14.383: Econometrics II

Derivation of FIML and LIML

Michael Powell
Department of Economics, MIT
October 12th, 2006

For all that follows, I will abstract from the issue of identification and assume that all equations are just identified or overidentified. There are two general classes of estimators for the parameters of interest in a system of equations. The first class is that of limited information estimators. These estimators use information only about a particular equation to estimate its parameters, in the sense that they do not take advantage of knowledge about the covariance structure across equations. Examples of limited information estimators are two-stage least squares (2SLS), optimal instrumental variables estimators (OIV), and limited information maximum likelihood (LIML).

Full information estimators, on the other hand, do in fact make use of information about the covariance structure of the entire system of equations. As a result, they are able to improve upon the efficiency of limited information estimators. Examples of full information estimators include three-stage least squares (3SLS) and full information maximum likelihood (FIML). The disadvantage of using full information estimators is that if at least one equation is misspecified, then these estimators will be inconsistent.

These notes will focus on deriving in explicit detail the FIML and LIML estimators. I will begin in section 1 with several mathematical preliminaries that are helpful in establishing several results. Section 2 is devoted to the derivation of FIML. LIML follows as a special case of FIML and will be the subject of section 3. Much of the material in these notes follows Alexis Leon's 2003 notes.

1 Mathematical Preliminaries

This section will be a cookbook presentation of a collection of mathematical propositions that will be helpful throughout. It should be viewed as a mathematical appendix and will thus have very little in the way of motivation. Understanding the proofs is not necessary for understanding the derivation of the estimators but I will include as many of the the proofs as I can for completeness.

1.1 Basic Linear Algebra and Statistics

Proposition 1 *Let A, B, and C be $m \times n$, $n \times k$, and $k \times m$ matrices respectively. Then*

$$tr\,(ABC) = tr\,(CAB) = tr\,(BCA)\,.$$

Proof. I will only establish that $tr\,(ABC) = tr\,(CAB)$. Establishing the second equality involves repeating the same argument. Here, we write

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},\ B = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix},\ C = \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{k1} & \cdots & c_{km} \end{bmatrix}.$$

Multiplying these matrices together, we get

$$\begin{aligned} ABC &= (AB)\,C = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \cdots & \sum_{i=1}^{n} a_{1i}b_{ik} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} a_{mi}b_{i1} & \cdots & \sum_{i=1}^{n} a_{mi}b_{ik} \end{bmatrix} \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{k1} & \cdots & c_{km} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{k}\sum_{i=1}^{n} a_{1i}b_{ij}c_{j1} & \cdots & \sum_{j=1}^{k}\sum_{i=1}^{n} a_{1i}b_{ij}c_{jm} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{k}\sum_{i=1}^{n} a_{mi}b_{ij}c_{j1} & \cdots & \sum_{j=1}^{k}\sum_{i=1}^{n} a_{mi}b_{ij}c_{jm} \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned}
CAB &= (CA)\,B = \begin{bmatrix} \sum_{i=1}^{m} c_{1i}a_{i1} & \cdots & \sum_{i=1}^{m} c_{1i}a_{in} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} c_{ki}a_{i1} & \cdots & \sum_{i=1}^{m} c_{ki}a_{in} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{n}\sum_{i=1}^{m} c_{1i}a_{ij}b_{j1} & \cdots & \sum_{j=1}^{n}\sum_{i=1}^{m} c_{1i}a_{ij}b_{jk} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n}\sum_{i=1}^{m} c_{ki}a_{ij}b_{j1} & \cdots & \sum_{j=1}^{n}\sum_{i=1}^{m} c_{ki}a_{ij}b_{jk} \end{bmatrix}.
\end{aligned}$$

Thus,

$$\begin{aligned}
tr\,(ABC) &= \sum_{\ell=1}^{m}\sum_{j=1}^{k}\sum_{i=1}^{n} a_{\ell i}b_{ij}c_{j\ell} \\
tr\,(CAB) &= \sum_{\ell=1}^{k}\sum_{j=1}^{n}\sum_{i=1}^{m} c_{\ell i}a_{ij}b_{j\ell} \\
&= \sum_{i=1}^{m}\sum_{\ell=1}^{k}\sum_{j=1}^{n} a_{ij}b_{j\ell}c_{\ell k} = tr\,(ABC)
\end{aligned}$$

which concludes the proof. ■

Proposition 2 *Let U have density $f_U(u)$. Suppose U is a function of Y. That is, $U = g(Y)$. Then the density of Y is given by*

$$f_Y(y) = f_U(g(y)) \left| \det\left(\frac{\partial g(Y)}{\partial Y}\right)\Big|_{Y=y} \right|.$$

Proposition 3 *Let A be an $n \times n$ matrix. Then* $\det A = \det A'$.

