14.383

Fall 2003

REVIEW

T.A.: Alexis León
aleon@mit.edu
E52-351.

14.383 : Revision for generals

Outline

→ Note if $Var(z) \to \infty$, then OLS (with endogenous vars in RHS) is consistent

1. Identification :

   - total $M(M-1)$ further restrictions after normalization
   - Rank condition (for single equation & linear restricts

     $\text{Rank}[\Phi A] = M$

     where $A = \begin{bmatrix} B \\ \Pi \end{bmatrix}$, Φ = matrix of restrictions on 1st equation

     $\Phi A_1 = \phi$

     Φ includes the normalization restriction

   - Order condition

     no. of restrictions $g \geq M$.

   - Interpretation → there must be enough "instruments" from other eqns, and these restrictions must be, in a sense, independent ie no other eqn should satisfy the same set of restrictions.

2. SF : $\underset{T\times M}{Y}\underset{M\times M}{B} + \underset{T\times K}{Z}\underset{K\times M}{\Gamma} = \underset{T\times M}{U}$; $y_t B + z_t \Gamma = u_t$

Assume : B is non singular

$\text{Rank}(Z) = k$ (full column rank)

$E(u_t' u_{t'}) = 0$ for $t \neq t'$

$E(\underset{M\times 1}{u_t'}\ \underset{1\times M}{u_t}) = \Sigma$

$$\text{plim } \frac{1}{T} Z'U = \text{plim } \frac{1}{T} \sum_{t=1}^{T} \underset{K\times 1}{z_t'}\ \underset{1\times M}{u_t} = 0$$

(Exogeneity of z's).

3. Predetermined variables :

y_i is predetermined in equation j if

$$\text{plim } \frac{1}{T} y_i' u_j = 0 \iff \text{plim } \frac{1}{T} v_i' u_j = 0$$

where V_i is the error term in the RF of equation i.

$$V = UB^{-1} \Rightarrow V_i = UB_i^{-1}$$

$$\Rightarrow \text{plim } \frac{1}{T} v_i' u_j = B_i^{-1\prime} \text{ plim } \frac{1}{T} U' u_j = B_j^{-1\prime} \Sigma_j$$

$\Rightarrow$ predeterminedness depends on the whole system; not just on eqns i and j.

4. Single equation IV :

We have k exclusion restrictions and q endogenous vars to instrument for. Choose a linear combination

$$\underset{T\times k}{W = Z} \underset{k\times q}{\hat{A}}$$ of the possible instruments. ($rk(\hat{A}) = q$)

$$\hat{\delta}_{IV} = (W'X)^{-1} W' y_1 = (\hat{A}' Z' X)^{-1} \hat{A}' Z' y_1$$

Under just id, there is no choice to make $\rightarrow \hat{\delta}_{IV}$ is unique.

$$\text{plim } \hat{\delta}_{IV} = \text{plim}(\hat{A}' Z' X)^{-1} \hat{A}' Z' X \delta + (\hat{A}' Z' X)^{-1} \hat{A}' Z' u$$

$$= \delta + \text{plim } \left(\frac{\hat{A}' Z' X}{T}\right)^{-1} \text{plim } \hat{A}' . \text{ plim } \frac{Z' u_1}{T}$$

$$= \delta$$

$$\sqrt{T}(\hat{\delta}_{IV} - \delta) = \sqrt{T} \left(\frac{\hat{A}' Z' X}{T}\right)^{-1} \hat{A}' . \frac{Z' u_1}{T}$$

$$\xrightarrow{d} N(0, V)$$

where $V = (A'G)^{-1} A' Q A (G'A)^{-1}$

$A = \text{plim } \hat{A}$, $G = \text{plim } \frac{Z'X}{T}$, $Q = \text{plim } \frac{Z' u_1}{T} \frac{u_1 Z}{T}$

$\hat{\delta}_{OIV} = \arg\min \ (y - X\delta)' Z \ J Z'(y - X\delta)$

Optimal J = Var

This is minimized iff $G\tilde{B} = QA$ for some $\tilde{B}$ non-singular

$\Rightarrow$ min. asymp. var. is $(G'Q^{-1}G)^{-1}$

or $A = Q^{-1}G\tilde{B}$

Wlg. take $\tilde{B} = I$, so we need $\boxed{A = Q^{-1}G}$

If we have conditional homescedasticity, then

$Q = \sigma^2 \text{plim} \dfrac{Z'Z}{T}$

$\Rightarrow \hat{A} = \left(\dfrac{Z'Z}{T}\right)^{-1} \dfrac{Z'X}{T} = (Z'Z)^{-1} Z'X$ would give us the optimal IV estimator.

