

第 8 回 : パネルデータ分析 (1)

北村 友宏

2020 年 11 月 20 日

本日の内容

1. パネルデータとは
2. 固定効果モデルの LSDV 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



この授業で分析している『青果物卸売市場調査報告』(平成 22 年版)の, 東京都中央卸売市場(計 9 市場, 12 ヶ月間)で取引されたみかんの数量と価格(市場別・月別)のデータも, パネルデータ.

分析用のパネルデータの作成方法

入手したデータを，パネルデータとして統計解析ソフトで読み込み，分析するには，以下の加工・整理をする。

- ▶ Excel ファイルの 1 行目は変数名
 - ▶ 個体識別番号の変数 (e.g., 市場番号など) と時点識別番号 (e.g., 月, 年度など) の変数を必ず含める。
- ▶ 2 行目には 1 番目の個体の第 1 時点目の各変数の数値, 3 行目には 1 番目の個体の第 2 時点目の各変数の数値, …
- ▶ 1 番目の個体の数値を全時点について並べたら, その次の行には 2 番目の個体の第 1 時点目の各変数の数値, 続く行には 2 番目の個体の第 2 時点目の各変数の数値, …

パネル分析用の Excel ファイルの形

	A	B	C	D	E
1	id	name	month	quantity	price
2	1	Tsukiji	1	2118	165
⋮	⋮	⋮	⋮	⋮	⋮
7	1	Tsukiji	6	80	1012
⋮	⋮	⋮	⋮	⋮	⋮
13	1	Tsukiji	12	3848	270
14	2	Ota	1	8281	173
⋮	⋮	⋮	⋮	⋮	⋮
19	2	Ota	6	630	904
⋮	⋮	⋮	⋮	⋮	⋮
25	2	Ota	12	15913	268



- ▶ この授業で分析している，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）のデータはすでに分析用のパネルデータの形になっているので，現在のデータセットを引き続き用いてパネルデータ分析を実行してみる。
- ▶ ここからは，モデルの説明変数と被説明変数の相互依存関係に起因する問題（同時性）は無視して，パネルデータ特有の手法での分析を解説する。
 - ▶ 同時性に対処する手法とパネルデータ分析の手法の組み合わせ（パネル操作変数法）もあるが，学部レベルを超えるため省略。

パネルデータのモデル

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it},$$
$$i = 1, 2, \dots, n,$$
$$t = 1, 2, \dots, T.$$

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
 - ▶ $E(\varepsilon_{it} | x_{it}) = 0$.
- ▶ 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)**
 - ▶ 2. **Within 推定** (次回の授業で説明)

※個別効果が説明変数と独立であることを仮定した**変量効果モデル (random effect model)** もあり、固定・変量のどちらを採択するかを検定する方法がある。これについては後の授業で説明する。

Least Squares Dummy Variable (LSDV)

例えば、個体が2個 ($n = 2$) のとき、

$$y_{1,t} = \beta_0 + \beta_1 x_{1,t} + \mu_1 + \varepsilon_{1,t},$$

$$y_{2,t} = \beta_0 + \beta_1 x_{2,t} + \mu_2 + \varepsilon_{2,t}.$$

書き換えると、

$$y_{1,t} = \underbrace{\beta_0 + \mu_1}_{=\beta_0^*} + \beta_1 x_{1,t} + \varepsilon_{1,t}$$

$$= \beta_0^* + \beta_1 x_{1,t} + \underbrace{(\mu_2 - \mu_1)}_{=\beta_2} \cdot 0 + \varepsilon_{1,t}$$

$$= \beta_0^* + \beta_1 x_{1,t} + \beta_2 \cdot 0 + \varepsilon_{1,t},$$

$$y_{2,t} = \beta_0 + \mu_1 + \beta_1 x_{2,t} + (\mu_2 - \mu_1) + \varepsilon_{2,t}$$

$$= \beta_0^* + \beta_1 x_{2,t} + \beta_2 \cdot 1 + \varepsilon_{2,t}.$$



$i = 1$ のとき 0, $i = 2$ のとき 1 となるダミー変数 d_{2i} を, モデルの説明変数に組み込み, OLS を適用すれば推定できる.

