LECTURE NOTE 12

MINIMUM DISTANCE

1. INTRODUCTION

We first encountered a minimum-distance estimator in Section 4 of Note 4. Working with the panel data model with complete conditioning, we obtained a least squares fit corresponding to the linear predictor of Y_{it} given $Z_{i1}, \ldots, Z_{iT}$ and other variables. Doing this for $t = 1, \ldots, T$ resulted in a matrix $\hat{\Pi}$ of least-squares coefficients, which were then arranged in a vector $\hat{\pi}$. The model imposed restrictions on the population values π, but the least-squares estimates in $\hat{\pi}$ did not impose these restrictions. The restrictions were imposed using a minimum-distance estimator. The input for the minimum-distance estimator is not the original data on Y_{it} and Z_{it}, but rather the statistic $\hat{\pi}$ formed from the original data.

We also used a minimum-distance estimator in Section 4 of Note 5. Working with an autoregression model for panel data, we formed the sample covariances corresponding to the population covariances $\mathrm{Cov}(Y_{is}, Y_{it})$ for $s, t = 1, \ldots, T$. These sample covariances were arranged in a vector $\hat{\sigma}$. The model imposed restrictions on the population values σ, but the sample covariances in $\hat{\sigma}$ did not impose these restrictions. The restrictions were imposed using a minimum-distance estimator. The input for the minimum-distance estimator is not the original data on Y_{it}, but rather the statistic $\hat{\sigma}$ formed from the original data.

This note sets up a general framework for minimum-distance estimation and provides a limit distribution that can be used for inference.

2. MINIMUM-DISTANCE ESTIMATOR

We are given a statistic $\hat{\pi}$ with a limit normal distribution centered at π:

$$\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} \mathcal{N}(0, \Omega). \tag{1}$$

We are also given a distance function $h(\,\cdot\,,\,\cdot\,)$, which is continuously differentiable. We assume that there is a unique point γ in some parameter space that satisfies the *key condition*:

$$h(\pi, \gamma) = 0. \tag{2}$$

Here h is $L \times 1$, γ is $K \times 1$, and $L \geq K$. In Note 4, Section 4, the form of the distance function is $h(\hat{\pi}, a) = \hat{\pi} - G \cdot a$, where G is a given, known matrix (consisting of zeros and ones).

Because $h(\pi, \gamma) = 0$ and $\hat{\pi}$ is a consistent estimate of π, there is motivation for obtaining an estimator $\hat{\gamma}$ from

$$\hat{\gamma} = \arg\min_{a} h(\hat{\pi}, a)' \hat{C} h(\hat{\pi}, a), \tag{3}$$

where $\hat{C}$ converges in probability to a $L \times L$ nonrandom matrix C, which is positive definite and symmetric. The first-order condition for the minimization in (3) is

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

So the estimator satisfies

$$\hat{D} h(\hat{\pi}, \hat{\gamma}) = 0, \tag{4}$$

with $\hat{D} = (\partial h(\hat{\pi}, \hat{\gamma})' / \partial a) \hat{C}$. The condition in (4) 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

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

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

$$D\frac{\partial h(\pi,\gamma)}{\partial a'} \quad \text{nonsingular.}$$

The derivation of the limit distribution is similar to the argument used for GMM in Note 11. Apply the mean value theorem:

$$0 = \hat{D}h(\hat{\pi},\hat{\gamma}) = \hat{D}[h(\pi,\gamma) + \frac{\partial h(\pi^*,\gamma^*)}{\partial \pi'}(\hat{\pi}-\pi) + \frac{\partial h(\pi^*,\gamma^*)}{\partial a'}(\hat{\gamma}-\gamma),$$

where (π^*,γ^*) is on the line segment connecting $(\hat{\pi},\hat{\gamma})$ and (π,γ). Solving for $(\hat{\gamma}-\gamma)$ gives

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

Because $(\hat{\pi},\hat{\gamma}) \xrightarrow{p} (\pi,\gamma)$ and (π^*,γ^*) is on the line segment connecting $(\hat{\pi},\hat{\gamma})$ and (π,γ), we have $(\pi^*,\gamma^*) \xrightarrow{p} (\pi,\gamma)$. Then Slutsky (i) implies that

$$[\hat{D}\frac{\partial h(\pi^*,\gamma^*)}{\partial a'}]^{-1}\hat{D}\frac{\partial h(\pi^*,\gamma^*)}{\partial \pi'} \xrightarrow{p} [D\frac{\partial h(\pi,\gamma)}{\partial a'}]^{-1}D\frac{\partial h(\pi,\gamma)}{\partial \pi'}.$$

Define

$$\alpha = [D\frac{\partial h(\pi,\gamma)}{\partial a'}]^{-1}D.$$

Then Slutsky (ii) implies that

$$\sqrt{n}(\hat{\gamma}-\gamma) \xrightarrow{d} -\alpha\frac{\partial h(\pi,\gamma)}{\partial \pi'}\mathcal{N}(0,\Omega) = \mathcal{N}(0,\alpha\Sigma\alpha'),$$

with

$$\Sigma = \frac{\partial h(\pi,\gamma)}{\partial \pi'}\Omega\frac{\partial h(\pi,\gamma)'}{\partial \pi}.$$

We have established

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

$$\alpha = [D\frac{\partial h(\pi,\gamma)}{\partial a'}]^{-1}D,$$

$$\Sigma = \frac{\partial h(\pi,\gamma)}{\partial \pi'}\Omega\frac{\partial h(\pi,\gamma)'}{\partial \pi}.$$

4. OPTIMAL WEIGHT MATRIX

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

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

The corresponding value for the weight matrix D is

$$D^* = \frac{\partial h(\pi,\gamma)'}{\partial a}\Sigma^{-1}.$$

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

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

5. DELTA METHOD

Suppose that $\gamma = g(\pi)$ for a given, known function g, which is continuously differentiable. As before we are given a statistic $\hat{\pi}$ with $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} \mathcal{N}(0, \Omega)$. By Slutsky (i), we can obtain a consistent estimate of γ from $\hat{\gamma} = g(\hat{\pi})$. Then the minimum-distance framework can be used to obtain a limit distribution for $\hat{\gamma}$. The moment function is

$$h(\hat{\pi}, a) = g(\hat{\pi}) - a.$$

The minimum distance estimator is $\hat{\gamma} = g(\hat{\pi})$, so

$$h(\hat{\pi}, \hat{\gamma}) = 0$$

(and we can set $D = I$). Since $\alpha = -I$, Claim 1 gives

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

with

$$\Sigma = \frac{\partial g(\pi)}{\partial \pi'}\Omega\frac{\partial g(\pi)'}{\partial \pi}.$$

This is known as the *delta method*.