9　周波数領域

周波数領域（Frequency Domain）：

1. スペクトラム (パワー・スペクトラム) の定義：

$$\begin{aligned} f(\lambda) &= (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \cos(\lambda\tau) \\ &= (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \exp(-i\lambda\tau) \end{aligned}$$

2. 三角関数と指数関数： $\exp(i\theta) = \cos(\theta) + i\sin(\theta)$

$\Longrightarrow$

$$\cos(\theta) = \frac{1}{2}\Big(\exp(i\theta) + \exp(-i\theta)\Big), \qquad \sin(\theta) = \frac{1}{2i}\Big(\exp(i\theta) - \exp(-i\theta)\Big)$$

加法定理： $\exp\Big(i(\theta_1+\theta_2)\Big) = \exp(i\theta_1)\exp(i\theta_2)$

$\Longrightarrow$

$$\exp\Big(i(\theta_1+\theta_2)\Big) = \cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)$$

$$\begin{aligned}\exp(i\theta_1)\exp(i\theta_2) &= \Big(\cos(\theta_1)+i\sin(\theta_1)\Big)\Big(\cos(\theta_2)+i\sin(\theta_2)\Big)\\ &= \cos(\theta_1)\cos(\theta_2)-\sin(\theta_1)\sin(\theta_2)+i\Big(\cos(\theta_1)\sin(\theta_2)+\sin(\theta_1)\cos(\theta_2)\Big)\end{aligned}$$

よって，

$$\begin{aligned}\cos(\theta_1+\theta_2) &= \cos(\theta_1)\cos(\theta_2)-\sin(\theta_1)\sin(\theta_2)\\ \sin(\theta_1+\theta_2) &= \cos(\theta_1)\sin(\theta_2)+\sin(\theta_1)\cos(\theta_2)\end{aligned}$$

が得られる。

3. y_t がホワイト・ノイズであれば， $f(\lambda) = (2\pi)^{-1}\sigma_\epsilon^2$

4. スペクトラムと自己相関関数との関係

$$\gamma(\tau) = \int_{-\pi}^{\pi} f(\lambda)\cos(\lambda\tau)\mathrm{d}\lambda$$

従って，スペクトラムは自己相関関数のすべての情報を持っている。

5. $\sum w_j^2 < \infty$ とする。

$$y_t = \sum_{j=-r}^{s} w_j x_{t-j}$$

$f_x(\lambda)$ を x_t のスペクとラムとする。$W(\lambda)$ を次のように定義する。

$$W(\lambda) = \sum_{j=-r}^{s} w_j e^{-i\lambda j}$$

このとき，y_t のスペクとラムは以下のようになる。

$$f_y(\lambda) = |W(\lambda)|^2 f_x(\lambda)$$

$|W(\lambda)|^2$ は伝達関数 (transfer function) と呼ばれ,

$$\begin{aligned} |W(\lambda)|^2 &= W(\lambda)\overline{W(\lambda)} \\ &= \sum_{j=-r}^{s} w_j e^{-i\lambda j} \sum_{j=-r}^{s} w_j e^{i\lambda j} \end{aligned}$$

$\overline{W(\lambda)}$ は $W(\lambda)$ の共役複素数とする。

6. MA(q) モデルの場合：

$$\begin{aligned} y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ &= (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t \\ &= \theta(L)\epsilon_t \end{aligned}$$

$y_t = \theta(L)\epsilon_t$ のとき，ϵ_t のパワー・スペクトラム $f_\epsilon(\lambda)$ から y_t のパワー・スペクトラム $f_y(\lambda)$ への変換：

$$\begin{aligned} f_y(\lambda) &= \theta(e^{-i\lambda})\theta(e^{i\lambda}) f_\epsilon(\lambda) \\ &= \theta(e^{-i\lambda})\theta(e^{i\lambda}) \frac{\sigma_\epsilon^2}{2\pi} \end{aligned}$$