Proposition 4 *Let A be an $n \times n$ matrix. Then* $\det A^{-1} = \frac{1}{\det A} = (\det A)^{-1}$.

1.2 Vector and Matrix Calculus

Proposition 5 *Let A be an $m \times n$ matrix and let t be a $n \times 1$ vector. Then*

$$\frac{\partial At}{\partial t'} = A.$$

Proof.

$$At = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{mn} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}t_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}t_i \end{bmatrix}$$

and

$$\begin{aligned}
\frac{\partial At}{\partial t'} &= \frac{\partial \begin{bmatrix} \sum_{i=1}^{n} a_{1i}t_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}t_i \end{bmatrix}}{\partial \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix}} = \begin{bmatrix} \frac{\partial \sum_{i=1}^{n} a_{1i}t_i}{\partial t_1} & \cdots & \frac{\partial \sum_{i=1}^{n} a_{1i}t_i}{\partial t_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sum_{i=1}^{n} a_{mi}t_i}{\partial t_1} & \cdots & \frac{\partial \sum_{i=1}^{n} a_{mi}t_i}{\partial t_n} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = A,
\end{aligned}$$

which completes the proof. ■

Proposition 6 *Let A be an $m \times n$ matrix and let B be an $n \times n$ matrix. Suppose A is a function of C. Then*

$$\frac{\partial tr\left(A'AB\right)}{\partial C} = 2\frac{\partial A'}{\partial C}AB.$$

Proposition 7 *Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix. Then*

$$\frac{\partial tr\left(AB\right)}{\partial A} = B$$

Proposition 8 *Let A be an $n \times n$ matrix. Then*

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

1.3 Selection Vectors

We often are interested in characterizing (or making use) of a single element or column of a matrix. Here, I will derive some tools that will allow us to think rigorously about the selection of such an element or column of a matrix.

Notation 9 *Let A be an $m \times n$ matrix. We denote the $j'th$ column of A by $A_{(j)}$. Thus, we can write A as $A = \begin{bmatrix} A_{(1)} & \cdots & A_{(n)} \end{bmatrix}$. Similarly, we denote the $i'th$ row of A by A_i. That is, $A = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix}$. Finally, we often denote the $(i,j)th$ element of A by $[A]_{ij}$.*

Notation 10 *Define the vector $e_j = \begin{bmatrix} 0 & \cdots & 0 & \underbrace{1}_{jth\ entry} & 0 & \cdots & 0 \end{bmatrix}$. So as to avoid dealing with excessive amounts of notation, I will avoid mention of the order of e_j and assume throughout that it is of the proper order so as to be conformable with whatever matrix.*

Proposition 11 *Let A be an $m \times n$ matrix. Then $A_{(j)} = Ae_j'$.*

Proof.

$$Ae_j' = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = A_{(j)},$$

which was what we wanted to show. ■

Proposition 12 *Let A be an $m \times n$ matrix. Then $A_i = e_i A$.*

Proof. Similar to the previous proposition. ■

Proposition 13 *Let A be an $m \times n$ matrix. Then $[A]_{ij} = e_i A e_j'$.*

Proof.

$$e_i A e_j' = e_i A_{(j)} = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{bmatrix} = a_{ij} = [A]_{ij},$$

as desired. ■

1.4 Vec Operator

Definition 14 *Let A be an $m \times n$ matrix. We define the **vec operator** by*

$$vec(A) = \begin{bmatrix} A_{(1)} \\ \vdots \\ A_{(n)} \end{bmatrix} = \begin{bmatrix} Ae_1' \\ \vdots \\ Ae_n' \end{bmatrix}.$$

Here, we see that $A_{(i)}$ is an $m \times 1$ vector for each i, so it follows that $vec(A)$ is an $mn \times 1$ vector.

