第 5 回：同時方程式モデルの推定（1）

北村　友宏

2020 年 10 月 30 日

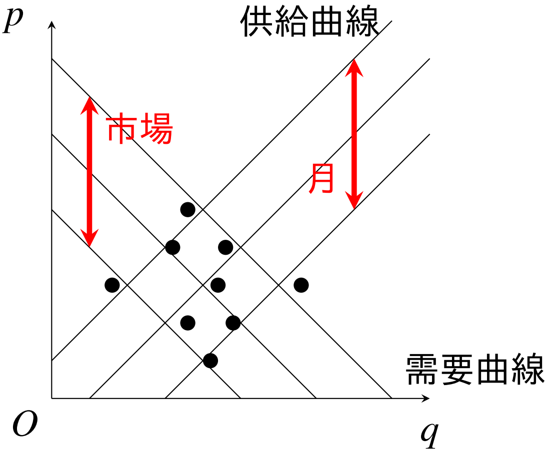
本日の内容

1. 需要関数と供給関数
2. 説明変数と誤差項の相関
3. 2 段階最小二乗法
4. みかんの需要曲線のより厳密な推定

需要関数と供給関数

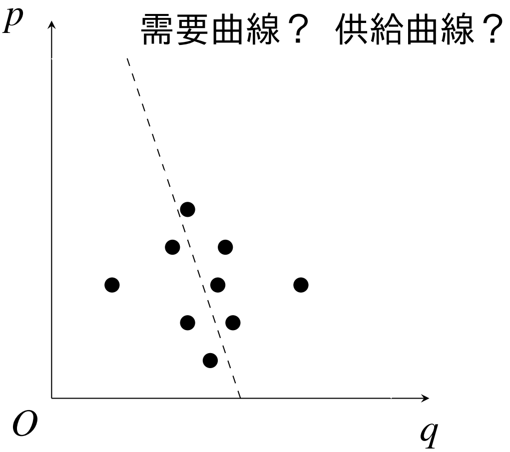
- 財の価格や数量を決定する関数
  ➡ 需要関数と供給関数
- 需要曲線の位置
  ➡ 諸要因（e.g., 消費者数や家計所得など）によって変化.
- 供給曲線の位置
  ➡ 諸要因（e.g., 季節や生産要素価格など）によって変化.
- グラフでの，各曲線からの乖離（誤差）
  ➡ 市場や時点によって変化.
- データとして観測されるのは，市場均衡での（誤差を含む）取引数量と価格の組み合わせ.

需要・供給曲線の位置



実際には需要・供給曲線は目に見えず，価格と数量の組み合わせだけが分かる．

価格と数量の散布図



数量を価格（と他の要因）に回帰しても，推定されたものは需要曲線でも供給曲線でもない．

$$\Downarrow$$

供給曲線の位置の変化を無視して需要関数を推定しようとすると、需要関数と供給関数を混合したものが推定される！

$$\Downarrow$$

第 3 回と第 4 回の授業で推定したものは需要関数ではなく，需要関数と供給関数を混合したもの.

線形同時方程式モデル

簡単化のため，定数項以外に説明変数が 1 つである式 2 本からなる線形同時方程式モデル（linear simultaneous equation model）を考える.

$$y_i = \beta_0 + \beta_1 x_i + u_i, \tag{1}$$

$$x_i = \alpha_0 + \alpha_1 y_i + w_i. \tag{2}$$

- 「$x_i \to y_i \to x_i \to y_i \to \cdots$」のように決定され，$x_i$ と y_i が相互依存関係にある.
  $\Rightarrow$ 需要・供給関数のモデルと同様.

もし，$E(u_i \mid x_i) = 0$ の仮定が正しければ，
$E(x_i u_i) = E(E(x_i u_i \mid x_i)) = E(x_i E(u_i \mid x_i)) = 0$ かつ
$E(u_i) = E(E(u_i \mid x_i)) = 0$ となり（繰り返し期待値の法則），

$$\mathrm{Cov}(x_i, u_i) = E(x_i u_i) - E(x_i)E(u_i) = 0.$$

「説明変数と誤差項は無相関」

このとき，

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1 = \beta_1 + \frac{\mathrm{Cov}(x_i, u_i)}{V(x_i)} = \beta_1.$$

「観測値数 n が十分に大きいとき，回帰係数の OLS 推定量は真の係数の値に確率収束する（一致推定量，consistent estimator になる）」
※ $\hat{\beta}_0$ についても，$\operatorname*{plim}_{n\to\infty} \hat{\beta}_0 = \beta_0$.