$\rightarrow$ 基本パターン

7. AR(p) モデルの場合:

$$\phi(L) y_t = \epsilon_t$$

$$y_t = \phi(L)^{-1} \epsilon_t$$

$\phi(L)y_t = \epsilon_t$ のとき，ϵ_t のパワー・スペクトラム $f_\epsilon(\lambda)$ から y_t のパワー・スペクトラム $f_y(\lambda)$ への変換：

$$\begin{aligned} f_y(\lambda) &= \frac{1}{\phi(e^{-i\lambda})\phi(e^{i\lambda})} f_\epsilon(\lambda) \\ &= \frac{1}{\phi(e^{-i\lambda})\phi(e^{i\lambda})} \frac{\sigma_\epsilon^2}{2\pi} \end{aligned}$$

8. ARMA(p, q) モデルの場合:

$$\phi(L)y_t = \theta(L)\epsilon_t$$

$$y_t = \phi(L)^{-1}\theta(L)\epsilon_t$$

$\phi(L)y_t = \theta(L)\epsilon_t$ のとき，ϵ_t のパワー・スペクトラム $f_\epsilon(\lambda)$ から y_t のパワー・スペクトラム $f_y(\lambda)$ への変換：

$$\begin{aligned} f_y(\lambda) &= \frac{\theta(e^{-i\lambda})\theta(e^{i\lambda})}{\phi(e^{-i\lambda})\phi(e^{i\lambda})} f_\epsilon(\lambda) \\ &= \frac{\theta(e^{-i\lambda})\theta(e^{i\lambda})}{\phi(e^{-i\lambda})\phi(e^{i\lambda})} \frac{\sigma_\epsilon^2}{2\pi} \end{aligned}$$

10 Generalized Method of Moments (GMM, 一般化積率法)

10.1 Method of Moments (MM, 積率法)

As $n \longrightarrow \infty$, we have the result: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathrm{E}(X) = \mu$.

$\Longrightarrow$ **Law of Large Number (大数の法則)**

$X_1, X_2, \cdots, X_n$ are n realizations of X.

[Review] Chebyshev's inequality (チェビシェフの不等式) is given by:

$$P(|X-\mu| > \epsilon) \le \frac{\sigma^2}{\epsilon^2} \quad \text{or} \quad P(|X-\mu| \le \epsilon) \ge 1 - \frac{\sigma^2}{\epsilon^2},$$

where $\mu = \mathrm{E}(X)$, $\sigma^2 = \mathrm{V}(X)$ and any $\epsilon > 0$.

Note that $P(|X-\mu| > \epsilon) + P(|X-\mu| \le \epsilon) = 1$.

Replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$ by $\overline{X}$, $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \dfrac{\sigma^2}{n}$.

As $n \longrightarrow \infty$,

$$P(|\overline{X}-\mu| \le \epsilon) \ge 1 - \frac{\sigma^2}{n\epsilon^2} \longrightarrow 1.$$

That is, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$.

[End of Review]

$\overline{X}$ is an approximation of $\mathrm{E}(X) = \mu$.

Therefore, $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ is taken as an estimator of μ.

$\Longrightarrow$ $\overline{X}$ is MM estimator of $\mathrm{E}(X) = \mu$.

MM is applied to the regression model as follows:

Regression model: $y_i = x_i\beta + u_i$, where x_i and u_i are assumed to be stochastic.

Familiar Assumption: $\mathrm{E}(x'u) = 0$, called the **orthogonality condition (直交条件)**, where x is a $1 \times k$ vector and u is a scalar.

We consider that $(x_1, x_2, \cdots, x_n)$ and $(u_1, u_2, \cdots, u_n)$ are realizations generated from random variables x and u, respectively.

