta_3$ から，元々存在した（2007 年における），病欠ありの人となしの人との間の生活満足度の差 β_1 を引いたものは，β_3.

$$\Downarrow$$

$shock_i y2_t$ すなわち「病欠ダミー × 2009 年ダミー」の係数 β_3 を推定すれば，それが「2008 年の病気やけがが生活満足度に与える影響」になる．

gretl での Within 推定

データセットをパネル・データとして読み込んだ状態で,

- メニューバーから「モデル」→「パネル」→「固定効果あるいは変量効果」と操作.
- ラジオボタンの中から「固定効果」を選ぶ.

⇒ (5) が推定され，Within 推定ができる.

- 「説明変数（回帰変数）」を複数選べば，定数項以外に説明変数が 2 つ以上あるモデルも推定できる.

パネル・データ分析における決定係数

gretl では，Within 推定などパネル・データ特有の方法でのモデル推定を行うと，決定係数が 2 種類表示される．

Within 推定の場合，2 種類の決定係数は次のとおり.

- LSDV R-squared: LSDV 推定での決定係数
  - 被説明変数の変動のうち，どの程度の割合を説明変数と各個体ダミー変数で説明できているかを表す.
- Within R-squared: 級内変動に基づく決定係数
  - 被説明変数の級内変動（個別効果を除いた部分の変動）のうち，どの程度の割合を説明変数の級内変動で説明できているかを表す.

※ 特に級内変動に基づく決定係数については，（説明変数が複数ある場合でも）自由度を修正しない場合が多い.

$$\Downarrow$$

- Within 推定において，モデルの当てはまりの良さの指標としては，級内変動（個別効果を除いた変動）を用いて計算した「級内変動に基づく決定係数」が適切．
- レポート・論文用の推定結果表に決定係数を載せる際，どの定義の決定係数なのかを明記する．

固定効果モデルの Within 推定結果

gretl: モデル1

ファイル 編集(E) 検定(T) 保存(S) グラフ(G) 分析(A) LaTeX

```
モデル 1: 固定効果モデル, 観測数: 6040
クロスセクションユニット数: 3020
時系列の長さ = 2
従属変数: life
完全な共線性のため次の変数を取り除きました: shock

              係数          標準誤差        t値          p値
  ------------------------------------------------------------
  const       2.48153       0.0453239      54.75      0.0000     ***
  y2          0.215365      0.0313550       6.869     7.84e-012  ***
  shock_y2   -0.140117      0.0484445      -2.892     0.0039     ***
  income      0.000223286   0.000161409     1.383     0.1667

Mean dependent var    2.621854   S.D. dependent var    0.952447
Sum squared resid     2584.365   回帰の標準誤差                0.925527
LSDV R-squared        0.528256   Within R-squared      0.018844
LSDV F(3022, 3017)    1.117939   P-value(F)            0.001099
Log-likelihood       -6006.637   Akaike criterion      18059.27
Schwarz criterion     38331.99   Hannan-Quinn          25097.00
rho                  -1.000000   Durbin-Watson         1.999449

名前の付けられた説明変数についての結合検定(Joint test on named regressors) -
  検定統計量: F(3, 3017) = 19.3148
  なお、p値(p-value) = P(F(3, 3017) > 19.3148) = 2.10651e-12

定数項がクロスセクションユニット (グループ) ごとに異なるかどうかの検定 -
  帰無仮説: 各クロスセクションユニットは共通の定数項を持つ
  検定統計量: F(3018, 3017) = 1.08665
  なお、p値(p-value) = P(F(3018, 3017) > 1.08665) = 0.0112508
```

