Advanced Econometrics

Lecture 11: Generalized Method of Moments (Hansen Chapter 11)

Instructor: Ma, Jun

Renmin University of China

December 20, 2018

Moment Equation Models

Moment Condition: $(W_1, \cdots, W_n)$ $p \times 1$

$g: \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}^l$ known $(g = (g_1, \cdots, g_l)')$

$E g(W_i, b) = \int g(w, b) f(w)\, dw$

$E g(W_i, \beta) = 0$, for some $\beta \in \mathbb{R}^k$

$\uparrow$ parameter of interest

We say the model is identified:

if $E g(W_i, \tilde{\beta}) = 0$. then $\tilde{\beta} = \beta$

$l \geq k$ ($l > k$ overidentified

$l = k$ just identified)

- All of the models that have been introduced so far can be written as **moment equation models**, where the population parameters solve a system of moment equations.
- Let $\boldsymbol{g}_i(\boldsymbol{\beta})$ be a known $\ell \times 1$ function of the i-th observation and the parameter $\boldsymbol{\beta}$. A moment equation model is

$$\mathbb{E}\left(\boldsymbol{g}_i(\boldsymbol{\beta})\right) = \mathbf{0}.$$

We know that the true parameter satisfies the system of equations.

$\{(Y_i, X_i, Z_i), i = 1, \cdots, n\}$

$\underset{l \times 1}{g_i(\beta)}$ 满足一个矩方程

$E(g(Y_i, X_i, Z_i, b)) = 0$

此时 b 就是待估参数 β

eg1. $E(X_i - b) = 0 \Rightarrow \mu - b = 0$

$\Rightarrow \beta = E(X_i)$

eg2. $E\begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix} = 0$

$g(X_i, \mu, \sigma^2) = \begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix}$

Moment Equation Models

$W_i = (Y_i, X_i', Z_i')$

$g(W_i, b) = Z_i(Y_i - X_i'b)$

$E Z_i(Y_i - X_i'b)$

Known $E \underbrace{Z_i Y_i}_{l\times 1} = E \underbrace{Z_i X_i'}_{l\times k}\beta$

if $E Z_i Y_i = E Z_i X_i' \tilde{\beta}$

then $\tilde{\beta} = \beta$ 有唯一的解

$Y_i = X_i'\beta + e_i$

$E e_i X_i \neq 0$

$E e_i Z_i = 0$ $\quad$ $\text{rank}(E(Z_i X_i')) = k$

- For example, in the instrumental variables model $\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$. 工具变量就是一个例子
- We say the parameter is identified if there is a unique mapping from the data distribution to $\boldsymbol{\beta}$. In other words, there is unique $\boldsymbol{\beta}$ solves the equations. A necessary condition for identification is $\ell \geq k$.
- $\ell = k$: just identified
- $\ell > k$: over-identified 过度识别

模型可以被识别的必要条件是：

方程个数大于等于未知数个数.

即 $l \geq k$

$g_i(\beta) = \underset{l\times 1}{Z_i}(Y_i - X_i'\underset{k\times 1}{\beta})$

$E[\underbrace{g_i(\beta)}_{l\times 1}] = 0$

$\Rightarrow E(Z_i Y_i) = E(Z_i X_i')\cdot\beta$, $\quad E(Z_i Y_i)$: $l\times 1$, $\quad \beta$: $k\times 1$

要有解：$l \geq k$

如果 $E(Z_i X_i')$ 达到满秩，

一定有唯一解.