From the law of large number, we have the following:

$$\frac{1}{n}\sum_{i=1}^{n} x_i' u_i = \frac{1}{n}\sum_{i=1}^{n} x_i'(y_i - x_i\beta) \longrightarrow \mathrm{E}(x'u) = 0.$$

Thus, the MM estimator of β, denoted by β_{MM}, satisfies:

$$\frac{1}{n}\sum_{i=1}^{n} x_i'(y_i - x_i\beta_{MM}) = 0.$$

Therefore, β_{MM} is given by:

$$\beta_{MM} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' y_i\right) = (X'X)^{-1}X'y,$$

which is equivalent to OLS and MLE.

Note that X and y are:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

- However, β_{MM} is inconsistent when $\mathrm{E}(x'u) \neq 0$, i.e.,

$$\beta_{MM} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u = \beta + \Big(\frac{1}{n}X'X\Big)^{-1}\Big(\frac{1}{n}X'u\Big) \not\longrightarrow \beta.$$

Note as follows:

$$\frac{1}{n}X'u = \frac{1}{n}\sum_{i=1}^{n} x_i'u_i \longrightarrow \mathrm{E}(x'u) \neq 0.$$

In order to obtain a consistent estimator of β, we find the instrumental variable z which satisfies $\mathrm{E}(z'u) = 0$.

Let z_i be the ith realization of z, where z_i is a $1 \times k$ vector.

Then, we have the following:

$$\frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i'u_i \longrightarrow \mathrm{E}(z'u) = 0.$$

The β which satisfies $\frac{1}{n}\sum_{i=1}^{n} z_i'u_i = 0$ is denoted by β_{IV}, i.e., $\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta_{IV}) = 0$.

Thus, β_{IV} is obtained as:

$$\beta_{IV} = \Big(\frac{1}{n}\sum_{i=1}^{n} z_i'x_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} z_i'y_i\Big) = (Z'X)^{-1}Z'y.$$

Note that $Z'X$ is a $k \times k$ square matrix, where we assume that the inverse matrix of $Z'X$ exists.

Assume that as n goes to infinity there exist the following moment matrices:

$$\frac{1}{n}\sum_{i=1}^{n} z_i' x_i = \frac{1}{n} Z'X \longrightarrow M_{zx},$$

$$\frac{1}{n}\sum_{i=1}^{n} z_i' z_i = \frac{1}{n} Z'Z \longrightarrow M_{zz},$$

$$\frac{1}{n}\sum_{i=1}^{n} z_i' u_i = \frac{1}{n} Z'u \longrightarrow 0.$$

As n goes to infinity, β_{IV} is rewritten as:

$$\begin{aligned} \beta_{IV} &= (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta + u) = \beta + (Z'X)^{-1}Z'u \\ &= \beta + (\frac{1}{n}Z'X)^{-1}(\frac{1}{n}Z'u) \longrightarrow \beta + M_{zx} \times 0 = \beta, \end{aligned}$$

Thus, β_{IV} is a consistent estimator of β.

- We consider the asymptotic distribution of β_{IV}.

By the central limit theorem,

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0, \sigma^2 M_{zz})$$

Note that $\mathrm{V}(\frac{1}{\sqrt{n}}Z'u) = \frac{1}{n}\mathrm{V}(Z'u) = \frac{1}{n}\mathrm{E}(Z'uu'Z) = \frac{1}{n}\mathrm{E}\Big(\mathrm{E}(Z'uu'Z|Z)\Big)$
$= \frac{1}{n}\mathrm{E}\Big(Z'\mathrm{E}(uu'|Z)Z\Big) = \frac{1}{n}\mathrm{E}(\sigma^2 Z'Z) = \mathrm{E}(\sigma^2\frac{1}{n}Z'Z) \longrightarrow \mathrm{E}(\sigma^2 M_{zz}) = \sigma^2 M_{zz}.$

We obtain the following asymmptotic distribution:

$$\sqrt{n}(\beta_{IV} - \beta) = (\frac{1}{n}Z'X)^{-1}(\frac{1}{\sqrt{n}}Z'u) \longrightarrow N(0, \sigma^2 M_{zx}^{-1} M_{zz} M_{zx}^{-1\prime})$$

Practically, for large n we use the following distribution:

$$\beta_{IV} \sim N\Big(\beta, s^2(Z'X)^{-1}Z'Z(Z'X)^{-1\prime}\Big),$$

where $s^2 = \frac{1}{n-k}(y - X\beta_{IV})'(y - X\beta_{IV})$.

