平成 25 年度 中級計量経済学・応用計量経済学
講義ノート 3: 操作変数法

操作変数法は、回帰変数と誤差項が相関しているときに、一致性のある推定量を得るために使われる手法である。例としては、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。この情報を使うことにより、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。

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。

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n. \tag{1}$$

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。

- 操作変数法がもっともよくつかわれている例は、教育の賃金への影響に関するものであろう。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素を含んだ量である。例えば、個人の仕事の能力などは u_i に含まれている。我々が興味があるのは、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影響である。その影響は上のモデルでは β_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 β_1 を一致推定できないおそれがある。例えば、仕事のできる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し、一方で仕事のあまりできない人はせめて教育をつけて仕事の役に立てるように努力すると考えると X_i と u_i とは負の相関をもつであろう。

しかし、次の性質を満たす操作変数 (IV: instrumental variable)Z_i が利用可能であるとする。ひとまず 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. 操作変数の関連性: $corr(Z_i, X_i) \neq 0.$
2. 操作変数の外生性: $corr(Z_i, u_i) = 0.$

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。Z_i を操作変数として使えるためには、外生変数であり、なおかつ関連性を持つ必要がある。

- 上で述べた教育と賃金の例で良く使われる操作変数としては、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ある。

操作変数推定量 (IV 推定量)

$$Y_i = \beta_0 + \beta_1 X_i + u_i. \tag{2}$$

操作変数法の考え方を簡単に示そう。モデル (2) と操作変数の外生性を使って、Z_i と Y_i の共分散を計算すると

$$\begin{aligned} Cov(Z_i, Y_i) &= Cov(Z_i, \beta_0 + \beta_1 X_i + u_i) \\ &= \beta_1 Cov(Z_i, X_i) + Cov(Z_i, u_i) \\ &= \beta_1 Cov(Z_i, X_i) \end{aligned}$$

を得る。従って、操作変数の関連性より

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

であるが、右辺の分子は標本共分散

$$s_{ZY} = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})(Y_i - \bar{Y}), \tag{3}$$

によってうまく推定することができる。分母も同様に $s_{ZX} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})(X_i - \bar{X})$ によって推定できる。従って、そこから得られる β_1 の推定量は

$$\hat{\beta}_1 = \frac{s_{ZY}}{s_{ZX}}$$

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見ていこう。直感的に理解しやすいように、操作変数の関連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。

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i \\ &= \beta_0 + \beta_1 E(X_i|Z_i) + \beta_1 \{X_i - E(X_i|Z_i)\} + u_i \\ &= \beta_0 + \beta_1 E(X_i|Z_i) + \epsilon_i \end{aligned}$$

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。$E(X_i|Z_i) = \pi_0 + \pi_1 Z_i$ という関係を想定して、以下のようにして 2 段階で β_1 を推定することが考えられる。

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 $\hat{\pi}_1$ を計算する。
2. 上の回帰から X の予測値を計算する: $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$.
3. Y を $\hat{X}$ に回帰する。$\tilde{\beta}_1$ を $\hat{X}$ の係数推定量とする。

$\tilde{\beta}_1$ は二段階最小二乗法 (2SLS) と呼ばれる。

2SLS 推定量の標本分布　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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})(Y_i - \bar{Y})}{\sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})^2} = \frac{\sum_{i=1}^n \hat{X}_i (Y_i - \bar{Y})}{\sum_{i=1}^n \hat{X}_i (\hat{X}_i - \bar{\hat{X}})} \tag{4}$$

となる。ここで、$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$ なので、

$$\sum_{i=1}^n \hat{X}_i (Y_i - \bar{Y}) = \hat{\pi}_0 \sum_{i=1}^n (Y_i - \bar{Y}) + \hat{\pi}_1 \sum_{i=1}^n Z_i (Y_i - \bar{Y}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) = \hat{\pi}_1 n s_{ZY} \tag{5}$$