- 任意の $c > 0$ について，

$$\lim_{n\to\infty} P\left(|\hat{\beta}_1 - \beta_1| < c\right) = 1$$

が成り立つとき，$\hat{\beta}_1$ は β_1 に確率収束する（converge in probability）という．

  - $\operatorname*{plim}_{n\to\infty} \hat{\beta}_1 = \beta_1$ や $\hat{\beta}_1 \xrightarrow{p} \beta_1$ と書く．
  - 観測値数 n が十分に大きいとき，$\hat{\beta}_1$ は β_1 からほぼ 100%，離れない．

しかし，

$$y_i = \beta_0 + \beta_1 x_i + u_i, \tag{1}$$
$$x_i = \alpha_0 + \alpha_1 y_i + w_i. \tag{2}$$

からなるモデルは，

- u_i が変化
  → (1) に沿って y_i が変化
  → (2) に沿って x_i が変化
  →⋯

⇒「(1) の説明変数と誤差項が相関する！」

$$\mathrm{Cov}(x_i, u_i) \neq 0$$

ならば $E(u_i \mid x_i) \neq 0.$

➥ OLS 推定における仮定の 1 つ「$E(u_i \mid x_i) = 0$」が満たされない！

このとき，$\mathrm{Cov}(x_i, u_i) \neq 0$ なので，

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1 = \beta_1 + \frac{\mathrm{Cov}(x_i, u_i)}{V(x_i)} \neq \beta_1.$$

「観測値数 n が十分に大きくても，回帰係数の OLS 推定量は真の係数の値と異なる値（偏った値）に確率収束する（一致推定量にならない）」
※ $\hat{\beta}_0$ や，(2) の $\hat{\alpha}_0, \hat{\alpha}_1$ についても一致推定量にならない.

$$\Downarrow$$

システムが複数の式からなり，各式の説明変数と被説明変数に相互依存関係がある場合（e.g., 需要・供給関数），各式の係数を OLS で推定すると偏り（バイアス）が生じ，正しく推定できない.

- 説明変数と被説明変数の相互依存関係によって生じる OLS 推定量の偏りを同時方程式バイアス（simultaneous equation bias）という．

- システムの外部で決まる変数を外生変数（exogenous variable）という．
  - e.g., 需要・供給モデルにおける市場規模，家計所得，生産要素価格（電力価格，職員賃金など），気候，法的規制，技術水準など

- 外生変数を所与として，システムの内部で決まる変数を内生変数（endogenous variable）という．
  - e.g., 需要・供給モデルにおける価格や数量

⇒ 需要関数や供給関数の説明変数には内生変数が含まれ，これにより同時方程式バイアスが生じる．

操作変数

同時方程式バイアスを緩和しつつ，説明変数と被説明変数に相互依存関係のある式を推定する方法を考える．

x_i を内生変数である説明変数，u_i を誤差項とする．

- x_i と相関し，かつ u_i と相関しない変数を操作変数（Instrumental Variable, IV）といい，z_i で表す．
  - $\mathrm{Cov}(z_i, u_i) = 0$ かつ $\mathrm{Cov}(z_i, x_i) \neq 0$.
- 外生変数であれば，「推定式の誤差項と相関しない」と考えることができる．

  ➡ 「内生説明変数と相関する外生変数」が，操作変数になりうる．
- 操作変数は 1 個とは限らず，複数個存在する場合もある．

2 段階最小二乗法

x_i が内生変数で，z_i が x_i と相関し，かつ u_i と相関していない場合，

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

を推定する際，2 段階最小二乗法（2-Stage Least Squares, 2SLS）を用いれば，同時方程式バイアスを緩和できる．

- z_i を操作変数として用いる．
- 2SLS は操作変数法（Instrumental Variable Method）の 1 つと考えることができる．
- 2SLS などの操作変数法は，観測値数が十分大きいときに使われる．

