# Econometrics II Tutorial 9: GMM Estimation

## Exercise 1

To show that the difference Var(any GMM estimator)-Var(efficient GMM estimator (with $\Xi_0 = \Lambda_0^{-1}$)), which is

$$(G_0'\Xi_0 G_0)^{-1}\ G_0'\Xi_0\Lambda_0\Xi_0 G_0\ (G_0'\Xi_0 G_0)^{-1} - (G_0'\Lambda_0^{-1}G_0)^{-1}$$

is positive semidefinite, it suffices to show that the difference between the inverses is negative semidefinite or that

$$(G_0'\Lambda_0^{-1}G_0) - (G_0'\Xi_0 G_0)\ (G_0'\Xi_0\Lambda_0\Xi_0 G_0)^{-1}(G_0'\Xi_0 G_0) \equiv \Delta_0$$

is positive semidefinite.

Define $D = \underset{(L\times P)}{\Lambda_0^{\frac{1}{2}}\Xi_0 G_0}$ and $D'D = G_0'\Xi_0\Lambda_0\Xi_0 G_0$ since $\Lambda_0 = \Lambda_0'$ and $\Xi_0 = \Xi_0'$.
$\Rightarrow G_0'\Xi_0 G_0 = G_0'\Lambda_0^{-\frac{1}{2}}\Lambda_0^{\frac{1}{2}}\Xi_0 G_0 = G_0'\Lambda_0^{-\frac{1}{2}}D$

and

$$\begin{aligned}
\Delta_0 &= G_0'\Lambda_0^{-1}G_0 - G_0'\Lambda^{-\frac{1}{2}}D(D'D)^{-1}D'\Lambda_0^{-\frac{1}{2}}G_0 \\
&= G_0'\Lambda_0^{-\frac{1}{2}}\underbrace{(I_L - D(D'D)^{-1}D')}_{M_D}\Lambda_0^{-\frac{1}{2}}G_0 \\
&= G_0'\Lambda_0^{-\frac{1}{2}}M_D\Lambda_0^{-\frac{1}{2}}G_0 \\
&= G_0'\Lambda_0^{-\frac{1}{2}}M_D'M_D\Lambda_0^{-\frac{1}{2}}G_0 \qquad \text{(property of } M_d\text{)} \\
&= (M_D\Lambda_0^{-\frac{1}{2}}G_0)'(M_D\Lambda_0^{-\frac{1}{2}}G_0)
\end{aligned}$$

This is a quadratic form which is by construction positive semidefinite. Q.E.D.

# Exercise 2

In general, we have

$$V_0 = A_0^{-1} B_0 A_0^{-1} = (G_0' \Xi_0 G_0)^{-1} (G_0' \Xi_0 \Lambda_0 \Xi_0 G_0)\ (G_0' \Xi_0 G_0)^{-1}$$

Just identifying: $L = P$, so that $G_{0LxP}$ is square and invertible. Moreover, $\Xi_0$ must be of full rank. Hence,

$$\begin{aligned} V_0 &= G_0^{-1} \Xi_0^{-1} (G_0')^{-1} G_0' \Xi_0 \Lambda_0 \Xi_0 G_0 G_0^{-1} \Xi_0^{-1} (G_0')^{-1} \\ &= G_0^{-1} \Lambda_0 (G_0')^{-1} \\ &= (G_0' \Lambda_0^{-1} G_0)^{-1} \end{aligned}$$

Note that just identifying restrictions imply that $Q_N(\theta)$ is in sample exactly zero because the moment conditions all attain zero:

$$\sum_{i=1}^{N} g(w_i, \hat{\theta}) = 0$$

Then weighting by $\Xi$ does not have an effect on the results. In this sense, all weights are optimal.

# Exercise 3

For a short repetition of binary choice models( $y_i$ is one or zero), see ML slides 7 and 52 and following. Furthermore, make sure that you can distinguish between $g(w, \theta), G_0$ and $G(...)$.