※ shock（病欠ダミー）の係数は自動的に取り除かれている．

- 病欠ダミーが時間を通じて変化しないことにより，固定効果モデルを仮定するとその係数が推定できないため．

- 病欠ダミー × 2009 年ダミーの係数
  - −0.140117
  - t 値は −2.892, p 値は 0.0039.
    ➡ 仮に「shock_y2 の係数が 0」だとすると，−2.892 という t 値は 0.39%の確率（1%を下回る確率）でしか出てこない．
    ➡ 有意水準 1%で，「係数は 0」の H_0 が棄却される（5%や 10%でも棄却される）．
    ➡ 病欠ダミー × 2009 年ダミーは生活満足度と統計的に有意に相関している．
    ➡ 年収を一定としたうえで，2008 年に病気やけがで休んだことがあると，そうでない場合に比べ，2009 年の生活満足度が 0.140117 ランク低くなる傾向がある．

- 決定係数
  - 級内変動に基づく決定係数は 0.018844
    ➥ 「3 つの説明変数の級内変動」の違いで，「生活満足度の級内変動」のバラつきが約 1.88%説明できる．

変量効果モデル

- 個別効果が説明変数と相関しないことを仮定したモデルを変量効果モデル（random effect model）という.

変量効果モデルは，

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{it} + \mu_i + \varepsilon_{it}, \\
E(\mu_i \mid x_{it}) &= 0, \\
E(\varepsilon_{it} \mid x_{it}) &= 0, \\
V(\mu_i \mid x_{it}) &= \sigma_\mu^2, \\
V(\varepsilon_{it} \mid x_{it}) &= \sigma_\varepsilon^2, \\
E(\mu_i \varepsilon_{it} \mid x_{it}) &= 0, \\
i &= 1, 2, \cdots, n, \\
t &= 1, 2, \cdots, T,
\end{aligned}
$$

のように仮定する．

固定効果モデルと変量効果モデルの大きな違い

- 固定効果モデル：$E(\mu_i \mid x_i) \neq 0.$
- 変量効果モデル：$E(\mu_i \mid x_i) = 0.$

変量効果モデルの推定方法

- 変量効果モデルでは個別効果 μ_i を誤差項から消去することを考えず，

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

  を誤差項として考える．
- 同じ個体でも，異なる時点同士の誤差項が相関する．
  - e.g., 1 人目の誤差項は他の人より大きく，2007 年も 2009 年も似たような値をとる．

⇒ OLS 推定の仮定の 1 つ（誤差項同士は無相関）が満たされず，変量効果モデルは実行可能な一般化最小二乗法（Feasible Generalized Least Squares, FGLS）で推定する（詳細な説明は省略）．

gretl での変量効果モデルの FGLS 推定

データセットをパネルデータとして読み込んだ状態で,

- メニューバーから「モデル」→「パネル」→「固定効果あるいは変量効果」と操作.
- ラジオボタンの中から「変量効果（ランダム効果)」を選ぶ.

⇒ 変量効果モデルの FGLS 推定ができる.

- 「説明変数（回帰変数)」を複数選べば，定数項以外に説明変数が 2 つ以上あるモデルも推定できる.

実証分析例：変量効果モデル

3020 人分の 2007 年と 2009 年のパネル・データを用い，病気やけがをすると生活の満足度が下がるかどうかを調べるモデルを FGLS 推定する．

$$
\begin{aligned}
life_{it} = \beta_0 + \beta_1 shock_i + \beta_2 y2_t + \beta_3 shock_i y2_t \\
+ \beta_4 income_{it} + \mu_i + \varepsilon_{it}
\end{aligned}
$$

- $life_{it}$: 生活満足度（不満を 0, 満足を 4 とする 5 段階のランク）
- $shock_i$: 病欠ダミー（2008 年に病気やけがで休んだことがある＝ 1, ない＝ 0）
- $y2_t$: 2009 年ダミー（2009 年＝ 1, 2007 年＝ 0）
- $income_{it}$: 年収（万円）
- i : 個人番号
- t : 年

