第 9 回：パネルデータ分析（2）

北村　友宏

2020 年 12 月 4 日

本日の内容

1. 固定効果モデルの Within 推定
2. 固定効果モデルの長所と短所

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., 市場番号
- n : 個体数
- t : 時点識別番号
  - e.g., 月
- 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., 市場の集客力
- ε_{it}：その他要因
  - $E(\varepsilon_{it} \mid x_{it}) = 0.$
- 説明変数と独立でない個別効果を固定効果（fixed effect）という．
- 説明変数と独立な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独立でない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)（前回の授業で説明）
  - 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，

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

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

$$\Downarrow$$

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,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}y_{it}\right)=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}(\beta_0+\beta_1x_{it}+\mu_i+\varepsilon_{it})\right]$$
$$\Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i=\frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1\bar{x}_i+\mu_i+\bar{\varepsilon}_i)$$
$$\Leftrightarrow \bar{\bar{y}}=\beta_0+\beta_1\bar{\bar{x}}+\bar{\mu}+\bar{\bar{\varepsilon}}. \tag{4}$$

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

(3) と (4) の左辺同士・右辺同士を足し合わせると,

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

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

$\Rightarrow$ 定数項が α として復活した！

$\tilde{\bar{x}}_{it}$ と $\tilde{\bar{\varepsilon}}_{it}$ はそれぞれ $\tilde{x}_{it}$ と $\tilde{\varepsilon}_{it}$ に定数を足したもの.

$\Rightarrow$ $\mathrm{Cov}(\tilde{\bar{x}}_{it}, \tilde{\bar{\varepsilon}}_{it}) = 0$ 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）.

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.

⇒ 前スライドの (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.

みかんの需要関数

固定効果モデルを仮定し，みかんの需要関数を，

$$q_{it} = \beta_0 + \beta_P p_{it} + \mu_i + \varepsilon_{it},$$

- q_{it}：取引数量
- p_{it}：価格
- μ_i：個別効果
- i：市場番号
- t：月（時点番号）

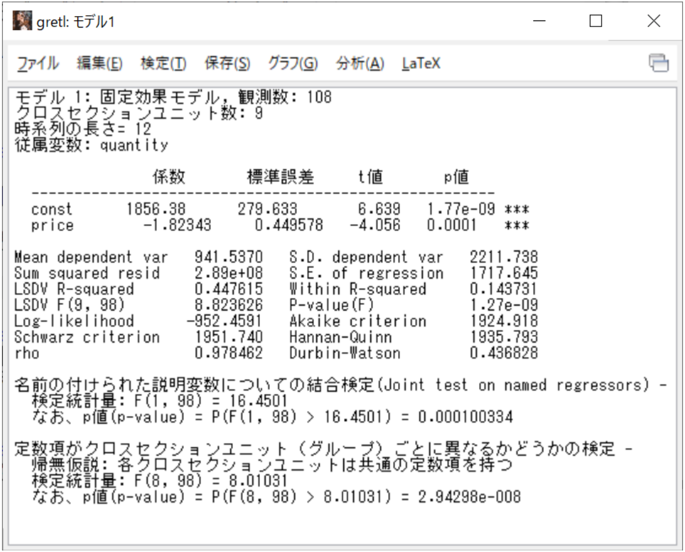
とする．

実習 1

東京都中央卸売市場で取引されたみかんの数量と価格（市場別・月別）のデータを用いて，みかんの需要関数の Within 推定を行う．

1. gretl を起動．
2. 「ファイル」→「データを開く」→「ユーザー・ファイル」と操作．
3. orangetokyo.gdt を選択し，「開く」をクリック．
4. gretl のメニューバーから「モデル」→「パネル」→「固定効果あるいは変量効果」と操作．
5. 出てきたウィンドウ左側の変数リストにある quantity をクリックし，3 つの矢印のうち上の青い右向き矢印をクリック．
    - 推定式の左辺の変数（被説明変数，従属変数）が quantity（みかんの取引数量）となる．

6. ウィンドウ左側の変数リストにある price をクリックし，3 つの矢印のうち真ん中の緑の右向き矢印をクリック．
   - 推定式の右辺の変数（説明変数，独立変数）が price（みかんの価格）となる．
   - 最初から説明変数リストに入っている const は推定式の切片（定数項）のこと．
7. 「頑健標準誤差を使用する」にチェックが入っていれば外す．今回も不具合防止のため，このオプションにはチェックしない．
   - デフォルトの標準誤差が計算される．
8. ラジオボタンの「固定効果」が選ばれていなければ，「固定効果」をクリック．
   - 固定効果モデルが仮定され，Within 推定が行われる．
9. 「OK」をクリックすると，結果が表示される．



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: 108
クロスセクションユニット数: 9
時系列の長さ= 12
従属変数: quantity

              係数        標準誤差      t値        p値
  ---------------------------------------------------------
  const       1856.38      279.633      6.639   1.77e-09 ***
  price         -1.82343     0.449578  -4.056   0.0001   ***

