

令和6年度 中級計量経済学  
講義ノート 3: 操作変数法

回帰分析においては、説明変数と誤差項が無関係（仮定1の意味で）であることが重要であった。操作変数法は、この仮定が成り立たない場合でも係数の一致推定量を得るために使われる手法である。計量経済学で問題となる典型的な例は、欠落変数がある場合、説明変数に測定誤差がある場合、連立方程式によってシステムが記述される場合等である。操作変数とは、誤差項とは無相関で、回帰変数の変動に関する情報をもたらす変数である。この情報を用いることにより、他の要素を一定としたときの、回帰変数の被説明変数に対する影響を調べることができる。

説明変数に測定誤差がある場合 次の線形回帰モデルを考える。

$$Y_i = \beta' X_i^* + u_i, \quad E(X_i^* u_i) = 0 \quad (1)$$

ただし、 $X_i^*$  そのものは観測できず、測定誤差を含んだ  $X_i = X_i^* + v_i$  がデータとして得られるとする。簡単のため、測定誤差  $v_i$  は期待値が0で、 $X_i^*, u_i$  と独立であるとしよう。そのとき、 $X_i^*$  の代わりに  $X_i$  を説明変数として用いる、つまり

$$Y_i = \beta' X_i + \epsilon_i, \quad i = 1, \dots, n. \quad (2)$$

として回帰分析を行うのは自然なことのように思われる。しかし、それは次のような理由によりうまくいかない。上の関係式から  $X_i^*$  を消去すると

$$Y_i = \beta' X_i + u_i - \beta' v_i, \quad i = 1, \dots, n. \quad (3)$$

となる。すなわち、 $\epsilon_i = u_i - \beta' v_i$  である。そのとき、 $\epsilon_i$  と  $X_i$  の相関を計算してみると、

$$E(X_i \epsilon_i) = E[(X_i^* + v_i)(u_i - \beta' v_i)] = -E(v_i v_i') \beta \quad (4)$$

となり、一般に0ではない。つまり、前の章で仮定した回帰モデルの4つの条件のうち、第一が満たされておらず、そのために不偏性と一致性が成り立たないのである。なお、もし被説明変数  $Y_i$  に観測誤差があっても、その誤差が  $X_i$  と相関を持たない限りこのような問題は生じない。

### 3.1 回帰変数が一つで操作変数も一つの場合の操作変数法

次の線形モデルを考える。回帰変数  $X_i$  はスカラーである。

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n. \quad (5)$$

回帰変数  $X_i$  が誤差項  $u_i$  と相関を持つ場合、最小二乗推定量は不偏性や一致性を持たない。これでは、何を推定したことになっているかわからず、推定値自体に意味がなくなってしまう。

- 操作変数法がよく使われる例は、教育が賃金に与える影響の分析である。 $Y_i$  を賃金（実証研究では対数賃金がよく使われる）、 $X_i$  を教育年数とする。 $u_i$  は教育年数以外で賃金に影響を与えるすべての要因を含む量である。例えば、個人の仕事の能力などは  $u_i$  に含まれている。我々が知りたいのは、教育年数を政策変更などによって増やした時の賃金への効果である。その影響は上のモデルでは  $\beta_1$  で表現されている。しかし、実際のデータ上では、 $X_i$  と  $u_i$  は相関している可能性があり、最小二乗法では  $\beta_1$  を一致推定できないおそれがある。例えば、仕事の能力の高い人は勉強もできるとすると、 $X_i$  と  $u_i$  とは正の相関をもつであろう。一方で仕事のあまりできない人はせめて教育を長く受けて仕事の役に立てるよう努力するなら  $X_i$  と  $u_i$  とは負の相関をもつであろう。

以下の2つの条件を満たす変数  $Z_i$  を操作変数 (IV: instrumental variable) という。まず  $Z_i$  がスカラーの場合を考える。

## 操作変数が満たすべき二つの条件

1. 関連性:  $\text{corr}(Z_i, X_i) \neq 0$ .
2. 外生性:  $\text{corr}(Z_i, u_i) = 0$ .

