Part III

Simultaneous Equation Models

Yoosoon Chang
Department of Economics
Indiana University

and

Joon Y. Park
Department of Economics
Indiana University
and
Department of Economics
Sungkyunkwan University

January 2012

In this chapter, we consider specification, identification and estimation of simultaneous equation models (SEM). For references, see Malinvaud (1980), Sargan (1988), Judge et.al. (1985) and Handbook of Econometrics (1983).

1. Specification

1.1 The Model

We consider the model given by

$$y_t'B + x_t'C = u_t'$$

for $t = 1, \ldots, n$, where $\{y_t\}$ and $\{x_t\}$ are ℓ- and m-dimensional, respectively, which we call *endogenous* and *exogenous* variables. The motivation for the distinction between $\{y_t\}$ and $\{x_t\}$ is the same as in the models previously considered. The only difference here is that we allow for the contemporaneous relationships among the endogenous variables. As before, $\{u_t\}$ represent errors, which are assumed to be serially uncorrelated unless specified otherwise, i.e., $\mathcal{E}u_t u_s' = \Sigma$ if $t = s$ and 0 otherwise. We will often let

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} B \\ C \end{pmatrix}$$

so that $z_t'A = y_t'B + x_t'C$.

We may write the model in matrix form as

$$ZA = YB + XC = U$$

where Z, Y, X and U are defined in the usual fashion. Assume
(a) B is nonsingular.
(b) X is of full column rank.
The assumption (a) is necessary for the model to be complete, and (b) is made as in standard regression models. We assume that the distribution of Y is completely determined by its first and second moments, which would really be the case under normality. All these assumptions will be maintained throughout the chapter.

1.2 Identification

If we allow the parameters $A = (B', C')'$ and Σ to be any $(\ell + m) \times \ell$ and $\ell \times \ell$ matrices in the model presented above, then different sets of values of A's and Σ's may imply the same distribution for Y. In this case, they are said to be *observationally equivalent*, and the model is not identified. To avoid this problem of the lack of identification, we restrict the parameter set by

$$R\,\overline{\text{vec}}A = r \quad \text{and} \quad \Phi\,\overline{\text{vec}}\,\Sigma = \varphi$$

To make precise our subsequent exposition on identification, the following conventions will be made:
(a) A *structure* is a specific value of (A, Σ), or of A if it is unnecessary to specify Σ.

(b) A *model* is the set of structures satisfying restrictions of the form given above. It will often be denoted by (R, r) or $((R, r), (\Phi, \varphi))$, depending upon whether only A or both A and Σ are restricted.
(c) A *structure* in a model is *identified* if there is no other observationally equivalent structure in the model. A *model* is *identified* if every structure in the model is identified. For instance, a structure A_0 is said to be in model (R, r) if $R\,\overline{\text{vec}}A_0 = r$. Similarly, (A_0, Σ_0) is in model $((R, r), (\Phi, \varphi))$ if $R\,\overline{\text{vec}}A_0 = r$ and $\Phi\,\overline{\text{vec}}\Sigma_0 = \varphi$.

1.3 RF and SF Models

When the parameter set is restricted by the restrictions $B = I$, the model can be written as

$$Y = X\Pi + V$$

Using our convention, the model may be specified by (R, r) with

$$R = I_\ell \otimes (I_\ell, 0) \quad \text{and} \quad r = \text{vec}\, I$$

since $B = (I_\ell, 0)A$ and $\overline{\text{vec}}\,(I_\ell, 0)A = (I_\ell \otimes (I_\ell, 0))\,\overline{\text{vec}}\,A$. The model (R_0, r_0) is called *reduced form* (RF) and, in contrast, the models specified by all the other R's are called *structural form* (SF). The models in SF allow for contemporaneous relationships among the endogenous variables as mentioned above, while those in RF do not. The latter are just multivariate regression models that we studied earlier. A structure (A_0, Σ_0) with $A_0 = (B_0', C_0')'$ in any SF model has a unique observationally equivalent structure in the RF model since B_0 is assumed to be nonsingular. The structure (Π_0, Ω_0) (or, $((I, -\Pi_0')', \Omega_0)$, more conformably with our previous definition) in the RF model is given by

$$\Pi_0 = -C_0 B_0^{-1} \quad \text{and} \quad \Omega_0 = B_0^{-1\prime}\Sigma_0 B_0^{-1}$$

Finding a structure in the RF model that is observationally equivalent to a structure in a SF model amounts to solving the model for Y.

1.4 Representation

Let

$$\overline{\text{vec}}\,A = s - S\alpha$$

where α is the vector of "free" parameters in A, and S and s can be obtained from R and r by

$$Rs = r \quad \text{and} \quad RS = 0$$

since $R\,\overline{\text{vec}}\,A = r$. Using $(I_\ell \otimes Z)\,\overline{\text{vec}}\,A = \overline{\text{vec}}\,U$, we may now write the model as

$$y_* = Z_*\alpha + u$$

with

$$y_* = (I_\ell \otimes Z)s\,, \quad Z_* = (I_\ell \otimes Z)S$$

and $u = \overline{\text{vec}}\, U$.

