操作変数と2段階最小二乗法

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
- $x_1 = \pi_0 + \pi_1 z + \pi_2 x_2 + \ldots + \pi_k x_k + v$

Ch.15 操作変数と2段階最小二乗法

15.1 モチベーション：欠落変数

- 内生性は社会科学/経済学固有の問題
  - 重要な個人的変数はほぼ観察不可能
    - これらはしばしば観察可能な説明変数情報と相関
  - 測定誤差も内生性につながる可能性
  - これまでの内生性問題の解決策：
    - 欠落変数を代理変数で代替（Ch.9）
    - 固定効果（ただし以下のケースのみ、Ch.14）
      - パネルデータ利用可
      - 内生性が時間に一定
      - 説明変数が時間に一定

15.1 モチベーション：欠落変数

- 操作変数法Instrumental variables method, IV
  - 例：賃金関数での教育収益率

    $\log(wage_i) = \beta_0 + \beta_1 educ_i + u_i$

    - 誤差項に教育と相関する先天的能力を含む
  - 操作変数の定義：
    1. 回帰式に含まれない（yへの部分効果を持たない）
    2. 内生変数xと高い相関
    3. 誤差項uと無相関
- 単回帰OLSでの再検討

  $y_i = \beta_0 + \beta_1 x_i + u_i$ & <u>仮定</u> $Cov(x_i, u_i) = 0$

15.1 モチベーション：欠落変数

- 外生性のもとでのOLSの一致性証明：

  $Cov(x_i, u_i) = 0$ （外生性）

  $\Leftrightarrow \; 0 = Cov(x_i, y_i - \beta_0 - \beta_1 x_i) = Cov(x_i, y_i) - \beta_1 Var(x_i)$

  $\Leftrightarrow \; \beta_1 = Cov(x_i, y_i)/Var(x_i)$

  $\Rightarrow \; \hat{\beta}_1 = \widehat{Cov}(x_i, y_i)/\widehat{Var}(x_i) \; \rightarrow \; Cov(x_i, y_i)/Var(x_i) = \beta_1$

  - 大数の法則LLNのように、$n\rightarrow\infty$に伴い標本の分散共分散が理論値に収束するようなデータであれば、一致性は保持
  - <u>xが外生的であればOLSEは基本的に一致性を満たす</u>

15.1 モチベーション：欠落変数

- 操作変数zの存在を仮定：

  $\boxed{Cov(z_i, u_i) = 0}$ （ただし $Cov(x_i, u_i) \neq 0$ ）

  - 操作変数は誤差項uと無相関（内生変数xと相関）

  $\Leftrightarrow \; 0 = Cov(z_i, y_i - \beta_0 - \beta_1 x_i) = Cov(z_i, y_i) - \beta_1 Cov(z_i, x_i)$

  $\Leftrightarrow \; \beta_1 = Cov(z_i, y_i)/Cov(z_i, x_i)$

  $\Rightarrow \; \hat{\beta}_{IV} = \widehat{Cov}(z_i, y_i)/\widehat{Cov}(z_i, x_i) \; \rightarrow \; Cov(z_i, y_i)/\boxed{Cov(z_i, x_i)} = \beta_1$

  - 操作変数は説明変数xと相関
- 操作変数推定量

  $$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$$

15.1 モチベーション：欠落変数

◆例15.1：操作変数「父親の教育年数」

- OLS推定

$$\widehat{\log}(wage) = -.185 + .109\, educ$$
$$\qquad\quad (.185) \quad (.014)$$
$$n = 428,\ R^2 = .118$$

  - （欠落変数バイアスのため）教育収益率は恐らく過大推定

- IV推定
  - 「父親の教育」変数はIVになり得るか？
    1. （元の式の）説明変数ではないこと
    2. 「（本人の）教育」変数と確かな相関があること
    3. （元の式の）誤差項と無相関であること*

15.1 モチベーション：欠落変数

◆例15.1：操作変数「父親の教育年数」（続き）

- IV（補助）推定
  - 父親の教育はIVになり得るか？

$$\widehat{educ} = 10.24 + .269\, \boxed{fatheduc}$$
$$\qquad (.28) \quad (.029) \qquad n = 428,\ R^2 = .173$$

    - 「父親の教育」変数は有意

- IV推定

$$\widehat{\log}(wage) = .441 + .059\, educ$$
$$\qquad (.446) \quad (.035)$$
$$n = 428,\ R^2 = 1 - RSS_{IV}/TSS = .093$$

  - 教育収益率はOLS推定値より減少
  - 標準誤差が増大→推定精度が低下

15.1 モチベーション：欠落変数

◆よく使用されるその他の「教育」の操作変数：

- 兄弟の数（→例15.2）
  - ①賃金決定要因でない、②家計の資源制約により教育と相関、③先天的能力と無相関
- 16歳時の大学の近在（→例15.4）
  - ①賃金決定要因でない、②大学が近在であれば教育増大の可能性、③先天的能力と無関係かは微妙
- 生年月日
  - ①賃金決定要因でない、②義務教育のために教育との相関、③誤差項と無関係

15.1 モチベーション：欠落変数

◆IV推定法の特性

- 操作変数が「非外生的」&「xと弱い相関」の場合、IVはOLSより非一致性の可能性（→例15.3）
  - OLS: $plim\ \hat{\beta}_{1,OLS} = \beta_1 + Corr(x,u) \cdot \frac{\sigma_u}{\sigma_x}$
  - IV: $plim\ \hat{\beta}_{1,IV} = \beta_1 + \frac{\boxed{Corr(z,u)}}{Corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x}$
    - 操作変数が外生的であれば問題ないが、そうでなければ漸近的バイアスはxとの相関が弱いほど大きくなる。
    - IVがOLSより悪化：

$$\frac{Corr(z,u)}{Corr(z,x)} > Corr(x,u) \qquad 例\ \frac{0.03}{0.2} > 0.1$$

15.2 重回帰モデルのIV推定

◆重回帰モデルのIV推定

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 \boxed{z_1} + \ldots + \beta_k \boxed{z_{k-1}} + u_1$$