以下の用語は頻繁に用いられる。

- 内生変数:  $u_i$  と相関がある変数。
- 外生変数:  $u_i$  と相関がない変数。

上の例の場合、 $X_i$  は内生変数であり、 $Z_i$  は外生変数である。

- 上で述べた教育と賃金の例では、親の教育年数や、育った土地の近くに大学があるかの指標、教育資金援助の利用しやすさの指標などが操作変数として使われることが多い。

**操作変数推定量 (IV 推定量)** 次の簡単なモデルを用いて操作変数法の考え方を示そう。

$$Y_i = \beta_0 + \beta_1 X_i + u_i. \quad (6)$$

モデル (2) と操作変数の外生性を用いて  $Z_i$  と  $Y_i$  の共分散を計算すると

$$\text{Cov}(Z_i, Y_i) = \text{Cov}(Z_i, \beta_0 + \beta_1 X_i + u_i) \quad (7)$$

$$= \beta_1 \text{Cov}(Z_i, X_i) + \text{Cov}(Z_i, u_i) \quad (8)$$

$$= \beta_1 \text{Cov}(Z_i, X_i) \quad (9)$$

を得る。従って、操作変数の関連性より

$$\beta_1 = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, X_i)} \quad (10)$$

が成り立つ。これらの変数に関する無作為標本  $(Y_i, X_i, Z_i), i = 1, \dots, n$  が得られたとすると、右辺の分子は標本共分散

$$s_{ZY} = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}), \quad (11)$$

によって一致推定することができる。また、分母も同様に  $s_{ZX} = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})$  によって推定できる。従って、そこから得られる  $\beta_1$  の推定量は

$$\tilde{\beta}_1 = \frac{s_{ZY}}{s_{ZX}} \quad (12)$$

である。これを操作変数推定量という。

**二段階最小二乗法 (2SLS, TSLS)** 次に、説明変数の数と操作変数の数が一致しない一般的な場合に拡張するための準備として、IV 推定量が別の考え方からも導出されることを示す。直感的に理解しやすいように、操作変数の外生性の仮定を少し強めて  $E(u_i|Z_i) = 0$  であるとしよう。(2) を次のように書き換える。

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (13)$$

$$= \beta_0 + \beta_1 E(X_i|Z_i) + \beta_1 \{X_i - E(X_i|Z_i)\} + u_i \quad (14)$$

$$= \beta_0 + \beta_1 E(X_i|Z_i) + \epsilon_i \quad (15)$$

ここで、 $\epsilon_i = \beta_1 \{X_i - E(X_i|Z_i)\} + u_i$  であるが、操作変数の外生性を用いると  $E(\epsilon_i|Z_i) = 0$  であることがわかる。また、説明変数  $E(X_i|Z_i)$  は  $Z_i$  の関数なので、最小二乗法の仮定 1 は満たされる。しかし、 $E(X_i|Z_i)$  のデータはない。そこで、 $E(X_i|Z_i) = \pi_0 + \pi_1 Z_i$  という関係を想定して、以下のようにまず  $E(X_i|Z_i)$  のデータを構成し、それを用いて  $\beta_1$  を推定するという 2 段階の方法を考えられる。

1.  $X$  を  $Z$  に回帰する。つまり、 $X_i = \pi_0 + \pi_1 Z_i + v_i$  という式を OLS 推定し、 $\hat{\pi}_0$  と  $\hat{\pi}_1$  を得て、この回帰から  $X$  の予測値を計算する:  $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$ .
2.  $Y$  を  $\hat{X}$  に回帰して、 $\hat{\beta}_1$  を  $\hat{X}$  の係数推定量とする。

この方法を二段階最小二乗法、 $\hat{\beta}_1$  を二段階最小二乗推定量 (2SLS) という。

次に、2SLS 推定量が操作変数推定量と一致することを示そう。OLS 推定量の式から

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})(Y_i - \bar{Y})}{\sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})^2} = \frac{\sum_{i=1}^n \hat{X}_i(Y_i - \bar{Y})}{\sum_{i=1}^n \hat{X}_i(\hat{X}_i - \bar{\hat{X}})} \quad (16)$$

となる。ここで、 $\bar{\hat{X}}$  は  $\hat{X}_i$  の平均である。また、 $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$  なので、

$$\sum_{i=1}^n \hat{X}_i(Y_i - \bar{Y}) = \hat{\pi}_0 \sum_{i=1}^n (Y_i - \bar{Y}) + \hat{\pi}_1 \sum_{i=1}^n Z_i(Y_i - \bar{Y}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) = \hat{\pi}_1 n s_{ZY} \quad (17)$$

となる。同様に、 $\sum_{i=1}^n \hat{X}_i(\hat{X}_i - \bar{\hat{X}}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(\hat{X}_i - \bar{\hat{X}})$ 。また、同じように

$$\hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(\hat{X}_i - \bar{\hat{X}}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})\hat{X}_i = \hat{\pi}_1^2 \sum_{i=1}^n (Z_i - \bar{Z})Z_i = \hat{\pi}_1^2 n s_{ZZ} \quad (18)$$

