Derivation of LIML*

Dong Woo Hahm†

Contents

1 Limited Information Maximum Likelihood

1.1 Setup

- Consider a linear model with endogeneity:

$$y_i = x_i'\beta + e_i$$

where $x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} \begin{matrix} k_1 \\ k_2 \end{matrix} \in \mathbb{R}^k$ where $k = k_1 + k_2$ and $E(x_{i1}e_i) = 0$, $E(x_{i2}e_i) \neq 0$.

Assume we have a vector of instruments $z_i = \begin{pmatrix} x_{i1} \\ z_{i2} \end{pmatrix} \begin{matrix} k_1 \\ l_2 \end{matrix} \in R^l$ where $l = k_1 + l_2$ that satisfy conditions for valid instruments.

- Limited Information Maximum Likelihood (LIML) is the maximum likelihood method for a structural equation for y_i combined with an unrestricted[1] reduced form equation for x_{2i}.

- Recall the structural equation and reduced form equation for x_{2i} are given by

$$y_i = x_{i1}'\beta_1 + x_{i2}'\beta_2 + e_i \tag{1}$$
$$x_{i2} = \Gamma_{12}' x_{i1} + \Gamma_{22}' z_{i2} + u_{i2} \tag{2}$$

- Let $Y_i = \begin{pmatrix} y_i \\ x_{i2} \end{pmatrix}$ and $\eta_i = \begin{pmatrix} e_i \\ u_{i2} \end{pmatrix}$ and stack the equations.

$$\underbrace{\begin{pmatrix} 1 & -\beta_2' \\ 0 & I \end{pmatrix}}_{A} \begin{pmatrix} y_i \\ x_{i2} \end{pmatrix} = \underbrace{\begin{pmatrix} \beta_1' & 0 \\ \Gamma_{12}' & \Gamma_{22}' \end{pmatrix}}_{B} \begin{pmatrix} x_{i1} \\ z_{i2} \end{pmatrix} + \eta_i$$
$$\iff AY_i = Bz_i + \eta_i$$

- LIML imposes an assumption that $\eta_i = \begin{pmatrix} e_i \\ u_{i2} \end{pmatrix}$ follows a multivariate normal conditional on z_i.

$$\eta_i | z_i \sim N(0, \Sigma_\eta)$$

*Columbia University Economics Ph.D. First Year 2018-2019. Professor Jushan Bai and Simon Lee.

†Department of Economics, Columbia University. dongwoo.hahm@columbia.edu. Please email me for any error.

[1] Unrestricted in the sense that it is not a cross-equation restrictions but instead, it can be constructed mechanically. If we also make use of the structural equation for x_{2i} (remember, it is the endogenous regressor that is determined *within* the system.) instead of the reduced form, it is called the Full Information Maximum Likelihood (FIML).

1.2 Derivation of LIML

Before derivation, I present some useful stuffs.

Lemma 1. The density function of a multivariate normal distribution $X \sim N(\mu, \Sigma)$ is

$$pdf_X(x) = (det(2\pi\Sigma))^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big)$$

Lemma 2. (Jacobian Transformation)
Let $X \in \mathbb{R}^k$ be a continuous random vector and let $g : \mathbb{R}^k \to \mathbb{R}^k$ be a one-to-one and onto function denoted by $g(x) = (g_1(x), \cdots, g_k(x))'$. Then for $Y = g(X)$, the density of Y is given by

$$f_Y(y) = f_X(h(y))|det(J_h(y))|$$

where h is the inverse function of g and $J_h(y)$ is the Jacobian of $h = g^{-1}$. We can also write $J_h(y) = [J_g(h(y))]^{-1}$ where $J_g(x)$ is the Jacobian of g.

Lemma 3. (Some Useful Matrix Differentiation Rules)

1. $\frac{d}{dA} \log det(A) = (A')^{-1}$
2. $\frac{d}{dA} tr(A^{-1}B) = -A^{-1}BA^{-1}$

Lemma 4. The solution to the problem

$$\begin{aligned} &\max_A \log(det(A)) + \sum_{i=1}^n b_i' A^{-1} b_i \\ &s.t.\ A = A' \end{aligned}$$

is given by $\sum_{i=1}^n b_i b_i'$.

