第 7 回：同時方程式モデル推定後の検定

北村　友宏

2020 年 11 月 13 日

本日の内容

1. 弱操作変数の検定
2. ハウスマン検定
3. gretl を用いた実習

弱操作変数

- 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には，例えば操作変数を用いて 2SLS などで推定する．
- 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，ここでは z_i とする．
    - $\mathrm{Cov}(z_i, u_i) = 0$ かつ $\mathrm{Cov}(z_i, x_i) \neq 0$.

- 操作変数 z_i と内生説明変数 x_i との相関が弱ければ，つまり $\mathrm{Cov}(z_i, x_i) \approx 0$ であれば，その操作変数を用いて 2SLS などで推定しても同時方程式バイアスは緩和されず，かえって OLS よりも強いバイアスが生じることもある．
- 内生説明変数との相関が弱い操作変数を**弱操作変数（weak instrument）**という．

$$\Downarrow$$

2SLS などを用いる際には，用いた操作変数が弱操作変数になっていないかを検定する必要がある．

弱操作変数の検定

ここでは，推定したい式に含まれる内生説明変数が1つである場合について説明する．

x_i を内生変数，z_{1i}, z_{2i}, z_{3i} を外生変数として，

$$y_i = \beta_0 + \beta_X x_i + \beta_Z z_{1i} + u_i$$

を 2SLS で推定することを考える．

$\Rightarrow$ x_i に対する操作変数は z_{2i} と z_{3i}.

- 2SLS の第 1 段階推定では，内生説明変数 x_i を，システムに登場する全ての外生変数に回帰．つまり，

$$x_i = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + v_i$$

  を OLS で推定．

- この場合，推定後，

$$H_0 : \pi_2 = 0 \text{ and } \pi_3 = 0 \quad \text{vs} \quad H_1 : \pi_2 \neq 0 \text{ or } \pi_3 \neq 0$$

  を検定すれば，操作変数 z_{2i}, z_{3i} と内生説明変数 x_i の相関を検定できる．
  - 一般的には，第 1 段階推定において，「全操作変数の係数がどれもゼロ」を帰無仮説，「少なくとも 1 つの操作変数の係数がゼロでない」を対立仮説として検定する．
- 複数の係数に関する仮説検定なので，F 検定を行う．
  - 操作変数が 1 つの場合も F 検定は可能．

Staiger=Stock の F 検定

内生説明変数が 1 つの場合，F 検定を行う際に計算する「第 1 段階推定の F 統計量の実現値（F statistic）」を見て，

第 1 段階推定の F 統計量が 10 を超えていれば

用いた操作変数が操作変数として機能している（弱操作変数の問題は発生していない）可能性がある，と判断するのが通例.

- p 値を見て統計的有意性を確認するのではない.
- 第 1 段階推定の F 統計量の実現値が 10 未満の場合，用いた操作変数は操作変数として機能していない可能性がある.

ハウスマン検定

- 一般的に，2SLS 推定量は OLS 推定量に比べ分散が大きい.

  $$\Downarrow$$

  2SLS 推定が不要な場合（推定式の説明変数と誤差項が無相関）は OLS を用いるのが望ましい.

  $$\Downarrow$$

  （OLS ではなく）2SLS などを用いる際には，その妥当性の有無を確認する.

- 「推定式の内生説明変数と誤差項が相関していない」ことを帰無仮説とする検定をハウスマン検定（Hausman test）という.

ハウスマン検定の考え方

x_i が内生変数で，x_i に対する操作変数 z_i が x_i と相関し，かつ u_i と相関していないとして，

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

を 2SLS で推定することを考える．

- 2SLS の第 1 段階推定では，内生説明変数 x_i を，システムに登場する全ての外生変数に回帰．つまり，

$$x_i = \underbrace{\pi_0 + \pi_1 z_i}_{u_i\text{と無相関}} + \underbrace{v_i}_{u_i\text{と相関}} \tag{2}$$

  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．

- 2SLS の第 1 段階推定では，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.
  ※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.