となる。 $\hat{\pi}_1 = s_{ZX}/s_{ZZ}$  なので、

$$\hat{\beta}_1 = \frac{\hat{\pi}_1 n s_{ZY}}{\hat{\pi}_1 n s_{ZX}} = \frac{s_{ZY}}{s_{ZX}} \quad (19)$$

と書ける。従って、 $\hat{\beta}_1 = \tilde{\beta}_1$  である。

**2SLS 推定量の統計的性質** 標本共分散は母共分散の一致推定量なので、2SLS 推定量の一致性は以下のように明らかである。

$$\hat{\beta}_1 = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} = \beta_1 \quad (20)$$

OLS 推定量の変形と同様にして

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \bar{Z}) u_i}{s_{ZX}} \quad (21)$$

と書けることから、 $\mu_Z = E(Z)$  として、2SLS 推定量の漸近分布は

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0, \frac{\text{var}((Z_i - \mu_Z)u_i)}{(\text{cov}(Z_i, X_i))^2}\right) \quad (22)$$

となる。つまり、標本数が大きいなら、 $\hat{\beta}_1$  の分布は、

$$\hat{\beta}_1 \sim_a N(\beta_1, \sigma_{\hat{\beta}_1}^2 / n) \quad (23)$$

で近似できる。なお、

$$\sigma_{\hat{\beta}_1}^2 = \frac{\text{var}((Z_i - \mu_Z)u_i)}{(\text{cov}(Z_i, X_i))^2} \quad (24)$$

である。また、 $\sigma_{\hat{\beta}_1}^2$  は、

$$\sigma_{\hat{\beta}_1}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2 \hat{u}_i^2}{s_{ZX}^2} \quad (25)$$

によって推定できる。

### 3.2 操作変数法の一般的な形

上で紹介した 2SLS 推定法を、内生変数が複数あり、外生変数も複数あり、さらに操作変数も複数ある一般的な場合に拡張する。

$$Y_i = \beta'_1 \mathbf{X}_i + \beta'_2 \mathbf{W}_i + u_i. \quad (26)$$

- $\mathbf{X}_i$ : 内生変数のベクトル ( $k$  次元)。
- $\mathbf{W}_i$ : 外生変数のベクトル。
- $\mathbf{Z}_i$ : 操作変数のベクトル ( $m$  次元)。

**識別 (identification)** モデル上で係数の値が一意に決まるとき、識別可能であるという。また、一意に決まらない時には識別不能であるという。一般論として、識別のできないパラメータを推定することはできない。操作変数モデルにおいて、どのような場合に識別可能か、上のモデルから外生変数を除いた次のモデルで考えてみよう。

$$Y_i = \beta'_1 \mathbf{X}_i + u_i. \quad (27)$$

説明変数がスカラーの場合の操作変数法の考え方 ((7)-(10) 式) と同様に、 $Y_i$  と  $\mathbf{Z}_i$  の共分散を計算すると

$$\text{Cov}(\mathbf{Z}_i, Y_i) = \text{Cov}(\mathbf{Z}_i, \mathbf{X}'_i \beta_1 + u_i) \quad (28)$$

$$= \text{Cov}(\mathbf{Z}_i, \mathbf{X}'_i) \beta_1 + \text{Cov}(\mathbf{Z}_i, u_i) \quad (29)$$

$$= \text{Cov}(\mathbf{Z}_i, \mathbf{X}'_i) \beta_1 \quad (30)$$

である。これを、両辺の共分散が与えられた時の  $\beta_1$  に関する連立方程式と見よう。もしも  $m = k$  で  $\text{Cov}(\mathbf{Z}_i, \mathbf{X}'_i)$  が正則であれば、この方程式の解は一つに定まる。もし  $m < k$  なら、方程式の数が未知数の個数よりも少ないために解は無限にあって一つに決まらない。逆に  $m > k$  なら、方程式の数が多すぎて解は存在しないかもしれない。これに対応して、次のような用語が用いられる。

- $m > k$ : 過剰識別 (over-identification)。
- $m = k$ : 過不足ない識別 (just-identified)。
- $m < k$ : 識別不能 (unidentifiable)。

前の節と同様に、過不足ない識別 ( $m = k$ ) の場合には  $\text{Cov}(\mathbf{Z}_i, \mathbf{X}_i)$  と  $\text{Cov}(\mathbf{Z}_i, Y_i)$  をデータから推定したものを用いて  $\beta_1$  を推定することができることがわかるだろう。