Mean dependent var   941.5370   S.D. dependent var   2211.738
Sum squared resid    2.89e+08   S.E. of regression   1717.645
LSDV R-squared       0.447615   Within R-squared     0.143731
LSDV F(9, 98)        8.823626   P-value(F)           1.27e-09
Log-likelihood      -952.4591   Akaike criterion     1924.918
Schwarz criterion    1951.740   Hannan-Quinn         1935.793
rho                  0.978462   Durbin-Watson        0.436828

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(1, 98) = 16.4501
  なお、p値(p-value) = P(F(1, 98) > 16.4501) = 0.000100334

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(8, 98) = 8.01031
  なお、p値(p-value) = P(F(8, 98) > 8.01031) = 2.94298e-008
```

このような画面が表示されれば成功.

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.

注：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 推定では,

- LSDV 推定において説明変数に含める各個体ダミー変数は，あくまで個別効果を表すものなので，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.

需要関数の Within 推定結果

- 価格の係数
  - −1.82343（符号は負）
    ➥ 前回行った LSDV 推定の結果と同じ値
  - 有意水準 1%で，係数ゼロの H_0 棄却．
    ➥ 価格は取引数量と統計的に有意に相関している．
    ➥ みかん 1kg 当たりの価格が 1 円高くなると，取引数量は 1.82343t 減少する．
    ⇒ 経済理論と整合的．
- 決定係数
  - 級内変動に基づく決定係数は 0.143731
    ➥ 「価格の級内変動」の違いで，「数量の級内変動」のバラつきが約 14.4%説明できる．

実習 2

1. 「gretl: モデル 1」のウィンドウのメニューバーから「ファイル」→「名前を付けて保存」と操作.
2. 「標準テキスト」を選び,「OK」をクリック。
3. 需要関数推定結果 4.txt という名前で「2020 ミクロデータ分析 2」フォルダに保存. すると, 表示された推定結果をそのままテキストファイルで保存できる.

固定効果モデルを仮定し，LSDV 推定や Within 推定をする場合の長所と短所

- 長所
  - 欠落変数バイアスを緩和できる場合がある.
- 短所
  - 時間を通じて一定の要因の影響を明らかにできない.

欠落変数バイアス

真のモデルは

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + w_{it}, \\
E(w_{it} \mid x_{1it}, x_{2it}) &= 0,
\end{aligned}
$$

であるが，x_{2it} が観測できないため，それを除外して

$$
y_{it} = \beta_0 + \beta_1 x_{1it} + u_{it},
$$

を推定する，つまり y_{it} を x_{1it} のみに回帰することを考える．

- $u_{it} = \beta_2 x_{2it} + w_{it}$.

もし x_{1it} と x_{2it} が相関していると，$u_{it} = \beta_2 x_{2it} + w_{it}$ なので x_{1it} と u_{it} も相関する．つまり，

$$\mathrm{Cov}(x_{1it}, u_{it}) \neq 0.$$

$$\Downarrow$$

y_{it} を x_{1it} のみに回帰すると，x_{1it} の係数の OLS 推定量 $\hat{\beta}_1$ は，

$$\operatorname*{plim}_{nT \to \infty} \hat{\beta}_1 = \beta_1 + \frac{\mathrm{Cov}(x_{1it}, u_{it})}{V(x_{1it})} \neq \beta_1.$$

$$\Downarrow$$

偏り（バイアス）が生じ，正しく推定できない（一致推定量が得られない）．

- このバイアスは，y_{it} に影響を与える x_{2it} が説明変数から欠落していることにより生じる．
- 説明変数の欠落によって生じる OLS 推定量の偏りを欠落変数バイアス（omitted variable bias）という．
- 欠落変数バイアスは，モデルに必要な説明変数が 1 つでも欠落している限り必ず発生する．
  $\Rightarrow$ 通常，避けられない．

長所：欠落変数バイアスの緩和

もし，x_{2it} が，各個体それぞれについて時間を通じてほとんど変化しないと考えられる変数であれば，x_{2it} の変動の一部を個別効果と考え，固定効果モデル

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

を仮定して LSDV 推定または Within 推定を適用すれば，欠落変数バイアスを緩和できる．

短所：時間不変要因の影響が検出不能

固定効果モデル

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

において，説明変数 x_{it} が，

- 時間を通じて全く変化しない場合
  ➡ その係数 β_1 は推定できない．
  - gretl では，そのような説明変数は自動的に落とされる．
- 時間を通じてほとんど変化しない場合
  ➡ その係数 β_1 の推定値がほぼ 0 となり，係数ゼロの帰無仮説が棄却されにくくなる．

$$\Downarrow$$

時間を通じて全く変化しない，またはほとんど変化しない変数が説明変数に含まれていると，LSDV 推定や Within 推定ではそのような説明変数の係数を正しく推定・検定できない.

➥ 時間を通じて一定の要因の影響を明らかにできない.

➥ このような場合は，固定効果モデルを仮定せずに通常の線形回帰モデル（プーリングモデル, pooling model）

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}$$

として推定する.

本日の作業はここまで.
今回は gretl のデータセットに変更を加えていないので，gretl のデータセット（orangetokyo.gdt）を上書き保存する必要はない.