* Today's topic ; [1] Subset Overidentification Test (C - test)
[2] Endogeneity Test.

$\Rightarrow$ We will use the form of J test to check a form of $E(z_i u_i) = 0$.

[1] Test of over-identifying restrictions. : C- test.

Revist J- test : $N \cdot \hat{g}(\hat{b}) \hat{\Omega}^{-1} \hat{g}(\hat{b}) \overset{?}{\simeq} 0$

$\Rightarrow$ When J-test reject the null hypothesis that $E(u_i z_i) = 0$, we don't know which of the J-moment conditions are violated.
(i.e., Rejection of J- test implies $E(u_i z_i) \neq 0$, We don't know which the problem is on).

Worse yet, we know from the construction of the test there is a k - dimension directions (corresponding to the F.O.C) for which the test does not have power.

(Prof. Ichimura said " The result could be contaminated by some of K so that we cannot clarify the credibility of J- test. ")

$\Downarrow$ To overcome this problem, use C- test.

C-test : to distinguish "Moment conditions" we are confident about. and those we are not.

J - IVs.

$J_1 (\geq K) \rightarrow$ confident IVs

$J - J_1 \rightarrow$ Not confident IVs.

Note < C-test >

① Estimate β optimally using J_1 IVs.
② Test the rest of the moment conditions using the estimated residual.

Let $J = J_1 + J_2$. where $J_1 \geq K$.

① Z_{1i} : $J_1 \times 1$ vector we are confident about. : $E(Z_{1i} u_i) = 0$.

Thus, we can estimate β optimally using J_1 IVs.

② Test $H_0 : E\{\underset{J_2 \times 1}{Z_{2i}} u_i\} = 0$, given a true model $Y_i = X_i'\beta + u_i$

- Moment condition : $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} (Y_i - X_i' \hat{\beta}_{gmm})$ where $\hat{\beta}$ is the optimal GMM using Z_{1i} as IV.

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} (X'\beta + u_i - X_i' \hat{\beta}_{gmm})$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} u_i - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} X_i' (\hat{\beta} - \beta)$$

$$= \boxed{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} u_i} - \left(\frac{1}{N} \sum_{i=1}^{N} Z_{2i} X_i'\right) \cdot \boxed{\sqrt{N}(\hat{\beta} - \beta)}$$

→ Need to study the joint distribution of $\begin{pmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{2i} u_i \\ \sqrt{N}(\hat{\beta} - \beta) \end{pmatrix}$

Note <Asymptotic property of GMM>

$$\hat{\beta}_{GMM} = (\mathbb{X}'\mathbb{Z} W \mathbb{Z}'\mathbb{X})^{-1} \mathbb{X}'\mathbb{Z} W \mathbb{Z}'\mathbb{Y}.$$

↓ When $W = \Omega^{-1}$, it is optimal

$$= (\mathbb{X}'\mathbb{Z} \hat{\Omega}^{-1} \mathbb{Z}'\mathbb{X})^{-1} \mathbb{X}'\mathbb{Z} \hat{\Omega}^{-1} \mathbb{Z}'\mathbb{Y} \quad \text{where } \hat{\Omega}^{-1} \xrightarrow{p} \Omega^{-1}$$

$$\sqrt{N}(\hat{\beta} - \beta) = \sqrt{N}(\mathbb{X}'\mathbb{Z} \hat{\Omega}^{-1} \mathbb{Z}'\mathbb{X})^{-1} \mathbb{X}'\mathbb{Z} \hat{\Omega}^{-1} \mathbb{Z}'\mathbb{U}$$

$$= \left(\frac{\mathbb{X}'\mathbb{Z}}{N} \hat{\Omega}^{-1} \frac{\mathbb{Z}'\mathbb{X}}{N}\right)^{-1} \frac{\mathbb{X}'\mathbb{Z}}{N} \hat{\Omega}^{-1} \frac{1}{\sqrt{N}} \mathbb{Z}'\mathbb{U}.$$