Proposition 15 *Let A, B, and C be $m \times n$, $n \times k$, and $k \times \ell$ matrices respectively. Then*

$$vec(ABC) = (C' \otimes A)\, vec(B).$$

Proof. Here, we have

$$\begin{aligned}
(C' \otimes A)\, vec(B) &= \begin{bmatrix} c_{11}A & \cdots & c_{k1}A \\ \vdots & \ddots & \vdots \\ c_{1\ell}A & \cdots & c_{k\ell}A \end{bmatrix} \begin{bmatrix} B_{(1)} \\ \vdots \\ B_{(k)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{k} c_{i1} AB_{(i)} \\ \vdots \\ \sum_{i=1}^{k} c_{i\ell} AB_{(i)} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{k} ABe_i' c_{i1} \\ \vdots \\ \sum_{i=1}^{k} ABe_i' c_{i\ell} \end{bmatrix} = \begin{bmatrix} AB\sum_{i=1}^{k} e_i' c_{i1} \\ \vdots \\ AB\sum_{i=1}^{k} e_i' c_{i\ell} \end{bmatrix}.
\end{aligned}$$

Next, note that

$$\sum_{i=1}^{k} e_i' c_{ij} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} c_{1j} + \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} c_{2j} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} c_{kj} = \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{kj} \end{bmatrix} = C_{(j)} = Ce_j',$$

so we have

$$(C' \otimes A)\, vec(B) = \begin{bmatrix} ABCe_1' \\ \vdots \\ ABCe_\ell' \end{bmatrix} = \begin{bmatrix} (ABC)\, e_1' \\ \vdots \\ (ABC)\, e_\ell' \end{bmatrix} = vec(ABC),$$

which is the desired result. ■

2 Full Information Maximum Likelihood

Suppose we have T observations of M endogenous variables and K exogenous variables. Our system of equations can be written compactly

$$\underset{T\times M}{Y}\underset{M\times M}{B} + \underset{T\times K}{Z}\underset{K\times M}{\Gamma} = \underset{T\times M}{U}$$

where $U = \begin{bmatrix} U_1 \\ \vdots \\ U_T \end{bmatrix}$ and $U_t = \begin{bmatrix} [U]_{t1} & \cdots & [U]_{tM} \end{bmatrix}$. Assume that $U_t' \sim N(0, \Sigma)$ and $E[u_{ti}u_{sj}] = \begin{cases} \sigma_{ij} & t = s \\ 0 & t \neq s \end{cases}$. For a single observation t, we have

$$\begin{aligned}
e_t Y B + e_t Z \Gamma &= e_t U \\
Y_t B + Z_t \Gamma &= U_t
\end{aligned}$$

or

$$U_t' = B' Y_t' + \Gamma' Z_t'.$$

2.1 Log-Likelihood Function

As with all maximum likelihood estimators, we proceed by writing down the likelihood function, taking logs, and then taking first order conditions. Since $U_t' \sim N(0, \Sigma)$, we have that

$$L(B, \Gamma, \Sigma; U) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{M}{2}} |\det(\Sigma)|^{\frac{1}{2}}} \exp\left\{-\frac{U_t \Sigma^{-1} U_t'}{2}\right\}.$$

Note that, by proposition 5,

$$\frac{\partial U_t'}{\partial Y_t} = \frac{\partial (B'Y_t' + \Gamma' Z_t')}{\partial Y_t} = \frac{\partial B'Y_t'}{\partial Y_t} = B'.$$

By proposition 2, we have that

$$\begin{aligned} L(B, \Gamma, \Sigma; Y) &= \prod_{t=1}^{T} f_Y(Y_t') = \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{M}{2}} |\det(\Sigma)|^{\frac{1}{2}}} \exp\left\{-\frac{(Y_t B + Z_t \Gamma)\Sigma^{-1}(Y_t B + Z_t \Gamma)'}{2}\right\} |\det B'| \\ &= \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{M}{2}} |\det(\Sigma)|^{\frac{1}{2}}} \exp\left\{-\frac{U_t \Sigma^{-1} U_t'}{2}\right\} |\det B|, \end{aligned}$$

where in the last step, I used proposition 3 to replace $\det B'$ with $\det B$. The log-likelihood function is then

$$\begin{aligned} \mathcal{L} &= \log L(B, \Gamma, \Sigma; Y) = -\frac{MT}{2}\log(2\pi) - \frac{T}{2}\log|\det(\Sigma)| - \frac{1}{2}\sum_{t=1}^{T} U_t \Sigma^{-1} U_t' + T\log|\det B| \\ &= c + T\log|\det B| - \frac{T}{2}\log|\det(\Sigma)| - \frac{1}{2}\sum_{t=1}^{T} U_t \Sigma^{-1} U_t'. \end{aligned}$$

