操作変数法†

担当:　長倉　大輔
（ながくらだいすけ）

† この資料は私の講義およびゼミにおいて使用するために作成した資料です。WEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし誤植、間違い、等があった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。誤植、間違いは発見次第、継続的に直していますが、常に存在する可能性があります。

操作変数法

- 説明変数が確率変数の場合

回帰式: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{K1} \\ \vdots & \cdots & \vdots \\ X_{1n} & \cdots & X_{KN} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

のOLSの性質を求める際には以下の仮定を置いていた

(A1) $E(\varepsilon_i) = 0$, for $i = 1,\ldots, n$.

(A2) (**無相関の仮定**) $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, $i, j = 1,\ldots, n$.

(A3) (**均一分散の仮定**) $\mathrm{var}\,(\varepsilon_i) = \sigma^2$, $i = 1,\ldots, n$.

(A4) (a) $\mathbf{X}$ は非確率的である。(b) $\mathrm{rank}(\mathbf{X}) = K < N$ である。

(c) $\lim_{n\to\infty} \dfrac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n} = \mathbf{Q}$ となる正則行列 $\mathbf{Q}$が存在する。

操作変数法

- 説明変数が確率変数の場合

これらの仮定の下で最小二乗推定量:

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

は仮定(A1)と(A4ab)が満たされれば**不偏性**、(A1)と(A4)が満たされれば**一致性**を持ち、仮定(A1)~(A4ab)が満たされれば**最良線形不偏推定量(Best Linear Unbiased Estimator; BLUE)**となる。
　また、仮定(A2)と(A3)が満たされない場合は**GLS推定量がBLUEとなる**。

という事であった。

操作変数法

- 説明変数が確率変数の場合

これらの仮定の内、仮定(A4a)が全ての性質に共通している。しかしながら現実の分析においては、説明変数が非確率的と考えられる場合は少ない。

もし説明変数も確率的な変数である場合にはOLSはどのような性質を持つであろうか？

操作変数法

- 説明変数が確率変数の場合

以下では(A4a)を仮定せず、仮定(A1)〜(A3)を以下の仮定に置き換える。

(A1') $E(\varepsilon_i \,|\mathbf{X}) = 0$, for $i = 1,\ldots, n$.
(A2') $\mathrm{cov}(\varepsilon_i, \varepsilon_j \,|\mathbf{X}) = 0$, $i \neq j$, $i, j = 1,\ldots, n$.
(A3') $\mathrm{var}\,(\varepsilon_i \,|\mathbf{X}) = \sigma^2$, $i = 1,\ldots, n$.

これはつまり、先ほどの仮定(A1) ~ (A3)は全て **X という条件付きでも**成立していると言い直すという事である。

操作変数法

- 説明変数が確率変数の場合

もし(A1)~(A3)の仮定を(A1')~(A3')で置き換えた場合、**Xが確率変数でも**、OLSは一致性、不偏性を満たす事を示すことができる。

よって、説明変数が確率変数かそうでないかは(A1')~(A3')が満たされている限りは重要な問題ではない。

操作変数法

- 内生性

しかしながら、説明変数が確率変数である時に、
経済学のデータでしばしば生じるのは、

「 仮定(A1') $E(\varepsilon|X) = 0$ 」が満たされない

という問題である。

操作変数法

- 内生性

この仮定が満たされないと、

説明変数と誤差項に相関が生じ、この時
OLSは一致推定量ではなくなる。

(厳密にいえば、ある特殊な状況では実は(A1')が満たされなくても誤差項と説明変数が無相関の時もあるが、特殊な状況なのでここでは扱わない。)

説明変数と誤差項に相関がある場合、その回帰モデルには**内生性**があるといい、そのような説明変数を**内生変数**と呼ぶ。

操作変数法

- 重複期待値の法則; Law of Iterated Expectation

2 つの確率変数 X と Y について**重複期待値の法則 (Law of Iterated Expectation)** とは

$$E(X) = E[E(X|Y)]$$

が成り立つことをいう。これを用いると、例えば

$E(X|Y) = \alpha Y$ であるとすると

$$E(X) = E[E(X|Y)] = E(\alpha Y) = \alpha E(Y)$$
$$E(XY) = E[E(X|Y)Y] = E(\alpha Y^2) = \alpha E(Y^2)$$