外生変数  $\mathbf{W}_i$  が右辺に含まれる一般の場合は、 $(\mathbf{Z}_i, \mathbf{W}_i)$  と  $Y_i$  の共分散を計算して上と同様に考えればよく、識別の条件は同じであることがわかる。

**一般的な場合の 2SLS** 一般的な場合でも推定の手続きは同じである。一段階目で内生変数を外生変数に回帰して予測値を作り、それを内生変数と置き換えて二段階目の回帰を行う。なお、一段階目の回帰で操作変数  $\mathbf{Z}_i$  のみならず外生変数  $\mathbf{W}_i$  も説明変数に含める必要がある。もし含めなければ一致性が失われてしまう。

- 一段階目の回帰:  $X$  に関する誘導系の式と呼ばれる次式を OLS 推定する。

$$\mathbf{X}_i = \pi'_1 \mathbf{Z}_i + \pi'_2 \mathbf{W}_i + v_i. \quad (31)$$

そして、 $X_i$  の予測値を計算する。

- 二段階目:  $Y_i$  を  $(\hat{\mathbf{X}}'_i, \mathbf{W}'_i)'$  に回帰する。

少し複雑になるが、この推定量の漸近的性質を調べるためにその表現を得ておこう。そのために、まず二段階目の回帰の説明変数  $(\hat{\mathbf{X}}'_i, \mathbf{W}'_i)'$  の形を求めよう。 $\mathbf{A}_i = (\mathbf{X}'_i, \mathbf{W}'_i)'$ 、 $\mathbf{B}_i = (\mathbf{Z}'_i, \mathbf{W}'_i)'$

とする。また、 $\pi' = (\pi'_1, \pi'_2)$  として、その OLS 推定量を  $\hat{\pi}$  とする。すると、(31) 式の最小二乗推定からの予測値は

$$\hat{\mathbf{X}}_i = \hat{\pi}' \mathbf{B}_i = \sum_{j=1}^n \mathbf{X}_j \mathbf{B}'_j (\sum_{j=1}^n \mathbf{B}_j \mathbf{B}'_j)^{-1} \mathbf{B}_i \quad (32)$$

である。ここで、和の添え字を  $j$  としたが、これは右辺の最後の  $\mathbf{B}_i$  の添え字の  $i$  と異なることを明確にするためである。また、 $\mathbf{W}_i = \gamma'_1 \mathbf{Z}_i + \gamma'_2 \mathbf{W}_i + v_i$  という回帰を考えてみると、左辺の  $\mathbf{W}_i$  を説明するモデルの右辺に  $\mathbf{W}_i$  が入っているので、当然完全にフィットしてしまい、その係数の OLS 推定量は

$$(\hat{\gamma}'_1, \hat{\gamma}'_2) = (\mathbf{0}, \mathbf{I}) \quad (33)$$

となるはずである。一方、形式的に  $(\hat{\gamma}'_1, \hat{\gamma}'_2) = \sum_{j=1}^n \mathbf{W}_j \mathbf{B}'_j (\sum_{j=1}^n \mathbf{B}_j \mathbf{B}'_j)^{-1}$  である。そのため

$$\mathbf{W}_i = \mathbf{0} \cdot \mathbf{Z}_i + \mathbf{I} \cdot \mathbf{W}_i = \hat{\gamma}'_1 \mathbf{Z}_i + \hat{\gamma}'_2 \mathbf{W}_i = \sum_{j=1}^n \mathbf{W}_j \mathbf{B}'_j (\sum_{j=1}^n \mathbf{B}_j \mathbf{B}'_j)^{-1} \mathbf{B}_i \quad (34)$$

が成り立つ。(32) と (34) を組み合わせると

$$(\hat{\mathbf{X}}'_i, \mathbf{W}'_i)' = \sum_{j=1}^n \mathbf{A}_j \mathbf{B}'_j (\sum_{j=1}^n \mathbf{B}_j \mathbf{B}'_j)^{-1} \mathbf{B}_i \quad (35)$$

と表されることがわかる。

その結果、2SLS 推定量は、次のように書ける。

$$\hat{\beta}^{2SLS} = \left( \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \right)^{-1} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \sum_{i=1}^n \mathbf{B}_i y_i. \quad (36)$$