This is nothing but 2SLS !

Generally we have $\hat{Q} = \dfrac{1}{T} \sum_{t=1}^{T} \hat{u}_{1t}^2 z_t' z_t$

$$\Rightarrow \boxed{\hat{\delta}_{OIV} = \left(X'Z\hat{Q}^{-1}Z'X\right)^{-1} X'Z\hat{Q}^{-1}Z'y_1}$$

Note that this $\hat{Q}$ is just White's estimator for the variance matrix. If we suspect autocorrelation as well, we could use the Newey-West estimator of the variance matrix. A consistent estimate of $\hat{u}_{1t}$'s can be obtained from 2SLS (~~$\Rightarrow$ this Optimal IV is exactly the same as the 3SLS discussed later~~ NO!).

5. Overid tests : to check orthogonality of instruments

ie. H_0: $\text{plim} \dfrac{Z'u_1}{T} = 0$ vs H_1: $\text{plim} \dfrac{Z'u_1}{T} \neq 0$.

Under just id $Z'\hat{u}_1 \equiv 0$, no test is possible.

Alternatively, there is only a unique IV estimate, so no comparison is possible.

USSIV

$\frac{K}{N}$ $\sigma^2_{\eta\varepsilon}$ π

$\varepsilon \to X_2$

$y_2 = X_2\beta + \varepsilon$
$X = Z\Pi + \eta$.

$\hat{X}_2 = Z_1\hat{\Pi}$
$\hat{\Pi} \to \eta_1$.

Overid test statistic is $\hat{u}_1' Z \hat{V}^{-1} Z' \hat{u}_1 \xrightarrow{d} \chi^2_{k-q}$

~~where $\hat{V} = \hat{\sigma}^2 Z'Z$ in the homoscedastic case~~

where $\hat{u}_1 = y_1 - X_1 \hat{\delta}_{OIV}$

In the homoscedastic case, $\hat{V} = \hat{\sigma}^2 (Z'Z)$

$\Rightarrow$ test statistic is $\dfrac{\hat{u}_1' P_Z \hat{u}_1}{\hat{\sigma}^2}$

where $\hat{\sigma}^2 = \dfrac{\hat{u}_1'\hat{u}_1}{T}$

This test can reject if (i) instruments are correlated with errors (ii) included exogenous vars are correlated with errors.

6. Hausman specification test: This again checks both the exogeneity of included exogenous vars + validity of instruments. ~~Under~~ (= specification). Under correct specification, plim $\hat{\delta}_{2SLS}$ = plim $\hat{\delta}_{OIV} = \delta$. and $\hat{\delta}_{OIV}$ is efficient. Under wrong specification, both $\hat{\delta}_{2SLS}$ and $\hat{\delta}_{OIV}$ are inconsistent and have different limits $\Rightarrow$

$$h = T(\hat{\delta}_{2SLS} - \hat{\delta}_{OIV})' \left[\hat{V}(\hat{\delta}_{2SLS}) - \hat{V}(\hat{\delta}_{OIV})\right] \times (\hat{\delta}_{2SLS} - \hat{\delta}_{OIV})$$

$$\xrightarrow{d} \chi^2_{(\dim(\delta))}$$

where $\sqrt{T}(\hat{\delta}_i - \delta) \longrightarrow N(0, V_i)$

Ques: What if we indeed have homoscedasticity? ~~Who~~ Would $\hat{\delta}_{OIV}$ & $\hat{\delta}_{2SLS}$ have different plims under Ho then? If not, then how to do this test?

7. SEEMINGLY UNRELATED REGRESSIONS (SUR).

Suppose we have T obs. on M endogenous vars. and we want to regress on purely exogenous vars. We allow contemporaneous correlation but not over time. (eg: Easterly & Levine on Africa → errors of each country are correlated, but not between countries).

~~situation~~

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mT} \end{pmatrix} = \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_m \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{pmatrix}$$

$$\Sigma_t = \Sigma$$

$$y_j = X_j \beta_j + \varepsilon_j \; ; \; j = 1, \ldots, m.$$

$$\Leftrightarrow \bar{y} = \bar{X} \bar{\beta} + \bar{\varepsilon}$$

$$E(\bar{\varepsilon} | \bar{X}) = 0 \;, \quad E(\varepsilon_j \varepsilon_k') = \sigma_{jk} I_T \;, \quad E(\varepsilon_j \varepsilon_j') = \sigma_{jj} I_T$$

$$\Rightarrow E(\bar{\varepsilon}, \bar{\varepsilon}') = \underset{M \times M}{\Sigma} \otimes \underset{T \times T}{I_T}$$ where Σ = matrix of contemporaneous correlation.