Next, note that

$$-\log|\det(\Sigma)| = \log\left(\frac{1}{|\det \Sigma|}\right) = \log\left|\frac{1}{\det(\Sigma)}\right| = \log\left|\det\left(\Sigma^{-1}\right)\right|$$

and using proposition 1 repeatedly,

$$\begin{aligned} \sum_{t=1}^{T} U_t \Sigma^{-1} U_t' &= tr\left(\sum_{t=1}^{T} U_t \Sigma^{-1} U_t'\right) = \sum_{t=1}^{T} tr\left(U_t \Sigma^{-1} U_t'\right) = \sum_{t=1}^{T} tr\left(\Sigma^{-1} U_t' U_t\right) \\ &= tr\left(\sum_{t=1}^{T} \Sigma^{-1} U_t' U_t\right) = tr\left(\Sigma^{-1} \sum_{t=1}^{T} U_t' U_t\right) = tr\left(\Sigma^{-1} U' U\right) \\ &= tr\left(U' U \Sigma^{-1}\right) \end{aligned}$$

so the log likelihood function becomes

$$\begin{aligned} \mathcal{L} &= c + T\log|\det B| + \frac{T}{2}\log\left|\det\left(\Sigma^{-1}\right)\right| - \frac{1}{2} tr\left(U' U \Sigma^{-1}\right) \\ &= c + T\log|\det B| + \frac{T}{2}\log\left|\det\left(\Sigma^{-1}\right)\right| - \frac{1}{2} tr\left(\Sigma^{-1} U' U\right). \end{aligned}$$

2.2 First-Order Conditions

Here, we take first order conditions only with respect to the elements of B, Γ, and Σ that we have not made a priori assumptions about. I will denote by $[A]^u$ the first order conditions with respect to only the unrestricted elements. (It will soon become clear why we need to keep track of this.) Let $\hat{B}, \hat{\Gamma}, \hat{\Sigma}^{-1}$ denote the matrices with the estimated coefficients in place of the unrestricted elements. (They also include any known elements of B, Γ, and Σ^{-1} respectively.)

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial B^u} &= \left[T \frac{\partial \log \left| \det \hat{B} \right|}{\partial B} - \frac{1}{2} \frac{\partial tr \left(\hat{U}' \hat{U} \hat{\Sigma}^{-1} \right)}{\partial B} \right]^u = 0 \\
\frac{\partial \mathcal{L}}{\partial \Gamma^u} &= \left[-\frac{1}{2} \frac{\partial tr \left(\hat{U}' \hat{U} \hat{\Sigma}^{-1} \right)}{\partial \Gamma} \right]^u = 0 \\
\frac{\partial \mathcal{L}}{\partial \left(\Sigma^{-1} \right)^u} &= \left[\frac{T}{2} \frac{\partial \log \left| \det \left(\hat{\Sigma}^{-1} \right) \right|}{\partial \Sigma^{-1}} - \frac{1}{2} \frac{\partial tr \left(\hat{\Sigma}^{-1} \hat{U}' \hat{U} \right)}{\partial \Sigma^{-1}} \right]^u = 0
\end{aligned}
$$

Recall that, by propositions $6, 7$, and 8, we have

$$
\begin{aligned}
\frac{\partial \log |\det B|}{\partial B} &= \left(B^{-1} \right)' \\
\frac{\partial tr \left(U' U \Sigma^{-1} \right)}{\partial B} &= 2 \frac{\partial U'}{\partial B} U \Sigma^{-1} = 2 Y' U \Sigma^{-1} \\
\frac{\partial tr \left(U' U \Sigma^{-1} \right)}{\partial \Gamma} &= 2 \frac{\partial U'}{\partial \Gamma} U \Sigma^{-1} = 2 Z' U \Sigma^{-1} \\
\frac{\partial tr \left(\Sigma^{-1} U' U \right)}{\partial \Sigma^{-1}} &= U' U.
\end{aligned}
$$

