LECTURE 4
SIMULTANEOUS EQUATIONS III: FULL INFORMATION ML (FIML)

Definition

Consider again the model defined in Lecture 2,

$$\begin{aligned} \Gamma_0 Y_i &= B_0 Z_i + U_i, \\ E Z_i U_i' &= 0, \end{aligned}$$

where subscript 0 is used to denote the true values of the coefficients. We assume that all m equations are identified through zero (and unity) restrictions on the elements of Γ and B. Thus, we have in fact that

$$\begin{aligned} \Gamma_0 &= \Gamma(\delta_0), \\ B_0 &= B(\delta_0), \end{aligned}$$

where $\delta = (\delta_1', \ldots, \delta_m')'$, and δ's are the coefficients in

$$\begin{aligned} y_{1i} &= X_{1,i}' \delta_{0,1} + u_{1i}, \\ &\ldots \\ y_{mi} &= X_{m,i}' \delta_{0,m} + u_{mi}, \end{aligned}$$

as defined in Lecture 3.

Assume that

$$U_i | Z_i \sim N(0, \Sigma_0)$$

and iid across i's. This implies that

$$Y_i | Z_i \sim N\left(\Gamma_0^{-1} B_0 Z_i, \Gamma_0^{-1} \Sigma_0 \left(\Gamma_0^{-1}\right)'\right),$$

and iid across the observations. For a square matrix A, let $|A|$ denote its determinant. The conditional density of Y_i given Z_i is

$$f(y|Z_i) = (2\pi)^{-m/2} \left|\Gamma_0^{-1}\Sigma_0\left(\Gamma_0^{-1}\right)'\right|^{-1/2} \exp\left(-\frac{1}{2}\left(y - \Gamma_0^{-1}B_0 Z_i\right)' \left(\Gamma_0^{-1}\Sigma_0\left(\Gamma_0^{-1}\right)'\right)^{-1}\left(y - \Gamma_0^{-1}B_0 Z_i\right)\right),$$

and the log-likelihood function for $(Y_1', \ldots, Y_n')'$ is then can be written as

$$\begin{aligned} Q_n(\Gamma, B, \Sigma) &= \frac{1}{n}\sum_{i=1}^n \log f(Y_i|Z_i) \\ &= -\frac{m}{2}\log(2\pi) + \frac{1}{2}\log\left|\Gamma^{-1}\Sigma\left(\Gamma^{-1}\right)'\right|^{-1} - \frac{1}{2n}\sum_{i=1}^n \left(Y_i - \Gamma^{-1}BZ_i\right)'\left(\Gamma^{-1}\Sigma\left(\Gamma^{-1}\right)'\right)^{-1}\left(Y_i - \Gamma^{-1}BZ_i\right) \\ &= -\frac{m}{2}\log(2\pi) + \log|\Gamma| + \frac{1}{2}\log\left|\Sigma^{-1}\right| - \frac{1}{2n}\sum_{i=1}^n (\Gamma Y_i - BZ_i)'\Sigma^{-1}(\Gamma Y_i - BZ_i). \qquad (1) \end{aligned}$$

In order to obtain the ML estimates of the parameters, the log-likelihood function must be maximized with respect to Σ and unknown elements of Γ and B. It is useful first to derive *concentrated* log-likelihood by maximizing Q_n with respect to Σ, taking Γ and B as fixed. Using

$$\begin{aligned} \frac{\partial \log|A|}{\partial A} &= (A')^{-1}, \text{ and} \\ \frac{\partial(c'Ac)}{\partial A} &= cc', \end{aligned}$$

we obtain:

$$2\frac{\partial Q_n(\Gamma, B, \Sigma)}{\partial \Sigma^{-1}} = \Sigma - n^{-1}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'.$$

Hence, given Γ and B, the ML estimator of Σ is

$$\widehat{\Sigma}(\Gamma, B) = n^{-1}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'. \tag{2}$$