- In the case where z_i is a $1 \times r$ vector for $r > k$, $Z'X$ is a $r \times k$ matrix, which is not a square matrix.

$\Longrightarrow$ **Generalized Method of Moments (GMM, 一般化積率法)**

10.2 Generalized Method of Moments (GMM, 一般化積率法)

In order to obtain a consistent estimator of β, we have to find the instrumental variable z which satisfies $\mathrm{E}(z'u) = 0$.

For now, however, suppose that we have z with $\mathrm{E}(z'u) = 0$.

Let z_i be the ith realization (i.e., the ith data) of z, where z_i is a $1 \times r$ vector and $r > k$.

Then, using the law of large number, we have the following:

$$\frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i' u_i = \frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta) \longrightarrow \mathrm{E}(z'u) = 0.$$

The number of equations (i.e., r) is larger than the number of parameters (i.e., k).

Let us define W as a $r \times r$ weight matrix, which is symmetric.

We solve the following minimization problem:

$$\min_{\beta} \left(\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta)\right)' W \left(\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta)\right),$$

which is equivalent to:

$$\min_{\beta} \Big(Z'(y - X\beta)\Big)' W \Big(Z'(y - X\beta)\Big),$$

i.e.,

$$\min_{\beta} \ (y - X\beta)'ZWZ'(y - X\beta).$$

Note that $\sum_{i=1}^{n} z_i'(y_i - x_i\beta) = Z'(y - X\beta)$.

W should be the inverse matrix of the variance-covariance matrix of $Z'(y - X\beta) = Z'u$.

Suppose that $\mathrm{V}(u) = \sigma^2\Omega$.

Then, $\mathrm{V}(Z'u) = \mathrm{E}\Big(Z'u(Z'u)'\Big) = \mathrm{E}(Z'uu'Z) = Z'\mathrm{E}(uu')Z = \sigma^2 Z'\Omega Z = W^{-1}$.

The following minimization problem should be solved.

$$\min_{\beta} \ (y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta).$$

The solution of β is given by the GMM estimator, denoted by β_{GMM}.

Remark: For the model: $y = X\beta + u$ and $u \sim (0, \sigma^2 \Omega)$, solving the following minimization problem:

$$\min_{\beta} (y - X\beta)' \Omega^{-1} (y - X\beta),$$

GLS is given by:

$$b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} y.$$

Note that b is the best linear unbiased estimator.

Remark: The solution of the above minimization problem is equivalent to the GLE estimator of β in the following regression model:

$$Z'y = Z'X\beta + Z'u,$$

where Z, y, X, β and u are $n \times r$, $n \times 1$, $n \times k$, $k \times 1$ and $n \times 1$ matrices or vectors.

Note that $r > k$.

$y^* = Z'y$, $X^* = Z'X$ and $u^* = Z'u$ denote $r \times 1$, $r \times k$ and $r \times 1$ matrices or vectors, where $r > k$.

Rewrite as follows:

$$y^* = X^*\beta + u^*,$$

$\Longrightarrow$ r is taken as the sample size.

u^* is a $r \times 1$ vector.

The elements of u^* are correlated with each other, beacuse each element of u^* is a function of u_1, u_2, $\cdots$, u_n.

