# Generalized Method of Moments

1. Population moment condition

$$E\left[f\left(\theta_0, z_t\right)\right] = \mathbf{0}$$

where $z_t$ is an $(r \times 1)$ vector of observable variables, $\theta_0 \in \Theta$ is a $(k \times 1)$ vector of true value of parameters and $f : \mathbb{R}^k \times \mathbb{R}^r \rightarrow \mathbb{R}^q$ vector valued function.

2. Sample analog

$$g\left(\theta, Z_T\right) = \frac{1}{T} \sum_{t=1}^{T} f\left(\theta, z_t\right)$$

where $Z_T = \left(z_T', z_{T-1}', \cdots, z_1'\right)$. A proper version of law of large numbers ensures that

$$plim \frac{1}{T} \sum_{t=1}^{T} f\left(\theta, z_t\right) = E\left[f\left(\theta, z_t\right)\right]$$

3. Traditional method of moments estimator; Suppose that $q = k$, i.e., we have $k$ moment conditions for $k$ unknown parameters. Then, we can find $\widehat{\theta}$ such that

$$g\left(\widehat{\theta}, Z_T\right) = \frac{1}{T} \sum_{t=1}^{T} f\left(\widehat{\theta}, z_t\right) = 0 \qquad ((*))$$

We expect $\widehat{\theta}$ to have good statistical properties such as consistency and asymptotic normality. Good news, under some conditions, it is indeed the case that

$$\widehat{\theta} \xrightarrow{p} \theta_0 \text{ and } \sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, V\right)$$

4. Generalized method of moments estimator; What if $q > k$? We have more equations than unknowns. There does not exist any solution which satisfies (*). A quick and easy solution is to discard $(q - k)$ equations - moment conditions- and apply the method of moments technique. Then, which to discard? The answer should be arbitrary. Moreover, the moment conditions discarded do include valuable information on the parameter. The generalized method of moments technique uses all $q$ moment conditions by

weighting them. Suppose that we have a sequence of $(q \times q)$ positive semi definite matrix $W_T$ converging to a positive definite matrix $W_0$. Then, GMM estimator is defined as

$$\begin{aligned}
\widehat{\theta}_{GMM} &= \arg\min_{\theta} J^{W_T}(\theta, Z_T) = \arg\min_{\theta} [g(\theta, Z_T)]' W_T [g(\theta, Z_T)] \\
&= \arg\min_{\theta} \left[\frac{1}{T}\sum_{t=1}^{T} f(\theta, z_t)\right]' W_T \left[\frac{1}{T}\sum_{t=1}^{T} f(\theta, z_t)\right]
\end{aligned}$$

Since $W_T$ is positive semi definite, $J(\theta, Z_T) \geq 0$. The minimization problem finds the solution which makes the value of the objective function $J(\theta, Z_T)$ as near to zero as possible.

5. Under some regularity conditions,

$$\begin{aligned}
& \widehat{\theta}_{GMM} \xrightarrow{p} \theta_0 \\
\sqrt{T} g(\theta_0, Z_T) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\theta_0, z_t) \xrightarrow{d} N(\mathbf{0}, \Gamma) \\
& \sqrt{T}\left(\widehat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} N(\mathbf{0}, V)
\end{aligned}$$

where

$$\begin{aligned}
\Gamma &= \operatorname{plim} \frac{\partial g(\theta_0, Z_T)}{\partial \theta'} = \operatorname{plim} \frac{1}{T}\sum_{t=1}^{T} f(\theta_0, z_t) \\
V &= [\Gamma' W_0 \Gamma]^{-1} [\Gamma' W_0 \Omega W_0 \Gamma] [\Gamma' W_0 \Gamma]^{-1}
\end{aligned}$$

with

$$\Omega = \lim_{j \to \infty} \sum_{i=-j}^{j} E\left([f(\theta_0, z_t)] [f(\theta_0, z_{t-i})]'\right)$$

6. Optimal weighting matrix; we can show that the optimal weighting matrix which minimizes the asymptotic variance of the estimator is given by

$$W_0 = \Omega^{-1}$$

Then,

$$\sqrt{T}\left(\widehat{\theta}_{GMM} - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left[\Gamma' \Omega^{-1} \Gamma\right]^{-1}\right)$$