などと計算できる。

操作変数法

- 内生性が生じる原因

内生変数が生じる例として、2つ例をあげる。

1. 観測誤差による内生性; Errors in Variable model)
2. 賃金関数における教育年数の効果の推定

操作変数法

- 観測誤差による内生性; Errors in Variable model)

Y_i を被説明変数、X_i を説明変数とし、これら 2 つの変数には(簡単化のために説明変数はX_i だけとする)

$$Y_i = \beta X_i + \varepsilon_i$$

という関係があるとしよう。ここで ε_i は X_i だけで説明しきれない誤差項であり、$E(\varepsilon_i \mid X_i) = 0$ であるとする。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

この時、重複期待値の法則より

$$E(\varepsilon_i X_i) = E[E(\varepsilon_i \mid X_i)X_i] = E(0\ X_i) = 0$$

である。(またこの時、再び重複期待値の法則より

$$\begin{aligned}\mathrm{cov}(\varepsilon_i, X_i) &= E(\varepsilon_i X_i) - E(\varepsilon_i)E(X_i)\\ &= 0 - E[E(\varepsilon_i \mid X_i)]E(X_i)\\ &= 0 - 0\ E(X_i) = 0\end{aligned}$$

である。つまり**内生性を持たない**事に注意)。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

この時、X_i を Y_i に回帰させて得られる最小二乗推定量は

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{N} X_i^2\right)^{-1} \sum_{i=1}^{N} X_i Y_i$$

となり大数の法則より、

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{N} X_i(\beta X_i + \varepsilon_i)}{\sum_{i=1}^{N} X_i^2} = \beta \frac{\sum_{i=1}^{N} X_i^2}{\sum_{i=1}^{N} X_i^2} + \frac{\sum_{i=1}^{N} X_i \varepsilon_i}{\sum_{i=1}^{N} X_i^2}$$

$$= \beta + \frac{N^{-1}\sum_{i=1}^{N} X_i \varepsilon_i}{N^{-1}\sum_{i=1}^{N} X_i^2} \xrightarrow{p} \beta + \frac{E(X_i \varepsilon_i)}{E(X_i^2)} = \beta$$

となり、**一致性を持つ**。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

今、説明変数 X_i は直接観測する事が出来ず、観測値には誤差 u_i が伴うとする。また、この誤差は

$$E(u_i) = 0,\ \ \mathrm{cov}(X_i, u_i) = 0,\ \text{および}\ \mathrm{cov}(u_i, \varepsilon_i) = 0$$

を満たすと仮定する。この時、実際に観測されるのは

$$X_i^* = X_i + u_i$$

である。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

今、$X_i = X_i^* - u_i$ であるから、Y_i と X_i^* との関係は

$$
\begin{aligned}
Y_i = \beta\,(X_i^* - u_i) + \varepsilon_i &= \beta X_i^* - \beta u_i + \varepsilon_i \\
&= \beta X_i^* - \beta u_i + \varepsilon_i \\
&= \beta X_i^* + \varepsilon_i^*
\end{aligned}
$$

という回帰モデルの関係になっている。ここで

$$\varepsilon_i^* = -\beta u_i + \varepsilon_i$$

である。この時 $E(\varepsilon_i^*) = -\beta E(u_i) + E(\varepsilon_i) = 0$ であることに注意。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

この時、

$$\begin{aligned}
\mathrm{cov}(X_i^*, \varepsilon_i^*) &= \mathrm{cov}(\, X_i + u_i\,,\ -\beta u_i + \varepsilon_i\,) \\
&= -\beta\,\mathrm{cov}(X_i, u_i) + \mathrm{cov}(X_i, \varepsilon_i) \\
&\quad -\beta\,\mathrm{cov}(u_i, u_i) + \mathrm{cov}(u_i, \varepsilon_i) \\
&= -\beta\,\mathrm{var}(u_i)
\end{aligned}$$

