- Note that 2SLS is equivalent to IV in the case of $Z = \hat{X}$, where this Z is different from the previous Z.

This Z is a $n \times k$ matrix, while the previous Z is a $n \times r$ matrix.

Z in the IV estimator is replaced by $\hat{X}$.

Then,

$$\beta_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'y = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y = \beta_{GMM}.$$

GMM is interpreted as the GLS applied to MM.

Serially Correlated Errors (Time Series Data):

- Suppose that $u_1, u_2, \cdots, u_n$ are serially correlated.

Consider the case where the subscript represents time.

Remember that $\beta_{GMM} \sim N\Big(\beta,\ \sigma^2(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}\Big)$,

We need to consider evaluation of $\sigma^2 Z'\Omega Z = \mathrm{V}(u^*)$, i.e.,

$$\begin{aligned}
\mathrm{V}(u^*) &= \mathrm{V}(Z'u) = \mathrm{V}(\sum_{i=1}^{n} z_i' u_i) = \mathrm{V}(\sum_{i=1}^{n} v_i) \\
&= \mathrm{E}\Big((\sum_{i=1}^{n} v_i)(\sum_{i=1}^{n} v_i)'\Big) = \mathrm{E}\Big((\sum_{i=1}^{n} v_i)(\sum_{j=1}^{n} v_j)'\Big) \\
&= \mathrm{E}(\sum_{i=1}^{n}\sum_{j=1}^{n} v_i v_j') = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}(v_i v_j')
\end{aligned}$$

where $v_i = z_i' u_i$ is a $r \times 1$ vector.

Define $\Gamma_\tau = \mathrm{E}(v_i v'_{i-\tau})$.

$\Gamma_0 = \mathrm{E}(v_i v'_i)$ represents the $r \times r$ variance-covariance matrix of v_i.

$\Gamma_{-\tau} = \mathrm{E}(v_{i-\tau} v'_i) = \mathrm{E}((v_i v'_{i-\tau})') = \left(\mathrm{E}(v_i v'_{i-\tau})\right)' = \Gamma'_\tau$.

$$\begin{aligned}
\mathrm{V}(u^*) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}(v_i v'_j) \\
&= \mathrm{E}(v_1 v'_1) + \mathrm{E}(v_1 v'_2) + \mathrm{E}(v_1 v'_3) + \cdots + \mathrm{E}(v_1 v'_n) \\
&\quad + \mathrm{E}(v_2 v'_1) + \mathrm{E}(v_2 v'_2) + \mathrm{E}(v_2 v'_3) + \cdots + \mathrm{E}(v_2 v'_n) \\
&\quad + \mathrm{E}(v_3 v'_1) + \mathrm{E}(v_3 v'_2) + \mathrm{E}(v_3 v'_3) + \cdots + \mathrm{E}(v_3 v'_n) \\
&\qquad\qquad \vdots \\
&\quad + \mathrm{E}(v_n v'_1) + \mathrm{E}(v_n v'_2) + \mathrm{E}(v_n v'_3) + \cdots + \mathrm{E}(v_n v'_n) \\
&= \Gamma_0 + \Gamma_{-1} + \Gamma_{-2} + \cdots + \Gamma_{1-n} \\
&\quad + \Gamma_1 + \Gamma_0 + \Gamma_{-1} + \cdots + \Gamma_{2-n}
\end{aligned}$$

$$
\begin{aligned}
&+ \Gamma_2 + \Gamma_1 + \Gamma_0 + \cdots + \Gamma_{3-n} \\
&\qquad\quad \vdots \\
&+ \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0 \\
&= \Gamma_0 + \Gamma'_1 + \Gamma'_2 + \cdots + \Gamma'_{n-1} \\
&+ \Gamma_1 + \Gamma_0 + \Gamma'_1 + \cdots + \Gamma'_{n-2} \\
&+ \Gamma_2 + \Gamma_1 + \Gamma_0 + \cdots + \Gamma'_{n-3} \\
&\qquad\quad \vdots \\
&+ \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0 \\
&= n\Gamma_0 + (n-1)(\Gamma_1 + \Gamma'_1) + (n-2)(\Gamma_2 + \Gamma'_2) + \cdots + (\Gamma_{n-1} + \Gamma'_{n-1}) \\
&= n\Gamma_0 + \sum_{i=1}^{n-1}(n-i)(\Gamma_i + \Gamma'_i)
\end{aligned}
$$

$$= n\Big(\Gamma_0 + \sum_{i=1}^{n-1}(1 - \frac{i}{n})(\Gamma_i + \Gamma_i')\Big)$$

$$\approx n\Big(\Gamma_0 + \sum_{i=1}^{q}(1 - \frac{i}{q+1})(\Gamma_i + \Gamma_i')\Big).$$

In the last line, $n-1$ is replaced by q, where $q < n-1$.

