Generalized Method of Moments

Suppose we have ℓ moment conditions

$$\underset{\ell\times 1}{E\, g(w_i, \theta_0)} = \underset{\ell\times 1}{0}, \quad \text{where } w_i := (y_i, x_i', z_i') \text{ is the data.}$$

Recall the criterion function

$$Q(\theta) = [E\, g(w_i,\theta)'] \underbrace{A'A}_{\substack{\text{pos def}\\ \text{and symmetric}}} [E\, g(w_i,\theta)]$$

the true parameter θ_0 minimizes $Q(\theta)$ for any choice of norm $A'A$. However, if $E\, g(w_i,\theta_0) \neq 0$, then different choices of $A'A$ yield a different minimizer. We refer to such value as the pseudo-true value.

GMM focuses on the sample version of the criterion function

$$\hat{\theta}_{GMM}(A_n) := \underset{\theta}{\operatorname{argmin}} \underbrace{\left[\frac{1}{n} \sum_{i=1}^{n} g(w_i,\theta)' \right] A_n' A_n \left[\frac{1}{n} \sum_{i=1}^{n} g(w_i,\theta) \right]}_{Q_n(\theta)}$$

In the linear IV model we assume

$$y_i = x_i' \underset{k\times 1}{\beta} + u_i$$

$$E \underset{\ell\times 1}{z_i}\, u_i = 0, \quad \ell > k.$$

Identification

Assume $\underset{}{E\, z_i x_i'}$ has full column rank K. To see the implications, let

$$\underset{1\times K}{x_i'} = \Big(\underbrace{\underset{1\times k_1}{x_{1i}'}}_{\substack{\text{included}\\ \text{endogenous}}} \quad \underbrace{\underset{1\times k_2}{x_{2i}'}}_{\substack{\text{included}\\ \text{exogenous}}} \Big), \quad \text{and let} \quad \underset{1\times l}{z_i'} = \Big(\underbrace{\underset{1\times l_1}{z_{1i}'}}_{\substack{\text{excluded}\\ \text{exogenous}}} \quad \underbrace{\underset{1\times k_2}{x_{2i}'}}_{\substack{\text{included}\\ \text{exogenous}}} \Big).$$

Let $\underset{k_1\times 1}{x_{1i}} = \underset{k_1\times l_1}{\Pi_1} \underset{l_1\times 1}{z_{1i}} + \underset{k_1\times k_2}{\Pi_2} \underset{k_2\times 1}{x_{2i}}$, which is a linear projection of the elements of x_{1i} onto (z_{1i}', x_{2i}'). Then

$$\underset{l\times K}{E\, z_i x_i'} = E \begin{pmatrix} z_{1i} \\ x_{2i} \end{pmatrix} \begin{pmatrix} x_{1i}' & x_{2i}' \end{pmatrix}$$

$$= \begin{pmatrix} E\, z_{1i} z_{1i}' \Pi_1' + E\, z_{1i} x_{2i}' \Pi_2' & E\, z_{1i} x_{2i}' \\ E\, x_{2i} z_{1i}' \Pi_1' + E\, x_{2i} x_{2i}' \Pi_2' & E\, x_{2i} x_{2i}' \end{pmatrix}.$$

Suppose there's a $\underset{K\times 1}{\theta} \neq 0$ such that $E\, z_i x_i' \theta = 0$. Then

$$E\, z_i x_i' \theta = \begin{pmatrix} E\, z_{1i} z_{1i}' \Pi_1' \theta_1 + E\, z_{1i} x_{2i}' \Pi_2' \theta_1 + E\, z_{1i} x_{2i}' \theta_2 \\ E\, x_{2i} z_{1i}' \Pi_1' \theta_1 + E\, x_{2i} x_{2i}' \Pi_2' \theta_1 + E\, x_{2i} x_{2i}' \theta_2 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} E\, z_{1i} z_{1i}' & E\, z_{1i} x_{2i}' \\ E\, x_{2i} z_{1i}' & E\, x_{2i} x_{2i}' \end{pmatrix}}_{\substack{\text{this second moment}\\ \text{matrix is p.d. iff}\\ \text{the exogenous variables}\\ \text{are all linearly indep}\\ \text{(this requires that } z_{1i} \text{ does}\\ \text{not form part of } x_{2i} \text{, i.e.}\\ \text{instruments are not in}\\ \text{the outcome model).}}} \underbrace{\begin{pmatrix} \Pi_1' \theta_1 \\ \Pi_2' \theta_1 + \theta_2 \end{pmatrix}}_{\substack{\text{Then this must}\\ \text{be } 0.}} = 0$$

$\Rightarrow$

Then

$$\Pi_2' \theta_1 + \theta_2 = 0$$
$$\Pi_1' \theta_1 = 0$$

Notice that if either θ_j is 0, the other must be zero as well, but this would contradict the fact that $\theta \neq 0$. Hence, suppose $\exists\, \theta_1 \neq 0$ such that $\Pi_1' \theta_1 = 0$. This means that the matrix Π_1' is rank deficient.