なぜなら、二段階目の説明変数を  $\hat{\mathbf{A}}_i = (\hat{\mathbf{X}}'_i, \mathbf{W}'_i)'$  とおくと、

$$\hat{\beta}^{2SLS} = (\sum_{i=1}^n \hat{\mathbf{A}}_i \hat{\mathbf{A}}'_i)^{-1} \sum_{i=1}^n \hat{\mathbf{A}}_i y_i \quad (37)$$

である。上で得た (35) 式（改めて  $j$  を  $i$  に置き換えて）を使うと

$$\sum_{i=1}^n \hat{\mathbf{A}}_i \hat{\mathbf{A}}'_i = \left( \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \right) \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \right) \quad (38)$$

$$= \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \quad (39)$$

となる。また、

$$\sum_{i=1}^n \hat{\mathbf{A}}_i y_i = \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \sum_{i=1}^n \mathbf{B}_i y_i \quad (40)$$

であるので、2SLS 推定量は上の式になることが分かる。

### 一般的な場合の操作変数の関連性と外生性

- 操作変数の関連性:  $(\hat{\mathbf{X}}'_i, \mathbf{W}'_i)$  は完全な多重共線性をもたない。つまり、 $(\hat{\mathbf{X}}, \mathbf{W})$  という行列はフルランクである。この条件は、 $m \geq k$  でないと満たされない。また、母集団では、

$$E \left( \begin{pmatrix} \mathbf{Z}_i \\ \mathbf{W}_i \end{pmatrix} (\mathbf{X}'_i \mathbf{W}'_i) \right) \quad (41)$$

という行列がフルランクであるという仮定に対応する。

- 操作変数の外生性:  $cov(\mathbf{Z}_i, u_i) = 0$ .

### 2SLS 推定量の漸近的性質 仮定:

- $E(u_i | \mathbf{W}_i) = 0$ .
- $(\mathbf{X}_i, \mathbf{W}_i, \mathbf{Z}_i, Y_i)$  は i.i.d. である。
- すべての変数の 4 次モーメントは、ゼロでなくかつ有界である。
- 操作変数は関連性があり、外生である。

2SLS 推定量の漸近分布は

$$\sqrt{n}(\hat{\beta}^{2SLS} - \beta) \quad (42)$$

$$= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \right)^{-1} \quad (43)$$

$$\times \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{B}_i u_i \quad (44)$$

であるので、

$$\sqrt{n}(\hat{\beta}^{2SLS} - \beta) \rightarrow_d N(0, V_{2SLS}), \quad (45)$$

$$(46)$$

である。ここで、

$$V_{2SLS} = \left( E(\mathbf{A}_i \mathbf{B}'_i) (E(\mathbf{B}_i \mathbf{B}'_i))^{-1} E(\mathbf{B}_i \mathbf{A}'_i) \right)^{-1} \quad (47)$$

$$\times E(\mathbf{A}_i \mathbf{B}'_i) (E(\mathbf{B}_i \mathbf{B}'_i))^{-1} E(u_i^2 \mathbf{B}_i \mathbf{B}'_i) (E(\mathbf{B}_i \mathbf{B}'_i))^{-1} E(\mathbf{B}_i \mathbf{A}'_i) \quad (48)$$

$$\times \left( E(\mathbf{A}_i \mathbf{B}'_i) (E(\mathbf{B}_i \mathbf{B}'_i))^{-1} E(\mathbf{B}_i \mathbf{A}'_i) \right)^{-1}. \quad (49)$$

漸近分散は、

$$\hat{V}_{2SLS} = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \right)^{-1} \quad (50)$$

$$\times \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \mathbf{B}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \quad (51)$$

$$\times \left( \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i \mathbf{B}'_i \left( \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{B}'_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{B}_i \mathbf{A}'_i \right)^{-1} \quad (52)$$

として推定できる。この漸近分散の式は、分散不均一の場合にも一致性をもつ頑健なものである。

- 計量経済学のソフトウェアを用いると、2SLS 推定量の標準誤差は自動的に計算される。一方、簡単な統計ソフトでも OLS 推定を 2 回行うことによって 2SLS 推定値を求めることができる。ただし、その場合の 2 段階目の OLS 推定で計算される標準誤差は  $\hat{V}_{2SLS}^{1/2}$  とは一致せず、正しい標準誤差にはなっていないため、検定や区間推定に用いてはいけない。

