<522A. Econometrics> 3/18.

- Order condition and rank condition for IV.

ex) $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$E(u_i)=0$, $E(X_{3i}u_i)=0$ but $E(X_{2i}u_i) \neq 0$

It says "$\exists$ correlation b/w X_{2i} and U_i"

→ It is possible that Endogeneity problems occur.

⇓

Assume $\exists$ Z_{1i}, Z_{2i} s.t $E(Z_{1i}u_i)=0$ and $E(Z_{2i}u_i)=0$.

Then, in this case, 1 is an IV candidate.

X_{3i} is an IV candidate

$\left.\begin{matrix} Z_{1i} \\ \& Z_{2i} \end{matrix}\right)$ are IV candidates.

$$E\left[\begin{pmatrix} 1 \\ X_{3i} \\ Z_{1i} \\ Z_{2i} \end{pmatrix} (1 \ X_{2i} \ X_{3i})\right] = E\begin{pmatrix} 1 & X_{2i} & X_{3i} \\ X_{3i} & X_{3i}X_{2i} & X_{3i}^2 \\ Z_{1i} & Z_{1i}X_{2i} & Z_{1i}X_{3i} \\ Z_{2i} & Z_{2i}X_{2i} & Z_{2i}X_{3i} \end{pmatrix} \circledast$$

⇒ The rank of this matrix needs to be at least 3.

※ Counting the number of IVs and that being greater than or equal to the number of regressors is called order condition. (Necessary condition).

※ But the condition that the above ⊛ matrix has the rank equal to or exceeding the number of regressors is called the rank condition.
(Necessary and sufficient condition for uniquely recovering β by IV moment conditions).

Note

$\boxed{\begin{array}{ll} \# \text{ of IVs} \geqq \# \text{ of regressors} & : \text{order condition.} \\ \text{rank of } \circledast \geqq \# \text{ of regressors} & : \text{rank condition.} \end{array}}$

- 2SLS in this context

1st stage : getting fitted values.

regress 1 on 1, X_{3i}, Z_{1i}, Z_{2i} and obtain fitted values.

$$1 = \pi_1^1 \cdot 1 + \pi_2^1 X_{3i} + \pi_3^1 Z_{1i} + \pi_4^1 Z_{2i} + V_{1i}$$

$\Rightarrow \hat{\pi}_1^1 = 1$ $\hat{\pi}_2^1 = 0$, $\hat{\pi}_3^1 = 0$, $\hat{\pi}_4^1 = 0$: yields perfect fit!

The fitted values are all 1. $\hat{1} = \hat{\pi}_1^1 1 + 0 \cdot X_{3i} + 0 \cdot Z_{1i} + 0 \cdot Z_{2i}$

$$\hat{1} = 1 \cdot 1$$

regress X_{2i} on 1, X_{3i}, Z_{1i}, Z_{2i} and obtain fitted values.

$$X_{2i} = \pi_1^2 \cdot 1 + \pi_2^2 \cdot X_{3i} + \pi_3^2 Z_{1i} + \pi_4^2 Z_{2i} + V_{2i}$$

$$\Rightarrow \hat{X}_{2i} = \hat{\pi}_1^2 \cdot 1 + \hat{\pi}_2^2 \cdot X_{3i} + \hat{\pi}_3^2 \cdot Z_{1i} + \hat{\pi}_4^2 \cdot Z_{2i}$$

regress X_{3i} on 1, X_{3i}, Z_{1i}, Z_{2i}

$$X_{3i} = \pi_1^3 \cdot 1 + \pi_2^3 \cdot X_{3i} + \pi_3^3 Z_{1i} + \pi_4^3 Z_{2i} + V_{3i}$$

$$\Rightarrow \hat{X}_{3i} = \hat{\pi}_1^3 \cdot 1 + \hat{\pi}_2^3 \cdot X_{3i} + \hat{\pi}_3^3 Z_{1i} + \pi_4^3 Z_{2i}$$

$= 0 \cdot 1 + 1 \cdot X_{3i} + 0 \cdot Z_{1i} + 0 \cdot Z_{2i}$: yield perfect fit!

$$\hat{X}_{3i} = X_{3i}$$

$\Downarrow$

2nd stage : regress y_i on fitted values $(1, \hat{X}_{2i}, X_{3i})$

: In 2SLS, if we have a regressor which is not correlated with the error terms, then at the 2nd stage the variable itself can be used as a regressor.

Note

$E(1 \cdot u_i) = 0$
$E(X_{3i} \cdot u_i) = 0$] 1, X_{3i} can be used as themselves.

$E(X_{2i} \cdot u_i) \neq 0$: X_{2i} should be changed by IVs.
For 2SLS, $\hat{X}_{2i}$ can be used at the 2nd stage.