$$\xrightarrow{d} (G'\Omega^{-1}G)^{-1} G'\Omega^{-1} \cdot N(0, E(u_i^2 Z_i Z_i'))$$

where $G = E(Z_i X_i')$ and $\Omega = E(u_i^2 Z_i Z_i')$

Therefore,

$$\sqrt{N}(\hat{\beta}-\beta) = \left(\frac{1}{N}X'Z\,\hat{\Omega}^{-1}\frac{1}{N}Z'X\right)^{-1}\frac{1}{N}X'Z\,\hat{\Omega}^{-1}\frac{1}{\sqrt{N}}Z'u.$$

$$= \left(\frac{1}{N}X'Z\,\hat{\Omega}^{-1}\frac{1}{N}Z'X\right)^{-1}\frac{1}{N}X'Z\,\hat{\Omega}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i u_i$$

$$= (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i u_i$$

$$+\Big\{\underbrace{\left(\frac{X'Z}{N}\hat{\Omega}^{-1}\frac{Z'X}{N}\right)^{-1}\frac{X'Z}{N}\hat{\Omega}^{-1} - (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}}_{①}\Big\}\underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i u_i}_{②}$$

① $\left(\frac{X'Z}{N}\hat{\Omega}^{-1}\frac{Z'X}{N}\right)^{-1}\frac{X'Z}{N}\hat{\Omega}^{-1} \xrightarrow{p} (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}$ where $G = E(Z_i X_i')$ (by WLLN.)

② $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i u_i \xrightarrow{d} N(0,\ E(u_i Z_i Z_i'))$ by CLT.

Thus, ① $= o_p(1)$ ② $= O_p(1)$.

$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}A Z_i u_i + o_p(1) \quad \text{where } A := (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}.$$

⇒ Following the above note,

when we use Z_{1i} to get $\hat{\beta}$, $\sqrt{N}(\hat{\beta}-\beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}A Z_{1i}u_i + o_p(1)$.

Thus, the joint distribution of $\begin{pmatrix}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_{2i}u_i \\ \sqrt{N}(\hat{\beta}-\beta)\end{pmatrix}$ is as follows:

$$\boxed{\begin{pmatrix}\sqrt{N}(\hat{\beta}-\beta) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_{2i}u_i\end{pmatrix}} = \begin{pmatrix}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}A Z_{1i}u_i \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_{2i}u_i\end{pmatrix} + o_p(1) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix}A Z_{1i} \\ Z_{2i}\end{pmatrix}u_i + o_p(1)$$

under the null hypothesis that $E(u_i Z_{2i}) = 0$

and maintaining the assumption $E(u_i Z_{1i}) = 0$.

→ point Using this joint distribution, check $E(u_i Z_{2i}) = 0$.

$$\begin{pmatrix} \sqrt{N}(\hat{\beta}-\beta) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}u_i \end{pmatrix} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} AZ_{1i} \\ Z_{2i}\end{pmatrix} u_i + o_p(1) \xrightarrow{d} \underline{N(0,\Sigma)} \text{ by CLT } \quad (**)$$

$$\text{where } \Sigma = \left(\begin{array}{c|c} A E(Z_{1i}Z_{1i}'u_i^2)A' & A E(Z_{1i}Z_{2i}'u_i^2) \\ \hline E(Z_{2i}Z_{1i}'u_i^2)A' & E(Z_{2i}Z_{2i}'u_i^2) \end{array}\right)$$

and $A E(Z_{1i}Z_{1i}'u_i^2)A' = (G'\Omega^{-1}G)^{-1}$ from step ①

⇒ Now, we get the joint distribution of $\begin{pmatrix} \sqrt{N}(\hat{\beta}-\beta) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}u_i \end{pmatrix}$ and then let's go back to the ⊛.