7. How to actually compute the GMM estimator;

    1. Construct the sample analog $g(\theta, Z_T)$ of the population moments $E[f(\theta, z_t)]$.
    2. Set $W_T = I$ and construct the objective function $J^I(\theta) = [g(\theta, Z_T)]' [g(\theta, Z_T)]$. Find
    $$\widehat{\theta}^{(1)}_{GMM} = \arg\min_{\theta} J^I(\theta)$$
    3. Calculate
    $$\widehat{\Omega}^{(2)} = \widehat{\Lambda}^{(2)}_0 + \sum_{h=1}^{l} \varpi(h, l) \left[\widehat{\Lambda}^{(2)}_h + \widehat{\Lambda}^{(2)\prime}_h\right]$$
    where
    $$\widehat{\Lambda}^{(2)}_h = \frac{1}{T} \sum_{t=h+1}^{T} \left[f\left(\widehat{\theta}^{(1)}_{GMM}, z_t\right)\right]' \left[f\left(\widehat{\theta}^{(1)}_{GMM}, z_t\right)\right]$$
    and $\varpi(h, l)$ is a kernel for weighs. For example, Newey-West (Bartlett) suggested $\varpi(h, l) = \left[1 - \frac{h}{l+1}\right]$. The number of lags included in the estimation of $\widehat{\Omega}^{(2)}$ should increase at a proper rate, say $O\left(T^{\frac{2}{3}}\right)$ ,as the sample size $T$ grows in order to ensure the consistency of $\widehat{\Omega}^{(2)}$.
    4. Construct next step objective function
    $$J^{\widehat{\Omega}^{(2)}}(\theta) = [g(\theta, Z_T)]' \left[\widehat{\Omega}^{(2)}\right]^{-1} [g(\theta, Z_T)]$$
    and find the minimum with $\widehat{\theta}^{(1)}_{GMM}$ as the starting value of iteration. Define
    $$\widehat{\theta}^{(2)}_{GMM} = \arg\min_{\theta} J^{\widehat{\Omega}^{(2)}}(\theta)$$
    5. Continue the previous two steps until convergence.
    6. Asymptotic variance can be calculated as
    $$\widehat{V} = \left[\left(\frac{\partial g\left(\widehat{\theta}_{GMM}, Z_T\right)}{\partial \theta'}\right)\right]' \left[\widehat{\Omega}\left(\widehat{\theta}_{GMM}\right)\right]^{-1} \left[\left(\frac{\partial g\left(\widehat{\theta}_{GMM}, Z_T\right)}{\partial \theta'}\right)\right]^{-1}$$

8. Example I ; OLS estimator with i.i.d. error

$$\begin{aligned} \varepsilon_t &= y_t - \beta' x_t \\ E(\varepsilon_t) &= 0,\ E\left(\varepsilon_t^2\right) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0\ t \neq s \end{aligned}$$

Then,

$$E[f(\theta_0, z_t)] = \mathbf{0} \Rightarrow E[x_t(y_t - \beta_0' x_t)] = 0$$

Therefore,

$$g(\theta, Z_T) = \frac{1}{T}\sum_{t=1}^{T} f(\theta, z_t) \ \Rightarrow \frac{1}{T}\sum_{t=1}^{T} [x_t(y_t - \beta' x_t)]' [x_t(y_t - \beta' x_t)] = 0$$

We have $k$ moment conditions in $k$ unknown parameters.

9. Example II; IV

$$E[f(\theta_0, z_t)] = \mathbf{0} \Rightarrow E[w_t(y_t - \beta_0' x_t)] = 0$$

In case of i.i.d.,

$$\min_{\beta} J = \left[\frac{1}{T}\sum_{t=1}^{T} [w_t(y_t - \beta' x_t)]\right]' \left[\frac{1}{T}\sum_{t=1}^{T} [w_t(y_t - \beta' x_t)]\right]$$

In case of heterogeneous and correlated data,

$$\min_{\beta} J = \left[\frac{1}{T}\sum_{t=1}^{T} [w_t(y_t - \beta' x_t)]\right]' \Omega^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} [w_t(y_t - \beta' x_t)]\right]$$

10. Example III; ML

$$\begin{aligned} E[s(\theta_0)] &= E\left[\frac{\dot{l}(\theta_0)}{l(\theta_0)}\right] = 0 \\ \min_{\theta} \left[\frac{1}{T}\sum_{t=1}^{T} s(\theta)\right]' \left[\frac{1}{T}\sum_{t=1}^{T} s(\theta)\right] &= 0 \end{aligned}$$

11. Example IV; intertemporal optimization

$$u'(c_t) = \beta E\left[(1 + r_{t+1})\, u'(c_{t+1}) \mid X_t\right]$$

where $X_t = \left(x_t', x_{t-1}' x_{t-2}', \cdots\right)$. Consider

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma} \qquad \gamma > 0 \text{ and } \gamma \neq 1$$

Then,

$$c_t^{-\gamma} = \beta E\left[(1 + r_{t+1})\, c_{t+1}^{-\gamma} \mid X_t\right]$$

$$E\left[1 - \beta(1 + r_{t+1})\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \mid X_t\right] = 0$$

parameter vector; $(\beta, \gamma)$, data set $(c_t, r_t)$.