- ▶ $y_{it} = \beta_0^* + \beta_1 x_{it} + \beta_2 d_{2i} + \varepsilon_{it}$ を推定.
- ▶ d_{2i} は, 2 番目の個体を表すダミー変数.
- ▶ $\beta_2 = \mu_2 - \mu_1$ は, 「1 番目の個体と比較して 2 番目の個体は被説明変数の値がどの程度異なる傾向があるか」を表す.

※個体が 3 個以上の場合, 各個体のダミー変数のうち 1 つの個体を除いてモデルの説明変数に組み込み, OLS を適用すればよい.

⇒ これは **Least Squares Dummy Variable (LSDV)** の手法.

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)

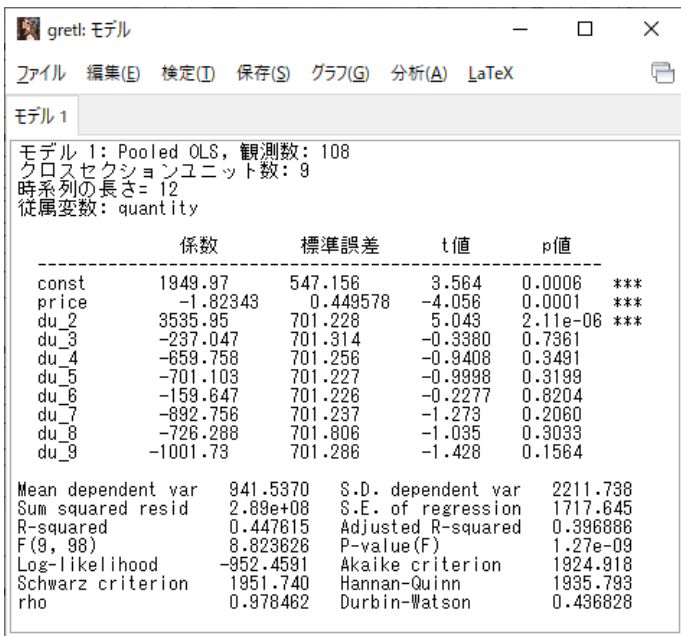
とする。

実習 1

みかんの需要関数の LSDV 推定を行う。個別効果 μ_i に各市場のダミー変数を組み込むので、基本的には第 4 回授業の実習で行った各市場のダミー変数を含む重回帰と同じ方法になる。

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」→「通常の最小二乗法」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. Ctrl キーを押しながら，ウィンドウ左側の変数リストにある price, du_2, du_3, du_4, du_5, du_6, du_7, du_8, du_9 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。du_1 はクリックしない。
 - ▶ 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，du_2 から du_9（築地を除く 8 市場のダミー変数）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。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. 「OK」をクリックすると，結果が表示される。



このような画面が表示されれば成功.

需要関数の LSDV 推定結果

▶ 価格の係数

- ▶ -1.82343 (符号は負)
- ▶ 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。

※レポート・論文用に作成する推定結果表には、論理展開上重要でない限り、個体ダミー変数の係数推定値やその t 値は載せなくてよい。

固定効果モデルを LSDV 推定する場合の 長所と短所

▶ 長所

- ▶ 各個体の個別効果の大きさを把握できる.

▶ 短所

- ▶ 個体数が非常に多い場合、統計解析ソフトでの推定結果表が縦に長くなり、見づらくなる.
- ▶ 個体数が非常に多い場合、統計解析ソフトを用いた推定計算に時間がかかる.

※ 「固定効果モデル推定」そのものの長所と短所は
次回の授業で説明する.

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.