Optimum is to do GLS

$$\hat{\beta}_{GLS} = [\bar{X}' (\Sigma \otimes I_T)^{-1} \bar{X}]^{-1} \bar{X}' (\Sigma \otimes I_T)^{-1} \bar{y}$$

But we don't know Σ, so we do FGLS:

$$\hat{\beta}_{FGLS} = [\bar{X}' (\hat{\Sigma} \otimes I_T)^{-1} \bar{X}]^{-1} \bar{X}' (\hat{\Sigma} \otimes I_T)^{-1} \bar{y}$$

where plim $\hat{\Sigma} = \Sigma$.

Then $\text{plim}\,(\hat{\hat{\beta}}_{GLS} - \hat{\beta}_{FGLS}) = 0$ and they have the same asymptotic distribution. One way to make $\hat{\beta}_{FGLS}$ is to do OLS equation-by-equation, get the residuals and form $\hat{\Sigma}$ and then make $\hat{\beta}_{FGLS}$. If we iterate this procedure, then this converges to the MLE $(\hat{\beta}_{GLS})$ under normality.

Drawback: If even one equation is misspecified, then all the estimates become inconsistent! Also this is good only when all the RHS vars are exogenous ie for RF. ~~We can make~~

Interesting property (for RF): If $X_j = X \ \forall j$, then $\hat{\hat{\beta}}_{GLS} \equiv \hat{\beta}_{OLS}$.

How? $\bar{X} = I_M \otimes X$

$$\hat{\beta}_{GLS} = \left[(I \otimes X')(\Sigma \otimes I)^{-1} (I \otimes X) \right]^{-1} (I \otimes X')(\Sigma \otimes I)^{-1} \bar{y}$$

$$= (\Sigma^{-1} \otimes X'X)^{-1} (\Sigma^{-1} \otimes X') \bar{y}$$

$$= (\Sigma \otimes (X'X)^{-1}) (\Sigma^{-1} \otimes X') \bar{y}$$

$$= \left[I \otimes (X'X)^{-1} X' \right] \bar{y} = \hat{\beta}_{OLS}.$$

So, for RF estimates, doing OLS eqn-by-eqn is efficient even in the presence of Σ non-diagonal.

We can make a Hausman test to check for exogeneity/misspecification H_0: $\text{plim}\, \frac{1}{T} X_j' \varepsilon_k = 0 \quad \forall j, k$.

$\hat{\hat{\beta}}_{GLS}$ and $\hat{\beta}_{OLS}$ are both consistent under H_0 & $\hat{\beta}_{GLS}$ is efficient; under H_1 $\hat{\hat{\beta}}_{GLS}$ is inconsistent, $\hat{\beta}_{OLS}$ is consistent.

8. SYSTEM IV estimators : 3SLS.

Let us have the same setup as in SUR

$$y_j = X_j \delta_j + u_j \quad ; \; j=1,\dots,M$$

but now X_j can contain endogenous vars. as well.

q_j = no. of RHS vars. in jth equation:

$\underset{T\times K}{Z}$ = set of exogenous vars.

Assume that each eqn. is identified through exclusion restrictions ie $k \geq q_j \; \forall$ eqns. $\Rightarrow$ $MK \geq \sum_j q_j$ and that possible rank condition is satisfied. Our matrix of instrument is now

$$\begin{bmatrix} Z & & 0 \\ & Z & \\ 0 & & \ddots \\ & & & Z \end{bmatrix} = I_M \otimes Z$$

As before, we choose a linear combination

$$\underset{TM\times \bar{q}}{\bar{W}} = \underset{MT\times mk}{(I \otimes Z)} \; \underset{mk\times\bar{q}}{\bar{A}} \qquad ; \quad \bar{q} = \sum_j q_j$$

$$\hat{\delta}_{IV} = \left[\bar{A}'(I \otimes Z)'\bar{X}\right]^{-1} \bar{A}'(I\otimes Z)'\bar{y}.$$

As before, optimal $\bar{A}$ is given by (2SLS formula $\rightarrow$ we assume homoscedasticity ie. $\Sigma_i = \Sigma$

$$\text{plim } \bar{A} = (\text{Cov. matrix of errors})^{-1} (\text{regression of } X \text{ on } Z)$$

$$= (\Sigma \otimes I)^{-1} (I \otimes (Z'Z)^{-1}Z')\bar{X}$$

$$= (\Sigma^{-1} \otimes I)(I \otimes (Z'Z)^{-1}Z')\bar{X}$$

$$\hat{A} = (\hat{\Sigma}^{-1} \otimes I)(I \otimes (Z'Z)^{-1}Z')\bar{X}$$

So we get the system IV as follows:

Stage 1 & 2 : Do 2SLS equation-by-equation.