- Intuitive Motivation for 2SLS.

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

When $E(X_{2i} u_i) \neq 0$, $E(X_{3i} u_i) = 0$ and $E(u_i) = 0$,

Regress X_{2i} on $1, X_{3i}, Z_{1i}, Z_{2i}$

$$X_{2i} = \underbrace{\hat{\pi}_1 + \hat{\pi}_2 X_{3i} + \hat{\pi}_3 Z_{1i} + \hat{\pi}_4 Z_{2i}}_{\hat{X}_{2i}} + \hat{V}_{2i} \Rightarrow \begin{cases} E(1 \mid V_{2i}) = 0 \\ E(X_{3i} \mid V_{2i}) = 0 \Rightarrow E(X_{3i} V_{2i}) = 0 \\ E(Z_{1i} \mid V_{2i}) = 0 \Rightarrow E(Z_{1i} V_{2i}) = 0 \\ E(Z_{2i} \mid V_{2i}) = 0 \Rightarrow E(Z_{2i} V_{2i}) = 0. \end{cases}$$

$\hat{V}_{2i}$: OLS residual of this regression.

$$\Downarrow$$

$$X_{2i} = \hat{X}_{2i} + \hat{V}_{2i}$$

Thus. $y_i = \beta_1 + \beta_2(\hat{X}_{2i} + \hat{V}_{2i}) + \beta_3 X_{3i} + u_i$

$$= \beta_1 \cdot 1 + \beta_2 \hat{X}_{2i} + \beta_3 X_{3i} + (u_i + \beta_2 \hat{V}_{2i})$$

$\hat{V}_{2i}$ and $(1, \hat{X}_{2i}, X_{3i})$ are orthogonal by construction.

: $\sum_{i=1}^{N} \hat{V}_{2i} = 0$, $\sum_{i=1}^{N} \hat{V}_{2i} X_{3i} = 0$ $\sum_{i=1}^{N} \hat{V}_{2i} Z_{1i} = 0$ $\sum_{i=1}^{N} \hat{V}_{2i} Z_{2i} = 0$.

※ The 2SLS exploits the variation in regressors that are not correlated with the residual term u_i.

※ When # of IVs = # of regressors. $\hat{\beta}_{2SLS} = \hat{\beta}_{IV}$.

: This result allows to study 2SLS and use the result to understand $\hat{\beta}_{IV}$.

(But, 2SLS are not useful under Heteroskedasticity).

- Generalized Method of Moment.

To motivate the objective function used in GMM, Need to know WLS.

$\mathbb{Y} = \mathbb{X}\beta + \mathbb{U}$ $\quad E(\mathbb{U}\mathbb{U}'|\mathbb{X}) = \sigma^2 I_N$ Under Homoskedasticity

+ No correlation across observations.

$\mathbb{Y} = \mathbb{X}\beta + \mathbb{U}$ $\quad E(\mathbb{U}\mathbb{U}'|\mathbb{X}) = \begin{pmatrix} \sigma^2(x_1) & 0 & \cdots & 0 \\ 0 & \sigma^2(x_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2(x_N) \end{pmatrix}$ Under Heteroskedasticity.

+ No correlation across observations.

$\Downarrow$

Now, we allow there exists correlation across observations.

$$E(\mathbb{U}\mathbb{U}'|\mathbb{X}) = \underset{N\times N}{\Omega} = \begin{pmatrix} \sigma^2(x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_N) \\ cov(x_2,x_1) & \ddots & & \vdots \\ \vdots & & \ddots & cov(x_{N-1},x_N) \\ cov(x_N,x_1) & \cdots & cov(x_N,x_{N-1}) & \sigma^2(x_N) \end{pmatrix}$$

Then, $\underbrace{\Omega^{-\frac{1}{2}}\mathbb{Y}}_{\tilde{\mathbb{Y}}} = \underbrace{\Omega^{-\frac{1}{2}}\mathbb{X}}_{\tilde{\mathbb{X}}}\beta + \underbrace{\Omega^{-\frac{1}{2}}\mathbb{U}}_{\tilde{\mathbb{U}}}$

$\tilde{\mathbb{Y}} = \tilde{\mathbb{X}}\beta + \tilde{\mathbb{U}}$ --- $(**)$

$$E(\tilde{\mathbb{U}}\tilde{\mathbb{U}}'|\mathbb{X}) = E\left(\Omega^{-\frac{1}{2}}\mathbb{U}\mathbb{U}'(\Omega^{-\frac{1}{2}})'|\mathbb{X}\right) = \Omega^{-\frac{1}{2}}E(\mathbb{U}\mathbb{U}'|\mathbb{X})(\Omega^{-\frac{1}{2}})'$$
$$= \Omega^{-\frac{1}{2}}\,\Omega\,(\Omega^{-\frac{1}{2}})' = I_N.$$