第 1 段階

- 内生説明変数を，システムに登場する全ての外生変数に回帰．つまり，

$$x_i = \underbrace{\pi_0 + \pi_1 z_i}_{u_i\text{と無相関}} + \underbrace{v_i}_{u_i\text{と相関}}$$

  を OLS で推定．
  - z_i は操作変数なので定義上，u_i と無相関．
  - π_0 は定数項なので変動せず，u_i と無相関．

  ⇒ x_i の変動を，u_i と無相関な部分と相関する部分に分割．
- x_i の予測値 $\hat{x}_i = \hat{\pi}_0 + \hat{\pi}_1 z_i$ を求める．

  ⇒ x_i の変動のうち，u_i と無相関な部分を抽出．

※内生説明変数が複数個あれば，各内生説明変数に対しこの作業を行う．

第 2 段階

- 推定したい式の内生説明変数 x_i を，第 1 段階で求めた予測値 $\hat{x}_i$ に変更した式

$$y_i = \beta_0 + \beta_1 \hat{x}_i + u_i$$

を OLS で推定.
    - $\hat{x}_i$ は u_i と無相関.

- β_1 の 2SLS 推定量は,

$$\hat{\beta}_{1,2SLS} = \frac{\sum_{i=1}^{n} \left(\hat{x}_i - \bar{\hat{x}}_i\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n} \left(\hat{x}_i - \bar{\hat{x}}_i\right)^2}.$$

※内生説明変数が複数個ある場合の係数ベクトルの 2SLS 推定量は，$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'\boldsymbol{y}$.

識別の次数条件

「推定したい式に含まれる内生説明変数の個数（G）」と，「推定したい式に含まれない外生変数の個数（K）」の大小関係によって，その式が 2SLS などで推定できる場合とできない場合がある．

- $G > K$
  - 過少識別（識別不能，under-identified）
  - 推定できない（操作変数を使わず無理に OLS 推定をするとバイアス発生）．
- $G = K$
  - ちょうど識別（just-identified）
  - 2SLS などで推定できる．
- $G < K$
  - 過剰識別（over-identified）
  - 2SLS などで推定できる．

$\Rightarrow G \leq K$ なら 2SLS などで推定できる．

みかんの需要・供給モデル

みかんの需要関数と供給関数をそれぞれ,

$$\text{需要}: q_{it} = \beta_{D0} + \beta_{DP} p_{it} + \sum_{m=2}^{9} \beta_m d_{mi} + u_{Dit},$$

$$\text{供給}: p_{it} = \beta_{S0} + \beta_{SQ} q_{it} + \beta_T t + \beta_{TT} t^2 + u_{Sit}.$$

- q_{it} : 取引数量
- p_{it} : 価格
- d_{mi} : 各市場ダミー
- i : 市場番号
- t : 月（時点番号）

とする.

この定式化における仮定

- 需要曲線の切片：各市場で異なる.
- 需要曲線の傾き：全 9 市場・各月で共通.

$$\Downarrow$$

需要曲線の位置は各市場で異なる，と仮定する（消費者の数などの条件が各市場で異なるから）.

- 供給曲線の切片：各月で異なる.
- 供給曲線の傾き：全 9 市場・各月で共通.

$$\Downarrow$$

供給曲線の位置は各月で異なる，と仮定する（季節・気象条件がみかんの限界費用に影響するから）.

