Advanced Econometrics

Lecture 10: Generalized Method of Moments (Hansen Chapter 11)

Instructor: Ma, Jun

Renmin University of China

December 13, 2021

2SLS for overidentified IV models

- We consider the simple model with endogeneity with the intercept known to be zero:

$$\begin{aligned} Y_i &= \alpha D_i + e_i \\ \mathbb{E}(e_i) &= 0 \\ \operatorname{Cov}(D_i, e_i) &\neq 0. \end{aligned}$$

- Suppose that we have ℓ IVs $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{i\ell})'$ which satisfies $\mathbb{E}(e_i Z_{ij}) = 0$, for $j = 1, 2, ..., \ell$.
- The first-stage (the reduced-form equation) of 2SLS uses the linear projection of D_i on 1 and $\boldsymbol{Z}_i$:

$$\begin{aligned} D_i &= \pi_0 + \pi_1 Z_{i1} + \cdots + \pi_\ell Z_{i\ell} + V_i \\ \mathbb{E}(V_i) &= 0 \\ \mathbb{E}(Z_{ij} V_i) &= 0, j = 1, 2, ..., \ell. \end{aligned}$$

- Then,

$$\begin{aligned} Y_i &= \alpha D_i + e_i \\ D_i &= \pi_0 + \pi_1 Z_{i1} + \cdots + \pi_\ell Z_{i\ell} + V_i \end{aligned} \implies$$

$$Y_i = \alpha\left(\pi_0 + \pi_1 Z_{i1} + \cdots + \pi_\ell Z_{i\ell}\right) + \alpha V_i + e_i.$$

  Regression of Y_i on $\pi_0 + \pi_1 Z_{i1} + \cdots + \pi_\ell Z_{i\ell}$ consistently estimates α.
- 2SLS replaces $(\pi_0, \pi_1, ..., \pi_\ell)$ with the first-stage OLS estimates.
- We know that the true parameter α satisfies the following equations simultaneously:

$$\begin{aligned} \mathbb{E}\left(e_i\right) = 0 &\implies \mathbb{E}\left(Y_i - \alpha D_i\right) = 0 \\ \mathbb{E}\left(e_i Z_{i1}\right) = 0 &\implies \mathbb{E}\left[\left(Y_i - \alpha D_i\right) Z_{i1}\right] = 0 \\ \vdots \quad &\vdots \quad \vdots \\ \mathbb{E}\left(e_i Z_{i\ell}\right) = 0 &\implies \mathbb{E}\left[\left(Y_i - \alpha D_i\right) Z_{i\ell}\right] = 0. \end{aligned}$$

Generalized method of moments

- By the method of moments principle, an estimator $\widehat{\alpha}$ can be constructed as the solution to the sample moment equations:

$$\frac{1}{n}\sum_{i=1}^{n} g_i^0\left(\widehat{\alpha}\right) = 0,\ \frac{1}{n}\sum_{i=1}^{n} g_i^1\left(\widehat{\alpha}\right) = 0, ..., \frac{1}{n}\sum_{i=1}^{n} g_i^\ell\left(\widehat{\alpha}\right) = 0,$$

where

$$\begin{array}{rcl} g_i^0\left(a\right) & = & Y_i - aD_i \\ g_i^1\left(a\right) & = & \left(Y_i - aD_i\right) Z_{i1} \\ \vdots & \vdots & \vdots \\ g_i^\ell\left(a\right) & = & \left(Y_i - aD_i\right) Z_{i\ell}. \end{array}$$

- There may not exist a solution to the sample moment equations.
- Alternatively we solve

$$\min_{a\in\mathbb{R}}\left\|\begin{pmatrix} n^{-1}\sum_{i=1}^{n} g_i^0(a) \\ n^{-1}\sum_{i=1}^{n} g_i^1(a) \\ \vdots \\ n^{-1}\sum_{i=1}^{n} g_i^\ell(a)\end{pmatrix}\right\|^2,$$

where $\left\|(x_1,...,x_\ell)'\right\|^2=\sum_{i=1}^{\ell}x_i^2$

- Let W_n be a $(\ell+1)\times(\ell+1)$ positive definite matrix, which means that $t'W_n t>0$ for all $(\ell+1)$-dimensional vectors $t\neq 0$. W_n is called a weighting matrix.
- The generalized method of moment (GMM) estimator $\widehat{\alpha}(W_n)$ is the solution to

$$\min_{a\in\mathbb{R}}\begin{pmatrix} n^{-1}\sum_{i=1}^{n} g_i^0(a) \\ n^{-1}\sum_{i=1}^{n} g_i^1(a) \\ \vdots \\ n^{-1}\sum_{i=1}^{n} g_i^\ell(a)\end{pmatrix}' W_n \begin{pmatrix} n^{-1}\sum_{i=1}^{n} g_i^0(a) \\ n^{-1}\sum_{i=1}^{n} g_i^1(a) \\ \vdots \\ n^{-1}\sum_{i=1}^{n} g_i^\ell(a)\end{pmatrix}.$$