We need to estimate Γ_τ as: $\hat{\Gamma}_\tau = \frac{1}{n}\sum_{i=\tau+1}^{n} \hat{v}_i \hat{v}_{i-\tau}'$, where $\hat{v}_i = z_i' \hat{u}_i$ for $\hat{u}_i = y_i - x_i \beta_{GMM}$.

As τ is large, $\hat{\Gamma}_\tau$ is unstable.

Therefore, we choose the q which is less than $n-1$.

Hansen's *J* Test: Is the model specification correct?

That is, is $\mathrm{E}(z'u) = 0$ for $y = x\beta + u$ correct?

H_0 : $\mathrm{E}(z'u) = 0$ (The model is correct. Or, the instrumental variables are appropriate.)

H_1 : $\mathrm{E}(z'u) \neq 0$

The number of equations is r, while the number of parameters is k.

The degree of freedom is $r - k$.

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i'\hat{u}_i\Big)'\Big(\widehat{\mathrm{V}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}} z_i\hat{u}_i)\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i'\hat{u}_i\Big) \longrightarrow \chi(r-k),$$

where $\hat{u}_i = y_i - x_i\beta_{GMM}$.

$\widehat{\mathrm{V}(\frac{1}{n}\sum_{i=1}^{n}} z_i'\hat{u}_i)$ indicates the estimate of $\mathrm{V}(\frac{1}{n}\sum_{i=1}^{n} z_i'u_i)$ for $u_i = y_i - x_i\beta$.

The *J* test is called a test for over-identifying restrictions (過剰識別制約).

Remark 1: $X_1, X_2, \cdots, X_n$ are mutually independent.

$X_i \sim N(\mu, \sigma^2)$ are assumed.

Consider $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Then, $\dfrac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1)$.

That is, $\sqrt{n}(\overline{X} - \mu) \longrightarrow N(0, \sigma^2)$.

Remark 2: $X_1, X_2, \cdots, X_n$ are mutually independent.

$X_i \sim N(\mu, \sigma^2)$ are assumed.

Then, $\Big(\dfrac{X_i - \mu}{\sigma^2}\Big)^2 \sim \chi^2(1)$ and $\sum_{i=1}^{n}\Big(\dfrac{X_i - \mu}{\sigma^2}\Big)^2 \sim \chi^2(n)$.

If μ is replaced by its estimator $\overline{X}$, then $\sum_{i=1}^{n}\Big(\dfrac{X_i - \overline{X}}{\sigma^2}\Big)^2 \sim \chi^2(n-1)$.

Note:

$$\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma^2}\right)^2=\begin{pmatrix} X_i-\overline{X} \\ X_i-\overline{X} \\ \vdots \\ X_n-\overline{X} \end{pmatrix}'\begin{pmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix}^{-1}\begin{pmatrix} X_i-\overline{X} \\ X_i-\overline{X} \\ \vdots \\ X_n-\overline{X} \end{pmatrix} \sim \chi^2(n-1)$$

In the case of GMM,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' u_i \longrightarrow N(0, \Sigma),$$

where $\Sigma = \mathrm{V}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' u_i\Big)$.

Therefore, we obtain: $\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' u_i\Big)'\Sigma^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' u_i\Big) \longrightarrow \chi^2(r)$.

In order to obtain $\hat{u}_i$, we have to estimate β, which is a $k \times 1$ vector.

Therefore, replacing u_i by $\hat{u}_i$, we have: $\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' \hat{u}_i\Big)'\Sigma^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' \hat{u}_i\Big) \longrightarrow \chi^2(r-k)$.

Moreover, from $\hat{\Sigma} \longrightarrow \Sigma$, we obtain: $\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' \hat{u}_i\Big)'\hat{\Sigma}^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i' \hat{u}_i\Big) \longrightarrow \chi^2(r-k)$,

where $\hat{\Sigma}$ is a consistent estimator of Σ.