(a) $\mathrm{E}(x_i' u_i) = \mathrm{E}[x_i'(y_i - G(x_i \theta_0))] = 0$
$\mathrm{E}(y_i^2 | x_i) = \mathrm{E}(y_i | x_i) = G(x_i \theta_0)$ (for binary choice)

$\Rightarrow g(w_i, \theta_0) = x_i'(y_i - G(x_i \theta_0)) = x_i' u_i$

$\Rightarrow G_0 = \mathrm{E}[\nabla_\theta g(w_i, \theta_0)] = \mathrm{E}[-x_i' x_i \cdot G'(x_i \theta_0)]$

$$\begin{aligned} \Lambda_0 &= \mathrm{E}[g(w_i, \theta_0) g(w_i, \theta_0)'] \\ &= \mathrm{E}[x_i' u_i u_i' x_i] = E(x_i' x_i u_i^2) = \mathrm{E}[\mathrm{E}(u_i^2 | x_i) x_i' x_i] \\ &= \mathrm{E}[\mathrm{E}[y_i^2 - 2 y_i G(x_i \theta_0) + G(x_i \theta_0)^2 | x_i] x_i' x_i] \\ &= \mathrm{E}[[\mathrm{E}(y_i^2 | x_i) - 2 G(x_i \theta_0)\, \mathrm{E}(y_i | x_i) + G(x_i \theta_0)^2] x_i' x_i] \\ &= \mathrm{E}[[G(x_i \theta_0) - G(x_i \theta_0)^2] x_i' x_i] \end{aligned}$$

Since we use just identifying restrictions, the variance is:

$$V_0 = \{G_0' \cdot \Lambda_0^{-1} \cdot G_0\}^{-1}$$

$$\Rightarrow V_0 = \{\mathrm{E}[G'(x_i\theta_0)x_i'x_i] \cdot \mathrm{E}[[G(x_i\theta_0) - G(x_i\theta_0)^2]x_i'x_i]^{-1} \cdot \mathrm{E}[G'(x_i\theta_0)x_i'x_i]\}^{-1}$$

(b) For logit model: $G(z) = \frac{e^z}{1+e^z}$

For logit it holds that $G'(z) = G(z) - G(z)^2$, since $G'(z) = \frac{e^z}{(1+e^z)^2}$ and $G'(z) = G(z) - G(z)^2 = \frac{e^z}{(1+e^z)^2}$

$$\begin{aligned}\Rightarrow V_{0,\mathrm{GMM}} &= \left[\mathrm{E}[G'(x_i\theta)x_i'x_i]\,\mathrm{E}[G'(x_i\theta)x_i'x_i]^{-1}\,\mathrm{E}[G'(x_i\theta)x_i'x_i]\right]^{-1} \\ &= \mathrm{E}(G'(x_i\theta)x_i'x_i)^{-1}\end{aligned}$$

Compare to Avar of CMLE (A is taken from Lecture ML, slide 47):

$$\begin{aligned}A_0(x_i,\theta) = \mathrm{E}[A(x_i,\theta)] &= \mathrm{E}\left[\frac{G'(x_i\theta)^2}{G(x_i\theta)(1-G(x_i\theta))}x_i'x_i\right] \\ &= \mathrm{E}\left[\frac{G'(x_i\theta)^2}{G'(x_i\theta)}x_i'x_i\right] = \mathrm{E}[G'(x_i\theta)x_i'x_i]^{-1} = V_{0,\mathrm{GMM}}\end{aligned}$$

Why?

The GMM approach under just identifying restrictions uses $\sum g(w_i,\hat{\theta}) = \sum x_i'\hat{u}_i \overset{!}{=} 0$ as FOC (FOC simplified due to meaningless weighting $\Xi$ under just identifying restrictions).

The CMLE approach uses $\sum s_i(\hat{\theta}) = \sum x_i'\hat{u}_i \overset{!}{=} 0$ as FOC.

$\Rightarrow$ Since both FOC are the same, both approaches lead to the same result!

(c) For general link function the CMLE FOC is

$$\sum s_i(\hat{\theta}) = \sum \frac{G'(x_i\hat{\theta})}{G(x_i\hat{\theta})(1-G(x_i\hat{\theta}))}x_i'\hat{u}_i \overset{!}{=} 0 \qquad \text{(see ML slides 42-44)}$$

The GMM uses the same FOC if it uses

$$g(w_i,\theta) = s_i(\theta) = \frac{G'(x_i\theta)}{G(x_i\theta)(1-G(x_i\theta_0))}x_i'u_i$$

$\Rightarrow$ Then again same Avars of both approaches since FOC identical!