The variance of u^* is:

$$\mathrm{V}(u^*) = \mathrm{V}(Z'u) = \sigma^2 Z'\Omega Z.$$

Go back to GMM:

$$
\begin{aligned}
&(y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta) \\
&= y'Z(Z'\Omega Z)^{-1}Z'y - \beta'X'Z(Z'\Omega Z)^{-1}Z'y - y'Z(Z'\Omega Z)^{-1}Z'X\beta + \beta'X'Z(Z'\Omega Z)^{-1}Z'X\beta \\
&= y'ZWZ'y - 2y'Z(Z'\Omega Z)^{-1}Z'X\beta + \beta'X'Z(Z'\Omega Z)^{-1}Z'X\beta.
\end{aligned}
$$

Note that $\beta'X'Z(Z'\Omega Z)^{-1}Z'y = y'Z(Z'\Omega Z)^{-1}Z'X\beta$ because both sides are scalars.

Remember that $\dfrac{\partial Ax}{x} = A'$ and $\dfrac{\partial x'Ax}{x} = (A + A')x$.
Then, we obtain the following derivation:

$$
\begin{aligned}
&\frac{\partial (y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta)}{\partial\beta} \\
&\qquad = -2(y'Z(Z'\Omega Z)^{-1}Z'X)' + \Big(X'Z(Z'\Omega Z)^{-1}Z'X + (X'Z(Z'\Omega Z)^{-1}Z'X)'\Big)\beta \\
&\qquad = -2X'Z(Z'\Omega Z)^{-1}Z'y + 2X'Z(Z'\Omega Z)^{-1}Z'X\beta = 0
\end{aligned}
$$

The solution of β is denoted by β_{GMM}, which is:

$$
\beta_{GMM} = (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'y.
$$

The mean of β_{GMM} is asymptotically obtained.

$$\begin{aligned}
\beta_{GMM} &= (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'(X\beta + u) \\
&= \beta + (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'u \\
&= \beta + \Big((\frac{1}{n}X'Z)(\frac{1}{n}Z'\Omega Z)^{-1}(\frac{1}{n}Z'X)\Big)^{-1}(\frac{1}{n}X'Z)(\frac{1}{n}Z'\Omega Z)^{-1}(\frac{1}{n}Z'u)
\end{aligned}$$

We assume that

$$\frac{1}{n}X'Z \longrightarrow M_{xz} \quad \text{and} \quad \frac{1}{n}Z'\Omega Z \longrightarrow M_{z\Omega z},$$

which are $k \times r$ and $r \times r$ matrices.

From the assumption of $\frac{1}{n}Z'u \longrightarrow 0$, we have the following result:

$$\beta_{GMM} \longrightarrow \beta + (M_{xz}M_{z\Omega z}^{-1}M'_{xz})^{-1}M_{xz}M_{z\Omega z}^{-1} \times 0 = \beta.$$

Thus, β_{GMM} is a consistent estimator of β (i.e., asymptotically unbiased estimator).

The variance of β_{GMM} is asymptotically obtained as follows:

$$
\begin{aligned}
\mathrm{V}(\beta_{GMM}) &= \mathrm{E}\Big((\beta_{GMM} - \mathrm{E}(\beta_{GMM}))(\beta_{GMM} - \mathrm{E}(\beta_{GMM}))'\Big) \approx \mathrm{E}\Big((\beta_{GMM} - \beta)(\beta_{GMM} - \beta)'\Big) \\
&= \mathrm{E}\Big((X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'u\big((X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'u\big)'\Big) \\
&= \mathrm{E}\Big((X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'uu'Z(Z'\Omega Z)^{-1}Z'X(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}\Big) \\
&\approx (X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}X'Z(Z'\Omega Z)^{-1}Z'\mathrm{E}(uu')Z(Z'\Omega Z)^{-1}Z'X(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1} \\
&= \sigma^2(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}.
\end{aligned}
$$

Note that $\beta_{GMM} \longrightarrow \beta$ implies $\mathrm{E}(\beta_{GMM}) \longrightarrow \beta$ in the 1st line.

$\approx$ in the 4th line indicates that Z and X are treated as exogenous variables although they are stochastic.

We assume that $\mathrm{E}(uu') = \sigma^2\Omega$ from the 4th line to the 5th line.