内生変数と外生変数

- 内生変数
  - 取引数量
  - 価格
- 外生変数
  - 各市場ダミー
    - 市場ごとの消費者の数などの条件を反映
    - 需要曲線のシフト要因
  - 月
    - 供給曲線のシフト要因
  - 月の二乗
    - 価格に対して二次関数の関係である可能性を考慮
    - 供給曲線のシフト要因

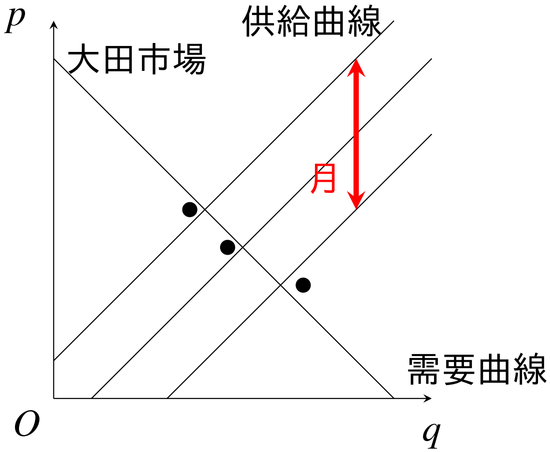
識別の次数条件の確認

- 需要関数の説明変数に含まれる内生変数：1 個
  - 価格 p_i
- 需要関数に含まれない外生変数：2 個
  - 月 t
  - 月の二乗 t^2

$$\Downarrow$$

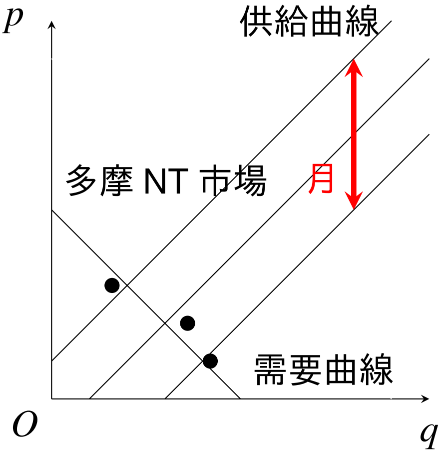
「需要関数の説明変数に含まれる内生変数」<「需要関数に含まれない外生変数」なので過剰識別.
$\Rightarrow$ 識別の次数条件を満たす.
$\Rightarrow$ 2SLS などで推定できる.
（この定式化の場合，供給関数も識別の次数条件を満たす.）

2SLS 推定のイメージ



需要曲線のシフト要因を固定し，月で供給曲線をシフトさせて需要曲線を炙り出す.

2SLS 推定のイメージ



需要曲線のシフト要因を固定し，月で供給曲線をシフトさせて需要曲線を炙り出す.

2SLS における決定係数

- モデルの説明変数と誤差項に相関がある場合，決定係数も自由度修正済み決定係数も適切に定義できない（証明は省略）.
- 2SLS を用いる目的
➡ モデルの説明変数と誤差項に相関がある場合に，より厳密な係数推定値を得るため（モデルの当てはまりの良さを高めるためではない.）

$$\Downarrow$$

2SLS の第 2 段階推定における R^2 や $\bar{R}^2$ は，解釈ができない.

- 参考：Wooldridge, J.M., 2013. *Introductory Econometrics: A Modern Approach*. Fifth ed., South-Western, Mason, OH, USA, p.523.

gretl での変数の二乗の作成方法

1. gretl の画面上で，二乗したい変数をクリックして選択.
2. gretl のメニューバーから「追加」→「選択された変数の二乗」と操作.
    - 新たに作成された，二乗の変数の名前の頭には sq_が付けられる.
    - sq は square（二乗）の略.
3. gretl のメニューバーから「ファイル」→「データを保存」と操作し，必ずデータセットを上書き保存.

gretl での 2 段階最小二乗法

メニューバーから「モデル」→「操作変数法」→「2 段階最小二乗法」と操作し，

- 「従属変数」には推定したい式の被説明変数を，
- 「説明変数（回帰変数）」には推定したい式の右辺に含まれる説明変数（内生変数と外生変数両方）を，
- 「操作変数」にはシステムに登場する全ての外生変数を，

それぞれ選び，標準誤差を設定して「OK」をクリックすればよい．

ただし，第 1 段階推定（内生説明変数を，システムに登場する全ての外生変数に回帰）の結果は表示されない.

$$\Downarrow$$

第 1 段階推定は，メニューバーから「モデル」→「通常の最小二乗法」と操作し，

- 「従属変数」には推定したい式の内生説明変数を，
- 「説明変数（回帰変数）」にはシステムに登場する全ての外生変数を，

それぞれ選んで結果を出力すればよい.