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

$$\hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(\hat{X}_i - \bar{\hat{X}}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z}) \hat{X}_i = \hat{\pi}_1^2 \sum_{i=1}^n (Z_i - \bar{Z}) Z_i = \hat{\pi}_1^2 \sum_{i=1}^n (Z_i - \bar{Z})^2 \tag{6}$$

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

$$\hat{\beta}_1 = \frac{\hat{\pi}_1 n s_{ZY}}{\hat{\pi}_1 n s_{ZX}} = \frac{s_{ZY}}{s_{ZX}} \tag{7}$$

と書ける。

2SLS 推定量の一致性は、標本共分散が母共分散の一致推定量であることから殆ど自明である。

$$\hat{\beta}^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \to_p \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)} = \beta_1 \tag{8}$$

となり、一致性が示せる。

2SLS 推定量の漸近分布は

$$\sqrt{N}(\hat{\beta}_1 - \beta_1) = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^n (Z_i - \bar{Z}) u_i}{s_{ZX}} \tag{9}$$

と書けることから、$\mu_Z = E(Z)$ として、

$$\sqrt{N}(\hat{\beta}_1 - \beta_1) \to_d N\left(0, \frac{var((Z_i - \mu_Z)u_i)}{(cov(Z_i, X_i))^2}\right) \tag{10}$$

となる。つまり、標本数が大きいなら、$\hat{\beta}_1^{2SLS}$ の分布は、

$$\hat{\beta} \sim_a N(\beta_1, \sigma^2_{\hat{\beta}_1^{2SLS}}/n) \tag{11}$$

で近似できる。なお、

$$\sigma^2_{\hat{\beta}_1} = \frac{var((Z_i - \mu_Z)u_i)}{(cov(Z_i, X_i))^2} \tag{12}$$

である。また、$\sigma^2_{\hat{\beta}_1}$ は、

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2 \hat{u}_i^2}{s_{ZY}^2} \tag{13}$$

によって推定できる。

3.2 操作変数法の一般的な形

上で紹介した操作変数法を、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場合に拡張する。

$$Y_i = \beta_1' \mathbf{X}_i + \beta_2' \mathbf{W}_i + u_i. \tag{14}$$

- $\mathbf{X}_i$: 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）。

用語

- $m > k$: 過剰識別。
- $m = k$: 過不足ない識別。
- $m < k$: 識別不能。

一般的な場合の 2SLS

- 一段階目の回帰: 次の式を推定する。

$$\mathbf{X}_i = \pi_1' \mathbf{Z}_i + \pi_2' \mathbf{W}_i + v_i. \tag{15}$$

この式は、時折、X に関する誘導系の式とも呼ばれる。そして、X_i の OLS 予測値を計算する。

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

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: $(\hat{\mathbf{X}}_i, \mathbf{W}_i)$ は完全に多重共線性になっていない。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、

$$E\left(\begin{pmatrix} \mathbf{Z}_i \\ \mathbf{W}_i \end{pmatrix} (\mathbf{X}_i' \mathbf{W}_i')\right) \tag{16}$$

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。

$\mathbf{A}_i = (\mathbf{X}_i', \mathbf{W}_i')'$ とし、また $\mathbf{B}_i = (\mathbf{Z}_i', \mathbf{W}_i')'$ とする。2SLS 推定量は、次のようにかける。

$$\hat{\beta}^{2SLS} = \left(\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{A}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{B}_i y_i. \tag{17}$$

上のようになるのは、以下の理由による。$\hat{\mathbf{A}}_i = (\hat{\mathbf{X}}_i', \mathbf{W}_i')'$ とすると、

$$\hat{\mathbf{A}}_i = \left(\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \right) \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \mathbf{B}_i \tag{18}$$

となる。したがって、

$$\begin{aligned} \sum_{i=1}^{n} \hat{\mathbf{A}}_i \hat{\mathbf{A}}_i' &= \left(\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \right) \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \right) && (19) \\ &= \sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{A}_i' && (20) \end{aligned}$$

となる。また、

$$\sum_{i=1}^{n} \hat{\mathbf{A}}_i y_i = \sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{B}_i y_i \tag{21}$$