Therefore, we showed that

$$\text{rank}\left(E z_i x_i' \right) = k \iff \text{rank}\left(\Pi_1 \right) = k_1 .$$

so that the rank condition is equivalent to another rank condition in terms of the first stage parameters (i.e sufficiently strong linear relationship between z_i' and x_i').

From the moment condition we obtain

$$E z_i y_i = E z_i x_i' \beta$$

$$A \, E z_i y_i = \underbrace{\underset{l \times l}{A} \; \underset{l \times k}{E z_i x_i'}} \; \beta$$

$\text{rank} \leq \min(\text{rank}(A), \text{rank}(E z_i x_i')) = k$

$\text{rank} \geq \text{rank}(A) + \text{rank}(E z_i x_i') - l = k$

therefore, the Gram matrix is invertible.

$$E x_i z_i' A' A \, E z_i y_i = E x_i z_i' A' A \, E z_i x_i' \beta$$

$$\Rightarrow \boxed{\beta(A) = \left(E x_i z_i' A' A \, E z_i x_i' \right)^{-1} E x_i z_i' A' A \, E z_i y_i}$$

Consistency

Assume

(i) $A_n \xrightarrow[p]{} A$, where A is a finite matrix.

(ii) $E z_i x_i'$ has rank k

(iii) $E\, x_{ij}^2 < \infty$, $j = 1, \dots, k$

$E\, z_{ij}^2 < \infty$, $j = 1, \dots, \ell$

Then $\hat{\beta}_n(A_n) \xrightarrow[p]{} \beta$, where $\hat{\beta}_n(A_n)$ is defined as follows.

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

Notice that $E\, z_i x_i' = O(1)$ because

$$E \| z_i x_i' \| \leq \left(E \| z_i \|^2 \; E \| x_i \|^2 \right)^{1/2}$$
$$= O(1) \text{ by (iii)}.$$

Notice that by (ii) we have that $E x_i z_i' A'A \, E z_i x_i'$ is p.d.

To see this

$$\underset{\ell \times \ell}{A} \; \underset{\ell \times k}{E z_i x_i'} \; \underset{k \times 1}{v} \neq 0 \quad \text{iff} \quad v \neq 0 \text{ by rank cond.}$$

Hence $v' E x_i' z_i \, A'A \, E z_i x_i' \, v > 0$.

The implication of this is that $(E x_i z_i A'A \, E z_i x_i')^{-1} = O(1)$ because the matrix is bounded away from zero.

Write

$$
\begin{aligned}
\hat{\beta}_n(A_n) &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \; A_n' A_n \; \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \; A_n' A_n \; \frac{1}{n}\sum_{i=1}^{n} z_i u_i \\
&= \beta + \left\{ \left[E x_i z_i' + o_p(1) \right] \left[A'A + o_p(1) \right] \left[E z_i x_i' + o_p(1) \right] \right\}^{-1} \left[E z_i x_i' + o_p(1) \right] \left[A'A + o_p(1) \right] o_p(1) \\
&= \beta + \left\{ E x_i z_i' \, A'A \, E z_i x_i' + o_p(1)\, O(1) \right\}^{-1} \left[E z_i x_i' \, A'A \, o_p(1) + o_p(1)\, O(1) \right] \\
&= \beta + \left\{ \left(E x_i z_i' \, A'A \, E z_i x_i' \right)^{-1} + o_p(1) \right\} \left[E z_i x_i' \, A'A \; o_p(1) + o_p(1) \right] \\
&= \beta + \left\{ O(1) + o_p(1) \right\} \left[O(1) \; o_p(1) + o_p(1) \right] \\
&= \beta + o_p(1) .
\end{aligned}
$$

Asymptotic Normality

Assume, in addition, the following

(iv) $E z_{ij}^4 < \infty$ for all $j = 1, \dots, l$.

