

10 説明変数と攪乱項の相関

10.1 確率的な説明変数

・ 仮定 1 「説明変数 X は非確率変数」 → 「説明変数 X は確率変数」

(1) X と u が独立である場合

(2) X と u が従属である場合 a. $Cov(X, u) = 0$

b. $Cov(X, u) \neq 0$

・ 不偏性
$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \beta + \frac{\sum (X_i - \bar{X})u_i}{\sum (X_i - \bar{X})^2} = \beta + \sum w_i u_i$$

仮定 1 のとき、 $E(\hat{\beta}) = E(\beta + \sum w_i u_i) = \beta + \sum w_i E(u_i) \Leftrightarrow \beta + E(\sum w_i u_i)$

仮定 3 のとき、 $E(\hat{\beta}) = \beta + \sum w_i \cdot 0 = \beta$

→ 仮定 1 と 3 が成立したときに、OLS 推定量は不偏性を持つ

・ X と u が独立 X と u は個別の確率変数に従うので、

仮定 3 「誤差項の期待値 $E(u_i) = 0$ 」 → 「 X を所与とする誤差項の期待値

$E(u_i | X_i) = 0$ 」と考えれば、OLS 推定量はこれまでと同じく不偏推定量になる

$$E(\hat{\beta}) = E_x [E(\hat{\beta} | X_i)] = \beta + E_x [\sum w_i E(u_i | X_i)] = \beta + E_x [\sum w_i \cdot 0] = \beta$$

10.2 X と u が従属である場合

・ 不偏性 $E(\hat{\beta}) = E(\beta + \sum w_i u_i) = \beta + E(\sum w_i u_i)$

→ 第 2 項は 0 にならず、不偏ではない

・ 一致性
$$\hat{\beta} = \beta + \frac{\sum (X_i - \bar{X})u_i}{\sum (X_i - \bar{X})^2} = \beta + \frac{(1/n) \sum (X_i - \bar{X})u_i}{(1/n) \sum (X_i - \bar{X})^2} \xrightarrow{p} \beta + \frac{Cov(X_i, u_i)}{Var(X_i)}$$

$Cov(X, u) = 0$ の場合、OLS 推定量は不偏性×、一致性○

$Y_i = \alpha + \beta X_i + \rho Y_{i-1} + u_i$ などのモデルが該当。この場合、OLS 推定量は大標本のもと

でこれまでと同じような性質を持つ。

$Cov(X, u) \neq 0$ の場合、OLS 推定量は不偏性×、一致性×

過小定式化、(説明変数の) 観測誤差、同時方程式などのモデルが該当。この場合、OLS で推定する根拠がなくなる

・ 推定量の偏り a. $Cov(X, u) > 0$ のとき、 $\hat{\beta} > \beta$

b. $Cov(X, u) < 0$ のとき、 $\hat{\beta} < \beta$

10.3 説明変数に観測誤差がある場合

・観測誤差 真のモデル $Y_i^* = \alpha + \beta X_i^* + u_i^*$ (10.22)

実際の観測値 $Y_i = Y_i^* + v_i$ (10.23)

$X_i = X_i^* + w_i$ (10.24)

誤差項の仮定	$E(v_i) = 0 \quad \text{Var}(v_i) = \sigma_v^2$ $E(w_i) = 0 \quad \text{Var}(w_i) = \sigma_w^2$ それぞれの誤差項は互いに独立と仮定
--------	---

実際の観測値を代入して整理すると、

$$Y_i - v_i = \alpha + \beta(X_i - w_i) + u_i^*$$

$$\rightarrow Y_i = \alpha + \beta X_i + u_i \quad (u_i = u_i^* + v_i - \beta w_i) \quad (10.27)$$

・Yの観測誤差 Xには観測誤差はないので、 $X_i = X_i^* (w_i = 0)$

$$\rightarrow Y_i = \alpha + \beta X_i + u_i \quad (u_i = u_i^* + v_i) \quad (10.28)$$

このとき説明変数Xと誤差項uは独立であるから、攪乱項の分散が大きくなる以外は観測上の誤差に影響を与えない。

・Xの観測誤差 Yには観測誤差はないので、 $Y_i = Y_i^* (v_i = 0)$

$$\rightarrow Y_i = \alpha + \beta X_i + u_i \quad (u_i = u_i^* - \beta w_i) \quad (10.29)$$

ここで $\text{Cov}(X_i, u_i) = \text{Cov}(X_i^* + w_i, u_i^* - \beta w_i) = -\beta \text{Var}(w_i) \neq 0$ より、説明変数Xと誤差項uは独立でなく、また相関があるため、 β は不偏性、一致性でもない。

→説明変数に観測誤差がある場合はOLSを適用できない

10.4 操作変数法 (IV)

