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.

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

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

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

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

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

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

(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.

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

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

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

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

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

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

S is estimatated as:

$$\hat{S} = \hat{\Gamma}_0 + \sum_{i=1}^{q}\Big(1 - \frac{i}{q+1}\Big)(\hat{\Gamma}_i + \hat{\Gamma}_i'),$$

$\Longrightarrow$ the Newey-West Estimator

Note that $\hat{S} \longrightarrow S$, because $\hat{\Gamma}_\tau \longrightarrow \Gamma_\tau$ as $n \longrightarrow \infty$.

Asymptotic Properties of GMM:

GMM is consistent and asymptotic normal as follows:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\Big(0, (D'S^{-1}D)^{-1}\Big),$$

where D is a $r \times k$ matrix, and $\hat{D}$ is an estimator of D, defined as:

$$D = \frac{\partial g(\theta; W)}{\partial \theta'}, \qquad \hat{D} = \frac{\partial g(\hat{\theta}; W)}{\partial \theta'}.$$

Proof of Asymptotic Normality:

Assumption 1: $\hat{\theta} \longrightarrow \theta$

Assumption 2: $\sqrt{n}g(\theta; W) \longrightarrow N(0, S)$, i.e., $S = \lim_{n\to\infty} \mathrm{V}\Big(\sqrt{n}g(\theta; W)\Big)$.

The first-order condition of GMM is:

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

The GMM estimator, denote by $\hat{\theta}$, satisfies the above equation.

Therefore, we have the following:

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

Linearize $g(\hat{\theta}; W)$ around $\hat{\theta} = \theta$ as follows:

$$g(\hat{\theta}; W) = g(\theta; W) + \frac{\partial g(\overline{\theta}; W)}{\partial \theta'}(\hat{\theta} - \theta) = g(\theta; W) + \overline{D}(\hat{\theta} - \theta),$$

where $\overline{D} = \dfrac{\partial g(\overline{\theta}; W)}{\partial \theta'}$, and $\overline{\theta}$ is between $\hat{\theta}$and θ.

$\Longrightarrow$ **Theorem of Mean Value (平均値の定理)**

Substituting the linear approximation at $\hat{\theta} = \theta$, we obtain:

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

which can be rewritten as:

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

Note that $\overline{D} = \dfrac{\partial g(\overline{\theta}; W)}{\partial \theta'}$, where $\overline{\theta}$ is between $\hat{\theta}$ and θ.

From Assumption 1, $\hat{\theta} \longrightarrow \theta$ implies $\overline{\theta} \longrightarrow \theta$

Therefore,

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

Accordingly , the GMM estimator $\hat{\theta}$ has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\Big(0, (D'S^{-1}D)^{-1}\Big).$$

Note that $\hat{D} \longrightarrow D$, $\overline{D} \longrightarrow D$, $\hat{S} \longrightarrow S$ and Assumption 2 are utilized.

Computational Procedure:

(1) Compute $\hat{S}^{(i)} = \hat{\Gamma}_0 + \sum_{i=1}^{q} \Big(1 - \frac{i}{q+1}\Big)(\hat{\Gamma}_i + \hat{\Gamma}_i')$, where $\hat{\Gamma}_\tau = \frac{1}{n} \sum_{i=\tau+1}^{n} h(\hat{\theta}; w_i) h(\hat{\theta}; w_{i-\tau})'$.

q is set by a researcher.

(2) Use the following iterative procedure:

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

(3) Repeat (1) and (2) until $\hat{\theta}^{(i+1)}$ is equal to $\hat{\theta}^{(i)}$.

In (2), remember that when S is given we take the following iterative procedure:

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

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

- If the assumption $\mathrm{E}\big(h(\theta; w)\big) = 0$ is violated, the GMM estimator $\hat{\theta}$ is no longer consistent.

Therefore, we need to check if $\mathrm{E}\big(h(\theta; w)\big) = 0$.