**Hausman 検定** ある回帰変数  $\mathbf{X}_i$  が内生か外生かわからない時、それを検定することができる。ただし、適切な操作変数  $\mathbf{Z}_i$  が利用可能である必要がある。

検定の基本的なアイデアは、OLS 推定量と 2SLS 推定量を比べることである。 $\mathbf{X}_i$  が外生なら OLS 推定量は一致性をもつが、 $\mathbf{X}_i$  が内生なら一致性はない。一方で、2SLS 推定量は、どちらの場合でも一致性がある。従って、 $\mathbf{X}_i$  が外生であるという帰無仮説が正しければ、この二つの推定量は同じような値をとる。しかし、対立仮説が正しければ（つまり  $\mathbf{X}_i$  が内生なら）、大きく違ってくるはずである。

Hausman 検定統計量は、次の式で与えられる。

$$n(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'(\hat{V}_{\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}})^{-1}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}). \quad (53)$$

ただし、分散の推定に用いる  $\hat{u}_i$  としては、2SLS 推定の残差を使う。分散均一の場合には、 $\hat{V}_{\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}} = \hat{V}_{2SLS} - \hat{V}_{OLS}$  となり、この式は多くのテキスト等に掲載されており、また計量経済ソフトウェアでもデフォルトで用いられていることが多い。しかし、分散不均一の場合には正しくない。

なお、この検定統計量は、 $E(\mathbf{X}_i u_i) = 0$  という仮説を検定する統計量と解釈することもできる。

- このような検定をする理由は、2SLS よりも OLS のほうが分散が小さいため、もし  $X_i$  が外生変数なら OLS の方を使いたいからである。
- しかし、一方で、Hausman 検定の結果によって使う推定量を変えるというのは、問題を起こす可能性がある事も指摘されている（Guggenberger (2010)）。

### 3.3 操作変数が適切かどうか調べる

#### 3.3.1 操作変数の関連性

操作変数は、関連性の条件を満たす必要がある。しかし、これまでの研究で操作変数の関連性が弱いと問題が生じ、十分に「強い」関連性が必要であることが指摘されている。

**弱い操作変数の問題** 弱い操作変数とは、内生変数との相関が(ゼロではないが) 小さい操作変数のことである。なぜ、弱い操作変数は問題になるのか？すでに見た通り、2SLS 推定量は

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}) u_i}{s_{ZX}} \quad (54)$$

となるので、 $cov(Z, X) \neq 0, cov(Z, u) = 0$  なら、漸近的には問題は生じない。しかし、実際にはデータ数は有限なので、関連性が弱い場合には右辺第二項の分母がゼロに近く、結果的に絶対値で大きな値なってしまうという問題が生じうる。

それでは、操作変数はどれほど強ければよいのだろうか。

内生変数が一つの場合には、次のような方法で操作変数が十分強いと言えるかどうかを調べることができる。

- 一段階目の回帰で、操作変数の係数が0であるという帰無仮説を検定する。つまり、

$$\mathbf{X}_i = \pi'_1 \mathbf{Z}_i + \pi'_2 \mathbf{W}_i + v_i \quad (55)$$

という式を OLS で推定し、 $\pi_1 = 0$  という帰無仮説を検定する。この  $F$  検定統計量が 10 以上であれば、操作変数  $\mathbf{Z}_i$  は十分に強いといえる (Stock and Yogo (2005))。

では、操作変数が弱い場合はどうしたらよいのか？

- 操作変数が弱くても比較的うまくいく推定量を使う。多くの推定量が提案されているが、中でも制限情報最尤推定量 (LIML) がよく用いられる。
- 操作変数が弱い場合は推定は難しくなるが、係数に関する仮説の検定ならうまくやる方法が提案されている。Anderson-Rubin (1949) 統計量や、Kleibergen (2002) 統計量などである。ただしこれらの方法については、この授業では取り扱わない。

### 3.3.2 操作変数の外生性

操作変数が満たすべきもう一つの条件は外生性である。過不足なく識別されているときには用いることができないが、過剰識別の場合には以下のようにして検定可能である。ただし、すべての操作変数の外生性を検定をすることはできず、推定する係数の数だけの操作変数は外生であるという仮定の下で、それ以外の操作変数の外生性を検定することができる。したがって、これは、過剰識別検定と呼ばれる。なお、ここでの帰無仮説は操作変数が外生性を満たすことである。

- 基本的なアイデアは、 $\hat{u}_i$  と  $\mathbf{Z}_i$  の相関を調べることである。
- 次の検定統計量を使用する。これは  $J$  検定統計量と呼ばれる。

$$n \left( \frac{1}{n} \sum_{i=1}^n \hat{u}_i \left( \begin{matrix} \mathbf{Z}_i \\ \mathbf{W}_i \end{matrix} \right) \right)' \left( \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2 \left( \begin{matrix} \mathbf{Z}_i \\ \mathbf{W}_i \end{matrix} \right) (\mathbf{Z}'_i \mathbf{W}'_i) \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \hat{u}_i \left( \begin{matrix} \mathbf{Z}_i \\ \mathbf{W}_i \end{matrix} \right) \right) \quad (56)$$