・操作変数法 説明変数Xと誤差項uの間に相関がある状況 ($\text{Cov}(X_i, u_i) = \sigma_{xu} \neq 0$) を考える。

→OLS推定量は不偏性、一致性を満たさない。

このとき、漸近的（大標本のとき）に誤差項uと相関がなく、説明変数Xと相関をもつ、操作変数Zを考える。

$$\text{plim} \frac{1}{n} \sum (Z_i - \bar{Z})(u_i - \bar{u}) = 0$$

$$\text{plim} \frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X}) = \sigma_{zx} \quad (\neq 0)$$

・IV推定量 $\hat{\beta}_{IV} = \frac{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})} = \beta + \frac{\sum (Z_i - \bar{Z})(u_i - \bar{u})}{\sum (Z_i - \bar{Z})(X_i - \bar{X})} \quad (10.39a) \quad (10.40)$

- 一致性
$$\text{plim } \hat{\beta}_{IV} = \beta + \text{plim } \frac{\frac{1}{n} \sum (Z_i - \bar{Z})(u_i - \bar{u})}{\frac{1}{n} \sum (Z_i - \bar{Z})(X_i - \bar{X})} = \beta + \frac{0}{\sigma_{ZX}} = \beta$$

 $\text{Cov}(\mathbf{X}, \mathbf{u}) \neq 0$ の場合、IV推定量はOLS推定量では満たさない一致性を満たす
(ただし $\mathbf{X}$ と $\mathbf{u}$ の相関から、IV 推定量は不偏性を満たすことはない)
- 操作変数 通常は、(系列相関がなければ) 説明変数 X_i のラグをとった X_{i-1} などを用いる
(操作変数は説明変数との相関が高いほどよい。したがって誤差項に系列相関がなければ、 $\text{Cov}(X_{i-1}, u_i) = 0$ であり、 $\text{Cov}(X_{i-1}, X_i) \neq 0$)

10.5 2段階最小2乗法 (2SLS)

- 操作変数 操作変数として X_{i-1} 以外にも、 X_{i-2} や W_i も候補として考えられる場合、

$$X_i = \gamma_0 + \gamma_1 X_{i-1} + \gamma_2 X_{i-2} + \gamma_3 W_i + v_i \quad (10.44)$$

を最小2乗法で推定し、 $\mathbf{X}$ の理論値を求め、これを操作変数とする。

$$\hat{X}_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_{i-1} + \hat{\gamma}_2 X_{i-2} + \hat{\gamma}_3 W_i \quad (10.45)$$
- 2SLS 推定量
$$\hat{\beta}_{2SLS} = \frac{\sum (\hat{X}_i - \bar{X})(Y_i - \bar{Y})}{\sum (\hat{X}_i - \bar{X})(X_i - \bar{X})} = \frac{\sum (\hat{X}_i - \bar{X})(Y_i - \bar{Y})}{\sum (\hat{X}_i - \bar{X})^2} \quad (10.46) \quad (10.47)$$

 $\uparrow Y_i = \alpha + \beta \hat{X}_i + u_i$ の OLS 推定量
- 2SLS の手順 ① $X_i = \gamma_0 + \gamma_1 X_{i-1} + \gamma_2 X_{i-2} + \gamma_3 W_i + v_i$ を OLS で推定し、理論値 $\hat{X}_i$ を求める
② $Y_i = \alpha + \beta \hat{X}_i + u_i$ を OLS で推定
- 統計的推論 2SLS 推定量について、この t 値は t 分布に従わないので注意を要する。
大標本の場合は標準正規分布を用い、小標本の場合は漸近的に t 分布に従うとすることが多い。

(参考)

- 内生性 $\text{Cov}(\mathbf{X}, \mathbf{u}) \neq 0$ の関係のとき、「内生性 endogeneity」の問題があるという。
内生性の問題がない場合、OLSE は IVE よりも有効である。(両者とも一致性)
- 内生性検定 理論からは内生性が不明な場合、Hausman-test を行う。
基本モデル $Y = \beta_0 + \beta_1 X + \beta_2 Z_1 + u \quad (a)$
操作変数 $X = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + v \quad (b)$
*ここで Z は外生変数 (u とは無相関) であるから、 X が内生変数であるかどうかは、 u が v の関数 ($u = \delta v + e \cdots (c)$) であるかどうか、すなわち (c) における δ の有意性による。したがって、
① (b) を推定し、その残差を v の推定値とする。
② (c) を (a) に代入したものを OLS で回帰し、 δ の有意性を検定する。
* δ が有意であれば、 u は v と相関、すなわち $\mathbf{X}$ と $\mathbf{u}$ の相関性を示す。