$E u_i^4 < \infty$

(v) $E u_i^2 z_i z_i'$ is positive definite.

Notice that (iv) implies that $E u_i^2 z_i z_i'$ is $O(1)$. To see this

$$
\begin{aligned}
E \| u_i^2 z_i z_i' \| &\le \left(E \|u_i^2\|^2 \; E \| z_i z_i' \|^2 \right)^{1/2} \\
&= \left(E u_i^4 \; E \| z_i \|^4 \right)^{1/2} \\
&= O(1).
\end{aligned}
$$

Therefore, by the CLT

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i \xrightarrow{d} N(0, \underbrace{E u_i^2 z_i z_i'}_{\Omega})
$$

Write

$$
\sqrt{n}(\hat{\beta}_n(A_n) - \beta_0) = \left(\frac{1}{n} \sum_{i=1}^n z_i x_i' A_n' A_n \frac{1}{n} \sum_{i=1}^n z_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_i x_i' A_n' A_n \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i
$$

$$
\begin{aligned}
&= \left[(Q' A' A Q)^{-1} Q' A' A + o_p(1) \right] \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i \\
&= (Q' A' A Q)^{-1} Q' A' A \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i + o_p(1) \, O_p(1) \\
&= (Q' A' A Q)^{-1} Q' A' A \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i u_i + o_p(1) \\
&\xrightarrow{d} N(0, (Q' A' A Q)^{-1} Q' A' A \Omega A' A Q (Q' A' A Q)^{-1}).
\end{aligned}
$$

Efficient GMM

The lower bound on the asymptotic variance is given by

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

To see why this is true we need to show that

$$\text{Asy Var}(A) - (Q' \Omega^{-1} Q)^{-1} \quad \text{p.s.d}$$

$$\Longleftrightarrow$$

$$Q' \Omega^{-1} Q - (\text{Asy Var}(A))^{-1} \quad \text{is p.s.d.}$$

Write

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

$$= Q' \Omega^{-1/2} [I - \underbrace{\Omega^{1/2} A' A Q}_{H} (Q' A' A \Omega A' A Q)^{-1} Q' A' A \Omega^{1/2}] \Omega^{-1/2} Q$$

$$= Q' \Omega^{-1/2} [\underbrace{I - H (H' H)^{-1} H'}_{\text{projection matrix = positive semi definite}}] \Omega^{-1/2} Q .$$

The efficient choice of $A = \Omega^{-1/2}$.

Suppose we use a matrix $B_{k\times l}$ that has rank k to linearly transform Z_i into a vector of k instruments $\underset{k\times 1}{W_i} := \underset{k\times l}{B}\, \underset{l\times 1}{Z_i}$. We then run IV with W_i.

$$\tilde{\beta}_n(B) = \left(\sum_{i=1}^n W_i X_i'\right)^{-1} \sum_{i=1}^n W_i y_i = (W'X)^{-1} W'Y$$

$$\circledast \quad W = \begin{pmatrix} W_1' \\ \vdots \\ W_n' \end{pmatrix} = \begin{pmatrix} Z_1' \\ \vdots \\ Z_n' \end{pmatrix} B' = \underset{n\times l}{Z}\, \underset{l\times k}{B'}$$

$$\tilde{\beta}_n(B) - \beta = \left(\frac{W'X}{n}\right)^{-1} \frac{W'U}{n}$$

$$\sqrt{n}\,(\tilde{\beta}_n(B) - \beta) = \left(\frac{W'X}{n}\right)^{-1} \frac{W'U}{\sqrt{n}} = \left(B\frac{Z'X}{n}\right)^{-1} B\frac{Z'U}{\sqrt{n}}$$

- $$\frac{Z'U}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n u_i Z_i \xrightarrow{d} N(0, E u_i^2 Z_i Z_i')$$

- $$\frac{Z'X}{n} \xrightarrow{p} E Z_i X_i' := Q$$

  by Slutsky's theorem $\left(B\frac{Z'X}{n}\right)^{-1} \xrightarrow{p} (BQ)^{-1}$