となり ($\beta \neq 0$の時) X_i^*は**内生変数**である。(またこの時

$$\begin{aligned}
E\,(\,X_i^*\varepsilon_i^*\,) &= \mathrm{cov}(X_i^*, \varepsilon_i^*) - E(X_i^*)\,E(\varepsilon_i^*) \\
&= \mathrm{cov}(X_i^*, \varepsilon_i^*) - E(X_i^*)\,0 \\
&= \mathrm{cov}(X_i^*, \varepsilon_i^*)
\end{aligned}$$

である事にも注意)。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

よってこの時、$Y_i = \beta X_i^* + \varepsilon_i^*$ のOLS推定量を $\hat{\beta}_{OLS}^*$ とすると、先ほどと同じく大数の法則により

$$\hat{\beta}_{OLS}^* = \left(\sum_{i=1}^{N} X_i^{*2}\right)^{-1} \sum_{i=1}^{N} X_i^* Y_i$$

$$= \beta + \frac{N^{-1}\sum_{i=1}^{N} X_i^* \varepsilon_i^*}{N^{-1}\sum_{i=1}^{N} X_i^{*2}}$$

$$\xrightarrow{p} \beta + \frac{E(X_i^* \varepsilon_i^*)}{E(X_i^{*2})} = \beta - \frac{\beta\,\mathrm{var}(u_i)}{E(X_i^{*2})}$$

となり、もはや β には確率収束しない、つまり β の**一致推定量ではない**という事になる。

操作変数法

- 観測誤差による内生性; Errors in Variable model)

これは誤差を伴った説明変数の観測値 X_i^* とY_i とX_i^* の回帰モデルの中の誤差項 ε_i^* が相関を持つ事により内生性が生じ、内生性が生じると*OLS*は一致性を失うという例である。

操作変数法

- 賃金関数における教育年数の効果の推定

労働経済学における問題の一つに、労働者の賃金がどのような変数の影響を受けて決定されているかという問題がある。

その中でよく論じられる問題に、教育年数 (もしくは学歴といっても良いかもしれない) は賃金にどのような影響を与えるかというものがあり、この問題について様々な学者が実証的に研究をしている。

操作変数法

- 賃金関数における教育年数の効果の推定

ここでの教育年数の効果というのは、**純粋に教育年数の効果**だけを問題にしており、教育によって個人の能力が高まる事による賃金への影響などは取り除いた時の教育年数の賃金への影響の事である。

(つまり、例え教育によって個人の能力がまったく向上しなくても教育年数が増えるだけで賃金に影響するかという問題である。もしそうなら、たとえ教育自体は能力を高めるのに意味がなくても、お金を払って教育を受ける価値がある事になろう)。

操作変数法

- 賃金関数における教育年数の効果の推定

個人 i の真の賃金 W_i は次のようなモデルによって決定されるとする。

$$\log(W_i) = \alpha + \beta X_i + \gamma Y_i + \varepsilon_i$$

ここで説明変数は単純化のために2つとして

X_i： 教育年数、 Y_i ： 個人 i の能力を表す数値

とし、ε_i は $\mathrm{cov}(X_i, \varepsilon_i) = \mathrm{cov}(Y_i, \varepsilon_i) = 0$ を満たす誤差項であるとする。今、興味があるのは β の値であり、これを推定したいとする。