Method of Moments Estimator

如果 $l=k$

$E z_i'(Y_i - X_i'\beta) = 0$

$\frac{1}{n}\sum_{i=1}^{n} z_i'(Y_i - X_i'\hat\beta^{MM}) = 0$

$\Rightarrow \hat\beta^{MM} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i X_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i Y_i$

$\min_b \|A \frac{1}{n}\sum_{i=1}^{n} z_i(Y_i - X_i'b)\|$

A_n是一个权重矩阵 $l\times l$ $\quad A_n \to_p A$

可以是与数据有关的.

eg. $A_n'A_n = \frac{1}{n}\sum_{i=1}^{n} z_i z_i'$

$\Leftrightarrow \min_b \|A_n \frac{1}{n}\sum_{i=1}^{n} g(w_i,b)\|^2 \quad l>k$

WLLN: $\forall b,\ \|A_n \frac{1}{n}\sum_{i=1}^{n} g(w_i,b)\|^2 \to \|AEg(w_i,b)\|^2$

+ CMT

$AEg(w_i,b) = 0$ if and only if $b=\beta$

$\|X\| \ge 0$
$\|X\| = 0$ iff $x = 0$

$\Leftrightarrow \|AEg(w_i,b)\|^2 \ge 0$; $\|AE(g(w_i,b))\|^2 = 0$ iff $b=\beta$

- We consider the just identified case: $\ell = k$
- The sample analogue of $\mathbb{E}(\boldsymbol{g}_i(\boldsymbol{\beta}))$:

$$\overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(\boldsymbol{\beta}).$$

- The method of moments estimator (MME) $\widehat{\boldsymbol{\beta}}_{\mathrm{mm}}$ for $\boldsymbol{\beta}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i\left(\widehat{\boldsymbol{\beta}}_{\mathrm{mm}}\right) = \mathbf{0}.$$

$\min_b \|A_n \frac{1}{n}\sum_{i=1}^{n} z_i(Y_i - X_i'b)\|^2$

$J(b) = n\left(\sum_{i=1}^{n} z_i(Y_i - X_i'b)\right)' \underbrace{A_n'A_n}_{W_n} \left(\sum_{i=1}^{n} z_i(Y_i - X_i'b)\right)$

$W_n \to_p W$

$l=k \quad \hat\beta^{GMM} = (X'ZW_nZ'X)^{-1}(X'ZW_nZ'Y)$

$= (Z'X)^{-1}W_n^{-1}(X'Z)^{-1}(X'Z)W_n(Z'Y)$

$= (Z'X)^{-1}Z'Y = \hat\beta_{IV}$

恰好识别时就是IV估计.

$\hat\beta^{GMM} = \arg\min_b \|A_n \frac{1}{n}\sum_{i=1}^{n} g(w_i,b)\|^2$

$\Updownarrow$

$\beta = \arg\min_b \|AEg(w_i,b)\|^2$

$x=0 \Leftrightarrow \|X\|=0$

过度识别条件下，$\|\bar{g}_n(b)\|=0$ 不可能.

只能求 $\min_b \|\bar{g}_n(b)\|$

$= \bar{g}_n(b)' W \bar{g}_n(b)$

任意的一个权重矩阵

Overidentified Moment Equations

- In the instrumental variables model $\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$.

$$\overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta}\right) = \frac{1}{n}\left(\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta}\right).$$

- We defined the MME for $\boldsymbol{\beta}$ to be the solution to $\overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = \boldsymbol{0}$. However, if the model is over-identified, there are more equations than parameters. The MME is not defined.
- We cannot find an estimator which sets $\overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = \boldsymbol{0}$ but we can try to find an estimator which makes $\overline{\boldsymbol{g}}_n(\boldsymbol{\beta})$ as close to zero as possible.

Overidentified Moment Equations

- Let $\boldsymbol{W}$ be an $\ell \times \ell$ positive definite weight matrix. The GMM criterion function is

$$J(\boldsymbol{\beta}) = n \cdot \overline{\boldsymbol{g}}_n(\boldsymbol{\beta})' \boldsymbol{W} \overline{\boldsymbol{g}}_n(\boldsymbol{\beta}).$$

- When $\boldsymbol{W} = \boldsymbol{I}_\ell$, $J(\boldsymbol{\beta}) = n \cdot \overline{\boldsymbol{g}}_n(\boldsymbol{\beta})' \overline{\boldsymbol{g}}_n(\boldsymbol{\beta}) = n \cdot \|\overline{\boldsymbol{g}}_n(\boldsymbol{\beta})\|^2$.

Definition
The Generalized Method of Moments estimator is
$\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \mathrm{argmin} J_n(\boldsymbol{\beta})$.