- We derive the asymptotic distribution of β_{GMM}.

From the central limit theorem,

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0,\ \sigma^2 M_{z\Omega z}).$$

Accordingly, β_{GMM} is asymptotically distributed as:

$$\begin{aligned}\sqrt{n}(\beta_{GMM}-\beta) &= \Big((\frac{1}{n}X'Z)(\frac{1}{n}Z'\Omega Z)^{-1}(\frac{1}{n}Z'X)\Big)^{-1}(\frac{1}{n}X'Z)(\frac{1}{n}Z'\Omega Z)^{-1}(\frac{1}{\sqrt{n}}Z'u) \\ &\longrightarrow N(0,\ \sigma^2(M_{xz}M_{z\Omega z}^{-1}M'_{xz})^{-1}).\end{aligned}$$

Practically, we use: $\beta_{GMM} \sim N\Big(\beta,\ s^2(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}\Big)$,

where $s^2 = \dfrac{1}{n-k}(y - X\beta_{GMM})'\Omega^{-1}(y - X\beta_{GMM})$.

We may use n instead of $n-k$.

Identically and Independently Distributed Errors:

- If $u_1, u_2, \cdots, u_n$ are mutually independent and u_i is distributed with mean zero and variance σ^2, the mean and variance of u^* are given by:

$$\mathrm{E}(u^*) = 0 \quad \text{and} \quad \mathrm{V}(u^*) = \mathrm{E}(u^* u^{*\prime}) = \sigma^2 Z'Z.$$

Using GLS, GMM is obtained as:

$$\beta_{GMM} = (X^{*\prime}(Z'Z)^{-1}X^*)^{-1}X^{*\prime}(Z'Z)^{-1}y^* = \Big(X'Z(Z'Z)^{-1}Z'X\Big)^{-1}X'Z(Z'Z)^{-1}Z'y.$$

- We derive the asymptotic distribution of β_{GMM}.

From the central limit theorem,

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0, \sigma^2 M_{zz}).$$

Accordingly, β_{GMM} is distributed as:

$$\sqrt{n}(\beta_{GMM} - \beta) = \Big((\frac{1}{n}X'Z)(\frac{1}{n}Z'Z)^{-1}(\frac{1}{n}Z'X)\Big)^{-1}(\frac{1}{n}X'Z)(\frac{1}{n}Z'Z)^{-1}(\frac{1}{\sqrt{n}}Z'u)$$
$$\longrightarrow N\Big(0,\ \sigma^2(M_{xz}M_{zz}^{-1}M'_{xz})^{-1}\Big).$$

Practically, for large n we use the following distribution:

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

where $s^2 = \dfrac{1}{n-k}(y - X\beta_{GMM})'(y - X\beta_{GMM})$.

- The above GMM is equivalent to 2SLS.

X: $n \times k$, Z: $n \times r$, $r > k$.

Assume:

$$\frac{1}{n}X'u = \frac{1}{n}\sum_{i=1}^{n} x_i'u_i \longrightarrow \mathrm{E}(x'u) \neq 0,$$

$$\frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i'u_i \longrightarrow \mathrm{E}(z'u) = 0.$$

Regress X on Z, i.e., $X = Z\Gamma + V$ by OLS, where Γ is a $r \times k$ unknown parameter matrix and V is an error term,

Denote the predicted value of X by $\hat{X} = Z\hat{\Gamma} = Z(Z'Z)^{-1}Z'X$, where $\hat{\Gamma} = (Z'Z)^{-1}Z'X$.

Review — IV estimator: Consider the regression model is:

$$y = X\beta + u,$$

Assumption: $\mathrm{E}(X'u) \neq 0$ and $\mathrm{E}(Z'u) = 0$.

The $n \times k$ matrix Z is called the instrumental variable (IV).

The IV estimator is given by:

$$\beta_{IV} = (Z'X)^{-1}Z'y,$$

- 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,

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

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.