$$\boxed{⊛ = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}(Y_i - X_i'\hat{\beta}_{gmm})} = \boxed{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}u_i} - \left(\frac{1}{N}\sum_{i=1}^{N} Z_{2i}X_i'\right)\cdot \boxed{\sqrt{N}(\hat{\beta}-\beta)}$$

$$= \underset{J_2\times K}{\left(-\frac{1}{N}\sum_{i=1}^{N} Z_{2i}X_i'\right)}\cdot \underset{K\times 1}{\sqrt{N}(\hat{\beta}-\beta)} + \underset{J_2\times 1}{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}u_i}$$

$$= \underbrace{\Big[\underset{J_2\times K}{-\frac{1}{N}\sum_{i=1}^{N} Z_{2i}X_i'} \;\Big|\; \underset{J_2\times J_2}{I}\Big]}_{\downarrow p} \underbrace{\begin{pmatrix} \sqrt{N}(\hat{\beta}-\beta) \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}u_i \end{pmatrix}}_{\downarrow d \text{ By } (**)} \text{ by continuous mapping theorem,}$$

$$\xrightarrow{d} \underbrace{[-E(Z_{2i}X_i') \;\vdots\; I]}_{:=H,\ J_2\times(K+J_2) = (J-J_1)\times(K+J-J_1)} \cdot N(0,\Sigma) = \boxed{N(0, \underbrace{H\Sigma H'}_{J_2\times(K+J_2)\cdot(K+J_2)\times(K+J_2)\cdot(K+J_2)\times J_2 = J_2\times J_2 = (J-J_1)\times(J-J_1)})}$$

Thus, we get the moment condition for C-test such that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}\hat{u}_i \sim N(0, H\Sigma H')$.

C-test Statistics : $\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}\hat{u}_i\right]' \widehat{(H\Sigma H')^{-1}} \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_{2i}\hat{u}_i\right] \sim \chi^2_{(J-J_1)}$

where $\hat{u}_i = y_i - x_i'\hat{\beta}_{GMM}$ such that $\hat{\beta}_{GMM}$ is the optimal GMM using Z_{1i} as IVs.

$\widehat{(H\Sigma H')^{-1}}$ is constructed as $(\hat{H}\hat{\Sigma}\hat{H}')^{-1}$.

$$\hat{H} = \left[-\frac{1}{N}\sum_{i=1}^{N} Z_{2i} X_i' \;\vdots\; I \right]_{(J-J_1)\times(K+J-J_1)}$$

$$\hat{\Sigma} = \begin{pmatrix} (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1} & \hat{A}\,E(Z_{1i}Z_{2i}'U_i^2) \\ \hat{E}(Z_{2i}Z_{1i}'U_i^2)\hat{A}' & \hat{E}(Z_{2i}Z_{2i}'U_i^2) \end{pmatrix}$$

$$\hat{A}\hat{E}(Z_{1i}Z_{1i}'U_i^2)\hat{A}' = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}\hat{\Omega}\hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}$$

$$\hat{G} = \frac{1}{N}\sum_{i=1}^{N} Z_{1i}X_i'$$

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N} Z_{1i}Z_{1i}'\hat{U}_i^2$$

$$\hat{A} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}$$

$$\hat{E}(Z_{1i}Z_{2i}'U_i^2) = \frac{1}{N}\sum_{i=1}^{N} Z_{1i}Z_{2i}'\hat{U}_i^2$$

$$\hat{E}(Z_{2i}Z_{2i}'U_i^2) = \frac{1}{N}\sum_{i=1}^{N} Z_{2i}Z_{2i}'\hat{U}_i^2$$

and use $\chi^2_{(J-J_1)}$

Typically, $J_1 = K$ is the case of "optimal GMM = IV"

so there is no need to estimate Ω under Homoskedasticity.

2 Endogeneity Test

; The test of overidentifying restrictions can be used to test endogeneity of regressors.

· If there is no endogeneity, one can use all regressors as IV.

(Note) < Endogeneity Test >

① Estimate β using reliable IVs.

② Test endogeneity of the variables you are concerned about.