- Suppose that $W_n \to_p W$ for some nonrandom positive definite matrix W.
- We can show that

$$\sqrt{n}\left(\widehat{\alpha}\left(W_n\right)-\alpha\right) \to_d \mathrm{N}\left(0, \sigma^2\left(W\right)\right)$$

and $\sigma^2\left(W\right) \geq \sigma^2\left(W^*\right)$, where

$$W^* = \left(\mathbb{E}\left[\begin{pmatrix} g_i^0\left(\alpha\right) \\ g_i^1\left(\alpha\right) \\ \vdots \\ g_i^\ell\left(\alpha\right) \end{pmatrix}\begin{pmatrix} g_i^0\left(\alpha\right) \\ g_i^1\left(\alpha\right) \\ \vdots \\ g_i^\ell\left(\alpha\right) \end{pmatrix}'\right]\right)^{-1}.$$

Efficient GMM

- The efficient GMM uses the weighting matrix $\widehat{W}^* \to_p W^*$, where

$$\widehat{W}^* = \left(\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} g_i^0(\widetilde{\alpha}) \\ g_i^1(\widetilde{\alpha}) \\ \vdots \\ g_i^\ell(\widetilde{\alpha}) \end{pmatrix} \begin{pmatrix} g_i^0(\widetilde{\alpha}) \\ g_i^1(\widetilde{\alpha}) \\ \vdots \\ g_i^\ell(\widetilde{\alpha}) \end{pmatrix}' \right)^{-1}$$

  and $\widetilde{\alpha}$ is a preliminary consistent estimator, which can be the 2SLS.
- The efficient GMM estimator $\widehat{\alpha}\left(\widehat{W}^*\right)$ has the lowest asymptotic variance.
- The 2SLS is an GMM estimator which uses

$$W_n^{2\text{sls}} = \left(\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ Z_{i1} \\ \vdots \\ Z_{i\ell} \end{pmatrix} \begin{pmatrix} 1 \\ Z_{i1} \\ \vdots \\ Z_{i\ell} \end{pmatrix}' \right)^{-1}.$$

Moment Equation Models

- 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

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

We know that the true parameter $\boldsymbol{\beta}$ satisfies the system of equations.

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

Method of Moments Estimator

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

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

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

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

Overidentified Moment Equations

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

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

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

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

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

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

Definition

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

- 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:

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

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

GMM Estimator

- The GMM criterion function:

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

- First-order conditions:

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

Theorem

For the overidentified IV model

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

Theorem

If $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ *then* $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \widehat{\boldsymbol{\beta}}_{\mathrm{2sls}}$.
Futhermore, if $k = l$ *then* $\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}} = \widehat{\boldsymbol{\beta}}_{\mathrm{iv}}$.

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

- Then,

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

Distribution of GMM Esttimator

Theorem

Asymptotic Distribution of GMM Estimator.

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

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., $\widehat{\boldsymbol{W}}=\left(n^{-1} \boldsymbol{Z}^{\prime} \boldsymbol{Z}\right)^{-1}$.

Efficient GMM

- The asymptotic distribution of the GMM estimator $\widehat{\boldsymbol{\beta}}_{\text{gmm}}$ 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:

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

- Feasible efficient GMM:

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

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

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

Theorem

Asymptotic Distribution of GMM with Efficient Weight Matrix

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

where

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

Theorem

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

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

where $\operatorname{avar}\left(\widehat{\boldsymbol{\beta}}_{\text{gmm}}\right)$ *denotes the asymptotic variances:* $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\text{gmm}} - \boldsymbol{\beta}\right) \to_d \mathrm{N}\left(\mathbf{0}, \operatorname{avar}\left(\widehat{\boldsymbol{\beta}}_{\text{gmm}}\right)\right)$.

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 $\widehat{\boldsymbol{W}} = \left(n^{-1}\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$.
- Since $\widehat{\boldsymbol{W}} \rightarrow_p \mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}$, the asymptotic distribution of 2SLS is the same as that of using the weight matrix $\boldsymbol{W} = \mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}$.
- 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: $\mathbb{E}\left(e_i^2 \mid \boldsymbol{Z}_i\right) = \sigma^2$. The efficient weight matrix is $\boldsymbol{W} = \mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\sigma^{-2}$.

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

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.

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

Theorem

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

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

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 $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$.
- 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

$$\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{R}'\boldsymbol{V}_{\boldsymbol{\beta}}\boldsymbol{R}$$
$$\boldsymbol{R} = \frac{\partial}{\partial \boldsymbol{b}}\boldsymbol{r}(\boldsymbol{b})'\bigg|_{\boldsymbol{b}=\boldsymbol{\beta}}.$$

- Estimator of the asymptotic variance matrix:

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\theta}} = \widehat{\boldsymbol{R}}'\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\widehat{\boldsymbol{R}}$$
$$\widehat{\boldsymbol{R}} = \frac{\partial}{\partial \boldsymbol{b}}\boldsymbol{r}(\boldsymbol{\beta})'\bigg|_{\boldsymbol{b}=\widehat{\boldsymbol{\beta}}_{\mathrm{gmm}}}.$$

- 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 $n \to \infty$,

$$W \to_d \chi_q^2.$$

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

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

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

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

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

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

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

  where

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

- 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 $\widehat{\boldsymbol{W}} = \boldsymbol{I}_\ell$.

Theorem

Distribution of Nonlinear GMM Estimator

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

where

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

with

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

$$\boldsymbol{Q} = \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{b}'}\boldsymbol{g}_i\left(\boldsymbol{b}\right)\bigg|_{\boldsymbol{b}=\boldsymbol{\beta}}\right).$$

If the efficient weight matrix is used then

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

- 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,

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