- 「u_i と相関する部分」である v_i は，第 1 段階推定式の誤差項（観測不能）.
  ➡ 第 1 段階推定式 (2) の残差 $\hat{v}_i$（v_i の推定値）を，推定したい元の式 (1) の説明変数に加えると，

$$y_i = \beta_0 + \beta_1 x_i + \underbrace{\delta \hat{v}_i + w_i}_{=u_i} . \tag{3}$$

  $\Rightarrow \delta \hat{v}_i$ は，元の式の誤差項 u_i の一部.

- v_i が u_i と本当に相関するなら，u_i と v_i は関係性を持ち，「v_i の推定値 $\hat{v}_i$」の係数 δ がゼロでないはず.

- 「v_i の推定値 $\hat{v}_i$」の係数 δ がゼロなら，u_i と v_i は関係性を持っていない（無相関）.
  ➡ x_i の変動全体が u_i と無相関.
  ➡ 「元の式の説明変数 x_i」と「元の式の誤差項 u_i」は無相関（元の式 (1) は OLS で一致推定できる）.

- 第 1 段階推定式の残差を元の式の説明変数に加えた (3) を推定，つまり y_i を x_i と $\hat{v}_i$ に回帰し，

$$H_0 : \delta = 0 \quad \text{vs} \quad H_1 : \delta \neq 0$$

を検定すれば，x_i と u_i の相関を検定できる．
    - H_0 棄却 ⇒ 元の式の推定には 2SLS などを使う必要あり（OLS ではバイアス発生）．
    - H_0 採択 ⇒ 元の式の推定に 2SLS などを使う必要性が支持されない．⇒OLS を用いる．
    - (3) の誤差項は w_i であり，u_i だった部分のうち，x_i と相関する（とされる）部分は全て $\hat{v}_i$ に吸収．
      ➥ (3) は OLS で推定できる．
    - 内生説明変数が複数個あれば，各内生説明変数を全ての外生変数に回帰してそれぞれの残差を求め，それら全ての残差を元の式の説明変数に加えて OLS 推定し，「全ての残差の係数がどれも 0」を検定する．

gretl での弱操作変数の検定とハウスマン検定

- 2SLS での推定（メニューバーから「モデル」→「操作変数法」→「2 段階最小二乗法」）を実行すると，出力結果の，

  「弱操作変数 (weak instrument) の検定」

  の部分に，第 1 段階推定の F 統計量の実現値などが表示される．

- さらに，出力結果の，

  「ハウスマン (Hausman) 検定」

  の部分に，ハウスマン検定統計量とその p 値が表示される．

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,

$$
\begin{aligned}
\text{需要}: q_{it} &= \beta_{D0} + \beta_{DP} p_{it} + \sum_{m=2}^{9} \beta_m d_{mi} + u_{Dit}, \\
\text{供給}: p_{it} &= \beta_{S0} + \beta_{SQ} q_{it} + \beta_T t + \beta_{TT} t^2 + u_{Sit}.
\end{aligned}
$$

- 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.

実習 1

みかんの需要関数を 2SLS で推定し，弱操作変数の検定とハウスマン検定の結果を確認する．その前に，復習として第 1 段階推定の結果を表示させる．

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 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．

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 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．
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．

gretl: モデル

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

```
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

                 係数         標準誤差        t値          p値
  --------------------------------------------------------------
  const        -275.092      112.298      -2.450     0.0161      **
  month         339.881       31.1652     10.91      1.50e-018  ***
  du_2            4.08333    104.420       0.03910   0.9689
  du_3          -24.7500     104.420      -0.2370    0.8131
  du_4          -14.5833     104.420      -0.1397    0.8892
  du_5           -3.66667    104.420      -0.03511   0.9721
  du_6           -2.00000    104.420      -0.01915   0.9848
  du_7            9.00000    104.420       0.08619   0.9315
  du_8          -63.5000     104.420      -0.6081    0.5445
  du_9          -20.4167     104.420      -0.1955    0.8454
  sq_month      -26.2072       2.33374   -11.23      3.05e-019  ***

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, 97)            12.68335   P-value(F)           7.54e-14
Log-likelihood      -746.2295   Akaike criterion     1514.459
Schwarz criterion    1543.962   Hannan-Quinn         1526.422
rho                  0.173824   Durbin-Watson        1.558617
```