The concentrated log-likelihood is then

$$\begin{aligned}
Q_n(\Gamma, B) &= -\frac{m}{2}\log(2\pi) + \log|\Gamma| + \frac{1}{2}\log\left|\widehat{\Sigma}^{-1}(\Gamma, B)\right| - \frac{1}{2n}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)'\,\widehat{\Sigma}^{-1}(\Gamma, B)(\Gamma Y_i - BZ_i) \\
&= -\frac{m}{2}\log(2\pi) + \log|\Gamma| + \frac{1}{2}\log\left|\widehat{\Sigma}^{-1}(\Gamma, B)\right| - \frac{1}{2n}\sum_{i=1}^{n} tr\left((\Gamma Y_i - BZ_i)'\,\widehat{\Sigma}^{-1}(\Gamma, B)(\Gamma Y_i - BZ_i)\right) \\
&= -\frac{m}{2}\log(2\pi) + \log|\Gamma| + \frac{1}{2}\log\left|\widehat{\Sigma}^{-1}(\Gamma, B)\right| - \frac{1}{2n}\sum_{i=1}^{n} tr\left(\widehat{\Sigma}^{-1}(\Gamma, B)(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'\right) \\
&= -\frac{m}{2}\log(2\pi) + \log|\Gamma| + \frac{1}{2}\log\left|\widehat{\Sigma}^{-1}(\Gamma, B)\right| - \frac{1}{2}tr\left(\widehat{\Sigma}^{-1}(\Gamma, B)\,n^{-1}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'\right) \\
&= -\frac{m}{2}(\log(2\pi) + 1) + \log|\Gamma| + \frac{1}{2}\log\left|\widehat{\Sigma}^{-1}(\Gamma, B)\right|, \\
&= -\frac{m}{2}(\log(2\pi) + 1) + \log|\Gamma| - \frac{1}{2}\log\left|n^{-1}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'\right| \qquad (3) \\
&= -\frac{m}{2}(\log(2\pi) + 1) - \frac{1}{2}\log\left|\Gamma^{-1}\right|^2 - \frac{1}{2}\log\left|n^{-1}\sum_{i=1}^{n}(\Gamma Y_i - BZ_i)(\Gamma Y_i - BZ_i)'\right| \\
&= -\frac{m}{2}(\log(2\pi) + 1) - \frac{1}{2}\log\left|n^{-1}\sum_{i=1}^{n}\left(Y_i - \Gamma^{-1}BZ_i\right)\left(Y_i - \Gamma^{-1}BZ_i\right)'\right|.
\end{aligned}$$

We can ignore the first summand in the above equation and redefine

$$Q_n(\Gamma, B) = \left|n^{-1}\sum_{i=1}^{n}\left(Y_i - \Gamma^{-1}BZ_i\right)\left(Y_i - \Gamma^{-1}BZ_i\right)'\right|.$$

The FIML estimator of (Γ, B) is defined as

$$\left(\widehat{\Gamma}, \widehat{B}\right) = \arg\min_{\Gamma, B} Q_n(\Gamma, B).$$

There exists no closed form solution for the FIML estimator of Γ and B, and the concentrated log-likelihood must be maximized numerically. Alternatively, since some elements of Γ and B are restricted to zero, we can define the FIML estimator of δ,

$$\widehat{\delta} = \arg\min_{\delta} Q_n(\Gamma(\delta), B(\delta)).$$

Consistency of FIML

The normality assumptions has been used to define the FIML estimator. It is not, however, required for consistency and asymptotic normality of FIML. If the underlying distribution of the data is not normal, FIML is actually a Quasi-ML estimator.

A Quasi-ML estimator is consistent if $Q_n(\Gamma, B) \to_p Q(\Gamma, B)$ uniformly in Γ and B, Γ_0 and B_0 uniquely minimize $Q(\Gamma, B)$ over some compact set, and Q is continuous.

Note that

$$Y_i - \Gamma^{-1} B Z_i = -\left(\Gamma^{-1} B - \Gamma_0^{-1} B_0\right) Z_i + \Gamma_0^{-1} U_i.$$

One can show that

$$n^{-1} \sum_{i=1}^{n} \left(Y_i - \Gamma^{-1} B Z_i\right)\left(Y_i - \Gamma^{-1} B Z_i\right)' \to_p \left(\Gamma^{-1} B - \Gamma_0^{-1} B_0\right) E Z_i Z_i' \left(\Gamma^{-1} B - \Gamma_0^{-1} B_0\right)' + \Gamma_0^{-1} \Sigma_0 \left(\Gamma_0'\right)^{-1},$$

and, therefore, by the Slutsky's Theorem,

$$\begin{aligned} Q_n(\Gamma, B) &\to_p Q(\Gamma, B) \\ &= \left| \left(\Gamma^{-1} B - \Gamma_0^{-1} B_0\right) E Z_i Z_i' \left(\Gamma^{-1} B - \Gamma_0^{-1} B_0\right)' + \Gamma_0^{-1} \Sigma_0 \left(\Gamma_0'\right)^{-1} \right|, \end{aligned}$$

and, in fact, convergence is uniform. Next, if A and B are two positive semi-definite matrices, then $|A + B| \geq |A|$. Assuming that both $E Z_i Z_i'$ and $\Sigma_0 = E U_i U_i'$ are positive definite, we have that

$$Q(\Gamma, B) \geq \left| \Gamma_0^{-1} \Sigma_0 \left(\Gamma_0'\right)^{-1} \right|.$$