Proof. The objective function can be rewritten as

$$\log(det(A)) + \sum_{i=1}^n b_i' A^{-1} b_i = \log(det(A)) + tr(\sum_{i=1}^n b_i' A^{-1} b_i) = \log(det(A)) + tr(A^{-1}\sum_{i=1}^n b_i b_i')$$

By the Lemma above, the first order condition becomes

$$\begin{aligned} &A'^{-1} - A^{-1}(\sum_{i=1}^n b_i b_i')A^{-1} = 0 \\ \Longrightarrow\ &A^{-1} = A^{-1}(\sum_{i=1}^n b_i b_i')A^{-1} \\ \Longrightarrow\ &A = \sum_{i=1}^n b_i b_i' \end{aligned}$$

Note that we implicitly require that $\sum_{i=1}^{n} b_i b_i'$ is of full-rank. □

Lemma 5. Let E be a matrix that can be partitioned into

$$E = \begin{pmatrix} E_{11} & E_{12} \\ E_{12}' & E_{22} \end{pmatrix}$$

Then the determinant of E is given by

$$\begin{aligned} |E| =& |E_{11}| \cdot |E_{22} - E_{12}' E_{11}^{-1} E_{12}| \\ =& |E_{22}| \cdot |E_{11} - E_{12} E_{22}^{-1} E_{12}'| \end{aligned}$$

Now we start deriving the LIML via MLE. Everything from now on will be "conditional on z_i" so I omit the conditional part for notational simplicity.

- First, let's derive the log-likelihood function.
  Since $\eta_i \sim N(0, \Sigma_\eta)$, the log-likelihood function for $\{\eta_i\}_{i=1}^{n}$ is

  $$-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\Sigma_\eta)) - \frac{1}{2}\sum_{i=1}^{n} \eta_i' \Sigma_\eta^{-1} \eta_i$$

  Next, since $Y_i = A^{-1}Bz_i + A^{-1}\eta_i$, using the Jacobian Transformation, the log-likelihood for $\{Y_i\}_{i=1}^{n}$ is given by

  $$-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\Sigma_\eta)) - \frac{1}{2}\sum_{i=1}^{n} \eta_i' \Sigma_\eta^{-1} \eta_i + n\log(det(A))$$

  Note that since $A = \begin{pmatrix} 1 & -\beta_2' \\ 0 & I \end{pmatrix}$, $det(A) = 1$ and the last term is equal to zero. Hence, our log-likelihood function to be maximized is

  $$l(A, B, \Sigma_\eta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\Sigma_\eta)) - \frac{1}{2}\sum_{i=1}^{n} \eta_i' \Sigma_\eta^{-1} \eta_i$$

- Now let's maximize $l(A, B, \Sigma_\eta)$. The way we proceed is to concentrate Σ_η out as a function of A, B, and then to maximize with respect to A and B only.

- First treating A,B as fixed, the maximizer of $l(A, B, \Sigma_\eta)$ w.r.t. Σ_η is given by $\tilde{\Sigma}_\eta = \frac{1}{n}\sum_{i=1}^{n} \eta_i \eta_i'$ by the Lemma above.

- Next, plug $\tilde{\Sigma}_\eta$ into $l(A, B, \Sigma_\eta)$ then the log-likelihood reduces to

$$\begin{aligned}
l(A,B) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\tilde{\Sigma}_\eta)) - \frac{1}{2}\sum_{i=1}^{n}\eta_i'\tilde{\Sigma_\eta}^{-1}\eta_i \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\tilde{\Sigma}_\eta)) - \frac{1}{2}\sum_{i=1}^{n}tr(\eta_i'\tilde{\Sigma_\eta}^{-1}\eta_i) \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\tilde{\Sigma}_\eta)) - \frac{1}{2}\sum_{i=1}^{n}tr(\tilde{\Sigma_\eta}^{-1}\eta_i\eta_i') \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\tilde{\Sigma}_\eta)) - \frac{1}{2}tr(\tilde{\Sigma_\eta}^{-1}\sum_{i=1}^{n}\eta_i\eta_i') \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(det(\tilde{\Sigma}_\eta)) - \frac{n}{2}
\end{aligned}$$