操作変数法

- 賃金関数における教育年数の効果の推定

通常、教育年数と個人の能力は(正の)相関関係があると考えるのが自然であろう。つまり

$$\mathrm{cov}\,(X_i,\ \ Y_i)\ \neq 0$$

であろう。

このようなモデルを推定する時に問題となるのが、X_i は観測できるが、個人の能力を表す Y_i は**直接観測する事ができない**という事である。

操作変数法

- 賃金関数における教育年数の効果の推定

この時、$\log(W_i)$ を X_i だけに回帰して β を推定しようとすると内生性の問題が生じる。これを確認しよう。

Y_i は観測されないので $\gamma Y_i + \varepsilon_i$ をまとめて

$$u_i = \gamma Y_i + \varepsilon_i$$

という誤差項とみなそう。すると

$$\log(W_i) = \alpha + \beta X_i + u_i$$

という回帰モデルを得る。このモデルに対してβ をOLSで推定するとどうなるだろうか？

操作変数法

- 賃金関数における教育年数の効果の推定

まず X_i と u_i に相関が生じてることを確認する。

$$\begin{aligned}\mathrm{cov}(X_i, u_i) &= \mathrm{cov}(X_i, \gamma Y_i + \varepsilon_i) \\ &= \gamma\, \mathrm{cov}(X_i, Y_i)\end{aligned}$$

であるので $\mathrm{cov}(X_i, Y_i) \neq 0$ であれば X_i と u_i は相関を持つ事になる。

操作変数法

- 賃金関数における教育年数の効果の推定

この時、β のOLSは

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})\log(W_i)}{\sum_{i=1}^{N} X_i^2 - N\overline{X}^2} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(\alpha + \beta X_i + u_i)}{\sum_{i=1}^{N} X_i^2 - N\overline{X}^2}$$

$$= \beta + \frac{N^{-1}\sum_{i=1}^{N}(X_i - \overline{X})(u_i - \overline{u})}{N^{-1}\sum_{i=1}^{N} X_i^2 - \overline{X}^2}$$

$$\xrightarrow{p} \beta + \gamma \frac{\operatorname{cov}(X_i, Y_i)}{\operatorname{var}(X_i)}$$

となり、推定したい値 β ではなく別の値に収束する(**一致性を持たない**)。

操作変数法

- 賃金関数における教育年数の効果の推定

この場合もOLSは推定したい値をきちんと推定してくれない。

今回も原因はやはり**説明変数と誤差項に相関がある**事である。

説明変数と誤差項の相関が疑われる時に、OLSのかわりに**操作変数法**と呼ばれる推定方法が用いられる。

演習問題(操作変数法)

問題1: 確率変数 X と Y は $E(X^2) = 1$、$E(X^3) = 0$、および $E(Y|X) = X^2$ を満たすとする。この時、重複期待値の法則によって $E(Y)$ および $E(XY)$ の値を求めよ。

問題2: 前述した賃金関数において Y_i は観測できないがその代理変数として個人の知能指数 I_i を使用する事にしよう。しかしながら、I_i は Y_i と完全にイコールではなく

$$I_i = Y_i + u_i$$

という関係があるとする。

(次ページに続く)

演習問題(操作変数法)

問題2(続き):

ここで u_i は

$$\mathrm{cov}(u_i, \varepsilon_i) = \mathrm{cov}(u_i, Y_i) = \mathrm{cov}(u_i, X_i) = 0$$

を満たす確率変数であるとする。賃金関数に $Y_i = I_i - u_i$ を代入した

$$\log(W_i) = \alpha + \beta X_i + \gamma I_i + e_i \quad (e_i = -\gamma u_i + \varepsilon_i \text{ である})。$$

という回帰モデルを考えたときに、この回帰モデルに内生性は生じているか？生じているならどの変数が内生変数か？答えなさい。

操作変数法

- 操作変数法

次の回帰モデルを考えよう。

$$Y_i = \alpha + \beta X_i + \varepsilon_i \ , \ i = 1, 2, \ldots, N$$

(簡単化のため説明変数は一つとする)
ここで推定したいパラメーターは β である。

最小二乗法による推定は誤差項と説明変数の間に相関が存在する時 (内生性がある時)、つまり $\mathrm{cov}(X_i, \varepsilon_i) \neq 0$ の時に一致性を持たなかった。