- $J$  検定統計量は、帰無仮説の元で、 $\chi^2_{m-k}$  に分布収束する。
- 注意点は、自由度が  $m - k$  であることである。これは、 $k$  個の操作変数は推定のために使われており、そのほかの操作変数の外生性しか検定できないことと対応している。なお、帰無仮説を棄却したときに、どの操作変数が内生であるかは、追加的な情報がない限り判断できない。
- 過不足なく識別されている場合には、 $\hat{u}_i$  と  $\mathbf{Z}_i$  の共分散は必ずゼロになるので、 $J$  を計算しても常にゼロになってしまい、この方法は使えない。

### 3.4 効果が不均一な場合の回帰分析と操作変数法による分析

ある変数  $X_i$  から  $Y_i$  への効果が、個人ごとに違う可能性がある状況を考えてみよう。このような状況は、プログラム効果分析の枠組みでよく見られる。もし、その違いが観察可能な変数のみで決まつくるならば、前の章で見たように、相互作用項をモデルに含めることによって、その違いを調べることができる。ここでは、観察できない違いがあるときに、OLS や 2SLS 推定量にどのような影響があるかを考える。

例として、教育年数  $X_i$  が賃金  $Y_i$  に与える影響を表す次のようなモデルを考える。

$$Y_i = \beta_{0i} + \beta_{1i}X_i + u_i. \quad (57)$$

$\beta_{1i}$  は個人  $i$  が教育年数  $X_i$  を 1 年増加させた時の賃金への効果で、各個人ごとに異なる値をとりうる確率変数であると考えることにする。添え字  $i$  によって、効果が個人ごとに異なることが表現されている。 $E(\beta_{1i})$  は教育年数が 1 年増えた時の平均的な賃金上昇額を表し、平均効果 (Average Treatment Effect, ATE) と呼ばれる。本来的にはすべての個人についての  $\beta_{1i}$  を別個に推定できれば一番よいが、それは何らかの追加的な仮定や情報がなければ不可能である。 $\beta_{1i}$  が個人ごとに異なることを無視して OLS や 2SLS を適用した時に、どういう量が推定されるのかを考えてみる。

$(\beta_{0i}, \beta_{1i})$  と  $X_i$  が独立であるかどうかによって、問題が変わってくる。

$(\beta_{0i}, \beta_{1i})$  と  $X_i$  が独立な場合 まず、 $X_i$  と  $u_i$  が無相関な場合を考える。説明変数と誤差項が無相関なので、OLS 推定を適用すれば良さそうである。実際、OLS 推定量は、

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \xrightarrow{p} \frac{\text{cov}(\beta_{0i} + \beta_{1i}X_i + u_i, X_i)}{\sigma_X^2} = \frac{\text{cov}(\beta_{1i}X_i, X_i)}{\sigma_X^2} = E(\beta_{1i}). \quad (58)$$

となり、ATE に収束する。

次に、 $X_i$  と  $u_i$  に相関があれば、操作変数法を使うべきであろう。適切な操作変数  $Z_i$  があるとして、2SLS を行う。このとき、次のように、一段階目の推定式のパラメータも以下のように個人ごとに異なり、パラメータ  $\pi_{0i}, \pi_{1i}$  は  $Z_i$  と独立であるとする。

$$X_i = \pi_{0i} + \pi_{1i}Z_i + v_i. \quad (59)$$

$(\beta_{0i}, \beta_{1i})$  と  $X_i$  が独立であるという条件を少し強めて、 $(\beta_{0i}, \beta_{1i})$  と  $\pi_{0i}, \pi_{1i}, Z_i, v_i$  が独立であるものとしよう。このとき、

$$\hat{\beta}_1^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})} = E(\beta_{1i}). \quad (60)$$

となり、やはり ATE を推定していることがわかる。

$\beta_{1i}$  と  $X_i$  が独立でない確率的係数モデル 次に、独立でない場合を考える。ただし、簡単にするために  $\beta_{0i}$  と  $X_i$  は独立であるとしよう。 $X_i$  と  $u_i$  に相関がないとき、OLS を適用すると、