  So the maximization problem reduces to

$$\max_{A,B} -\frac{n}{2}\log(det(\tilde{\Sigma}_\eta))$$

- Partition the matrices A and B such that

$$\begin{aligned}
A &= \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} (& 1 & -\beta_2' &) \\ (& 0 & I &) \end{pmatrix} \\
B &= \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} (& \beta_1' & 0 &) \\ (& \Gamma_{12}' & \Gamma_{22}' &) \end{pmatrix}
\end{aligned}$$

  for notational simplicity. Note that A_2 is not an unknown. Using these notations we can write

$$\begin{aligned}
e' &= A_1 Y' - B_1 Z' \\
u_2' &= A_2 Y' - B_2 Z'
\end{aligned}$$

- Note that

$$\begin{aligned}
det(\tilde{\Sigma_\eta}) &= det(\frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_i') \\
&= det(\frac{1}{n}\sum_{i=1}^{n}(AY_i - Bz_i)(AY_i - Bz_i)') \\
&= det(\frac{1}{n}(YA' - ZB')(YA' - ZB')')
\end{aligned}$$

  and that

$$(YA' - ZB')(YA' - ZB')' = \begin{bmatrix} e'e & e'u_2 \\ u_2'e & u_2'u_2 \end{bmatrix}$$

  where $Y = \begin{pmatrix} - & Y_1' & - \\ & \vdots & \\ - & Y_n' & - \end{pmatrix}$, $Z = \begin{pmatrix} - & z_1' & - \\ & \vdots & \\ - & z_n' & - \end{pmatrix}$, $e = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix}'$, $u_2 = \begin{pmatrix} - & u_{12}' & - \\ & \vdots & \\ - & u_{n2}' & - \end{pmatrix}$.

- So the maximization problem is

$$\max_{A_1,B_1,B_2} -\log(e'e) - \log(det(u_2'(I_n - e(e'e)^{-1}e')u_2)) \tag{1}$$

  using Lemma 5.

- We again use the concentration method. First, let's see the FOC for B_2. By chain rule and Lemma above for derivatives of matrices,

$$B_2 : \quad 2(u_2' M_e u_2)^{-1} u_2' M_e Z = 0$$

where $M_e = I_n - e(e'e)^{-1}e'$. Using that $YA_2' = X_2$, rearranging terms yield

$$\hat{B}_2 = X_2' M_e Z (Z' M_e Z)^{-1}$$

and hence

$$\begin{aligned} \hat{u_2}' &= A_2 Y' - \hat{B}_2 Z' \\ &= X_2' - X_2' M_e Z (Z' M_e Z)^{-1} Z' \end{aligned}$$

- Concentrate the objective function in (1) by replacing u_2 by $\hat{u}_2$.

$$\begin{aligned} &- \log(e'e) - \log(det((X_2' - X_2' M_e Z (Z' M_e Z)^{-1} Z') M_e (X_2' - X_2' M_e Z (Z' M_e Z)^{-1} Z')') \\ =& - \log(e'e) - \log(det(X_2'(I_n - M_e Z (Z' M_e Z)^{-1} Z') M_e (I_n - Z (Z' M_e Z)^{-1} Z' M_e) X_2) \\ =& - \log(e'e) - \log(det(X_2' M_e (I - M_e Z (Z' M_e Z)^{-1} Z' M_e) M_e X_2) \end{aligned}$$

- Note that the second term involves the determinant of the moment matrix of the residuals from the regression of $M_e X_2$ on $M_e Z$, or equivalently of X_2 on e and Z, or again equivalently of $M_Z X_2$ on $M_Z e$ where $M_Z = I_n - Z(Z'Z)^{-1}Z'$ by Frisch-Waugh-Lovell. Hence it once more reduces to

$$- \log(e'e) - \log(det(X_2' M_Z X_2 - X_2' M_Z e (e' M_Z e)^{-1} e' M_Z X_2))$$

- From Lemma 5 for $E = \begin{pmatrix} E_{11} & E_{12} \\ E_{12}' & E_{22} \end{pmatrix}$,

$$|E_{11}| \cdot |E_{22} - E_{12}' E_{11}^{-1} E_{12}| = |E_{22}| \cdot |E_{11} - E_{12} E_{22}^{-1} E_{12}'|$$

so that

$$\begin{aligned} & - \log|E_{11}| - \log|E_{22} - E_{12}' E_{11}^{-1} E_{12}| = - \log|E_{22}| - \log|E_{11} - E_{12} E_{22}^{-1} E_{12}'| \\ \implies & - \log|E_{22} - E_{12}' E_{11}^{-1} E_{12}| = - \log|E_{22}| - \log|E_{11} - E_{12} E_{22}^{-1} E_{12}'| + \log|E_{11}| \end{aligned}$$