From Assumption 2, note as follows:

$$J = \Big(\sqrt{n} g(\hat{\theta}; W) \Big)' \hat{S}^{-1} \Big(\sqrt{n} g(\hat{\theta}; W) \Big) \longrightarrow \chi^2(r-k),$$

which is called Hansen's J test.

Because of r equations and k parameters, the degree of freedom is given by $r-k$.

If J is small enough, we can judge that the specified model is correct.

Testing Hypothesis:

Remember that the GMM estimator $\hat{\theta}$ has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\Big(0, (D'S^{-1}D)^{-1}\Big).$$

Consider testing the following null and alternative hypotheses:

- The null hypothesis: $H_0 : R(\theta) = 0,$
- The alternative hypothesis: $H_1 : R(\theta) \neq 0,$

where $R(\theta)$ is a $p \times 1$ vector function for $p \leq k$.

p denotes the number of restrictions.

$R(\theta)$ is linearized as: $R(\hat{\theta}) = R(\theta) + R_{\overline{\theta}}(\hat{\theta} - \theta)$, where $R_{\overline{\theta}} = \dfrac{\partial R(\overline{\theta})}{\partial \theta'}$, which is a $p \times k$ matrix.

Note that $\overline{\theta}$ is bewteen $\hat{\theta}$ and θ. If $\hat{\theta} \longrightarrow \theta$, then $\overline{\theta} \longrightarrow \theta$ and $R_{\overline{\theta}} \longrightarrow R_\theta$.

Under the null hypothesis $R(\theta) = 0$, we have $R(\hat{\theta}) = R_{\overline{\theta}}(\hat{\theta} - \theta)$, which implies that the distribution of $R(\hat{\theta})$ is equivalent to that of $R_{\overline{\theta}}(\hat{\theta} - \theta)$.
The distribution of $\sqrt{n}R(\hat{\theta})$ is given by:

$$\sqrt{n}R(\hat{\theta}) = \sqrt{n}R_{\overline{\theta}}(\hat{\theta} - \theta) \longrightarrow N\Big(0, R_\theta(D'S^{-1}D)^{-1}R'_\theta\Big).$$

Therefore, under the null hypothesis, we have the following distribution:

$$nR(\hat{\theta})\Big(R_\theta(D'S^{-1}D)^{-1}R'_\theta\Big)^{-1}R(\hat{\theta})' \longrightarrow \chi^2(p).$$

Practically, replacing θ by $\hat{\theta}$ in R_θ, D and S, we use the following test statistic:

$$nR(\hat{\theta})\Big(R_{\hat{\theta}}(\hat{D}'\hat{S}^{-1}\hat{D})^{-1}R'_{\hat{\theta}}\Big)^{-1}R(\hat{\theta})' \longrightarrow \chi^2(p).$$

$\Longrightarrow$ Wald type test

Examples of $h(\theta; w)$:

1. **OLS:**

   Regression Model: $\quad y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ is taken as:

   $$h(\theta; w_i) = x_i'(y_i - x_i\beta).$$

2. **IV (Instrumental Variable, 操作変数法):**

   Regression Model: $\quad y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) \neq 0, \quad \mathrm{E}(z_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ is taken as:

   $$h(\theta; w_i) = z_i'(y_i - x_i\beta),$$

   where z_i is a vector of instrumental variables.

When z_i is a $1 \times k$ vector, the GMM of β is equivalent to the instrumental variable (IV) estimator.

When z_i is a $1 \times r$ vector for $r > k$, the GMM of β is equivalent to the two-stage least squares (2SLS) estimator.

3. **NLS (Nonlinear Least Squares, 非線形最小二乗法)**:

Regression Model: $\quad f(y_i, x_i, \beta) = \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) \neq 0, \quad \mathrm{E}(z_i'\epsilon_i) = 0$

$h(\theta; w_i)$ is taken as:

$$h(\theta; w_i) = z_i' f(y_i, x_i, \beta)$$

where z_i is a vector of instrumental variables.