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \xrightarrow{p} \frac{\text{cov}(\beta_{0i} + \beta_{1i}X_i + u_i, X_i)}{\sigma_X^2} = \frac{E(\beta_{1i}X_i(X_i - \mu_X))}{E(X_i(X_i - \mu_X))} \quad (61)$$

となり、 $\beta_{1i}$  に重み  $X_i(X_i - \mu_X)$  をつけた重みつき平均を推定していることがわかる。特に、 $X_i$  の値が絶対値で大きい個人を大きな重みをかけた平均を見ていることがわかる。

$X_i$  と  $u_i$  に相関があって操作変数法を使う場合は、上と同等の仮定の下で

$$\hat{\beta}_1^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E(\beta_{1i}X_i(Z_i - \mu_Z))}{E(X_i(Z_i - \mu_Z))} \quad (62)$$

であるので、やはり、重みつき平均を推定していることがわかる。特に、操作変数の  $X_i$  に与える影響の大きい個人を中心とした平均を見ていることがわかる。またこの議論から、 $X_i$  の  $y_i$  に与える影響を操作変数法によって推定すると、使用する操作変数によって推定値が変わってくる可能性があるということがわかる。

なお、 $\beta_{1i}$  と  $X_i$  に相関があるかどうかは検定可能である。Heckman, Schimeirer and Urzua (2010) によると、検定法は二つある。そのうちの一つは、過剰識別の検定を使うというものである。このことは、過剰識別検定の解釈に問題を起こす。つまり、もし、相関のある確率係数モデルを考えないなら、過剰識別検定は、操作変数の外生性を検定しているが、操作変数の外生性を仮定したときにはそれは相関のある確率係数モデルの検定と考えることができるということである。つまり、ある仮定なしには、過剰識別検定の解釈をすることはできない。もう一つの方法はここでは取り扱わない。

### 3.5 内生性の問題に対する対処法

適切な操作変数を見つけるのが、実証研究においては、最も重要で難しい問題である。今まで使用されていなかった適切な操作変数をみつければ、それだけで、重要な貢献になりうる。

現在の経済実証分析では、内生性に対する対処法には主に次の二つの方法がある。

- 操作変数法：経済理論から、操作変数を導く。例：需要曲線と供給曲線の推定。需要に影響を与えるが、供給に影響を与えない変数を使い、供給曲線を推定する。
- 自然実験になっている状況をうまく使う。

**自然実験** 回帰変数  $X_i$  の変化が、少なくとも部分的には実験のようにランダムに決まっている（つまり外生的）と考えられる状況のこと。準実験とも呼ばれる。

自然実験には、二つのタイプがある。

1.  $X_i$  の値が、無作為に決まるとみなせる場合。このときには、単に OLS を使って推定をすればよい。
2. 無作為な変化分は、 $X_i$  の値を部分的にしか決めない場合。このときには、操作変数法を使い推定する。 $X_i$  の無作為な変化をもたらすものを操作変数として使う。

例：

- Levitt (1996): 犯罪率と刑務所への収監率の関係を調べる。（人は刑務所にいる限り再犯できないので刑務所に多くの人を収監すると犯罪率が下がると思われる。刑法学や犯罪学でいうところの無力化理論である。）刑務所の混雑を避ける目的で行われた法決定を操作変数として使う。
- Hoxby (2000): 学級の大きさと、学力テスト成績の関係を調べる。児童数の長期傾向からの乖離部分を操作変数として使う。
- Card (1990): 移民増はその地域の賃金を下げる効果があるかを調べる。キューバ革命時のマイアミへの移民がおこった事件からの情報を使う

- Angrist (1990): 従軍が民間での給与に与える影響（徴兵政策の公平性）。徴兵くじの番号を使用する。

1970～72年のアメリカにおいて、19歳になった人を対象に誕生日ごとにランダムに数字を割り当てて、その数字が国防省が定めたある数値以下であれば徴兵対象となった。実際には種々の理由で対象者でも兵役につかなかった人や非対象者でも志願して兵役についた人もいた。そのため、くじ番号は実際の兵役とは一致しないが、高い相関を持ち、またくじ番号は完全に外生的なので操作変数として有効である。

- 他の例は、Rosenzweig and Wolpin (2000) や Angrist and Krueger (2001) といった概説論文にある。

なお、自然実験は、完全な対照実験とはならないので、グループ分けに偏りが出る可能性がある等の理由から、 $X_i$  が異なるもの同志の違いを適切に調節する必要がある。特に、 $X_i$  の値を決める前から存在する違いをあらわす変数を、モデルに含める必要がある。