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

$$
\begin{aligned}
\mathrm{V}(u^*) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}(v_i v'_j) \\
&= \mathrm{E}(v_1 v'_1) + \mathrm{E}(v_1 v'_2) + \mathrm{E}(v_1 v'_3) + \cdots + \mathrm{E}(v_1 v'_n) \\
&+ \mathrm{E}(v_2 v'_1) + \mathrm{E}(v_2 v'_2) + \mathrm{E}(v_2 v'_3) + \cdots + \mathrm{E}(v_2 v'_n) \\
&+ \mathrm{E}(v_3 v'_1) + \mathrm{E}(v_3 v'_2) + \mathrm{E}(v_3 v'_3) + \cdots + \mathrm{E}(v_3 v'_n) \\
&\qquad \vdots \\
&+ \mathrm{E}(v_n v'_1) + \mathrm{E}(v_n v'_2) + \mathrm{E}(v_n v'_3) + \cdots + \mathrm{E}(v_n v'_n) \\
&= \Gamma_0 + \Gamma_{-1} + \Gamma_{-2} + \cdots + \Gamma_{1-n} \\
&+ \Gamma_1 + \Gamma_0 + \Gamma_{-1} + \cdots + \Gamma_{2-n} \\
&+ \Gamma_2 + \Gamma_1 + \Gamma_0 + \cdots + \Gamma_{3-n}
\end{aligned}
$$

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

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:

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

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.

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

(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: $\quad \hat{S} = \dfrac{1}{n}\sum_{i=1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_i)' \longrightarrow S.$

- 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.

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

$$
S = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\Big(h(\theta; w_i) h(\theta; w_j)'\Big)
$$

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

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:

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

$\Longrightarrow$ 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:

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

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 $\hat{S}^{(i)} = \hat{\Gamma}_0 + \sum_{i=1}^{q}\Big(1 - \dfrac{i}{q+1}\Big)(\hat{\Gamma}_i + \hat{\Gamma}_i')$, where $\hat{\Gamma}_\tau = \dfrac{1}{n}\sum_{i=\tau+1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_{i-\tau})'$. 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: $y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ 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: $y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) \neq 0, \quad \mathrm{E}(z_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ 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: $f(y_i, x_i, \beta) = \epsilon_i$, $\mathrm{E}(x_i'\epsilon_i) \neq 0$, $\mathrm{E}(z_i'\epsilon_i) = 0$

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

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

where z_i is a vector of instrumental variables.

Example: Demand function using STATA

二人以上の世帯のうち勤労者世帯（全国）

```
year
y   ＝ 実収入（一月当たり，実質データ）
q1  ＝ 穀類支出額（一年当たり，実質データ）
p1  ＝ 穀類価格　（相対価格＝穀類 CPI ／総合 CPI）
p2  ＝ 魚介類価格（相対価格＝魚介類 CPI ／総合 CPI）
p3  ＝ 肉類価格　（相対価格＝肉類 CPI ／総合 CPI）
```