Note also that $\Gamma_0^{-1} \Sigma_0 \left(\Gamma_0'\right)^{-1}$ does not depend on Γ and B and can be attained by choosing $\Gamma = \Gamma_0$ and $B = B_0$. Thus, Γ_0 and B_0 minimize $Q(\Gamma, B)$. Further, since $\Pi_0 = \Gamma_0^{-1} B_0$, a minimizer of $Q(\Gamma, B)$ must be a solution to

$$\Gamma \Pi_0 = B.$$

The rank identification condition guarantees that Γ_0 and B_0 is the only solution to the above equation. Thus,

$$\left(\hat{\Gamma}, \hat{B}\right) \to_p (\Gamma_0, B_0).$$

Since Γ and B are composed of the elements of δ, the above result can be written as well as

$$\widehat{\delta} \to_p \delta_0.$$

Asymptotic equivalence of FIML and 3SLS

One can show further that the FIML estimator of δ is asymptotically normal and has the same asymptotic variance as that of the 3SLS estimator (under conditional homoskedasticity). We won't provide a formal prove of asymptotic normality, but will illustrate the reason for asymptotic equivalence of FIML and 3SLS.

The FIML estimator solves the ML first-order conditions. Consider (1). Note that some elements of $\Gamma(\delta)$ and $B(\delta)$ are restricted to ones or zeros. Therefore, we cannot simply set to zero the derivative of the log-likelihood function Q_n with respect to Γ and B. The solution is to set to zero only the partial derivatives that correspond to the unrestricted elements of Γ and B. Let's introduce the following partitions:

$$\begin{aligned} \Gamma &= \begin{pmatrix} \Gamma_1' \\ \vdots \\ \Gamma_m' \end{pmatrix}, \\ B &= \begin{pmatrix} B_1' \\ \vdots \\ B_m' \end{pmatrix}, \\ \Sigma^{-1} &= \begin{pmatrix} \sigma^{11} & \dots & \sigma^{1m} \\ \dots & \dots & \dots \\ \sigma^{m1} & \dots & \sigma^{mm} \end{pmatrix}, \end{aligned}$$

where $\left(\Gamma_j', B_j'\right)'$ is the vector of structural parameters corresponding to equation j, and σ^{ij} is a scalar. The last summand in (1) can be written as

$$-\frac{1}{2n}\sum_{i=1}^{n}\sum_{s=1}^{m}\sum_{t=1}^{m}\sigma^{st}\left(\Gamma_s' Y_i - B_s' Z_i\right)\left(\Gamma_t' Y_i - B_t' Z_i\right).$$

Its derivative with respect to B_s' (the coefficients of the exogenous variables in equation s) is given by

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\sigma^{st}\left(\Gamma_t' Y_i - B_t' Z_i\right) Z_i' \\ = \;& n^{-1}\sum_{i=1}^{n}\left[\Sigma^{-1}\right]_s'\left(\Gamma Y_i - B Z_i\right) Z_i', \end{aligned}$$

where $\left[\Sigma^{-1}\right]_s'$ is row s of Σ^{-1}. Hence, the derivative of $-(2n)^{-1}\sum_{i=1}^{n}\left(\Gamma Y_i - B Z_i\right)'\Sigma^{-1}\left(\Gamma Y_i - B Z_i\right)$ with respect to B is given by

$$\frac{\partial Q_n\left(\Gamma, B, \Sigma\right)}{\partial B} = n^{-1}\sum_{i=1}^{n}\Sigma^{-1}\left(\Gamma Y_i - B Z_i\right) Z_i'.$$

The derivative corresponding to equation j is given by

$$n^{-1}\sum_{i=1}^{n}\left[\Sigma^{-1}\right]_j'\left(\Gamma Y_i - B Z_i\right) Z_i'.$$

In equation j only l_j elements of B are unrestricted. Hence, we have

$$\left[\widehat{\Sigma}^{-1}\right]_j'\sum_{i=1}^{n}\left(\widehat{\Gamma} Y_i - \widehat{B} Z_i\right) Z_{ji}' = 0, \tag{4}$$

where $\widehat{\Gamma} = \Gamma\left(\widehat{\delta}\right)$, $\widehat{B} = B\left(\widehat{\delta}\right)$, $\widehat{\delta}$ is the FIML estimate of δ, and $\widehat{\Sigma} = \widehat{\Sigma}\left(\widehat{\Gamma}, \widehat{B}\right)$.