Find residuals $\hat{u}_j$. Compute $\hat{\sigma}_{jk} = \dfrac{\hat{u}_j' \hat{u}_k}{T}$

Stage 3 : Compute $\bar{W} = (I \otimes Z)\hat{A} = (\hat{\Sigma}^{-1} \otimes P_Z)\bar{X}$

$$\hat{\delta}_{3SLS} = \left[\bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)\bar{X}\right]^{-1} \bar{X}'(\hat{\Sigma}^{-1} \otimes P_Z)\bar{y}$$

Note : $\hat{\delta}_{3SLS} = \left[\hat{\bar{X}}'(\hat{\Sigma}^{-1} \otimes I)\hat{\bar{X}}\right]^{-1} \hat{\bar{X}}'(\hat{\Sigma}^{-1} \otimes I)\bar{y}$

where $\hat{\bar{X}} = P_Z \bar{X}$

$\Rightarrow$ $\hat{\delta}_{3SLS}$ is the SUR estimate of the model $y_j = \hat{X}_j \delta_j + v_j$ using $\hat{\Sigma}$ as the estimated variance matrix

9. Since $\hat{\delta}_{3SLS}$ is optimal system IV, it is in general better than doing 2SLS equation by equation. However, if all the equations are just id, then $3SLS \equiv 2SLS$ (same feature as $OIV \equiv 2SLS$). Similarly, if a subset of equations is exactly identified, then 3SLS on the <u>remaining</u> equations gives the same estimates as 2SLS. Intuition : 3SLS uses info from <u>other</u> equations to fine-tune our instruments (here single-equation $OIV = 2SLS$), but if a set of eqns. is just id, there is no information "left over" to exploit.

Implication : 3SLS is a purely <u>system</u> method ; it is no good for estimating a single eqn. if other eqns. are specified only in reduced form.

10. Maximum Likelihood Estimation : FIML.

$Y_t B + Z_t \Gamma = u_t$.

Assume $u_t' \sim N(0, \Sigma)$. Then log-likelihood for B, and Σ_1^{-1} in terms of $\underline{\underline{Y}}$ can be written as

$$\mathcal{L}(B, \Gamma, \Sigma^{-1}) = \text{const.} + T \underbrace{\log |\det B|}_{\text{Jacobian term}} + \frac{T}{2} \log |\det(\Sigma^{-1})|$$
$$- \frac{1}{2} \operatorname{tr}(\Sigma^{-1} \underbrace{U'U}_{YB+Z\Gamma})$$

FOC wrt unknown (unrestricted) elements are:

$$\frac{\partial \mathcal{L}}{\partial B^u} = [T B^{-1\prime} - Y' U \Sigma^{-1}]^u = 0 \quad (1)$$

$$\frac{\partial \mathcal{L}}{\partial \Gamma^u} = [-Z' U \Sigma^{-1}]^u = 0 \quad (2)$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1u}} = \left[\frac{T}{2} \Sigma - \frac{1}{2} U'U\right]^u = 0 \quad (3)$$

Usually Σ is unrestricted $\overset{(3)}{\Rightarrow} \hat{\Sigma}_{MLE} = \frac{U'U}{T}$. —(*)

FOC for B_{jk}^u :$^{(1)}$ $\Pi_{(j)}'(Z'U)\Sigma_{(k)}^{-1'} = 0$

FOC for Γ_{jk}^u :$^{(2)}$ $Z_{(j)}' U \Sigma_{(k)}^{-1'} = 0 \iff S_j'(Z'U)\Sigma_{(k)}^{-1} = 0$

where S_j is the selection matrix which picks out the jth column

$\Rightarrow D_k' Z'U . \Sigma_{(k)}^{-1} = 0$ for the kth equation

where $D_k = [\Pi_k \; S_k]$

$$\Rightarrow \begin{bmatrix} D_1' & & 0 \\ & \ddots & \\ 0 & & D_m' \end{bmatrix} \begin{bmatrix} Z'U\,\Sigma_{(1)}' \\ \vdots \\ Z'U\,\Sigma_{(m)}' \end{bmatrix} = 0$$

$$\Rightarrow \quad \bar{D}'(\Sigma^{-1} \otimes Z')\, \mathrm{Vec}(U) = 0$$

$$\Rightarrow \quad \bar{D}'(\Sigma^{-1} \otimes Z')(\bar{X}\delta_{MLE} - \bar{y}) = 0 \qquad \text{———} \quad (**)$$