Then

$$\sqrt{n}\,(\tilde{\beta}_n(B) - \beta) \xrightarrow{d} N(0, (BQ)^{-1} B\Omega B' (Q'B')^{-1})$$

$$= N(0, (Q'B'(B\Omega B')^{-1}BQ)^{-1})$$

$\circledast$ $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

Recall that the Asy-var of the efficient GMM is $V^* = (Q'\Omega^{-1}Q)^{-1}$.
Then we want to show

$$V(B) - V^* \text{ p.s.d} \iff V^{*-1} - V(B)^{-1} \text{ p.s.d}$$

$$= Q'\Omega^{-1}Q - Q'B'(B\Omega B')^{-1}BQ$$

$$= \underset{k\times l}{Q'}\, \underset{l\times l}{\Omega^{-1/2}} \underbrace{\left(I_l - \underbrace{\Omega^{1/2}B'}_{H} \underbrace{(B\Omega B')^{-1}}_{H'H} \underbrace{B\Omega^{1/2}}_{H'}\right)}_{\text{projection matrix}} \underset{l\times l}{\Omega^{-1/2}}\, \underset{l\times k}{Q}$$

Therefore, it is p.s.d.

Notice that we can find the B such that it attains the lower bound V^*.

$$\underbrace{Q' B' (B \Omega B')^{-1}}_{\text{we want this to cancel each other}} \underbrace{B}_{\text{this should be this}} Q = Q' \Omega^{-1} Q$$

$$\Rightarrow \boxed{B^* = Q' \Omega^{-1}}$$ and plug it to confirm :

$$Q' \cancel{\Omega^{-1}} Q \, (Q' \cancel{\Omega^{-1} \Omega \Omega^{-1}} Q)^{-1} Q' \Omega^{-1} Q$$
$= V^*$ as desired.

We can estimate it as

$$B_n^* = \frac{X'Z}{n} \hat{\Omega}_n^{-1} = \sum_{i=1}^{n} \frac{x_i z_i'}{n} \hat{\Omega}_n^{-1}$$

where $\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i' \hat{\beta}_n)^2 z_i z_i'$

↳ some consistent estimator of β

Finally, we can write

$$\hat{\beta}_n (B_n^*) = \left(B_n \sum_{i=1}^{n} z_i x_i' \right)^{-1} B_n^* \sum_{i=1}^{n} z_i y_i$$

$$= \left(\sum_{i=1}^{n} x_i z_i' \hat{\Omega}_n^{-1} \sum_{i=1}^{n} z_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i z_i' \hat{\Omega}_n^{-1} \sum_{i=1}^{n} z_i y_i = \hat{\beta}^{2step}.$$

Hypothesis Testing

We consider the test $H_0: \beta = \beta_0$

$H_1: \beta = \beta_0 + \delta$, $\delta \in \mathbb{R}^K$ such that $\delta \neq 0$.

The proposed statistic is

$$W_n = n(\hat{\beta}_n - \beta_0)' \hat{V}_n^{-1}(A_n)(\hat{\beta}_n - \beta_0)$$

$$= \sqrt{n}(\hat{\beta}_n - \beta_0)' \hat{V}_n^{-1}(A_n) \sqrt{n}(\hat{\beta}_n - \beta_0)$$

$$= \left[V_n^{-1/2}(A_n)\sqrt{n}(\hat{\beta}_n - \beta_0)\right]' \left[V_n^{-1/2}(A_n)\sqrt{n}(\hat{\beta}_n - \beta_0)\right]$$

Under H_0

$$W_n \xrightarrow{d} \chi^2_K \quad \text{and we reject if} \quad W_n(\beta_0) > \chi^2_{1-\alpha,K}.$$

Under H_1 $\quad \hat{\beta}_n - \beta_0 = \delta + o_p(1)$, then

$$\Pr_{H_1}(W_n > \chi^2_{1-\alpha,K}) = \Pr_{H_1}\left(\frac{W_n}{n} > \frac{\chi^2_{1-\alpha,K}}{n}\right)$$

$$= \Pr_{H_1}(\delta'(V(A))^{-1}\delta + o_p(1) > o(1))$$

$$= \Pr_{H_1}(\delta'(V(A))^{-1}\delta + o_p(1) > 0)$$

Notice that $\delta'(V(A))^{-1}\delta > 0$ because $V(A)$ is pos def and bounded. Hence by taking limits

$$\lim_{n\to\infty} \Pr_{H_1}(\delta'(V(A))^{-1}\delta + o_p(1) > 0) = 1.$$

In other words, in the limit the test always rejects if H_0 is not true.