year	y	q1	p1	p2	p3
2000	567865	7087.0	1.043390	0.884965	0.818365
2001	561722	6993.1	1.032520	0.886179	0.822154
2002	553768	6934.4	1.031800	0.891282	0.834872
2003	539928	6816.8	1.050410	0.876543	0.843621
2004	547006	6651.6	1.089510	0.865226	0.868313
2005	541367	6615.8	1.020640	0.862745	0.887513
2006	540863	6523.7	1.000000	0.878601	0.891975
2007	543994	6680.5	0.994856	0.886831	0.908436
2008	541821	6494.7	1.043610	0.894523	0.932049
2009	533154	6477.3	1.066870	0.898148	0.934156
2010	539577	6458.2	1.040420	0.889119	0.924352
2011	529750	6448.4	1.025960	0.894081	0.925234
2012	538988	6377.6	1.057170	0.904366	0.917879
2013	542018	6360.7	1.047620	0.909938	0.916149
2014	523953	6174.6	1.016130	0.971774	0.960685
2015	525669	6268.0	1.000000	1.000000	1.000000
2016	527501	6244.8	1.018020	1.019020	1.017020

```
2017 531693 6106.6 1.027890 1.066730 1.025900


. tsset year
        time variable:  year, 2000 to 2017
                delta:  1 unit

. reg q1 y p1 p2 p3 if year>2000.5

      Source |       SS           df       MS      Number of obs   =        17
-------------+----------------------------------   F(4, 12)        =     25.83
       Model |  913640.443         4  228410.111   Prob > F        =    0.0000
    Residual |  106100.077        12  8841.67308   R-squared       =    0.8960
-------------+----------------------------------   Adj R-squared   =    0.8613
       Total |  1019740.52        16  63733.7825   Root MSE        =     94.03

------------------------------------------------------------------------------
          q1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           y |   .0067843   .0045443     1.49   0.161     -.003117    .0166856
          p1 |  -1128.834   998.7698    -1.13   0.280    -3304.966    1047.299
          p2 |   356.8095   806.2301     0.44   0.666    -1399.815    2113.434
          p3 |  -3442.221   1130.078    -3.05   0.010    -5904.448   -979.9931
       _cons |   6850.563   3179.316     2.15   0.052    -76.57278     13777.7
------------------------------------------------------------------------------

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(y p1
> p2 p3)
```

```
Step 1
Iteration 0:   GMM criterion Q(b) =   42400764
Iteration 1:   GMM criterion Q(b) =  6.781e-12
Iteration 2:   GMM criterion Q(b) =  6.781e-12  (backed up)

Step 2
Iteration 0:   GMM criterion Q(b) =  1.966e-15
Iteration 1:   GMM criterion Q(b) =  1.963e-15  (backed up)
convergence not achieved
    The Gauss-Newton stopping criterion has been met but missing standard errors
    indicate some of the parameters are not identified.

GMM estimation

Number of parameters =   5
Number of moments    =   5
Initial weight matrix: Unadjusted                 Number of obs   =         17
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   6850.563   17645.71     0.39   0.698     -27734.4    41435.53
         /b1 |   .0067843   .0282325     0.24   0.810    -.0485504     .062119
         /b2 |  -1128.834   1057.915    -1.07   0.286    -3202.309    944.6415
         /b3 |   356.8095    1565.86     0.23   0.820    -2712.219    3425.838
         /b4 |  -3442.221   5085.561    -0.68   0.498    -13409.74    6525.296
------------------------------------------------------------------------------
```

```
Instruments for equation 1: y p1 p2 p3 _cons
Warning: convergence not achieved

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(p1 p2
> p3 l.p1 l.p2 l.p3)

Step 1
Iteration 0:   GMM criterion Q(b) =   42404066
Iteration 1:   GMM criterion Q(b) =  2790.3146
Iteration 2:   GMM criterion Q(b) =  2790.3146

Step 2
Iteration 0:   GMM criterion Q(b) =   .3201826
Iteration 1:   GMM criterion Q(b) =   .2469289
Iteration 2:   GMM criterion Q(b) =   .2469289

GMM estimation

Number of parameters =   5
Number of moments    =   7
Initial weight matrix: Unadjusted                 Number of obs   =         17
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |  -1192.466   4669.012    -0.26   0.798    -10343.56     7958.63
         /b1 |   .0186312   .0067682     2.75   0.006     .0053657    .0318967
```

```
         /b2 |  -1016.864    780.979    -1.30   0.193    -2547.554    513.8271
         /b3 |  -905.5585   598.0885    -1.51   0.130     -2077.79    266.6734
         /b4 |  -499.8064   1147.985    -0.44   0.663    -2749.815    1750.202
------------------------------------------------------------------------------
Instruments for equation 1: p1 p2 p3 L.p1 L.p2 L.p3 _cons

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(2) = 4.19779 (p = 0.1226)
```