$\Rightarrow$ $(**)$ Satisfies condition for the OLS estimator to be BLUE

$$\hat{\beta}_{GLS} = (\tilde{\mathbb{X}}'\tilde{\mathbb{X}})^{-1}\tilde{\mathbb{X}}'\tilde{\mathbb{Y}} = (\mathbb{X}'(\Omega^{-\frac{1}{2}})'\Omega^{-\frac{1}{2}}\mathbb{X})^{-1}\mathbb{X}'(\Omega^{-\frac{1}{2}})'\Omega^{-\frac{1}{2}}\mathbb{Y}$$
$$= (\mathbb{X}'\Omega^{-1}\mathbb{X})^{-1}\mathbb{X}'\Omega^{-1}\mathbb{Y} \quad \left(\because (\Omega^{-\frac{1}{2}})'(\Omega^{-\frac{1}{2}}) = \left(\Omega^{\frac{1}{2}}(\Omega^{\frac{1}{2}})'\right)^{-1} = \Omega^{-1}\right)$$

$\Downarrow$ Let's check a simple case.

$y_1 = X_1\beta_1 + u_1$
$y_2 = X_2\beta_2 + u_2$ $\Rightarrow$ $\Omega = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$

Under Homoskedasticity,
but $\exists$ correlation across regressors

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \sigma^2 \begin{pmatrix} a^2 & ab \\ ab & b^2+c^2 \end{pmatrix}$$

$a^2 = 1$ $\quad ab = \rho$
$b^2 + c^2 = 1$
$\Downarrow$
$a = 1,\ b = \rho,\ c = \sqrt{1-\rho^2}$

$$\Omega = \sigma^2 \Gamma\Gamma' = \sigma^2 \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}\begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix}$$

Then, $\Gamma^{-1} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}^{-1} = \dfrac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 \\ -\rho & 1 \end{pmatrix}$

Assume $\sigma^2 = 1$. and then $\Omega^{-\frac{1}{2}} = \Gamma^{-1}$.

$$\Omega^{-\frac{1}{2}} \mathbb{Y} = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 \\ -\rho & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ \frac{1}{\sqrt{1-\rho^2}}(y_2 - \rho y_1) \end{pmatrix}$$

$y_1 = X_1\beta_1 + u_1$
$y_2 = X_2\beta_2 + u_2$ $\Big\} \Rightarrow \mathbb{Y} = \mathbb{X}\beta + \mathbb{U}$

$$\Omega^{-\frac{1}{2}} \mathbb{X} = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 \\ -\rho & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ \frac{1}{\sqrt{1-\rho^2}}(X_2 - \rho X_1) \end{pmatrix}$$

$\Rightarrow$ We regress $\begin{pmatrix} y_1 \\ \frac{1}{\sqrt{1-\rho^2}}(y_2 - \rho y_1) \end{pmatrix}$ on $\begin{pmatrix} X_1 \\ \frac{1}{\sqrt{1-\rho^2}}(X_2 - \rho X_1) \end{pmatrix}$

↳ Weighted by $\Omega^{-\frac{1}{2}}$, we can eliminate correlation across regressors, and then taking OLS regression is the idea of $\boxed{\text{GLS}}$

- The idea of GMM

: the same as the method of moments allowing for the possibility that there are more than enough moment conditions to recover unknown parameters.

⇒ In our context, the moment conditions can be written as

$$\hat{g}(b) \equiv E\{\underbrace{\underset{J\times 1}{Z_i}\,(\underset{1\times 1}{y_i - X_i'\hat{\beta}})}\} = 0$$

↳ $J\times 1$ moment conditions for estimating K unknown parameters.

⇓

General form for Non-linear cases.

$$E\{\underbrace{g(w_i;\theta)}_{r\times 1}\} = 0$$

(Note) A sample analog of the moment conditions is

$$\frac{1}{N}\sum_{i=1}^{N}\underbrace{Z_i(Y_i - X_i'\overset{K\times 1}{b})}_{J\times 1} = 0$$

- If $J<K$, impossible.
- In general, Moment of Method would solve for the solution to be fine the estimator, but that approach fail if $J>K$. (overidentification).

⇒ For MoM, we need the condition $J=K$.

⇒ Define the metric between vectors by the norm;

$\|V\|_A = \sqrt{V'AV}$ for a given positive definite matrix A.

※ The GMM estimator of β is defined as the minimizer of

$\min_{b\in\mathbb{R}^K} \hat{g}(b)'A\,\hat{g}(b)$ for a given A. $(\hat{\beta}_{GMM} = \text{argmin}\ \hat{g}(b)'A\,\hat{g}(b))$

How to choose A? ⇒ "Optimal of $(\hat{g})^{-1}$"

//