Under local alternatives $\beta = \beta_0 + \delta/\sqrt{n}$ we have that

$$\sqrt{n}\,(\hat{\beta}_n - \beta_0) = \delta + (Q' A' A\, Q)^{-1} Q' A' A \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i + o_p(1)$$

$$\xrightarrow{d} N(\delta, V(A))$$

Therefore

$$W_n(\beta_0) \xrightarrow[d]{H_1} Z_\delta' Z_\delta \sim \chi^2_K(\delta' V^{-1}(A)\,\delta)$$

The power function indexed by δ is given by

$$\underbrace{\alpha}_{\|\delta\| \to 0} \leq \Pr\big(\underbrace{\chi^2_K(\delta' V(A)^{-1}\delta) > \chi^2_{1-\alpha,K}}_{\pi(\delta' V(A)^{-1}\delta)}\big) \leq \underbrace{1}_{\|\delta\| \to \infty}$$

Overid Test

Suppose we want to test $\underset{l\times 1}{E z_i u_i} = 0$.

We may use the criterion function

$$J_n = n\left(\frac{1}{n}\sum_{i=1}^n z_i \hat{u}_i\right)' \hat{\Omega}_n^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i \hat{u}_i\right)$$

⊛ we need $\hat{u}_i$ so we'll lose degrees of freedom

$$= \underbrace{\left(\Omega^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \hat{u}_i\right)'}_{\bar{g}_n(\hat{\beta}_n)} \Omega^{1/2}\hat{\Omega}_n^{-1}\Omega^{1/2}\left(\Omega^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \hat{u}_i\right)$$

where

$$\bar{g}_n(\beta) = \Omega^{-1/2}\frac{Z'u}{\sqrt{n}} - \Omega^{-1/2}Z'X\sqrt{n}(\hat{\beta}_n - \beta)$$

$$= \Omega^{-1/2}\frac{Z'u}{\sqrt{n}} - \Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}\frac{Z'u}{\sqrt{n}} + o_p(1),$$

$$= \left\{I_l - \Omega^{-1/2}Q(Q'\Omega^{-1}Q)^{-1}Q\,\Omega^{-1}\Omega^{1/2}\right\}\Omega^{-1/2}\frac{Z'u}{\sqrt{n}} + o_p(1).$$

$$= \left\{I_l - \underbrace{\Omega^{-1/2}Q}_{R}(\underbrace{Q'\Omega^{-1/2}}_{R'}\underbrace{\Omega^{-1/2}Q}_{R})^{-1}\underbrace{Q\,\Omega^{-1/2}}_{R'}\right\}\Omega^{-1/2}\frac{Z'u}{\sqrt{n}} + o_p(1).$$

$$= \underbrace{\left\{I_l - \underset{l\times k}{R}\,\underset{k\times k}{(R'R)^{-1}}\,\underset{k\times l}{R'}\right\}}_{\text{idempotent and symmetric}}\Omega^{-1/2}\frac{Z'u}{\sqrt{n}} + o_p(1).$$

= projection

Write

$$J_n = Z' \left[I_\ell - R(R'R)^{-1}R' \right] \left[I_\ell + o_p(1) \right] \left[I_\ell - R(R'R)^{-1}R' \right] Z$$

$$= Z' \left[I_\ell - R(R'R)^{-1}R' \right] Z + o_p(1)$$

$$= Z' \left(I_\ell - R(R'R)^{-1}R' \right) Z + o_p(1)$$

$$= Z' (H \Lambda H') Z + o_p(1)$$

H orthonormal, i.e. pure rotation (normality is preserved)

$$= \tilde{Z}' \Lambda \tilde{Z} + o_p(1)$$

Λ diagonal of 0 and 1s. $\mathrm{Tr}(I_\ell - R(R'R)^{-1}R') = \ell - k$

$$= \sum_{j=1}^{\ell-k} \tilde{Z}_j + o_p(1)$$

$$\xrightarrow{d} \chi^2_{\ell-k} .$$

Under local alternatives we get normals around $\Omega^{-1/2} \delta$.
Hence the non centrality parameter is

$$\delta' \Omega^{-1/2} \left(I_\ell - R(R'R)^{-1}R' \right) \Omega^{-1/2} \delta$$

$$= \delta' \left(\Omega^{-1} - \Omega^{-1} \Omega^{-1/2} \left(\Omega^{-1/2} \Omega^{-1} \Omega^{-1/2} \right)^{-1} \Omega^{-1/2} \Omega^{-1} \right) \delta$$

if $\delta = \Omega^{-1/2} \pi$ then the non-centrality parameter collapses to zero (trivial power).