例えば，実習で推定するみかんの需要曲線の場合は，

- 推定したい式の被説明変数
  - 取引数量
- 推定したい式の右辺に含まれる説明変数
  - 価格（内生）
  - 各市場ダミー（外生）
- システムに登場する全ての外生変数
  - 各市場ダミー
  - 月
  - 月の二乗

実習 1

まず，月の二乗の変数を作成する．

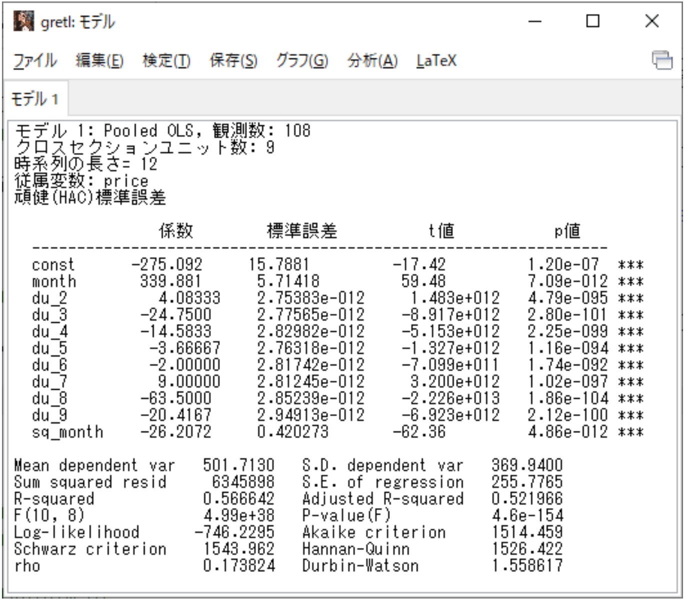
1. gretl を起動．
2. 「ファイル」→「データを開く」→「ユーザー・ファイル」と操作．
3. orangetokyo.gdt を選択し，「開く」をクリック．
4. gretl の画面上で，month をクリックして選択．
5. gretl のメニューバーから「追加」→「選択された変数の二乗」と操作．
   - 15 番目の変数として，「sq_month」が出現し，それが「month の二乗」という定義の変数となる．
6. メニューバーから「ファイル」→「データを保存」と操作．これでデータセットが上書き保存される．

実習 2

先に説明した定式化の下で，みかんの需要関数を 2SLS で推定する．gretl で 2 段階最小二乗法の操作をすると，第 2 段階推定の結果のみが表示されるので，その前に第 1 段階推定（内生説明変数を，システムに登場する全ての外生変数に回帰）の結果を表示させておく．

1. gretl のメニューバーから「モデル」→「通常の最小二乗法」と操作．
2. 出てきた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 price（みかんの価格）となる．

3. 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 month，du_2，du_3，du_4，du_5，du_6，du_7，du_8，du_9，sq_month 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 など，他の変数はクリックしない．
    - 推定式の右辺の変数（説明変数，独立変数）が month（月）と，du_2 から du_9（築地を除く 8 市場のダミー変数）と，sq_month（月の二乗）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．
4. 「頑健標準誤差を使用する」にチェック．
    - 誤差項のクラスター構造に対して頑健な，Arellano の標準誤差が計算され，推定式の誤差項 u_i の分散に関する仮定が誤っていても，より厳密な分析ができるようになる．
5. 「OK」をクリックすると，結果が表示される．




```
モデル 1: Pooled OLS, 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: price
頑健(HAC)標準誤差

               係数        標準誤差          t値             p値
  ----------------------------------------------------------------
  const     -275.092      15.7881        -17.42          1.20e-07  ***
  month      339.881      5.71418         59.48          7.09e-012 ***
  du_2         4.08333    2.75383e-012    1.483e+012     4.79e-095 ***
  du_3       -24.7500     2.77565e-012   -8.917e+012     2.80e-101 ***
  du_4       -14.5833     2.82982e-012   -5.153e+012     2.25e-099 ***
  du_5        -3.66667    2.76318e-012   -1.327e+012     1.16e-094 ***
  du_6        -2.00000    2.81742e-012   -7.099e+011     1.74e-092 ***
  du_7         9.00000    2.81245e-012    3.200e+012     1.02e-097 ***
  du_8       -63.5000     2.85239e-012   -2.226e+013     1.86e-104 ***
  du_9       -20.4167     2.94913e-012   -6.923e+012     2.12e-100 ***
  sq_month   -26.2072     0.420273       -62.36          4.86e-012 ***

Mean dependent var   501.7130   S.D. dependent var   369.9400
Sum squared resid     6345898   S.E. of regression   255.7765
R-squared            0.566642   Adjusted R-squared   0.521966
F(10, 8)             4.99e+38   P-value(F)           4.6e-154
Log-likelihood      -746.2295   Akaike criterion     1514.459
Schwarz criterion    1543.962   Hannan-Quinn         1526.422
rho                  0.173824   Durbin-Watson        1.558617
```