であるので、2SLS 推定量は上の式になることが分かる。

2SLS 推定量の漸近分布は

$$\begin{aligned} & \sqrt{n}(\hat{\beta}^{2SLS} - \beta) && (22) \\ = & \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{A}_i' \right)^{-1} && (23) \\ & \times \frac{1}{n} \sum_{i=1}^{n} \mathbf{A}_i \mathbf{B}_i' \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{B}_i \mathbf{B}_i' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{B}_i u_i && (24) \end{aligned}$$

であるので、

$$\begin{aligned} \sqrt{n}(\hat{\beta}^{2SLS} - \beta) \quad &\to_d \quad N(0, V_{2SLS}), && (25) \\ & && (26) \end{aligned}$$

である。ここで、

$$
\begin{aligned}
V_{2SLS} \quad = \quad & \left(E(\mathbf{A}_i\mathbf{B}_i')\left(E(\mathbf{B}_i\mathbf{B}_i')\right)^{-1}E(\mathbf{B}_i\mathbf{A}_i')\right)^{-1} & (27)\\
& \times E(\mathbf{A}_i\mathbf{B}_i')\left(E(\mathbf{B}_i\mathbf{B}_i')\right)^{-1}E(u_i^2\mathbf{B}_i\mathbf{B}_i')\left(E(\mathbf{B}_i\mathbf{B}_i')\right)^{-1}E(\mathbf{B}_i\mathbf{A}_i') & (28)\\
& \times\left(E(\mathbf{A}_i\mathbf{B}_i')\left(E(\mathbf{B}_i\mathbf{B}_i')\right)^{-1}E(\mathbf{B}_i\mathbf{A}_i')\right)^{-1}. & (29)
\end{aligned}
$$

漸近分散は、

$$
\begin{aligned}
\hat{V}_{2SLS} = \quad = \quad & \left(\frac{1}{n}\sum_{i=1}^n \mathbf{A}_i\mathbf{B}_i'\left(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{B}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{A}_i'\right)^{-1} & (30)\\
& \times\frac{1}{n}\sum_{i=1}^n \mathbf{A}_i\mathbf{B}_i'\left(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{B}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2\mathbf{B}_i\mathbf{B}_i'\left(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{B}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{A}_i' & (31)\\
& \times\left(\frac{1}{n}\sum_{i=1}^n \mathbf{A}_i\mathbf{B}_i'\left(\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{B}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n \mathbf{B}_i\mathbf{A}_i'\right)^{-1} & (32)
\end{aligned}
$$

として推定できる。この漸近分散の式は、分散不均一の元でも通用する頑健なものである。

- 計量経済学のソフトウェアを用いると、2SLS 推定が可能で、その標準誤差も自動的に計算される。一方、2SLS 推定量を OLS 推定を 2 回行うことによって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ことに注意する必要がある。

Hausman 検定　ある回帰変数 $\mathbf{X}_i$ が内生が外生かどうか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。OLS 推定量は、$\mathbf{X}_i$ が外生なら一致性がある。しかし、$\mathbf{X}_i$ 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。つまり、$\mathbf{X}_i$ が外生であるという帰無仮説のもとでは、この二つの推定量は同じような値をとるはずだであるが、対立仮説のもとでは、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。

$$n(\hat{\beta}_{2SLS}-\hat{\beta}_{OLS})'(\hat{V}_{\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}})^{-1}(\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}). \tag{33}$$

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。分散均一性を仮定すると、$\hat{V}_{\hat{\beta}_{2SLS}-\hat{\beta}_{OLS}} = \hat{V}_{2SLS}-\hat{V}_{OLS}$ となり、この式は色々な文献に出ているため有名であるが、これは、分散均一性が成立する場合のみ正しい。

また、この検定統計量は、$E(\mathbf{X}_iu_i) = 0$ という仮説を検定する統計量としても書くことができる。

- なぜこの検定をするのかというと、OLS のほうが分散が小さいため、もしできるなら、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、これまでの研究で分かってきたことは、操作変数は、単に関連性を持つだけでなく、十分に「強い」関連性を持つべきであるということである。