• Test of Endogeneity Using 2SLS. [Telser's Method].

Under Homoskedasticity,

(*) ··· $y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_i$ and when $E(x_{2i}u_i) \neq 0$,

$x_{2i} = x_{1i}'\Pi_{21} + Z_{2i}'\Pi_{22} + v_i$: Reduced form for x_{2i}

x_{1i} → Valid IV ($\because E(x_{1i}u_i)=0$)

Z_{2i} → Additional IV ($E(Z_{2i}u_i)$ should be zero).

• Endogeneity : u_i, v_i correlated. (This implies x_{2i}, u_i correlated such that $E(x_{2i}u_i) \neq 0$)

Thus, Let's think of u_i as two parts.

$$u_i = \underbrace{v_i'\alpha}_{\text{correlated part}} + \underbrace{\varepsilon_i}_{\text{Not correlated part with } u_i}$$

⇒ In fact, we can guess v_i by the above reduced form.

Thus, $\alpha = (v_i v_i')^{-1} E(v_i u_i)$

From ✱,

$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + V_i'\alpha + \varepsilon_i$ where ε_i is not correlated with x_{1i}, x_{2i}, and V_i ⋯ ✱✱

$H_0: \alpha = 0$ is the test of Endogeneity.

reject the null : $\exists$ Endogeneity
cannot reject H_0 : No Endogeneity

$\Rightarrow$ The idea is to include estimated V_i, $\hat{V}_i = X_{2i} - X_{1i}'\hat{\Gamma}_{21} - Z_{2i}'\hat{\Gamma}_{22}$ in the regression of y_i on x_{1i} & x_{2i}.

: $\hat{\beta}_1$ & $\hat{\beta}_2$ obtained in this way is the 2SLS.

$$\mathbb{Y} = \mathbb{X}_1\hat{\beta}_1 + \mathbb{X}_2\hat{\beta}_2 + \hat{\mathbb{V}}\hat{\alpha} + \hat{\mathbb{\varepsilon}}$$ where $\hat{\mathbb{\varepsilon}}$: OLS residual from the regression including $\hat{\mathbb{V}}$.

$$\hat{\mathbb{V}} = \mathbb{X}_2 - \underbrace{\mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{X}_2}_{=\hat{\mathbb{X}}_2}$$

$x_{2i} = x_{1i}'\Gamma_{21} + Z_{2i}'\Gamma_{22} + V_i$ and let $x_{1i} = Z_{1i}$ then $Z_i = (Z_{1i}\ Z_{2i})$

$$x_{2i} = (x_{1i}\ x_{2i})'\underbrace{\begin{pmatrix}\Gamma_{21}\\ \Gamma_{22}\end{pmatrix}}_{:=\Gamma} + V_i = (Z_{1i}\ Z_{2i})'\begin{pmatrix}\Gamma_{21}\\ \Gamma_{22}\end{pmatrix} + V_i$$

$$\mathbb{X}_2 = \mathbb{Z}\Gamma + \mathbb{V} \longrightarrow \hat{\Gamma} = (\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{X}_2$$ Then, $\mathbb{X}_2 = \mathbb{Z}\hat{\Gamma} + \hat{\mathbb{V}}$

$$\hat{\mathbb{V}} = \mathbb{X}_2 - \mathbb{Z}\cdot\hat{\Gamma} = \mathbb{X}_2 - \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{X}_2$$

Note

$$\mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\cdot\mathbb{Z} = \mathbb{Z}$$

$$[\mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'[\mathbb{Z}_1\ \mathbb{Z}_2]] = [\mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{Z}_1 \vdots \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{Z}_2] = [\mathbb{Z}_1\ \mathbb{Z}_2]$$

$$\hat{\mathbb{V}} = (I - \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}')\mathbb{X}_2 = (I - P_Z)\mathbb{X}_2$$