このような画面が表示されれば成功.「gretl: モデル」のウィンドウはまだ閉じない！

指数表示

gretl では数値の桁数や小数点以下桁数が大きい場合，指数で表示される．

- 例えば 1.483e+012 は，

$$1.483 \times 10^{12}$$

  という意味．
  - 変数 du_2（大田市場ダミー）の係数の t 値は 1.483×10^{12}.
- 例えば 4.79e-095 は，

$$4.79 \times 10^{-95}$$

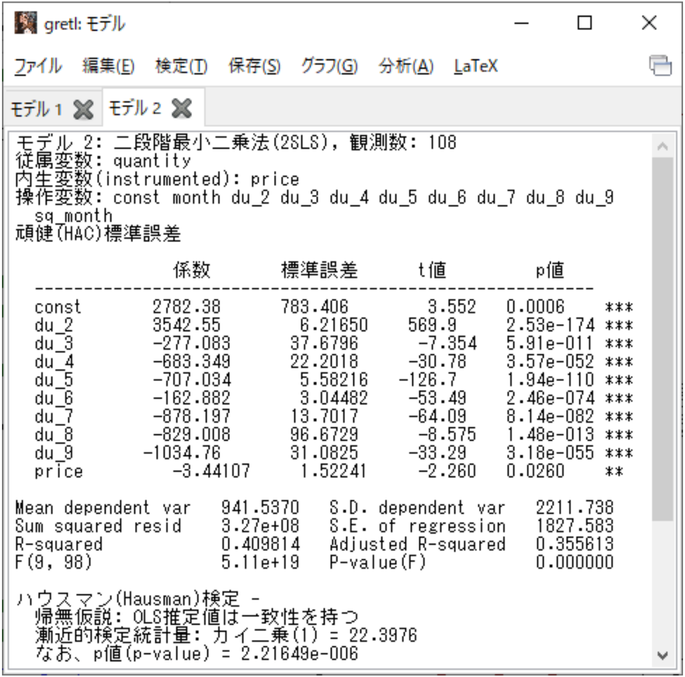
  という意味．
  - e.g., 変数 du_2（大田市場ダミー）の係数の p 値は 4.79×10^{-95}.

続いて，第 2 段階推定（推定したい式の内生説明変数を，第 1 段階で求めた予測値に変更した式を OLS で推定）の結果を表示させる．

6. gretl のメニューバーから「モデル」→「操作変数法」→「2 段階最小二乗法」と操作．
7. 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，5 つの矢印のうち一番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．
8. 説明変数（回帰変数）のリストに入っている変数のうち，month と sq_month の 2 つをドラッグして選択し，5 つの矢印のうち上から 3 番目の赤い左向き矢印をクリック．
   - 選択した変数が説明変数リストから消去される．