操作変数法

- 操作変数法

次のような変数 Z_i を考えよう。

$$\mathrm{cov}(\varepsilon_i, Z_i) = 0, \quad \mathrm{cov}(X_i, Z_i) \neq 0$$

内生変数 X_i に対するこのような変数 Z_i を**操作変数**という。

操作変数法

- 操作変数推定量

この時、Y_i と Z_i の共分散は

$$\begin{aligned} \mathrm{cov}\,(Y_i, Z_i) &= \mathrm{cov}(\alpha + \beta X_i + \varepsilon_i,\ Z_i) \\ &= \beta\,\mathrm{cov}(X_i, Z_i) \end{aligned}$$

となる。これより β は

$$\beta = \frac{\mathrm{cov}(Y_i, Z_i)}{\mathrm{cov}(X_i, Z_i)}$$

を満たすことがわかる。

操作変数法

- 操作変数推定量

この場合、β の自然な推定量はY_i と Z_i の標本共分散と X_i と Z_i の標本共分散であろう。すなわち

$$\hat{\beta}_{IV} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(Z_i - \bar{Z})}{\sum_{i=1}^{N}(X_i - \bar{X})(Z_i - \bar{Z})}$$

である。これを β の**操作変数推定量**という。
操作変数推定量はβ の**一致推定量**となる。

操作変数法

- 2段階最小二乗法

この場合の操作変数推定量は以下の
2段階最小二乗法 (2 stage least square method; 以後2SLSと略す)　と呼ばれる方法によっても得ることができる。

操作変数法

- 2段階最小二乗法

(1段階目) X_i を Z_i に回帰し

$$X_i = \hat{\gamma} + \hat{\delta} Z_i + \hat{e}_i$$

を得る。ここで $\hat{\gamma}$, $\hat{\delta}$ はZ_i を X_i に回帰させた時の定数項と係数の最小二乗推定値、$\hat{e}_i$ はその回帰式の残差である。

この $\hat{\gamma}$ と $\hat{\delta}$ を使って、Z_i が与えられた下でのX_i の予測値を

$$\hat{X}_i = \hat{\gamma} + \hat{\delta} Z_i$$

とする。

操作変数法

- 2段階最小二乗法

(2段階目) この $\hat{X}_i$ を Y_i に(最小二乗法によって)回帰させて

$$Y_i = \hat{\alpha}_{2sls} + \hat{\beta}_{2sls}\hat{X}_i + \hat{\varepsilon}_i$$

を得る。このように得た2段階最小二乗推定量 $\hat{\beta}_{2ls}$ は先ほどの操作変数推定量と一致する(数値的に同じになる)事を示すことができる(演習問題)。よってもちろん一致推定量である。

操作変数法

- 2段階最小二乗法

2段階最小二乗法が一致性を持つのは以下のように考えると直観的にわかる。まず

$$
\begin{aligned}
Y_i &= \alpha + \beta X_i + \varepsilon_i \\
&= \alpha + \beta \hat{X}_i + \varepsilon_i + \beta (X_i - \hat{X}_i) \\
&= \alpha + \beta \hat{X}_i + u_i
\end{aligned}
$$

と書き換える。ここで $u_i = \varepsilon_i + \beta (X_i - \hat{X}_i)$ である。2段階目でこの回帰式を推定している。ここで $\hat{X}_i$ と u_i が無相関であればこの式のOLSは一致性を持つが $\hat{X}_i$ と u_i は漸近的に無相関である事を示すことができる。

操作変数法

- 操作変数法 (内生変数が複数の場合)

次の回帰式を考えよう

$$Y_i = \beta_1 X_{i1} + \ldots + \beta_K X_{iK} + \alpha_1 W_{i1} + \ldots + \alpha_J W_{iJ} + \varepsilon_i ,$$

ここで

$$\mathrm{cov}(X_{ik}, \varepsilon_i) \neq 0, \quad k = 1, \ldots, K, \quad E(\varepsilon_i \mid W_{i1}, \ldots, W_{iJ}) = 0$$

であるとする。すなわち X_{ik} は**内生変数**、W_{ij} は**外生変数**である(定数項が入る場合は$W_{i1} = 1$ であるとする)。

操作変数法

- 操作変数法 (内生変数が複数の場合)