For the representation of the i-th equation, we let the i-th column a_i of A be restricted by

$$R_i a_i = r_i$$

and define S_i and s_i by $R_i s_i = r_i$ and $R_i S_i = 0$, similarly as above, so that $a_i = s_i - S_i \alpha_i$ with the free parameter α_i. Then the i-th equation can be represented as

$$y_i = Z_i \alpha_i + u_i$$

where $y_i = Z s_i$, $Z_i = Z S_i$ and u_i is the i-th column of U.

The restrictions on A are often composed exclusively of those for normalization and exclusion, for which R becomes a matrix of zeros and ones. If this is the case,

$$Z_i = (Y_i, X_i)$$

where Y_i and X_i represent ℓ_i-endogenous and m_i-exogenous variables, respectively. Moreover, y_i is simply the endogenous variable with normalization restriction. We have in this case

$$y_* = \begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix}, \quad Z_* = \begin{pmatrix} Z_1 & & \\ & \ddots & \\ & & Z_\ell \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_\ell \end{pmatrix}$$

and the model is given in SUR form.

2. Identification

2.1 Characterization of Observational Equivalence

Since the distribution of Y is assumed to be completely determined by the first two moments, two structures are observationally equivalent if they yield the same mean and variance for Y. It can be deduced that

Lemma 1 *Two structures* (A_1, Σ_1) *and* (A_2, Σ_2) *are observationally equivalent if and only if*

$$A_2 = A_1 T \quad \text{and} \quad \Sigma_2 = T' \Sigma_1 T$$

for a nonsigular matrix T.

Proof Let $A_1 = (B_1', C_1')'$ and $A_2 = (B_2', C_2')'$. Notice that (A_1, Σ_1) and (A_2, Σ_2) are observationally equivalent if and only if $C_1 B_1^{-1} = C_2 B_2^{-1}$ and $B_1^{-1\prime} \Sigma_1 B_1^{-1} = B_2^{-1\prime} \Sigma_2 B_2^{-1}$. If this condition holds, then $A_2 = A_1 T$ and $\Sigma_2 = T' \Sigma_1 T$ with $T = B_1^{-1} B_2$. Conversely, if $A_2 = A_1 T$ and $\Sigma_2 = T' \Sigma_1 T$, then $C_2 B_2^{-1} = C_1 T T^{-1} B_1 = C_1 B_1$ and $B_2^{-1\prime} \Sigma_2 B_2^{-1} = B_1^{-1\prime} T^{-1\prime} \Sigma_2 T^{-1} B_1^{-1} = B_1^{-1\prime} \Sigma_1 B_1^{-1}$, as was to be shown. ■

All the observationally equivalent systems of equations may therefore be generated by taking independent linear combinations of the equations in a given system.

2.2 First-Order Identification

Let a model be specified by the restrictions only on A and consequently denoted by (R, r). Accordingly, structures are designated by the values on A only. Identification conditions can now be easily deduced from Lemma 1.

Theorem 2 (Rank Condition) *A necessary and sufficient condition for identification of A_0 in (R, r) is that*

$$\operatorname{rank} R(I_\ell \otimes A_0) = \ell^2$$

i.e., $R(I_\ell \otimes A_0)$ must have full column rank.

Proof From Lemma 1, A_0 is identified in (R, r) if and only if there is no A in (R, r) of the form $A_0 T$ with $T \neq I$, or equivalently,

$$R \,\overline{\operatorname{vec}}\, A_0 T = R(I_\ell \otimes A_0) \,\overline{\operatorname{vec}}\, T = r$$

has the unique solution $T = I$. The condition holds when and only when $R(I \otimes A_0)$ has full column rank. ■

The RF model is identified, since for $R = I_\ell \otimes (I_\ell, 0)$ and $A_0 = (I_\ell, -P_0')'$ we have $R(I_\ell \otimes A_0) = I_{\ell^2}$, and the rank condition is satisfied for all of its structures.

Let R be $q \times \ell(\ell + m)$, i.e., q be the number of restrictions. Since $R(I_\ell \otimes A_0)$ is $q \times \ell^2$, it is obvious that

Corollary 3 (Order Condition) *For identification of A_0 in (R, r), it is necessary that*

$$q \geq \ell^2$$

If the order condition is satisfied, then we may normally expect any given structure to be identified since the set of unidentified structures is a lower dimensional subset of the parameter set. For this reason, we often say (incorrectly, but commonly) that model (R, r) is identified when $q \geq \ell^2$. It is said to be *just identified* (or exactly identified) if $q = \ell^2$, and *over identified* if $q > \ell^2$. When $q < \ell^2$, it is called *unidentified* (or under identified).