4.3 Generalized Method of Moments (GMM, 一般化積率法) II — Nonlinear Case —

Consider the general case:

$$\mathrm{E}(h(\theta; w)) = 0,$$

which is the orthogonality condition.

A $k \times 1$ vector θ denotes a parameter to be estimated.

$h(\theta; w)$ is a $r \times 1$ vector for $r \geq k$.

Let $w_i = (y_i, x_i)$ be the ith observed data, i.e., the ith realization of w.

Define $g(\theta; W)$ as:

$$g(\theta; W) = \frac{1}{n} \sum_{i=1}^{n} h(\theta; w_i),$$

where $W = \{w_n, w_{n-1}, \cdots, w_1\}$.

$g(\theta; W)$ is a $r \times 1$ vector for $r \geq k$.

Let $\hat{\theta}$ be the GMM estimator which minimizes:

$$g(\theta; W)' S^{-1} g(\theta; W),$$

with respect to θ.

- Solve the following first-order condition:

$$\frac{\partial g(\theta; W)'}{\partial \theta} S^{-1} g(\theta; W) = 0,$$

with respect to θ. There are r equations and k parameters.

Computational Procedure:

Linearizing the first-order condition around $\theta = \hat{\theta}$,

$$\begin{aligned}
0 &= \frac{\partial g(\theta; W)'}{\partial \theta} S^{-1} g(\theta; W) \\
&\approx \frac{\partial g(\hat{\theta}; W)'}{\partial \theta} S^{-1} g(\hat{\theta}; W) + \frac{\partial g(\hat{\theta}; W)'}{\partial \theta} S^{-1} \frac{\partial g(\hat{\theta}; W)}{\partial \theta'} (\theta - \hat{\theta}) \\
&= \hat{D}' S^{-1} g(\hat{\theta}; W) + \hat{D}' S^{-1} \hat{D} (\theta - \hat{\theta}),
\end{aligned}$$

where $\hat{D} = \dfrac{\partial g(\hat{\theta}; W)}{\partial \theta'}$, which is a $r \times k$ matrix.

Note that in the second term of the second line the second derivative is ignored and omitted.

Rewriting, we have the following equation:

$$\theta - \hat{\theta} = -(\hat{D}' S^{-1} \hat{D})^{-1} \hat{D}' S^{-1} g(\hat{\theta}; W).$$

Replacing θ and $\hat{\theta}$ by $\hat{\theta}^{(i+1)}$ and $\hat{\theta}^{(i)}$, respectively, we obtain:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - (\hat{D}^{(i)\prime} S^{-1} \hat{D}^{(i)})^{-1} \hat{D}^{(i)\prime} S^{-1} g(\hat{\theta}^{(i)}; W),$$

where $\hat{D}^{(i)} = \dfrac{\partial g(\hat{\theta}^{(i)}; W)}{\partial \theta'}$.

Given S, repeat the iterative procedure for $i = 1, 2, 3, \cdots$, until $\hat{\theta}^{(i+1)}$ is equal to $\hat{\theta}^{(i)}$.

How do we derive the weight matrix S?

- In the case where $h(\theta; w_i)$, $i = 1, 2, \cdots, n$, are mutually independent, S is:

$$\begin{aligned}
S &= \mathrm{V}\Big(\sqrt{n}g(\theta; W)\Big) = n\mathrm{E}\Big(g(\theta; W)g(\theta; W)'\Big) \\
&= n\mathrm{E}\Big(\Big(\frac{1}{n}\sum_{i=1}^{n} h(\theta; w_i)\Big)\Big(\frac{1}{n}\sum_{j=1}^{n} h(\theta; w_j)\Big)'\Big) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\Big(h(\theta; w_i)h(\theta; w_j)'\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\Big(h(\theta; w_i)h(\theta; w_i)'\Big),
\end{aligned}$$

which is a $r \times r$ matrix.

Note that

(i) $\mathrm{E}\Big(h(\theta; w_i)\Big) = 0$ for all i and accordingly $\mathrm{E}\Big(g(\theta; W)\Big) = 0$,

(ii) $g(\theta; W) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} h(\theta; w_i) = \dfrac{1}{n}\sum_{j=1}^{n} h(\theta; w_j)$,

(iii) $\mathrm{E}\Big(h(\theta; w_i)h(\theta; w_j)'\Big) = 0$ for $i \neq j$.

The estimator of S, denoted by $\hat{S}$ is given by: $\quad \hat{S} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_i)' \longrightarrow S$.

