LECTURE NOTE 11

GENERALIZED METHOD OF MOMENTS (GMM)

1. INTRODUCTION

Generalized method of moments is a framework that extends the orthogonality condition framework in Note 9. Let $Data_i$ denote the variables observed for the i^{th} cross-section unit and assume random sampling: $Data_i$ i.i.d. for $i = 1, \ldots, n$. We are given a moment function $\psi(\,\cdot\,,\,\cdot\,)$. We assume that there is a unique point γ in some parameter space that satisfies the *key condition*:

$$E[\psi(Data_i, \gamma)] = 0. \tag{1}$$

Here ψ is $L \times 1$, γ is $K \times 1$, and $L \geq K$. We want to make inferences on γ.

This framework contains the orthogonality condition framework in Note 9 as a special case. The moment function is

$$\psi(Data_i, a) = W_i(Q_i - R_i a).$$

In Note 9, we had

$$Q_i = R_i\gamma + V_i, \quad E(W_i V_i) = 0.$$

So

$$E[\psi(Data_i, \gamma)] = E[W_i(Q_i - R_i\gamma)] = E(W_i V_i) = 0,$$

and the key condition is satisfied. The special aspect of the moment function $\psi(Data_i, a) = W_i Q_i - W_i R_i a$ is that it depends upon the parameter a in a linear way. (For a given value of the first argument, $\psi(\,\cdot\,,\,\cdot\,)$ is an affine function of the second argument.) So GMM extends the framework to allow for moment functions that are nonlinear in the parameter. We shall need this extension in our discussion of likelihood methods in Note 13.

2. GMM ESTIMATOR

Define

$$g(a) = \frac{1}{n}\sum_{i=1}^{n} \psi(Data_i, a).$$

Then by the law of large numbers,

$$g(\gamma) \xrightarrow{p} E[\psi(Data_i, \gamma)] = 0 \quad \text{as} \quad n \to \infty.$$

This suggests obtaining an estimator $\hat{\gamma}$ from

$$\hat{\gamma} = \arg\min_a g(a)'\hat{C}g(a), \tag{2}$$

where $\hat{C}$ converges in probability to a nonrandom $L \times L$ matrix C, which is positive definite and symmetric. The first-order condition for the minimization in (2) is

$$(\partial g(\hat{\gamma})'/\partial a)\hat{C}g(\hat{\gamma}) = 0.$$

So the estimator satisfies

$$\hat{D}g(\hat{\gamma}) = 0, \tag{3}$$

with $\hat{D} = (\partial g(\hat{\gamma})'/\partial a)\hat{C}$. The condition in (3) will be very useful in obtaining the limit distribution of the estimator.

3. LIMIT DISTRIBUTION

Assume that $\hat{\gamma}$ is a consistent estimate of γ: $\hat{\gamma} \xrightarrow{p} \gamma$, and that it satisfies

$$\hat{D}g(\hat{\gamma}) = 0,$$

where $\hat{D}$ converges in probability to a $K \times L$ nonrandom matrix D, which satisfies the rank condition

$$DE[\frac{\partial \psi(Data_i, \gamma)}{\partial a'}] \quad \text{nonsingular}.$$

Then we have

$$0 = \hat{D}g(\hat{\gamma}) = \hat{D}[g(\gamma) + \frac{\partial g(\gamma^*)}{\partial a'}(\hat{\gamma} - \gamma)], \tag{4}$$

where, by the mean value theorem, this expansion holds for some point γ^* on the line segment connecting $\hat{\gamma}$ and γ. (There is a different point γ^* for each component of g.) By the central limit theorem,

$$\sqrt{n}g(\gamma) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \text{Cov}(\psi(Data_i, \gamma)) = E[\psi(Data_i, \gamma)\psi(Data_i, \gamma)'].$$

From (4),

$$\sqrt{n}(\hat{\gamma} - \gamma) = -[\hat{D}\frac{\partial g(\gamma^*)}{\partial a'}]^{-1}\hat{D}\sqrt{n}g(\gamma).$$

Because γ^* is on the line segment connecting $\hat{\gamma}$ and γ, and $\hat{\gamma} \xrightarrow{p} \gamma$, we have $\gamma^* \xrightarrow{p} \gamma$. Because at any fixed value a, $(\partial g(a)/\partial a')$ converges in probability to $E[\partial\psi(Data_i, a)/\partial a']$, under regularity conditions we can obtain

$$[\hat{D}\frac{\partial g(\gamma^*)}{\partial a'}]^{-1}\hat{D} \xrightarrow{p} \left[DE[\frac{\partial\psi(Data_i, \gamma)}{\partial a'}]\right]^{-1} D \equiv \alpha.$$

Then Slutsky (ii) implies that

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} -\alpha \cdot \mathcal{N}(0, \Sigma) = \mathcal{N}(0, \alpha\Sigma\alpha').$$

We have provided a heuristic argument for the following

Claim 1. $\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(0, \Lambda)$ with $\Lambda = \alpha\Sigma\alpha'$ and

$$\alpha = \left[DE[\frac{\partial\psi(Data_i, \gamma)}{\partial a'}]\right]^{-1} D,$$
$$\Sigma = \text{Cov}(\psi(Data_i, \gamma)) = E[\psi(Data_i, \gamma)\psi(Data_i, \gamma)'].$$

4. OPTIMAL WEIGHT MATRIX

Following the argument in Note 10, it can be shown that the optimal choice for the weight matrix C is

$$C^* = \Sigma^{-1}.$$

The corresponding value for the weight matrix D is

$$D^* = E[\frac{\partial \psi(Data_i, \gamma)'}{\partial a}]\Sigma^{-1}.$$

With the optimal weight matrix, the asymptotic covariance matrix for $\hat{\gamma}$ is

$$\Lambda^* = \Big[E[\frac{\partial \psi(Data_i, \gamma)'}{\partial a}]\Sigma^{-1}E[\frac{\partial \psi(Data_i, \gamma)}{\partial a'}]\Big]^{-1}.$$