Overidentified Moment Equations

- When the moment equations are linear in the parameters then we have explicit solutions for the estimates.
- We focus on the over-identified IV model:

$$\boldsymbol{g}_i(\boldsymbol{\beta}) = \boldsymbol{Z}_i \left(Y_i - \boldsymbol{X}_i' \boldsymbol{\beta} \right),$$

  where $\boldsymbol{Z}_i$ is $\ell \times 1$ and $\boldsymbol{X}_i$ is $k \times 1$.

GMM Estimator

- The GMM criterion function:

$$J(\boldsymbol{\beta}) = n\left(\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{W}\left(\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta}\right).$$

- First-order conditions:

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial\boldsymbol{\beta}} J\left(\widehat{\boldsymbol{\beta}}\right) \\
&= 2\frac{\partial}{\partial\boldsymbol{\beta}}\overline{\boldsymbol{g}}_n\left(\widehat{\boldsymbol{\beta}}\right)'\boldsymbol{W}\overline{\boldsymbol{g}}_n\left(\widehat{\boldsymbol{\beta}}\right) \\
&= -2\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\right).
\end{aligned}$$

GMM Estimator

Theorem
For the overidentified IV model

$$\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{Y}\right)$$

广义矩估计量是一组估计量.
不是一个.

Theorem
If $\boldsymbol{W} = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$ *then* $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \widehat{\boldsymbol{\beta}}_{\mathrm{2sls}}$. — $A_n = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-\frac{1}{2}}$
Futhermore, if $k = l$ *then* $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \widehat{\boldsymbol{\beta}}_{\mathrm{iv}}$.

两阶段最小二乘是广义矩法的一个特例.

$$\hat{\beta}^{GMM} = \left(\sum_{i=1}^{n} x_i z_i' W_n \sum_{i=1}^{n} z_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i z_i' W_n \sum_{i=1}^{n} z_i Y_i\right) \qquad W_n \xrightarrow{p} W$$

$$W_n = \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1} \quad W = \left(E z_i z_i'\right)^{-1} \qquad Y_i = X_i'\beta + e_i$$

$$\hat{\beta}^{GMM} = \beta + \left(\frac{1}{n}\sum x_i z_i' W_n \frac{1}{n}\sum z_i x_i'\right)^{-1}\left(\frac{1}{n}\sum x_i z_i' W_n \frac{1}{n}\sum z_i e_i\right) \xrightarrow{p} \beta$$

$E z_i e_i = 0$

$\frac{1}{n}\sum z_i e_i \xrightarrow{p} 0$

Distribution of GMM Esttimator

- Denote

$$\boldsymbol{Q} = \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right)$$

$$\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{Z}_i' e_i^2\right) = \mathbb{E}\left(\boldsymbol{g}_i \boldsymbol{g}_i'\right)$$

where $\boldsymbol{g}_i = \boldsymbol{Z}_i e_i$.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i e_i \xrightarrow{d} N(0, \Omega)$$

- Then,

$$\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right) \xrightarrow{p} \boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q}$$

$$\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}\left(\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{e}\right) \xrightarrow{d} \boldsymbol{Q}'\boldsymbol{W} \cdot \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{\Omega}\right).$$

$$\sqrt{n}\left(\hat{\beta}^{GMM} - \beta\right)$$

$$= \left(\frac{1}{n}\sum x_i z_i' W_n \frac{1}{n}\sum z_i x_i'\right)^{-1} \left(\underbrace{\frac{1}{n}\sum x_i z_i' W_n}_{\xrightarrow{p} Q'W} \underbrace{\frac{1}{\sqrt{n}}\sum z_i e_i}_{\xrightarrow{d} N(0,\Omega)}\right)$$

$$\xrightarrow{p} (Q'WQ)^{-1}$$

$$\xrightarrow[d]{\text{slutsky}} Q'W N(0, \Omega)$$

$$= N(0, Q'W\Omega WQ)$$

$$\xrightarrow[d]{\text{slutsky}} N\left(0, (Q'WQ)^{-1} Q'W\Omega WQ (Q'WQ)^{-1}\right)$$

Distribution of GMM Esttimator

Theorem
Asymptotic Distribution of GMM Estimator.