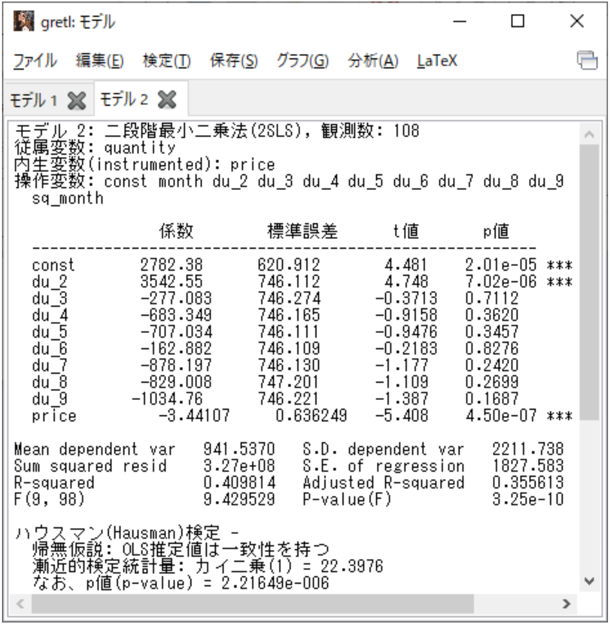
このような画面が表示されれば成功.

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．

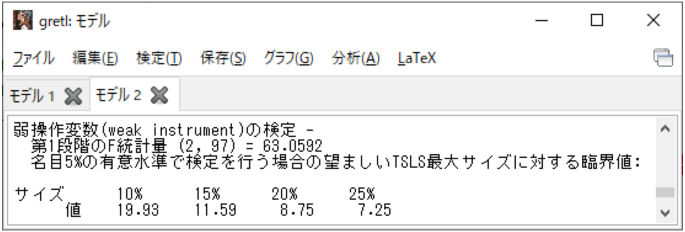
9. gretl のメニューバーから「モデル」→「操作変数法」→「2 段階最小二乗法」と操作．
10. 出てきたウィンドウ左側の変数リストにある quantity をクリックし，5 つの矢印のうち一番上の青い右向き矢印をクリック．
    - 推定式の左辺の変数（被説明変数，従属変数）が quantity（みかんの取引数量）となる．
11. 説明変数（回帰変数）のリストに入っている変数のうち，month と sq_month の 2 つをドラッグして選択し，5 つの矢印のうち上から 3 番目の赤い左向き矢印をクリック．
    - 選択した変数が説明変数リストから消去される．

12. 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．
13. 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．

14. 「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．
15. 「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.



弱操作変数の検定結果は，下のほうに表示されている．

「gretl: モデル」のウィンドウはまだ閉じない！

弱操作変数の検定結果

- 第 1 段階の F 統計量の実現値
  - 63.0592（10 を大きく超えている）
  - 「第 1 段階の F 統計量 (2, 97)」のカッコ内は F 分布の自由度.

$\Rightarrow$ 用いた操作変数（「月」と「月の二乗」）は，内生説明変数（価格）と強く相関しており，操作変数として機能している（弱操作変数でない）可能性がある.

ハウスマン検定結果

- カイ二乗統計量の実現値
  - 22.3976（「漸近的検定統計量」に表示）
  - 「カイ二乗 (1)」のカッコ内はカイ二乗分布の自由度.
  - 帰無仮説の「OLS 推定値は一致性を持つ」とは，「(2SLS を使わず）OLS で推定しても観測値数が十分に大きければ真の係数に近い推定値が得られる」という意味.
  - p 値は 2.21649×10^{-6}.
  - 有意水準 1%で，需要関数の係数の OLS 推定値は一致性を持つ（第 1 段階推定の残差の係数はゼロ，需要関数の価格と誤差項は無相関）という H_0 棄却.

⇒ 需要関数の説明変数である価格と需要関数の誤差項が相関している可能性があり，2SLS による推定が支持される.

レポート・論文での弱操作変数の検定結果とハウスマン検定結果の提示

- 弱操作変数の検定結果
  - 第 1 段階の F 統計量の実現値を，推定結果の表中または本文中で示し，それが 10 を超えていることを説明すればよい．
- ハウスマン検定結果
  - 検定統計量の実現値と，その p 値を，推定結果の表中または本文中で示せばよい．

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.