変量効果モデルの FGLS 推定結果

gretl: モデル2

ファイル 編集(E) 検定(T) 保存(S) グラフ(G) 分析(A) LaTeX

```
モデル 2: 変量効果モデル(GLS), 観測数: 6040
クロスセクションユニット数: 3020
時系列の長さ = 2
従属変数: life

                係数         標準誤差        z          p値
  ------------------------------------------------------------
  const       2.45439       0.0263360     93.20     0.0000     ***
  shock       0.0153078     0.0349948      0.4374   0.6618
  y2          0.212996      0.0310253      6.865    6.64e-012  ***
  shock_y2   -0.139777      0.0484373     -2.886    0.0039     ***
  income      0.000303324   5.20242e-05    5.830    5.53e-09   ***

Mean dependent var   2.621854   S.D. dependent var   0.952447
Sum squared resid    5393.587   回帰の標準誤差              0.945288
Log-likelihood      -8228.545   Akaike criterion     16467.09
Schwarz criterion    16500.62   Hannan-Quinn         16478.73
rho                 -1.000000   Durbin-Watson        1.999449

'Between' 分散 = 0.0372258
'Within' 分散 = 0.856601
theta used for quasi-demeaning = 0.0408154
corr(y,yhat)^2 = 0.0154661

名前の付けられた説明変数についての結合検定(Joint test on named regressors)
  漸近的検定統計量: カイ二乗(4) = 95.628
  なお、p値(p-value) = 8.37905e-20

ブロイシュ=ペーガン(Breusch-Pagan)検定 -
  帰無仮説: 個別誤差(unit-specific error)の分散は0である
  漸近的検定統計量: カイ二乗(1) = 5.22889
  なお、p値(p-value) = 0.0222147

ハウスマン(Hausman)検定 -
  帰無仮説: GLS推定値は一致性を持つ
  漸近的検定統計量: カイ二乗(1) = 0.27467
  なお、p値(p-value) = 0.600216
```

※ 変量効果モデルの場合，時間を通じて変化しない変数の係数も推定できる．
➡ shock（病欠ダミー）の係数の推定結果も出力されている．

- 病欠ダミー × 2009 年ダミーの係数
  - –0.139777
  - z 値は –2.886, p 値は 0.0039.
    ➡ 仮に「shock_y2 の係数が 0」だとすると，–2.886 という z 値は 0.39%の確率（1%を下回る確率）でしか出てこない．
    ➡ 有意水準 1%で，「係数は 0」の H_0 が棄却される（5%や 10%でも棄却される）．
    ➡ 病欠ダミー × 2009 年ダミーは生活満足度と統計的に有意に相関している．
    ➡ 年収を一定としたうえで，2008 年に病気やけがで休んだことがあると，そうでない場合に比べ，2009 年の生活満足度が 0.139777 ランク低くなる傾向がある．

固定効果モデルと変量効果モデルでの，「病欠ダミー × 2009 年ダミー」の係数推定値の違い

- 固定効果モデルでは，係数が −0.140117 なので，年収を一定としたうえで，2008 年に病気やけがで休んだことがあると，そうでない場合に比べ，2009 年の生活満足度が 0.140117 ランク低くなる傾向がある，という解釈となる．
- 変量効果モデルでは，係数が −0.139777 なので，年収を一定としたうえで，2008 年に病気やけがで休んだことがあると，そうでない場合に比べ，2009 年の生活満足度が 0.139777 ランク低くなる傾向がある，という解釈となる．

$$\Downarrow$$

- 固定効果モデルも変量効果モデルも，病欠ダミー × 2009 年ダミーの係数推定値が近い.
- 推定結果の画面より，ハウスマン検定では「（変量効果モデルを仮定した F）GLS 推定値は一致性を持つ」という H_0 が有意水準 10%でも棄却されないため，この場合は固定効果モデルを仮定して Within 推定を行う意味がなく，変量効果モデルを仮定した FGLS 推定が支持される（詳細な説明は省略）.

今日のキーワード

パネル・データ，個別効果，固定効果，変量効果，固定効果モデル，級内変動，変量効果モデル

次回までの準備

- 今回の講義スライドを読み直す.
- 「提出課題 9」に取り組む.
- 教科書第 9 章第 2 節を読む.