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}}=\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{Q}\right)^{-1}\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{W} \boldsymbol{Q}\right)\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{Q}\right)^{-1}$$

- The theorem carries over to the case where the weight matrix $\widehat{\boldsymbol{W}}$ is random (depends on the data) so long as it converges in probability to some positive definite limit $\boldsymbol{W}$. E.g., $\widehat{\boldsymbol{W}}=\left(n^{-1} \boldsymbol{Z}^{\prime} \boldsymbol{Z}\right)^{-1}$.

$W>0$ 严格正定

if $W=I \Rightarrow$

$V_\beta=(Q'Q)^{-1}Q'\Omega Q(Q'Q)^{-1}$

if $W_n=(\frac{1}{n}Z'Z)^{-1}$ (两阶段最小二乘)

$W=(EZ_iZ_i')^{-1}$

$\Rightarrow V_\beta = V_\beta^{2SLS}$

$$V_\beta(W) = (Q'WQ)^{-1} Q'W\Omega WQ(Q'WQ)^{-1}$$

Theorem: $V_\beta(W) - V_\beta(\Omega^{-1})$ 是半正定的 $\forall W$ 正定

$$V_\beta(\Omega^{-1}) = (Q'\Omega^{-1}Q)^{-1}$$

$$(Q'\Omega^{-1}Q)^{-1} - (Q'WQ)^{-1}(Q'W\Omega WQ)(Q'WQ)^{-1} \leq 0$$

Fact: $A^{-1} - B^{-1} \leq 0 \iff A - B \geq 0$

要证 $Q'\Omega^{-1}Q - (Q'WQ)(Q'W\Omega WQ)^{-1}(Q'WQ) \geq 0$

Ω 正定 $\iff \Omega^{-1}$ 正定

$$\Omega^{-1} = G\Lambda G' = \underbrace{G\Lambda^{\frac{1}{2}}}_{C'}\underbrace{\Lambda^{\frac{1}{2}}G'}_{C} \qquad \Omega = C^{-1}(C')^{-1}$$

令 $H = (C')^{-1}WQ$

$$\Rightarrow Q'C'CQ - (Q'WQ)(Q'WC^{-1}(C')^{-1}WQ)^{-1}(Q'WQ) \geq 0$$

$$= Q'C'(I - (C')^{-1}WQ(Q'WC^{-1}(C')^{-1}WQ)^{-1}Q'WC^{-1})CQ = Q'C'(\underbrace{I - H(H'H)^{-1}H'}_{\text{幂等}})CQ \text{ 半正定}$$

Efficient GMM — 即 Optimal

$$\overbrace{\min_{W>0} (Q'WQ)^{-1}(Q'W\Omega WQ)(Q'WQ)^{-1}}^{V(W)}$$

find W^*, $V(W)-V(W^*)\geq 0$ $\forall W>0$ $\quad W^*=\Omega^{-1}$

$\widehat{W}^* \xrightarrow{p} W^*$

最优的广义矩估计是 $\widehat{\Omega}^{-1}$

$\Omega = E z_i z_i' e_i^2$

估计 $\frac{1}{n}\sum_{i=1}^{n} z_i z_i' e_i^2$

$=\frac{1}{n}\sum_{i=1}^{n} z_i z_i' (Y_i - X_i'\hat\beta)^2$

广义矩估计还没得到.
只需要找一个一致估计量.
用 $\hat\beta_{2SLS}$

- The asymptotic distribution of the GMM estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}$ depends on the weight matrix $\boldsymbol{W}$ through the asymptotic variance $\boldsymbol{V}_{\boldsymbol{\beta}}$.
- The asymptotically optimal weight matrix $\boldsymbol{W}_0$ is one which minimizes $\boldsymbol{V}_{\boldsymbol{\beta}}$. This turns out to be $\boldsymbol{W}_0 = \boldsymbol{\Omega}^{-1}$.
- The efficient GMM:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{Y}\right).$$