Thus, $\mathbb{Y} = \mathbb{X}_1\hat{\beta}_1 + \mathbb{X}_2\hat{\beta}_2 + (I - P_Z)\mathbb{X}_2\cdot\hat{\alpha} + \hat{\mathbb{\varepsilon}}$

Multiplying $\mathbb{X}'P_Z$,

$$\mathbb{X}'P_Z\mathbb{Y} = \mathbb{X}'P_Z\mathbb{X}_1\hat{\beta}_1 + \mathbb{X}'P_Z\mathbb{X}_2\hat{\beta}_2 + \mathbb{X}'P_Z(I-P_Z)\mathbb{X}_2\cdot\hat{\alpha} + \mathbb{X}'P_Z\hat{\mathbb{e}}$$

$$= (\mathbb{X}'P_Z\mathbb{X}_1 \quad \mathbb{X}'P_Z\mathbb{X}_2)\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} + \mathbb{X}'\underbrace{P_Z(I-P_Z)}_{=0 \text{ by orthogonality.}}\mathbb{X}_2\cdot\hat{\alpha} + \mathbb{X}'P_Z\hat{\mathbb{e}}$$

$$= \mathbb{X}'P_Z[\mathbb{X}_1 \quad \mathbb{X}_2]\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} + \mathbb{X}'P_Z\hat{\mathbb{e}} = \mathbb{X}'P_Z\mathbb{X}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} + \underset{\odot}{\mathbb{X}'P_Z\hat{\mathbb{e}}}$$

$\odot$ check it!

$\mathbb{X} = [\mathbb{X}_1 \quad \mathbb{X}_2] = [\mathbb{Z}_1 \quad \mathbb{X}_2]$ since $x_{1i} = z_i$

$$\mathbb{X}'P_Z = \begin{pmatrix}\mathbb{Z}_1'\\ \mathbb{X}_2'\end{pmatrix}P_Z = \begin{pmatrix}\mathbb{Z}_1'P_Z\\ \mathbb{X}_2'P_Z\end{pmatrix} = \begin{pmatrix}\mathbb{Z}_1'\\ \mathbb{X}_2'P_Z\end{pmatrix}$$

$\hat{\mathbb{e}}$ is orthogonal to $\underset{\mathbb{Z}_1}{\overset{\|}{\mathbb{X}_1}}$, $\mathbb{X}_2$, and $\underset{\mathbb{X}_2-\hat{\mathbb{X}}_2}{\overset{\|}{\hat{\mathbb{V}}}}$ from (**)

i.e., $\mathbb{Z}_1'\hat{\mathbb{e}}=0$, $\mathbb{X}_2'\hat{\mathbb{e}}=0$, $\hat{\mathbb{V}}'\hat{\mathbb{e}} = \mathbb{X}_2'\hat{\mathbb{e}} - \hat{\mathbb{X}}_2'\hat{\mathbb{e}} = 0$.

Thus, $\hat{\mathbb{X}}_2'\hat{\mathbb{e}}=0$.

Therefore, $\mathbb{X}'P_Z\hat{\mathbb{e}} = [\mathbb{X}_1 \quad \mathbb{X}_2]'P_Z\hat{\mathbb{e}}$

$$= \begin{bmatrix}\mathbb{Z}_1'P_Z\hat{\mathbb{e}}\\ \mathbb{X}_2'P_Z\hat{\mathbb{e}}\end{bmatrix} = \begin{bmatrix}\mathbb{Z}_1'\hat{\mathbb{e}}\\ \hat{\mathbb{X}}_2'\hat{\mathbb{e}}\end{bmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$

$$= \mathbb{X}'P_Z\mathbb{X}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} + \begin{pmatrix}0\\ 0\end{pmatrix} = \mathbb{X}'P_Z\mathbb{X}\cdot\hat{\beta}$$

Therefore, $\hat{\beta} = (\mathbb{X}'P_Z\mathbb{X})^{-1}\mathbb{X}'P_Z\mathbb{Y} = \hat{\beta}_{2SLS}$.