9. ウィンドウ左側の変数リストにある price をクリックし，5 つの矢印のうち上から 2 番目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 du_2 から du_9（築地を除く 8 市場のダミー変数）と，price（みかんの価格）となる．
10. 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 month，du_2，du_3，du_4，du_5，du_6，du_7，du_8，du_9，sq_month をクリックし，5 つの矢印のうち上から 4 番目の緑の右向き矢印をクリック．du_1 など，他の変数はクリックしない．
    - システムに登場する全ての外生変数が，month（月）と，du_2 から du_9（築地を除く 8 市場のダミー変数）と，sq_month（月の二乗）と定義される．

11. 「頑健標準誤差を使用する」にチェック.
   - 誤差項のクラスター構造に対して頑健な，Arellano の標準誤差が計算され，推定式の誤差項 u_i の分散に関する仮定が誤っていても，より厳密な分析ができるようになる.
12. 「OK」をクリックすると，結果が表示される.




```
モデル 2: 二段階最小二乗法(2SLS), 観測数: 108
従属変数: quantity
内生変数(instrumented): price
操作変数: const month du_2 du_3 du_4 du_5 du_6 du_7 du_8 du_9
  sq_month
頑健(HAC)標準誤差

                係数          標準誤差       t値          p値
  -------------------------------------------------------------
  const        2782.38       783.406        3.552    0.0006     ***
  du_2         3542.55         6.21650    569.9      2.53e-174  ***
  du_3         -277.083       37.6796      -7.354    5.91e-011  ***
  du_4         -683.349       22.2018     -30.78     3.57e-052  ***
  du_5         -707.034        5.58216   -126.7      1.94e-110  ***
  du_6         -162.882        3.04482    -53.49     2.46e-074  ***
  du_7         -878.197       13.7017     -64.09     8.14e-082  ***
  du_8         -829.008       96.6729      -8.575    1.48e-013  ***
  du_9        -1034.76        31.0825     -33.29     3.18e-055  ***
  price          -3.44107      1.52241     -2.260    0.0260     **

Mean dependent var   941.5370   S.D. dependent var   2211.738
Sum squared resid    3.27e+08   S.E. of regression   1827.583
R-squared            0.409814   Adjusted R-squared   0.355613
F(9, 98)             5.11e+19   P-value(F)           0.000000

ハウスマン(Hausman)検定 -
  帰無仮説: OLS推定値は一致性を持つ
  漸近的検定統計量: カイ二乗(1) = 22.3976
  なお、p値(p-value) = 2.21649e-006
```

第 1 段階推定結果

- 月の係数
  - 339.881
  - 有意水準 1%で，係数ゼロの H_0 棄却．
- 月の二乗の係数
  - −26.2072
  - 有意水準 1%で，係数ゼロの H_0 棄却．

⇒「需要関数に含まれない外生変数」とした「月」も「月の二乗」も，「需要関数の説明変数に含まれる内生変数」の「価格」と統計的に有意に相関している．

⇒「月」も「月の二乗」も操作変数として機能している可能性がある．

- より一般的な判断方法は，後の授業で説明する．

第 2 段階（需要関数）推定結果

- 価格の係数
  - –3.44107（符号は負）
  - 有意水準 5%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，取引数量は平均して 3.44107t 減少する傾向がある．
    ⇒ 経済理論と整合的．
- OLS と 2SLS での，価格の係数推定値の違い
  - OLS（前回の授業で推定）では，価格の係数は –1.82343.
    ➥ 価格と数量の相互依存関係により，OLS 推定値に同時方程式バイアスが生じた（需要関数と供給関数を混合したものが推定された）可能性がある．
  - このバイアスが生じているかを検定する方法は，後の授業で説明する．

補足

- 「冬になるとみかんを食べたくなる」消費者がみかん市場の多数派である場合，需要曲線も「月」によってシフトしてしまう．
  ➡ 「月」は需要曲線の説明変数にも含まれる．
  ➡ その場合，「月」は「推定したい式の説明変数に含まれない外生変数」にならないので操作変数として使えない．
  - 参考：鹿野繁樹（2015）『新しい計量経済学ーデータで因果関係に迫る』日本評論社．
- 企業・団体・事業者レベルのデータを用いて需要関数や供給関数を推定する場合，生産要素価格（電力価格，職員賃金など）を供給曲線のシフト要因（需要関数の価格に対する操作変数）として使うことが考えられる．