- Feasible efficient GMM:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \left(\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Z}'\boldsymbol{Y}\right),$$

  where $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$.
- We find:

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}$$
$$= \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}.$$

Efficient GMM

Theorem

Asymptotic Distribution of GMM with Efficient Weight Matrix

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}$$

Theorem

If $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}$ is the efficient GMM estimator and $\widetilde{\boldsymbol{\beta}}_{\mathrm{gmm}}$ is another GMM estimator, then

$$\mathrm{avar}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}\right) \leq \mathrm{avar}\left(\widetilde{\boldsymbol{\beta}}_{\mathrm{gmm}}\right)$$

Efficient GMM versus 2SLS

- We have introduced the GMM estimator which includes 2SLS as a special case. Is there a context where 2SLS is efficient?
- The 2SLS estimator is GMM given the weight matrix $\widehat{\boldsymbol{W}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ or equivalently $\widehat{\boldsymbol{W}} = (n^{-1}\boldsymbol{Z}'\boldsymbol{Z})^{-1}$.
- Since $\widehat{\boldsymbol{W}} \to_p \mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i')^{-1}$, the asymptotic distribution of 2SLS is the same as that of using the weight matrix $\boldsymbol{W} = \mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i')^{-1}$.
- The efficient weight matrix takes the form $\mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'e_i^2\right)^{-1}$.
- Suppose that the error e_i is conditionally homoskedastic: $\mathbb{E}\left(e_i^2|\boldsymbol{Z}_i\right) = \sigma^2$. The efficient weight matrix is $\boldsymbol{W} = \mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i')^{-1}\sigma^{-2}$.

Theorem

Under $\mathbb{E}\left(e_i^2 \mid \boldsymbol{Z}_i\right) = \sigma^2$ *then* $\widehat{\boldsymbol{\beta}}_{\text{2sls}}$ *is efficient GMM.*

两阶段最小二乘是广义矩法的一个特例.

那么什么时候最优的广义矩估计是2SLS估计呢?

$\Omega = E(z_i z_i' e_i^2)$

$= E(z_i z_i' E(e_i^2|z_i))$

$= \sigma^2 E(z_i z_i') \leftarrow$ 同方差

$\hat{\Omega} = \hat{\sigma}^2 \frac{1}{n}\sum_{i=1}^{n} z_i z_i' = \hat{\sigma}^2 \cdot \frac{1}{n} Z'Z$

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

Estimation of the Efficient Weight Matrix

- To construct the efficient GMM estimator we need a consistent estimator of $\boldsymbol{W}_0 = \boldsymbol{\Omega}^{-1}$.
- The **two-step GMM** estimator uses a consistent estimate of $\boldsymbol{\beta}$ to construct the weight matrix estimator $\widehat{\boldsymbol{W}}$.
- In the linear IV model, the natural one-step estimator for $\boldsymbol{\beta}$ is the 2SLS estimator $\widehat{\boldsymbol{\beta}}_{\text{2sls}}$.
- Set $\widetilde{e}_i = Y_i - \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}_{\text{2sls}}$, $\widetilde{\boldsymbol{g}}_i = \boldsymbol{g}_i\left(\widetilde{\boldsymbol{\beta}}\right) = \boldsymbol{Z}_i\widetilde{e}_i$ and $\overline{\boldsymbol{g}}_n = n^{-1}\sum_{i=1}^n \widetilde{\boldsymbol{g}}_i$.
- Two moment estimators of $\boldsymbol{\Omega}$ are

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^n \widetilde{\boldsymbol{g}}_i\widetilde{\boldsymbol{g}}_i'$$

$$\widehat{\boldsymbol{\Omega}}^* = \frac{1}{n}\sum_{i=1}^n \left(\widetilde{\boldsymbol{g}}_i - \overline{\boldsymbol{g}}_n\right)\left(\widetilde{\boldsymbol{g}}_i - \overline{\boldsymbol{g}}_n\right)'.$$

Either estimator is consistent.