- Taking into account serial correlation of $h(\theta; w_i)$, $i = 1, 2, \cdots, n$, S is given by:

$$
\begin{aligned}
S &= \mathrm{V}\Big(\sqrt{n} g(\theta; W)\Big) = n\mathrm{E}\Big(g(\theta; W) g(\theta; W)'\Big) \\
&= n\mathrm{E}\Big(\Big(\frac{1}{n}\sum_{i=1}^{n} h(\theta; w_i)\Big)\Big(\frac{1}{n}\sum_{j=1}^{n} h(\theta; w_j)\Big)'\Big) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\Big(h(\theta; w_i) h(\theta; w_j)'\Big).
\end{aligned}
$$

Note that $\mathrm{E}\Big(\sum_{i=1}^{n} h(\theta; w_i)\Big) = 0$.

Define $\Gamma_\tau = \mathrm{E}\Big(h(\theta; w_i) h(\theta; w_{i-\tau})'\Big) < \infty$, i.e., $h(\theta; w_i)$ is stationary.

Stationarity:

(i) $\mathrm{E}\Big(h(\theta; w_i)\Big)$ does not depend on i,

(ii) $\mathrm{E}\Big(h(\theta; w_i) h(\theta; w_{i-\tau})'\Big)$ depends on time difference τ.

$$
\Longrightarrow \mathrm{E}\Big(h(\theta; w_i) h(\theta; w_{i-\tau})'\Big) = \Gamma_\tau
$$

$$
\begin{aligned}
S &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\Big(h(\theta; w_i)h(\theta; w_j)'\Big) \\
&= \frac{1}{n}\Big(\mathrm{E}\Big(h(\theta; w_1)h(\theta; w_1)'\Big) + \mathrm{E}\Big(h(\theta; w_1)h(\theta; w_2)'\Big) + \cdots + \mathrm{E}\Big(h(\theta; w_1)h(\theta; w_n)'\Big) \\
&\qquad \mathrm{E}\Big(h(\theta; w_2)h(\theta; w_1)'\Big) + \mathrm{E}\Big(h(\theta; w_2)h(\theta; w_2)'\Big) + \cdots + \mathrm{E}\Big(h(\theta; w_2)h(\theta; w_n)'\Big) \\
&\qquad \vdots \\
&\qquad \mathrm{E}\Big(h(\theta; w_n)h(\theta; w_1)'\Big) + \mathrm{E}\Big(h(\theta; w_n)h(\theta; w_2)'\Big) + \cdots + \mathrm{E}\Big(h(\theta; w_n)h(\theta; w_n)'\Big)\Big) \\
&= \frac{1}{n}(\Gamma_0 \quad + \Gamma_1' \quad + \Gamma_2' \quad + \cdots + \Gamma_{n-1}' \\
&\qquad \Gamma_1 \quad + \Gamma_0 \quad + \Gamma_1' \quad + \cdots + \Gamma_{n-2}' \\
&\qquad \vdots \\
&\qquad \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\Big(n\Gamma_0 + (n-1)(\Gamma_1 + \Gamma_1') + (n-2)(\Gamma_2 + \Gamma_2') + \ \cdots\ + (\Gamma_{n-1} + \Gamma_{n-1}')\Big) \\
&= \Gamma_0 + \sum_{i=1}^{n-1} \frac{n-i}{n}(\Gamma_i + \Gamma_i') = \Gamma_0 + \sum_{i=1}^{n-1}\Big(1 - \frac{i}{n}\Big)(\Gamma_i + \Gamma_i') \\
&= \Gamma_0 + \sum_{i=1}^{q}\Big(1 - \frac{i}{q+1}\Big)(\Gamma_i + \Gamma_i').
\end{aligned}
$$

Note that $\Gamma_\tau' = \mathrm{E}\Big(h(\theta; w_{i-\tau})h(\theta; w_i)'\Big) = \Gamma(-\tau)$, because $\Gamma_\tau = \mathrm{E}\Big(h(\theta; w_i)h(\theta; w_{i-\tau})'\Big)$.

In the last line, n is replaced by $q+1$, where $q < n$.

We need to estimate Γ_τ as: $\quad \hat{\Gamma}_\tau = \dfrac{1}{n}\displaystyle\sum_{i=\tau+1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_{i-\tau})'$.

As τ is large, $\hat{\Gamma}_\tau$ is unstable.

Therefore, we choose the q which is less than n.