Similarly, the derivative of $-(2n)^{-1}\sum_{i=1}^{n}\left(\Gamma Y_i - B Z_i\right)'\Sigma^{-1}\left(\Gamma Y_i - B Z_i\right)$ with respect to Γ is

$$-n^{-1}\sum_{i=1}^{n}\Sigma^{-1}\left(\Gamma Y_i - B Z_i\right) Y_i',$$

and, since

$$\frac{\partial \log|A|}{\partial A} = \left(A'\right)^{-1},$$

we have

$$\frac{\partial Q_n\left(\Gamma, B, \Sigma\right)}{\partial \Gamma} = \left(\Gamma'\right)^{-1} - n^{-1}\sum_{i=1}^{n}\Sigma^{-1}\left(\Gamma Y_i - B Z_i\right) Y_i'. \tag{5}$$

Next, from (2) we have

$$\begin{aligned} \left(\Gamma'\right)^{-1} &= \hat{\Sigma}^{-1} n^{-1}\sum_{i=1}^{n}\left(\Gamma Y_i - B Z_i\right)\left(\Gamma Y_i - B Z_i\right)'\left(\Gamma'\right)^{-1} \\ &= n^{-1}\sum_{i=1}^{n}\hat{\Sigma}^{-1}\left(\Gamma Y_i - B Z_i\right)\left(Y_i - \Pi Z_i\right)'. \end{aligned} \tag{6}$$

Combining (5) and (6), we obtain

$$\frac{\partial Q_n\left(\Gamma, B, \hat{\Sigma}\right)}{\partial \Gamma} = n^{-1}\sum_{i=1}^{n}\hat{\Sigma}^{-1}\left(\Gamma Y_i - B Z_i\right)\left(\Pi Z_i\right)'.$$

The derivative corresponding to equation j is given by

$$n^{-1}\sum_{i=1}^{n}\left[\hat{\Sigma}^{-1}\right]_j'\left(\Gamma Y_i - BZ_i\right)\left(\Pi Z_i\right)'.$$

By the same argument as in (4), we have,

$$\left[\widehat{\Sigma}^{-1}\right]_j'\sum_{i=1}^{n}\left(\widehat{\Gamma}Y_i-\widehat{B}Z_i\right)\widehat{Y}_{ji}'=0, \tag{7}$$

where $\widehat{Y}_i = \Pi\left(\widehat{\delta}\right)Z_i$.

Let's define

$$\widehat{X}_{j,i}=\begin{pmatrix} Z_{ji} \\ \widehat{Y}_{ji}\end{pmatrix}.$$

Then, after transposing, (4) and (7) can be written together as

$$\begin{aligned}
0 &= \sum_{i=1}^{n}\widehat{X}_{j,i}\left(\widehat{\Gamma}Y_i-\widehat{B}Z_i\right)'\left[\widehat{\Sigma}^{-1}\right]_j \\
&= \widehat{X}_j'\begin{pmatrix} y_1 - X_1\widehat{\delta}_1 & \dots & y_m - X_m\widehat{\delta}_m\end{pmatrix}\left[\widehat{\Sigma}^{-1}\right]_j \\
&= \widehat{X}_j'\left(\sum_{s=1}^{m}\left(y_s - X_s\widehat{\delta}_s\right)\widehat{\sigma}^{js}\right) \\
&= \sum_{s=1}^{m}\widehat{\sigma}^{js}\widehat{X}_j'\left(y_s - X_s\widehat{\delta}_s\right),
\end{aligned}$$

where the matrix notation as in Lecture 3. Collecting the results for all m equations, we obtain

$$\begin{aligned}
0 &= \begin{pmatrix} \widehat{\sigma}^{11}\widehat{X}_1' & \dots & \widehat{\sigma}^{1m}\widehat{X}_1' \\ \dots & \dots & \dots \\ \widehat{\sigma}^{m1}\widehat{X}_m' & \dots & \widehat{\sigma}^{mm}\widehat{X}_m'\end{pmatrix}\begin{pmatrix} y_1 - X_1\widehat{\delta}_1 \\ \vdots \\ y_m - X_m\widehat{\delta}_m\end{pmatrix} \\
&= \begin{pmatrix} \widehat{X}_1' & & 0 \\ & \dots & \\ 0 & & \widehat{X}_m'\end{pmatrix}\left(\widehat{\Sigma}^{-1}\otimes I_n\right)\begin{pmatrix} y_1 - X_1\widehat{\delta}_1 \\ \vdots \\ y_m - X_m\widehat{\delta}_m\end{pmatrix} \\
&= \widehat{X}'\left(\widehat{\Sigma}^{-1}\otimes I_n\right)\left(Y - X\widehat{\delta}\right).
\end{aligned}$$

Hence, the FIML estimator must satisfy

$$\widehat{\delta}=\left(\widehat{X}'\left(\widehat{\Sigma}^{-1}\otimes I_n\right)X\right)^{-1}\widehat{X}'\left(\widehat{\Sigma}^{-1}\otimes I_n\right)Y.$$

Note that this is not a closed-form expression for $\widehat{\delta}$, since $\widehat{X}$ and $\widehat{\Sigma}$ both depend on $\widehat{\delta}$. However, it is similar to the expression for 3SLS in Lecture 3. Together with consistency of $\widehat{\delta}$, it explains the asymptotic equivalence of 3SLS and FIML.