内生変数　外生変数

- 操作変数z_kの条件
  1. 元の回帰式に現れない
  2. 誤差項と無相関
  3. 内生的説明変数と相関
- 誘導形reduced form regression

$$y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

  - 内生的説明変数をすべての外生変数に回帰する式において、操作変数の係数は非ゼロでなければならない。

15.2 重回帰モデルのIV推定

◆重回帰モデルのIV推定：

- 前提条件
  - 操作変数の外生性　$Cov(z_j, u_1) = 0,\ j = 1, \ldots, k$
  - 誤差項の平均ゼロ　$E(u_1) = 0$
- 両条件の標本表現

$$n^{-1}\sum_{i=1}^{n} z_{ij}\hat{u}_{i1} = \widehat{Cov}(z_j, \hat{u}_1) = 0,\ j = 1, \ldots, k$$

$$n^{-1}\sum_{i=1}^{n}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \ldots - \hat{\beta}_k z_{ik-1}) = n^{-1}\sum_{i_1}^{n}\hat{u}_{i1} = 0$$

  - ここではk+1本の方程式& k + 1個の推定値$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$

15.3 2段階最小二乗法2SLS

◆2段階最小二乗法(2SLS)

- IV推定量の直感的解釈は以下の手順と同じ
  - 構造形(元のモデル)

    $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + u_1$
  - 誘導形(第1段階)

    $\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \ldots + \hat{\pi}_k z_{k-1} + \hat{\pi}_k z_k$
    - 内生説明変数y_2を<u>外生変数情報のみ</u>で予測値を算出
    - z_kは新たに付加された外生変数(=操作変数)
  - 第2段階

    $y_1 = \beta_0 + \beta_1 \hat{y}_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + error$
    - 内生変数y_2を<u>外生変数情報のみの予測値で代替し</u>OLS推定

15.3 2段階最小二乗法2SLS

◆2段階最小二乗法はなぜ有効なのか？

- 第2段階回帰式の全変数は外生的
  - 内生的なy_2を外生的要素のみの予測値に置換
    - この置換によりy_2は(誤差項に相関する)内生的部分を除去

◆2段階最小二乗法の特性

- 実は第2段階回帰式のOLS標準誤差は間違い
  - 正しい標準誤差の計算は容易
- 2SLSは複数の内生&操作変数の場合も使用可
  - 内生&操作変数が1つずつの場合、2SLS = IV

15.3 2段階最小二乗法2SLS

◆例15.5:就業女性の教育収益率

- 第1段階(教育を全外生変数に回帰)

$$\widehat{educ} = \underset{(.27)}{8.37} + \underset{(.026)}{.085}\, exper - \underset{(.001)}{.002}\, exper^2 + \underset{(.024)}{.185}\, fatheduc + \underset{(.026)}{.186}\, motheduc$$

  - 両親の教育は本人の教育と有意に(偏)相関
- 2SLS推定結果

$$\widehat{log}(wage) = \underset{(.400)}{.048} + \underset{(.031)}{\boxed{.061}}\, educ + \underset{(.013)}{.044}\, exper - \underset{(.0004)}{.0009}\, exper^2$$

$$n = 428, R^2 = 0.136$$

  - 教育収益率はOLS推定値(10.8%)より低い

15.4 測定誤差と2SLS/IV

◆Error-in-variable

  - 観測変数=真の値+測定誤差 $x_1 = x_1^* + e_1$
  - 母集団回帰 $y = \beta_0 + \beta_1 x_1^* + \ldots + \beta_k x_k + u$
  - 標本回帰⇒ $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + (u - \beta_1 e_1)$

    $\Rightarrow Cov(x_1, u - \beta_1 e_1) = -\beta_1 Cov(x_1, e_1) = -\beta_1 \sigma_{e_1}^2$
    - MLR.4に反するため、OLSEは不偏性&一致性を満たさない
- 2SLSでの対処法
  - 第2の観測変数$z = x_1^* + e_z$
    - ここで測定誤差e_zはe_1やuと無相関
  - 真の変数x_1^*の第2の観測変数zが利用可能な場合、変数x_1の操作変数として使用可能

15.5 内生性検定と過剰識別制約

◆内生性の検定

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + u_1$

    - 説明変数y_2:内生性が疑われる変数
  - 誘導形方程式

    $y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_{k-1} + \pi_k z_k + v_2$
    - v_2とu_1に相関がなければ、変数y_2は外生的

    $u_1 = \delta_1 v_2 + e_1$
    - 「v_2とu_1は無相関」→「係数δ_1はゼロ」であればy_2は外生的
- 検定式

  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \ldots + \beta_k z_{k-1} + \delta_1 \hat{v}_2 + e_1$
    - 係数δ_1が有意にゼロと異なれば、「y_2は外生的」という仮説を棄却

15.6-8 2SLSの応用

◆2SLS/IV推定の統計的特性

- OLSとほぼ同じ仮定のもとで、2SLS/IV推定量は一致性&漸近的正規分布
  - 変数xより変数zに対する諸条件
- 2SLS/IVの推定精度はOLSにかなり劣る
  - 操作変数の存在により、第2段階回帰式ではより多重共線的&より小さい説明変数の変動性
- 不均一分散/系列相関の対処法はOLSと同じ
- 2SLS/IVは時系列&パネルデータに容易に拡張