LECTURE NOTE 10

OPTIMAL WEIGHT MATRIX

## 1. INTRODUCTION

As in Note 9, we shall work with the following general framework:

$$Q_i = R_i\gamma + V_i, \tag{1}$$

$$E(W_iV_i) = 0, \tag{2}$$

where we observe $(Q_i, R_i, W_i)$ for $i = 1, \ldots, n$. As in Note 6, we shall assume random sampling, so that $(Q_i, R_i, W_i)$ are independent and identically distributed (i.i.d.) from some joint distribution. The system aspect is that $Q_i$ can be a vector: $Q_i$ is $H \times 1$, $R_i$ is $H \times K$, the parameter vector $\gamma$ is $K \times 1$, and the vector of errors $V_i$ is $H \times 1$. Estimation of $\gamma$ is based on the orthogonality between $W_i$ and the error $V_i$. The matrix $W_i$ is $L \times H$, so $W_iV_i$ is a $L \times 1$ vector, and $E(W_iV_i) = 0$ provides $L$ orthogonality conditions to estimate the $K$ components of $\gamma$. We need $L \geq K$.

Note 9, equation $(26')$ has the following moment equation:

$$S_{WQ} = S_{WR}\gamma + S_{WV},$$

which leads to the minimum distance estimator:

$$\begin{aligned}\hat{\gamma} &= \arg\min_a (S_{WQ} - S_{WR}a)'\hat{C}(S_{WQ} - S_{WR}a) \\ &= (S'_{WR}\hat{C}S_{WR})^{-1}S'_{WR}\hat{C}S_{WQ}.\end{aligned}$$

Here $\hat{C}$ is a positive definite, symmetric matrix that converges in probability to a nonrandom matrix $C$, which is positive definite and symmetric. By random sampling, $W_iV_i$ is i.i.d., and so

$$\mathrm{Cov}(S_{WQ} - S_{WR}\gamma) = \mathrm{Cov}(S_{WV}) = \frac{1}{n}\mathrm{Cov}(W_i V_i) = \frac{1}{n}\Sigma.$$

Suppose that $\Sigma$ is proportional to an identity matrix. Then symmetry suggests that we simply use Euclidean distance, so that $\hat{C} = C = I$. We shall show in Section 2 that $C = I$ is in fact optimal in this special case. So in this special case, we should obtain $\hat{\gamma}$ from a least-squares fit of $S_{WQ}$ on $S_{WR}$.

Now let $\Sigma$ be a general, positive definite, symmetric matrix. It has a square root:

$$\Sigma = \Sigma^{1/2}\Sigma^{1/2},$$

where $\Sigma^{1/2}$ is positive definite, symmetric. Define

$$\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$$

and use it to transform the moment equation:

$$\Sigma^{-1/2}S_{WQ} = (\Sigma^{-1/2}S_{WR})\gamma + \Sigma^{-1/2}S_{WV},$$

which we can write as

$$\tilde{S}_{WQ} = \tilde{S}_{WR}\gamma + \tilde{S}_{WV}.$$

Then

$$\mathrm{Cov}(\tilde{S}_{WV}) = \Sigma^{-1/2}(\frac{1}{n}\Sigma)\Sigma^{-1/2} = \frac{1}{n}I_L.$$

Now we are in the special case and should obtain $\hat{\gamma}$ from a least-squares fit of $\tilde{S}_{WQ}$ on $\tilde{S}_{WR}$:

$$\begin{aligned}
\hat{\gamma} &= (\tilde{S}'_{WR}\tilde{S}_{WR})^{-1}\tilde{S}'_{WR}\tilde{S}_{WQ} \\
&= (S'_{WR}\Sigma^{-1/2}\Sigma^{-1/2}S_{WR})^{-1}S'_{WR}\Sigma^{-1/2}\Sigma^{-1/2}S_{WQ} \\
&= (S'_{WR}\Sigma^{-1}S_{WR})^{-1}S'_{WR}\Sigma^{-1}S_{WQ}.
\end{aligned}$$

So the optimal choice for the weight matrix in the general case is $C = \Sigma^{-1}$, and $\hat{C}$ should be a consistent estimate of $\Sigma^{-1}$:

$$\hat{C} = \hat{\Sigma}^{-1} \xrightarrow{p} \Sigma^{-1} = C.$$

An alternative is to apply a weight matrix $\hat{D}$ directly to the moment equation, as in Note 9, equations (30), (31), and (32). The optimal choice is

$$\hat{D} = S'_{WR}\hat{\Sigma}^{-1} \xrightarrow{p} [E(W_i R_i)]'\Sigma^{-1},$$

with

$$\hat{\gamma} = (\hat{D}S_{WR})^{-1}\hat{D}S_{WQ}.$$

## 2. OPTIMAL WEIGHT MATRIX WHEN $\Sigma = I$

Note 9, Section 5 shows in Claim 2 that the limit distribution is

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

with $\Lambda = \alpha\Sigma\alpha'$ and

$$\alpha = [DE(W_i R_i)]^{-1}D, \quad \Sigma = E(W_i V_i V_i' W_i').$$

Define $B = E(W_i R_i)$. When $\Sigma = I$, we have

$$\Lambda = (DB)^{-1}DD'(B'D')^{-1}.$$

If $D = B'$, this reduces to

$$\Lambda^* = (B'B)^{-1}.$$

So we need to show that for any choice of $D$ such that $DB$ is nonsingular, $\Lambda \geq \Lambda^*$. Here the inequality means that $\Lambda - \Lambda^*$ is a positive semidefinite matrix, so that $l'(\Lambda - \Lambda^*)l \geq 0$ for any $K \times 1$ vector $l$.

*Claim 1*. For any $K \times L$ matrix $D$ such that $DB$ is nonsingular,

$$(DB)^{-1}DD'(B'D')^{-1} \geq (B'B)^{-1}.$$

*Proof*. First note that

$$I_L - B(B'B)^{-1}B' \geq 0,$$

because for any $L \times 1$ vector $a$,

$$a'(I_L - B(B'B)^{-1}B')a = a'(a - \hat{a}) = (a - \hat{a})'(a - \hat{a}) \geq 0,$$

with

$$\hat{a} = B(B'B)^{-1}B'a.$$

(This says that the sum of squared residuals from a least-squares fit of $a$ on $B$ is non-negative, and we have used $(a - \hat{a})$ orthogonal to $\hat{a}$.) If an $L \times L$ matrix $G$ is positive semidefinite, then so is $HGH'$ for any $K \times L$ matrix $H$, because

$$l'(HGH')l = a'Ga \geq 0,$$

where the $L \times 1$ vector $a = H'l$. So with

$$H = (DB)^{-1}D$$

(so $HB = I_K$), we have

$$H(I_L - B(B'B)^{-1}B')H' = (DB)^{-1}DD'(B'D')^{-1} - (B'B)^{-1} \geq 0. \quad \diamond$$