The first order conditions become

$$
\left[T \left(\hat{B}^{-1} \right)' - Y' \hat{U} \hat{\Sigma}^{-1} \right]^u = 0 \tag{1}
$$

$$
\left[-Z' \hat{U} \hat{\Sigma}^{-1} \right]^u = 0 \tag{2}
$$

$$
\left[T \hat{\Sigma} - \hat{U}' \hat{U} \right]^u = 0. \tag{3}
$$

Our next goal is to write these first order conditions in a more tractable form. Note that (3) becomes

$$
\hat{\Sigma} = \frac{\hat{U}' \hat{U}}{T}.
$$

Since $\hat{\Sigma} \hat{\Sigma}^{-1} = \frac{\hat{U}' \hat{U}}{T} \hat{\Sigma}^{-1} = I$ and $\hat{U} = Y \hat{B} + Z \hat{\Gamma}$, we have that (1) becomes

$$
\begin{aligned}
0 &= \left[T \left(\hat{B}^{-1} \right)' \frac{\hat{U}' \hat{U}}{T} \hat{\Sigma}^{-1} - Y' \hat{U} \hat{\Sigma}^{-1} \right]^u \\
&= \left[\left(\hat{B}^{-1} \right)' \left(\hat{B}' Y' + \hat{\Gamma}' Z' \right) \hat{U} \hat{\Sigma}^{-1} - Y' \hat{U} \hat{\Sigma}^{-1} \right]^u \\
&= \left[\left(Y' + \left(\hat{B}^{-1} \right)' \hat{\Gamma}' Z' - Y' \right) \hat{U} \hat{\Sigma}^{-1} \right]^u \\
&= \left[-\hat{\Pi}' Z' \hat{U} \hat{\Sigma}^{-1} \right]^u,
\end{aligned}
$$

or $\left[\hat{\Pi}' Z' \hat{U} \hat{\Sigma}^{-1} \right]^u = 0$, where $\hat{\Pi} = -\hat{\Gamma} \hat{B}^{-1}$. (Note that I also used the fact that $\left(\hat{B}^{-1} \right)' = \left(\hat{B}' \right)^{-1}$.) Let us now focus our attention on equation j.

If $[B]_{ij}$ is an unrestricted element of B, then, from the first order conditions, $\left[\frac{\partial\mathcal{L}}{\partial B^u}\right]_{ij}=0$. That is

$$\begin{aligned}
0 &= e_i\left[\frac{\partial\mathcal{L}}{\partial B^u}\right]e_j' = e_i\hat{\Pi}'Z'\hat{U}\hat{\Sigma}^{-1}e_j' \\
&= \left(\hat{\Pi}e_i'\right)'Z'\hat{U}\hat{\Sigma}^{-1}e_j' \\
&= \hat{\Pi}_{(i)}'Z'\hat{U}\hat{\Sigma}_{(j)}^{-1}.
\end{aligned}$$

Working towards thinking about all the elements of B that are unrestricted, consider the coefficients for the $j'th$ equation $B_{(j)}=\begin{bmatrix}B_{1j}\\ \vdots\\ B_{mj}\end{bmatrix}$. Suppose for all $i\in R(j)=\left\{i_1,\ldots,i_{r_j}\right\}$, B_{ij} is unrestricted and for all $i\notin R(j)$, B_{ij} is restricted. Then we have that

$$\begin{aligned}
\hat{\Pi}_{(i_1)}'Z'\hat{U}\hat{\Sigma}_{(j)}^{-1} &= 0 \\
&\vdots \\
\hat{\Pi}_{\left(i_{r_j}\right)}'Z'\hat{U}\hat{\Sigma}_{(j)}^{-1} &= 0.
\end{aligned}$$

In a recycling of notation, create the following matrix which collects all the relevant first order conditions for equation j:

$$\hat{\Pi}_j'=\begin{bmatrix}\hat{\Pi}_{(i_1)}'\\ \vdots\\ \hat{\Pi}_{\left(i_{r_j}\right)}'\end{bmatrix}.$$

Then our first order conditoions become

$$\hat{\Pi}_j'Z'\hat{U}\hat{\Sigma}_{(j)}^{-1}=0.$$

We can proceed in a similar fashion for the first order condition with respect to Γ^u. Consider the coefficient matrix for the exogenous variables in equation j, $\Gamma_{(j)}=\begin{bmatrix}\Gamma_{1j}\\ \vdots\\ \Gamma_{kj}\end{bmatrix}$. Suppose for all $i\in S(j)=\left\{i_1,\ldots,i_{s(j)}\right\}$, Γ_{ij} is unrestricted and for all $i\notin S(j)$, Γ_{ij} is restricted. Then we have that

$$\begin{aligned}
Z_{(i_1)}'\hat{U}\hat{\Sigma}_{(j)}^{-1} &= 0 \\
&\vdots \\
Z_{\left(i_{s_j}\right)}'\hat{U}\hat{\Sigma}_{(j)}^{-1} &= 0.
\end{aligned}$$