- Applying to the second term, we have the objective function now as,

$$- \log(e'e) - \log(det(X_2' M_Z X_2)) - \log(det(e' M_Z e - e' M_Z X_2 (X_2' M_Z X_2)^{-1} X_2' M_Z e)) + \log(det(e' M_Z e))$$

- Recall that $e' = A_1 Y' - B_1 Z'$ so

$$e' M_Z = A_1 Y' M_Z$$

and substituting the expressions for $e' M_Z$ and e' back,

$$\begin{aligned} \max_{A_1, B_1} & - \log((A_1 Y' - B_1 Z')(A_1 Y' - B_1 Z')') - \log(det(X_2' M_Z X_2)) \\ & - \log(det(A_1 Y' M_Z Y A_1' - A_1 Y' M_Z X_2 (X_2' M_Z X_2)^{-1} X_2' M_Z Y A_1')) + \log(det(A_1 Y' M_Z Y A_1')) \end{aligned}$$

- Note that in the first term

$$A_1 Y' - B_1 Z' = A_1 Y' - \beta_1' X_1'$$

and the second term doesn't involve A_1, B_1 and the third term is

$$\begin{aligned}
&-\log(det(A_1 \begin{bmatrix} y' \\ X_2' \end{bmatrix} M_Z M_{M_Z X_2} M_Z \begin{bmatrix} y & X_2 \end{bmatrix} A_1'))\\
=&-\log(det(A_1 \begin{bmatrix} y' M_Z \\ X_2' M_Z \end{bmatrix} M_{M_Z X_2} \begin{bmatrix} M_Z y & M_Z X_2 \end{bmatrix} A_1'))\\
=&-\log(det(A_1 \begin{bmatrix} y' M_Z \\ 0 \end{bmatrix} \begin{bmatrix} M_Z y & 0 \end{bmatrix} A_1'))\\
=&-\log(det(y' M_Z M_{M_Z X_2} M_Z y))
\end{aligned}$$

where $M_{M_Z X_2} = I_n - M_Z X_2 (X_2' M_Z X_2)^{-1} X_2' M_Z$ and also doesn't involve A_1, B_1.

- Hence, the problem is now,

$$\max_{A_1, \beta_1} -\log((A_1 Y' - \beta_1' X_1')(A_1 Y' - \beta_1' X_1')') + \log(det(A_1 Y' M_Z Y A_1'))$$

- FOC w.r.t β_1:

$$\begin{aligned}
&\frac{1}{(A_1 Y' - \beta_1' X_1')(A_1 Y' - \beta_1' X_1')'}(-2X_1' Y A_1' + 2X_1' X_1 \hat{\beta}_1) = 0\\
\implies & X_1' Y A_1' = X_1' X_1 \hat{\beta}_1\\
\implies & \hat{\beta}_1 = (X_1' X_1)^{-1} X_1' Y A_1' = (X_1' X_1)^{-1} X_1' (y - X_2 \beta_2)
\end{aligned}$$

- Concentrate β_1 out:

$$\max_{A_1} -\log(det(A_1 Y' M_{X_1} Y A_1')) + \log(det(A_1 Y' M_Z Y A_1'))$$

Hence the maximizer $\hat{A}_1$ is the solution to[2]

$$\min_{A_1} \frac{A_1 Y' M_{X_1} Y A_1'}{A_1 Y' M_Z Y A_1'}$$

which is equivalent to the smallest generalized eigenvalue of $Y' M_{X_1} Y$ with respect to $Y' M_Z Y$.

- The solution $\hat{A}_1 = [1 \quad -\hat{\beta}_2']$ together with $\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' Y A_1' = (X_1' X_1)^{-1} X_1' (y - X_2 \beta_2)$, we can derive that

$$\hat{\beta}_{LIML} = (X'(I_n - \hat{\kappa} M_Z) X)^{-1} X' (I_n - \hat{\kappa} M_Z) y$$

where $\hat{\kappa} = \hat{A}_1$. See Hansen's book page 414 for the rest of the calculation.

[2] Note that both $A_1 Y' M_{X_1} Y A_1'$ and $A_1 Y' M_Z Y A_1'$ are scalars and the determinants are simply themselves.