MLE jointly solves (*) and (**). This is the FIML estimator:

$$\boxed{\bar{D}'(\hat{\Sigma}^{-1}_{FIML} \otimes Z')(\bar{X}\hat{\delta}_{FIML} - \bar{y}) = 0}$$

(11). Compare with 3SLS :

$$\bar{X}'(\hat{\Sigma}^{-1}_{3SLS} \otimes P_Z)(\bar{X}\hat{\delta}_{3SLS} - \bar{y}) = 0$$

$Z\hat{A}$

Note that both are of the form $\bar{W}'(\bar{X}\hat{\delta} - \bar{y}) = 0$ ie both ~~can~~ can be looked upon as system IV estimates using different sets of instruments. ~~(note that in~~
For an IV estimator $\bar{W} = Z\hat{A}$, the asymptotic properties depend only on plim $\hat{A}$.

For 3SLS, $\hat{A}'_{3SLS} = \bar{X}'(\hat{\Sigma}^{-1}_{3SLS} \otimes Z(Z'Z)^{-1})$

$$\Rightarrow \text{plim } \hat{A}_{3SLS} = \text{plim } (\hat{\Sigma}^{-1} \otimes I)(I \otimes (Z'Z)^{-1}Z')\,.\,\bar{X}$$

$$= (\Sigma \otimes I)\,[\Pi_k \cdot S_k] = (\Sigma \otimes I)\,\bar{D}$$

$$\text{plim } \hat{A}_{FIML} = \text{plim } (\hat{\Sigma}^{-1} \otimes I)\,\bar{D} = (\Sigma \otimes I)\,\bar{D}$$

$\Rightarrow$ 3SLS is asymptotically equivalent to FIML.

This means that 3SLS is asymptotically efficient under normality. Ques: Should we ever do FIML?
Note: Since FIML is an IV estimator, it is always consistent under correct specification.

12. LIML : MLE for single equation.
Do FIML on
$$y_1 = Y_1 \beta_1 + Z_1 \gamma_1 + U_1$$
$$Y_1 = Z\Pi_1 + V_1 \longrightarrow \text{rf for all other endogenous vars.}$$

FOC leads to $\boxed{D_1' \; Z' U = 0}$

$\Rightarrow$ LIML is also an IV estimator.
Since the remaining eqns are exactly identified,
$$\hat{\delta}_{1,2SLS} \equiv \hat{\delta}_{1,3SLS} \overset{\text{asymp.}}{=\!=} \hat{\delta}_{1,FIML} \overset{\text{defn}}{=\!=} \hat{\delta}_{1,LIML}$$

$\Rightarrow$ LIML is asymptotically equivalent to 2SLS.
But note that $\hat{\Pi}_{1,LIML} \neq (Z'Z)^{-1} Z' Y_1$ because 1st eqn is overid.
Let $\tilde{X} = X - u \cdot \frac{X'u}{u'u}$. Then LIML FOC is $\tilde{X}' P_Z (y - X\hat{\delta}) = 0$. FOC for 2SLS is $X' P_Z (y - X\hat{\delta}) = 0$. Since LIML has taken out some part of X correlated with u's, it has smaller bias. Tradeoff is only the computational diffic.

13. System overid test:
$$LR: \; 2\left[\log L(\hat{\Pi}^U, \hat{\Omega}^U) - \log L(\hat{\Pi}^R, \hat{\Omega}^R)\right] \xrightarrow{d} \chi^2_p$$
degree of ove

$\hat{\Pi}^U, \hat{\Omega}^U$ = normal reduced form estimates
$$\hat{\Pi}^R = -\hat{\Gamma}_{FIML} \hat{B}^{-1}_{FIML} \; ; \; \hat{\Omega}^R = \hat{B}'^{-1}_{FIML} \hat{\Sigma}_{FIML} \hat{B}^{-1}_{FIML}.$$