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 (ゼロでなくとも) 弱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。推定量の式をみると

$$\hat{\beta}_1^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \to_p \frac{cov(Z,Y)}{cov(Z,X)} = \beta_1 + \frac{cov(Z,u)}{cov(Z,X)} \tag{34}$$

となり、もし、$cov(Z,X) = 0$ なら、この確率極限は、うまく定義できない。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、操作変数が弱いかどうかを調べる方法。

- 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、

$$\mathbf{X}_i = \pi_1' \mathbf{Z}_i + \pi_2' \mathbf{W}_i + v_i \tag{35}$$

  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、操作変数は十分に強いといえる (Stock and Yogo (2005))。

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検定不可能であるが、過剰識別の場合には検定可能である。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。

- 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。

$$n\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\begin{pmatrix}\mathbf{Z}_i\\ \mathbf{W}_i\end{pmatrix}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\begin{pmatrix}\mathbf{Z}_i\\ \mathbf{W}_i\end{pmatrix}(\mathbf{Z}_i'\mathbf{W}_i')\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\begin{pmatrix}\mathbf{Z}_i\\ \mathbf{W}_i\end{pmatrix}\right) \tag{36}$$

- J 検定統計量は、帰無仮説の元で、χ^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。

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 の効果は、個人ごとに違ってくる可能性がある。もし、その違いが観察可能な変数のみで決まってくるならば、前の章で見たように、相互作用項をモデルにいれることで、違いを見ることができる。ここでは、観察できない違いが、推定量にどのような影響をもたらすかを考える。

次のようなモデルを考える。

$$Y_i = \beta_{0i} + \beta_{1i} X_i + u_i. \tag{37}$$

β_{1i} は個人 i における X_i の効果である。添え字 i を付けているのは、効果が個人ごとに異なることを表現するためである。

β_{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。

β_{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。このとき、OLS 推定量は、

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \rightarrow_p \frac{cov(\beta_{0i} + \beta_{1i} X_i + u_i, X_i)}{\sigma_X^2} = \frac{cov(\beta_{1i} X_i, X_i)}{\sigma_X^2} = E(\beta_{1i}). \tag{38}$$

となり、平均効果 (Average Treatment Effect, ATE) に収束する。

また操作変数法を考える。このとき、次のように、一段階目の推定式を書くとわかりやすい。

$$X_i = \pi_{0i} + \pi_{1i} Z_i + v_i. \tag{39}$$

β_{1i} と X_i が独立である場合の例として、β_{1i} と X_i の式に出てくる変数とが独立である場合を考えることができる。このとき、

$$\hat{\beta}_1^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \rightarrow_p \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})} = E(\beta_{1i}). \tag{40}$$

となり、平均効果を推定していることがわかる。

β_{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。このとき、OLS は、

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} \rightarrow_p \frac{cov(\beta_{0i} + \beta_{1i} X_i + u_i, X_i)}{\sigma_X^2} = \frac{E(\beta_{1i} X_i (X_i - \mu_X))}{E(X_i(X_i - \mu_X))} \tag{41}$$

となり、β_{1i} の重みつき平均を推定していることがわかる。特に、X_i の値が大きい個人を主にした平均を見ていることがわかる。

操作変数法の場合は、

$$\hat{\beta}_1^{2SLS} = \frac{s_{ZY}}{s_{ZX}} \rightarrow_p \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E(\beta_{1i}\pi_{1i})}{E(\pi_{1i})} \tag{42}$$

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に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、β_{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　操作変数の見つけ方

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。いま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。

主に次の二つの方法がある。

- 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。
- 自然実験を使う。

自然実験　回帰変数 X_i の変化が、実験のようにランダムに決まってくるように見える状況のこと。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): 移民の労働市場への影響を調べる。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。徴兵くじの番号を使用する。
- 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、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。