Estimation of the Efficient Weight Matrix

$\sqrt{n}(\hat\beta^{GMM} - \beta) \to_d N(0, (Q'\Omega^{-1}Q)^{-1})$

$(\hat Q' \hat\Omega^{-1} \hat Q)^{-1}$

- We set $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Omega}}^{-1}$. Then construct the two-step GMM estimator using the weight matrix $\widehat{\boldsymbol{W}}$.

$\hat V_\beta \to_p V_\beta$

$\hat Q = \frac{1}{n}\sum_{i=1}^{n} z_i x_i'$

Theorem

If $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Omega}}^{-1}$ *or* $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Omega}}^{*-1}$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}$$

Covariance Matrix Estimation

- For the one-step GMM estimator the covariance matrix estimator is

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)^{-1}\left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)\left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)^{-1}$$

  where

$$\widehat{\boldsymbol{Q}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{X}_i'.$$

- For the two-step efficient GMM estimator, the covariance matrix estimator is

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{Q}}\right)^{-1} = \left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\widehat{\boldsymbol{\Omega}}^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right)^{-1}.$$

Wald Test

- Given $\boldsymbol{r} : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q$, the parameter of interest is $\boldsymbol{\theta} = \boldsymbol{r}\left(\boldsymbol{\beta}\right)$.
- A natural estimator is $\widehat{\boldsymbol{\theta}}_{\mathrm{gmm}} = \boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}\right)$.
- $\widehat{\boldsymbol{\theta}}_{\mathrm{gmm}}$ is asymptotically normal with covariance matrix

$$\begin{aligned} \boldsymbol{V}_{\boldsymbol{\theta}} &= \boldsymbol{R}'\boldsymbol{V}_{\boldsymbol{\beta}}\boldsymbol{R} \\ \boldsymbol{R} &= \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{r}\left(\boldsymbol{\beta}\right)'. \end{aligned}$$

- Estimator of the asymptotic variance matrix:

$$\begin{aligned} \widehat{\boldsymbol{V}}_{\boldsymbol{\theta}} &= \widehat{\boldsymbol{R}}'\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{R}} \\ \widehat{\boldsymbol{R}} &= \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}\right)'. \end{aligned}$$

Wald Test

- We are interested in testing

$$\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$$
$$\mathbb{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

- The Wald statistic:

$$W = n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \widehat{\boldsymbol{V}}_{\boldsymbol{\theta}}^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

Theorem

If $\boldsymbol{r}(\boldsymbol{\beta})$ is continuously differentiable at $\boldsymbol{\beta}$, and $\mathbb{H}_0$ holds, then as $n \to \infty$,

$$W \xrightarrow{d} \chi_q^2.$$

For c satisfying $\alpha = 1 - G_q(c)$,

$$\Pr(W > c \mid \mathbb{H}_0) \to \alpha$$

so the test "Reject $\mathbb{H}_0$ if $W > c$" has asymptotic size α.

Continuously-Updated GMM

$$\min_b n \left(\frac{1}{n}\sum_{i=1}^{n} z_i'(Y_i - X_i'b)\right)\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'b)^2 z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} z_i'(Y_i - X_i'b)\right)$$

- An alternative to the two-step GMM estimator can be constructed by letting the weight matrix be an explicit function of $\boldsymbol{\beta}$:

$$J(\boldsymbol{\beta}) = n \cdot \overline{\boldsymbol{g}}_n(\boldsymbol{\beta})' \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\beta})\, \boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\beta})'\right)^{-1} \overline{\boldsymbol{g}}_n(\boldsymbol{\beta}).$$

- The $\widehat{\boldsymbol{\beta}}$ which minimizes this function is the CU-GMM estimator.
- Minimization requires numerical methods.
- We have:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{cu-gmm}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right)$$

  where

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}.$$

GMM: The General Case

- The general moment equation model:

$$\mathbb{E}\left(\boldsymbol{g}_i\left(\boldsymbol{\beta}\right)\right) = \mathbf{0}.$$