(操作変数)　さらに X_{ik} のそれぞれについて

$$\mathrm{cov}(X_{ik}, Z_{ik}) \neq 0, \qquad \mathrm{cov}(\varepsilon_i, Z_{ik}) = 0$$

である**操作変数** Z_{ik} , $k = 1, \ldots, K$ が存在するとする。

操作変数と内生変数の数は同じであるとする。

操作変数法

- 操作変数法 (内生変数が複数の場合)

この回帰式を

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W}_i \boldsymbol{\alpha} + \varepsilon_i$$

と書き換えよう。ここで

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} & \dots & X_{iK} \end{bmatrix}, \qquad \mathbf{W}_i = \begin{bmatrix} W_{i1} & \dots & W_{iJ} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{bmatrix}, \text{ および } \mathbf{Z}_i = \begin{bmatrix} Z_{i1} & \dots & Z_{iK} \end{bmatrix}$$

とする。

操作変数法

- 操作変数法 (内生変数が複数の場合)

条件より

$$\mathrm{cov}(W_{ij}, \varepsilon_i) = E(W_{ij}\,\varepsilon_i) = 0, \quad j = 1,\ldots, J$$

$$\mathrm{cov}(Z_{ik}, \varepsilon_i) = E(Z_{ik}\,\varepsilon_i) = 0, \quad k = 1,\ldots, K$$

であるので

$$E\left(\varepsilon_i\,[Z_{i1}, Z_{i2}, \ldots, Z_{iK}, W_{i1}\ W_{i2}\ .,\ldots, W_{iJ}]\right) = \mathbf{0}$$

$$\Rightarrow \quad E\,[\,\varepsilon_i\,[\,\mathbf{Z}_i\ \mathbf{W}_i\,]\,] = \mathbf{0}$$

$$\Rightarrow \quad E\left(\begin{bmatrix}\mathbf{Z}_i^{\mathrm{T}} \\ \mathbf{W}_i^{\mathrm{T}}\end{bmatrix}\varepsilon_i\right) = \mathbf{0}$$

が成り立つ。

操作変数法

- 操作変数法 (内生変数が複数の場合)

よって $\varepsilon_i = Y_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{W}_i\boldsymbol{\alpha} = Y_i - \begin{bmatrix}\mathbf{X}_i & \mathbf{W}_i\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{bmatrix}$

である事に注意すると

$$E\left(\begin{bmatrix}\mathbf{Z}_i^{\mathrm{T}}\\ \mathbf{W}_i^{\mathrm{T}}\end{bmatrix}\varepsilon_i\right) = \mathbf{0} \quad \Rightarrow \quad E\left[\begin{bmatrix}\mathbf{Z}_i^{\mathrm{T}}\\ \mathbf{W}_i^{\mathrm{T}}\end{bmatrix}\left(Y_i - \begin{bmatrix}\mathbf{X}_i & \mathbf{W}_i\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{bmatrix}\right)\right] = \mathbf{0}$$

$$\Rightarrow \quad E\left(\begin{bmatrix}\mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i & \mathbf{Z}_i^{\mathrm{T}}\mathbf{W}_i\\ \mathbf{W}_i^{\mathrm{T}}\mathbf{X}_i & \mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{bmatrix} = E\left(\begin{bmatrix}\mathbf{Z}_i^{\mathrm{T}}\\ \mathbf{W}_i^{\mathrm{T}}\end{bmatrix}Y_i\right)$$

$$\Rightarrow \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{bmatrix} = \begin{bmatrix}E(\mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i) & E(\mathbf{Z}_i^{\mathrm{T}}\mathbf{W}_i)\\ E(\mathbf{W}_i^{\mathrm{T}}\mathbf{X}_i) & E(\mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i)\end{bmatrix}^{-1}\begin{bmatrix}E(\mathbf{Z}_i^{\mathrm{T}}Y_i)\\ E(\mathbf{W}_i^{\mathrm{T}}Y_i)\end{bmatrix}$$

である。

操作変数法

- 操作変数法 (内生変数が複数の場合)