Define the matrix Z_j by

$$Z_j'=\begin{bmatrix}Z_{(i_1)}'\\ \vdots\\ Z_{\left(i_{s_j}\right)}'\end{bmatrix}.$$

Then our first order conditions become

$$Z_j'\hat{U}\hat{\Sigma}_{(j)}^{-1}=0.$$

It will be useful to define the matrix C_j by

$$Z_j=ZC_j.$$

With this definition, the first order conditions become

$$\begin{aligned} \hat{\Pi}_j' Z' \hat{U} \hat{\Sigma}_{(j)}^{-1} &= 0 \\ C_j' Z' \hat{U} \hat{\Sigma}_{(j)}^{-1} &= 0. \end{aligned}$$

Define the matrix D_j by

$$D_j' = \begin{bmatrix} \hat{\Pi}_j' \\ C_j' \end{bmatrix}$$

so we can write our first order conditions as

$$D_j' Z' \hat{U} \hat{\Sigma}_{(j)}^{-1} = \begin{bmatrix} \hat{\Pi}_j' Z' \hat{U} \Sigma_{(j)}^{-1} \\ C_j' Z' \hat{U} \Sigma_{(j)}^{-1} \end{bmatrix} = 0$$

Now we can write the first order conditions for all our equations as

$$\begin{bmatrix} D_1' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_m' \end{bmatrix} \begin{bmatrix} Z' \hat{U} \hat{\Sigma}_{(1)}^{-1} \\ \vdots \\ Z' \hat{U} \hat{\Sigma}_{(m)}^{-1} \end{bmatrix} = 0$$

Define $\bar{D} \equiv \begin{bmatrix} D_1' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_m' \end{bmatrix}$ and note that we can write

$$\begin{bmatrix} Z' \hat{U} \hat{\Sigma}_{(1)}^{-1} \\ \vdots \\ Z' \hat{U} \hat{\Sigma}_{(m)}^{-1} \end{bmatrix} = \begin{bmatrix} Z' \hat{U} \hat{\Sigma}^{-1} e_1' \\ \vdots \\ Z' \hat{U} \hat{\Sigma}^{-1} e_m' \end{bmatrix} = vec\left(Z' \hat{U} \hat{\Sigma}^{-1} \right).$$

By proposition 15,

$$\begin{aligned} vec\left(Z' \hat{U} \hat{\Sigma}^{-1} \right) &= \left(\left(\hat{\Sigma}^{-1} \right)' \otimes Z' \right) vec\left(\hat{U} \right) \\ &= \left(\hat{\Sigma}^{-1} \otimes Z' \right) vec\left(\hat{U} \right). \end{aligned}$$

Our first order conditions now become

$$\bar{D} \left(\hat{\Sigma}^{-1} \otimes Z' \right) vec\left(\hat{U} \right) = 0.$$

However, we can do even more. Recall that we can write our system of equations in another way (as per the SUR model):

$$\begin{aligned} \underset{T \times 1}{Y_{(1)}} &= Y^1 \beta^1 + Z^1 \gamma^1 + U_{(1)} \\ &\vdots \\ Y_{(M)} &= Y^M \beta^M + Z^M \gamma^M + U_{(M)}, \end{aligned}$$

where Y^j denotes the right-hand side included endogenous variables for equation j and Z^j denotes the included exogenous variables for equation j. Let $X^1 = \begin{bmatrix} Y^1 & Z^1 \end{bmatrix}$. Then,

$$\begin{aligned} Y_{(1)} &= X^1 \delta^1 + U_{(1)} \\ &\vdots \\ Y_{(M)} &= X^M \delta^M + U_{(M)} \end{aligned}$$

or

$$\begin{bmatrix} Y_{(1)} \\ \vdots \\ Y_{(M)} \end{bmatrix} = \begin{bmatrix} X^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X^M \end{bmatrix} \begin{bmatrix} \delta^1 \\ \vdots \\ \delta^M \end{bmatrix} + \begin{bmatrix} U_{(1)} \\ \vdots \\ U_{(M)} \end{bmatrix}$$
$$vec(Y) = \bar{X}\bar{\delta} + vec(U)$$
$$\bar{Y} = \bar{X}\bar{\delta} + vec(U)$$