- The GMM estimator minimizes

$$J\left(\boldsymbol{\beta}\right) = n \cdot \overline{\boldsymbol{g}}_n\left(\boldsymbol{\beta}\right)' \widehat{\boldsymbol{W}} \overline{\boldsymbol{g}}_n\left(\boldsymbol{\beta}\right),$$

where

$$\overline{\boldsymbol{g}}_n\left(\boldsymbol{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_i\left(\boldsymbol{\beta}\right).$$

- The efficient GMM estimator can be constructed by setting

$$\widehat{\boldsymbol{W}} = \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{g}}_i \widehat{\boldsymbol{g}}_i' - \overline{\boldsymbol{g}}_n \overline{\boldsymbol{g}}_n'\right)^{-1},$$

with $\widehat{\boldsymbol{g}}_i = \boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}\right)$ constructed using a preliminary consistent estimator obtained by first setting $\widehat{\boldsymbol{W}} = \boldsymbol{I}_\ell$.

对于非线性模型（考试不涉及）

$E\, g(w_i, \beta) = 0$ $W_n = A_n' A_n$ $W_n \xrightarrow{p} W$

$\hat\beta^{GMM} = \underset{b}{\operatorname{argmin}}\ n\left(\frac{1}{n}\sum g(w_i, b)\right)' W_n \left(\frac{1}{n}\sum g(w_i, b)\right)$

$Q = E\left(\left.\frac{\partial g(w_i, b)}{\partial b'}\right|_{b=\beta}\right)$

$\Omega = E\, g(w_i, \beta)\, g(w_i, \beta)'$

$g(w_i, \beta) = Z_i(Y_i - X_i'\beta)$

$\sqrt{n}\left(\hat\beta^{GMM} - \beta\right) \xrightarrow{d} N(0, V_\beta)$

$V_\beta = (Q'WQ)^{-1} Q'W \Omega W Q (Q'WQ)^{-1}$

↳ minimized at $W = \Omega^{-1}$

$\tilde\beta$ 是 $W = I$ 时

step1: $\hat\Omega = \frac{1}{n}\sum_{i=1}^{n} g(w_i, \tilde\beta)\, g(w_i, \tilde\beta)'$

step2: $\min_b n\left(\frac{1}{n}\sum g(w_i, b)\right)'\left(\frac{1}{n}\sum g(w_i, \tilde\beta)\, g(w_i, \tilde\beta)'\right)^{-1} \frac{1}{n}\sum g(w_i, b)$

GMM: The General Case

Theorem

Distribution of Nonlinear GMM Estimator

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right).$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}}=\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{Q}\right)^{-1}\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{\Omega} \boldsymbol{W} \boldsymbol{Q}\right)\left(\boldsymbol{Q}^{\prime} \boldsymbol{W} \boldsymbol{Q}\right)^{-1}$$

with

$$\boldsymbol{\Omega}=\mathbb{E}\left(\boldsymbol{g}_i \boldsymbol{g}_i^{\prime}\right)$$

$$\boldsymbol{Q}=\mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\beta}^{\prime}} \boldsymbol{g}_i(\boldsymbol{\beta})\right).$$

If the efficient weight matrix is used then

$$\boldsymbol{V}_{\boldsymbol{\beta}}=\left(\boldsymbol{Q}^{\prime} \boldsymbol{\Omega}^{-1} \boldsymbol{Q}\right)^{-1}.$$

GMM: The General Case

- The asymptotic covariance matrices can be estimated by sample counterparts of the population matrices.
- For the case of a general weight matrix,

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)^{-1}\left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)\left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}}\right)^{-1}$$

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{g}_i\left(\widehat{\boldsymbol{\beta}}\right) - \overline{\boldsymbol{g}}\right)\left(\boldsymbol{g}_i\left(\widehat{\boldsymbol{\beta}}\right) - \overline{\boldsymbol{g}}\right)'$$

$$\overline{\boldsymbol{g}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{g}_i\left(\widehat{\boldsymbol{\beta}}\right)$$

$$\widehat{\boldsymbol{Q}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{g}_i\left(\widehat{\boldsymbol{\beta}}\right).$$

- For efficient weight matrix,

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{Q}}\right)^{-1}.$$