先ほどの式で期待値を対応する標本平均で置き換えると

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_{IV} \\ \hat{\boldsymbol{\alpha}}_{IV} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{N} \mathbf{Z}_i^{\mathrm{T}}\mathbf{X}_i & \frac{1}{n}\sum_{i=1}^{N} \mathbf{Z}_i^{\mathrm{T}}\mathbf{W}_i \\ \frac{1}{n}\sum_{i=1}^{N} \mathbf{W}_i^{\mathrm{T}}\mathbf{X}_i & \frac{1}{n}\sum_{i=1}^{N} \mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{N} \mathbf{Z}_i^{\mathrm{T}} Y_i \\ \frac{1}{n}\sum \mathbf{W}_i^{\mathrm{T}} Y_i \end{bmatrix}$$

となる。これが**内生変数が複数の場合の操作変数推定量**である。

操作変数法

- 操作変数法 (内生変数が複数の場合)

先ほどの推定量は

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_{IV} \\ \hat{\boldsymbol{\alpha}}_{IV} \end{bmatrix} = \left(\frac{1}{n} \sum_{i=1}^{N} \begin{bmatrix} \mathbf{Z}_i^{\mathrm{T}} \\ \mathbf{W}_i^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_i & \mathbf{W}_i \end{bmatrix} \right)^{-1} \frac{1}{n} \sum_{i=1}^{N} \begin{bmatrix} \mathbf{Z}_i^{\mathrm{T}} \\ \mathbf{W}_i^{\mathrm{T}} \end{bmatrix} Y_i$$

$$= \left(\sum_{i=1}^{N} [\mathbf{Z}_i \ \ \mathbf{W}_i]^{\mathrm{T}} [\mathbf{X}_i \ \ \mathbf{W}_i] \right)^{-1} \sum_{i=1}^{N} [\mathbf{Z}_i \ \ \mathbf{W}_i]^{\mathrm{T}} Y_i$$

$$= \left([\mathbf{Z} \ \ \mathbf{W}]^{\mathrm{T}} [\mathbf{X} \ \ \mathbf{W}] \right)^{-1} [\mathbf{Z} \ \ \mathbf{W}]^{\mathrm{T}} \mathbf{Y}$$

と書くこともできる。ここで

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ \vdots & \cdots & \vdots \\ X_{N1} & \cdots & X_{NK} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} W_{11} & \cdots & W_{1J} \\ \vdots & \cdots & \vdots \\ W_{N1} & \cdots & W_{NJ} \end{bmatrix}, \ \mathbf{Z} = \begin{bmatrix} Z_{11} & \cdots & Z_{1K} \\ \vdots & \cdots & \vdots \\ Z_{N1} & \cdots & Z_{NK} \end{bmatrix},$$

である。

操作変数法

- 2段階最小二乗法 (内生変数が複数ある場合)

(1段階目) $\mathbf{X}$ の k 番目の列 (これを$\mathbf{X}^{(k)}$ と表そう) を $\mathbf{Z}$ と $\mathbf{W}$ に回帰し

$$\mathbf{X}^{(k)} = \mathbf{Z}\hat{\boldsymbol{\gamma}}_{ols}^{(k)} + \mathbf{W}\hat{\boldsymbol{\delta}}_{ols}^{(k)} + \hat{\mathbf{e}}^{(k)}$$
$$\hat{\mathbf{X}}^{(k)} = \mathbf{Z}\hat{\boldsymbol{\gamma}}_{ols}^{(k)} + \mathbf{W}\hat{\boldsymbol{\delta}}_{ols}^{(k)}$$

を得る。ここで $\hat{\boldsymbol{\gamma}}_{ols}^{(k)}$ と $\hat{\boldsymbol{\delta}}_{ols}^{(k)}$ は $\mathbf{X}^{(k)}$ を $\mathbf{Z}$ と $\mathbf{W}$ に回帰させた時の係数ベクトルの最小二乗推定値、$\hat{\mathbf{e}}^{(k)}$ はその回帰式の残差ベクトルである。

操作変数法

- 2段階最小二乗法 (内生変数が複数ある場合)

