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.

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

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

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

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.

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:

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

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