Thus, $vec\left(\hat{U}\right) = \bar{Y} - \bar{X}\hat{\delta}_{FIML}$, and our first order conditions become

$$\bar{D}\left(\hat{\Sigma}^{-1} \otimes Z'\right)\left(\bar{Y} - \bar{X}\hat{\delta}_{FIML}\right) = 0.$$

Rearranging to solve for $\hat{\delta}_{FIML}$, we have

$$\bar{D}\left(\hat{\Sigma}^{-1} \otimes Z'\right)\bar{X}\hat{\delta}_{FIML} = \bar{D}\left(\hat{\Sigma}^{-1} \otimes Z'\right)\bar{Y}$$

or

$$\hat{\delta}_{FIML} = \left(\bar{D}\left(\hat{\Sigma}^{-1} \otimes Z'\right)\bar{X}\right)^{-1}\bar{D}\left(\hat{\Sigma}^{-1} \otimes Z'\right)\bar{Y}.$$

3 Limited Information Maximum Likelihood

As a special case of FIML, we have LIML. Let us look closely at a single equation. Without loss of generality, let us examine equation 1 :

$$\underset{T\times 1}{Y_{(1)}} = \underset{T\times r_1}{Y^1}\underset{r_1\times 1}{\beta^1} + \underset{T\times s_1}{Z^1}\underset{s_1\times 1}{\gamma^1} + \underset{T\times 1}{U_{(1)}}.$$

Further, consider the first stage regression for Y^1:

$$\underset{T\times r_1}{Y^1} = \underset{T\times K}{Z}\underset{K\times r_1}{\Pi_1} + \underset{T\times r_1}{V_1}.$$

Our system of equations here becomes

$$Y_{(1)} = Y^1\beta^1 + Z^1\gamma^1 + U_{(1)} = X^1\delta^1 + U_{(1)} \tag{1}$$
$$Y^1 = Z\Pi_2 + V_{(1)}. \tag{2}$$

LIML on equation 1 is simply FIML on this system of equations. Let

$$\Psi = Var\begin{pmatrix} U_{t1} \\ V_{t1} \end{pmatrix} = \begin{bmatrix} \sigma_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}.$$

It will be useful to note the dimensions of the following objects: $\underset{1\times 1}{\sigma_{11}}, \underset{1xr_1}{\psi_{12}}, \underset{r_1\times 1}{\psi_{21}}, \underset{r_1\times r_1}{\psi_{22}}$.

Assuming homoskedasticity within equations, we have

$$Var\begin{pmatrix} U_{(1)} \\ V_{(1)} \end{pmatrix} = \Psi \otimes I_T.$$

Define $\Psi^{-1} = \begin{bmatrix} \psi^{11} & \psi^{12} \\ \psi^{21} & \psi^{22} \end{bmatrix}$ by $\Psi\Psi^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & I_{r_1} \end{bmatrix}$. Recall that our first order conditions for FIML are given by

$$\left(\beta^1\right) : \hat{\Pi}_1' Z'\hat{U}\hat{\Sigma}_{(1)}^{-1} = 0$$
$$\left(\gamma_1\right) : Z_1'\hat{U}\hat{\Sigma}_{(1)}^{-1} = 0$$
$$\left(\Pi_2\right) : Z'\hat{U}\hat{\Sigma}_{(2)}^{-1} = 0$$

for $j \in \{1,2\}$. For (1), we have

$$\begin{aligned} 0 &= \hat{\Pi}_1' Z' \hat{U} \hat{\Sigma}_{(1)}^{-1} = \hat{\Pi}_1' Z' \left[\begin{array}{cc} \hat{U}_{(1)} & \hat{V}_{(1)} \end{array} \right] \left[\begin{array}{c} \hat{\psi}^{11} \\ \hat{\psi}^{21} \end{array} \right] \\ &= \hat{\Pi}_1' Z' \left[\hat{U}_{(1)} \hat{\psi}^{11} + \hat{V}_{(1)} \hat{\psi}^{21} \right] \end{aligned}$$

and

$$\begin{aligned} 0 &= Z_1' \left[\begin{array}{cc} \hat{U}_{(1)} & \hat{V}_{(1)} \end{array} \right] \left[\begin{array}{c} \hat{\psi}^{11} \\ \hat{\psi}^{21} \end{array} \right] \\ &= Z_1' \left[\hat{U}_{(1)} \hat{\psi}^{11} + \hat{V}_{(1)} \hat{\psi}^{21} \right]. \end{aligned}$$