この $\hat{\boldsymbol{\gamma}}_{ols}^{(k)}$ と $\hat{\boldsymbol{\delta}}_{ols}^{(k)}$ を使って、$\mathbf{Z}$ と $\mathbf{W}$ が与えられた下での $\mathbf{X}^{(k)}$ の予測値を

$$\hat{\mathbf{X}}^{(k)} = \mathbf{Z}\hat{\boldsymbol{\gamma}}_{ols}^{(k)} + \mathbf{W}\hat{\boldsymbol{\delta}}_{ols}^{(k)}$$

とする。これを $k=1$ から K まで順に並べて $\mathbf{X}$ の予測値

$$\hat{\mathbf{X}} = [\hat{\mathbf{X}}^{(1)}, \hat{\mathbf{X}}^{(2)}, \ldots, \hat{\mathbf{X}}^{(K)}]$$

を作る。

操作変数法

- 2段階最小二乗法 (内生変数が複数ある場合)

(2段階目) この $\hat{\mathbf{X}}$ を $\mathbf{Y}$ に(最小二乗法によって)回帰させて

$$\mathbf{Y} = \hat{\mathbf{X}}\hat{\boldsymbol{\beta}}_{2sls} + \mathbf{W}\hat{\boldsymbol{\alpha}}_{2sls} + \hat{\boldsymbol{\varepsilon}}$$

2段階最小二乗推定値を得る。

操作変数法

- 操作変数と2段階最小二乗法

2段階最小二乗法が操作変数と等しい事を示す。
簡単化のため外生変数 **W** がない場合のみを説明する。この時、操作変数推定量は(行列表記の方で)

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y}$$

と表せる。

操作変数法

- 操作変数と2段階最小二乗法

2段階最小二乗法の$\mathbf{X}^{(k)}$についての1段階のOLS は

$$\hat{\gamma}_{ols}^{(k)} = (\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}^{(k)}$$

であるので $\hat{\mathbf{X}}$ は

$$\begin{aligned}\hat{\mathbf{X}} &= [\mathbf{Z}\hat{\gamma}_{ols}^{(1)}, \mathbf{Z}\hat{\gamma}_{ols}^{(2)}, \ldots, \mathbf{Z}\hat{\gamma}_{ols}^{(K)}] \\ &= \mathbf{Z}[\hat{\gamma}_{ols}^{(1)}, \hat{\gamma}_{ols}^{(2)}, \ldots, \hat{\gamma}_{ols}^{(K)}] \\ &= \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}[\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(K)}] \\ &= \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}\end{aligned}$$

となる。

操作変数法

- 操作変数と2段階最小二乗法

2段階目のOLS は

$$\hat{\boldsymbol{\beta}}_{2sls} = (\hat{\mathbf{X}}^{\mathrm{T}}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^{\mathrm{T}}\,\mathbf{Y}$$

となる。これに先ほどの結果を代入すると

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{2sls} &= (\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X})\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\
&= (\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}\,\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\
&= (\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})(\mathbf{X}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\
&= (\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\
&= (\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y}
\end{aligned}$$

となり操作変数推定量と等しい事がわかる。

操作変数法

- 操作変数法の問題点

1. 操作変数は存在すること前提としているが、どのように操作変数を見つければいいかは全く不明

2. 例え、操作変数の仮定を満たすような変数 Z_i が存在したとしても(ε_i と無相関、X_i と相関を持つ) X_i との相関が十分に大きくないと推定量は小標本では非常にバイアスを持つ。

3. そもそも操作変数を使う必要がある(ε_i と X_i が相関を持つ)のかは確かめようがなく、しばしば経済理論などから「疑わしい」という理由で使われる。

演習問題(操作変数法)

問題3：回帰式:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

に対して X_i の操作変数として Z_i を使って2段階最小二乗推定法で β を推定し、それが操作変数推定量と等しい事を示しなさい(ヒント: 定数項(外生変数)の操作変数として定数項そのものを使ってみる。これは操作変数の仮定を満たす)。