For (2), we have

$$\begin{aligned} 0 &= Z' \left[\begin{array}{cc} \hat{U}_{(1)} & \hat{V}_{(1)} \end{array} \right] \left[\begin{array}{c} \hat{\psi}^{12} \\ \hat{\psi}^{22} \end{array} \right] \\ &= Z' \left(\hat{U}_{(1)} \hat{\psi}^{12} + \hat{V}_{(1)} \hat{\psi}^{22} \right). \end{aligned}$$

Thus, the first order conditions have become

$$\hat{\Pi}_1' Z' \left[\hat{U}_{(1)} \hat{\psi}^{11} + \hat{V}_{(1)} \hat{\psi}^{21} \right] = 0 \tag{3}$$

$$Z_1' \left[\hat{U}_{(1)} \hat{\psi}^{11} + \hat{V}_{(1)} \hat{\psi}^{21} \right] = 0 \tag{4}$$

$$Z' \left(\hat{U}_{(1)} \hat{\psi}^{12} + \hat{V}_{(1)} \hat{\psi}^{22} \right) = 0. \tag{5}$$

We can rearrange (5) as follows:

$$\begin{aligned} Z' \hat{V}_{(1)} \hat{\psi}^{22} &= -Z' \hat{U}_{(1)} \hat{\psi}^{12} \\ Z' \hat{V}_{(1)} &= -Z' \hat{U}_{(1)} \hat{\psi}^{12} \left(\hat{\psi}^{22} \right)^{-1}. \end{aligned} \tag{6}$$

Plugging (6) into (3) gives us

$$\begin{aligned} \hat{\Pi}_1' Z' \left[\hat{U}_{(1)} \hat{\psi}^{11} - \hat{U}_{(1)} \hat{\psi}^{12} \left(\hat{\psi}^{22} \right)^{-1} \hat{\psi}^{21} \right] &= 0 \\ \hat{\Pi}_1' Z' \hat{U}_{(1)} \left[\underset{1\times 1}{\hat{\psi}^{11}} - \underset{1\times r_1}{\hat{\psi}^{12}} \underset{r_1\times r_1}{\left(\hat{\psi}^{22} \right)^{-1}} \underset{r_1\times 1}{\hat{\psi}^{21}} \right] &= 0 \end{aligned}$$

or

$$\hat{\Pi}_1' Z' \hat{U}_{(1)} = 0.$$

Rewriting (4), using the fact that C_1 can be defined such that $Z_1 = ZC_1$, and then plugging in (6),

$$\begin{aligned} 0 &= C_1' Z' \left[\hat{U}_{(1)} \hat{\psi}^{11} + \hat{V}_{(1)} \hat{\psi}^{21} \right] \\ &= C_1' Z' \hat{U}_{(1)} \left[\hat{\psi}^{11} - \hat{\psi}^{12} \left(\hat{\psi}^{22} \right)^{-1} \hat{\psi}^{21} \right], \end{aligned}$$

or

$$C_1' Z' \hat{U}_{(1)} = 0.$$

Thus, our first order conditions become

$$\begin{aligned}\hat{\Pi}_1' Z' \hat{U}_{(1)} &= 0 \\ C_1' Z' \hat{U}_{(1)} &= 0.\end{aligned}$$

Define $D_1' = \begin{bmatrix} \hat{\Pi}_1' \\ C_1' \end{bmatrix}$. Then we have

$$D_1' Z' \hat{U}_{(1)} = 0.$$

If we plug in $\hat{U}_{(1)} = Y_{(1)} - X^1 \hat{\delta}_{LIML}$, we get

$$D_1' Z' \left(Y_{(1)} - X^1 \hat{\delta}_{LIML} \right) = 0,$$

which we can rearrange

$$\begin{aligned}D_1' Z' X^1 \hat{\delta}_{LIML} &= D_1' Z' Y_{(1)} \\ \hat{\delta}_{LIML} &= \left(D_1' Z' X^1 \right)^{-1} D_1' Z' Y_{(1)}.\end{aligned}$$