2.3 Second-Order Identification

Now assume both A and Σ are restricted. It follows from Lemma 1 that (A_0, Σ_0) is identified in $((R, r), (\Phi, \varphi))$ if and only if there is no (A, Σ) in $((R, r), (\Phi, \varphi))$ of the form $A_0 T$ and $T' \Sigma_0 T$ with $T \neq I$. Or, equivalently, $T = I$ must be the only solution for

$$\begin{aligned} R(I_\ell \otimes A_0) \,\overline{\operatorname{vec}}\, T &= r \\ \Phi \,\overline{\operatorname{vec}}\, T' \Sigma_0 T &= \varphi \end{aligned}$$

The second equation is quadratic in T, which leads us to look at a condition for local identification instead of the global one. We define a structure (A_0, Σ_0) in $((R, r), (\Phi, \varphi))$ to be *locally* identified if there is a neighborhood of (A_0, Σ_0) where that no other structure is observationally equivalent to (A_0, Σ_0). Of course, a locally identified structure may not be identified (or, *globally* identified).

Define the restriction matrix Φ for Σ more precisely as follows: First we write the restrictions as $\Phi_* \text{vech}\Sigma = \varphi$, where "vech" vectorizes as "vec" nonredundant elements of Σ. Then let $\Phi = \Phi_*(D'D)^{-1}D'$ with the *duplication matrix* D such that $\text{vec}\,\Sigma = D\,\text{vech}\,\Sigma$. For Φ constructed as such, we have $\Phi\,\text{vec}\,(\cdot) = \Phi\,\overline{\text{vec}}\,(\cdot)$.

Theorem 4 *A necessary and sufficient condition for local identification of* (A_0, Σ_0) *in* $((R, r), (\Phi, \varphi))$ *is that*

$$\text{rank} \begin{pmatrix} R(I_\ell \otimes A_0) \\ \Phi(I_\ell \otimes \Sigma_0) \end{pmatrix} = \ell^2$$

i.e., the matrix must be of full column rank.

Proof Clearly, (A_0, Σ_0) is locally identified if and only if $T = I$ is the only solution to the tangent plane at $T = I$ of the equations introduced above. The first equation itself represents the the tangent plane at any T, since it is linear. To get the tangent plane at $T = I$ for the second equation, we totally differentiate the equation, set it equal to zero, and let $T = I$ and $dT = T - I$. Consequently, we have

$$\Phi(I_\ell \otimes \Sigma_0)\,\overline{\text{vec}}\,T = \varphi$$

and the result follows directly. ∎

The condition in the above theorem is only a necessary condition for (global) identification. Notice also that, for the condition to hold, it is necessary that there must be at least ℓ^2 restrictions on A and Σ together. We say that a model is just-, over- and un-identified based on the number of restrictions, as in the first-order identification.

2.4 Identification of Sub-structures

Let a model (R, r) be given, and let

$$A = (A_1, A_2)$$

where A_1 and A_2 represent the parameters of the 1st and 2nd subsystems, which are assumed to consist of ℓ_1 and ℓ_2 equations, respectively. We only consider restrictions imposed exclusively on A_1, which are written as

$$R_1\,\overline{\text{vec}}\,A_1 = r_1$$

For $A_0 = (A_1^0, A_2^0)$ in (R, r), it is said that the sub-structure A_1^0 is identified if there is no other observationally equivalent structure of A_0 with the first sub-structure satisfying the restriction. The sub-model (R_1, r_1) is said to be identified if the corresponding sub-structure of any structure in model (R, r) is identified.

Corollary 5 *In the model given above, it is necessary and sufficient that*

$$\operatorname{rank} R_1(I_{\ell_1} \otimes A_0) = \ell_1 \ell$$

for A_1^0 to be identified.

Proof A_1^0 is identified if and only if there is no A such that $A = A_0 T$ and $A_1 \neq A_1^0$. Let $T = (T_1, T_2)$, where the partition is made conformably with A, so that $A_1 = A_0 T_1$. Then the condition holds if and only if

$$R_1 \,\overline{\text{vec}}\, A_0 T_1 = R_1(I_{\ell_1} \otimes A_0)\,\overline{\text{vec}}\, T_1 = r_1$$

has the unique solution $T_1 = (I_{\ell_1}, 0)'$. ■

The implied order condition for identification of a sub-structure should be evident. A sub-model is said to be just-, over- and un-identified accordingly. Notice that our result here is directly applicable for a model with *identities*.

Suppose there are no restrictions across the two subsystems, and the restrictions on A_2 are given by

$$R_2 \,\overline{\text{vec}}\, A_2 = r_2$$

so that $R = \text{diag}(R_1, R_2)$ and $r = (r_1', r_2')'$. Then it follows that A_0 in (R, r) is identified if and only if both A_1^0 and A_2^0 are identified. This is because $R(I_\ell \otimes A_0) = \text{diag}(R_1(I_{\ell_1} \otimes A_0), R_2(I_{\ell_2} \otimes A_0))$ in this case.

If the result in Corollary 5 is applied to a single equation, the rank condition for identification of the i-th equation becomes

$$\operatorname{rank}(R_i A_0) = \ell$$

where R_i is the restriction matrix for the i-th equation, with the order condition

$$q_i \geq \ell$$

if R_i is $q_i \times \ell$. If we further concentrate on an equation with only normalization and exclusion restrictions, then the order condition becomes $m - m_i \geq \ell_i$ i.e., must be excluded at least as many exogenous variables as the number of endogenous variables in the right-hand side. Notice that the number of restrictions for the i-th equation is $1 + (\ell - (\ell_i + 1)) + (m - m_i)$, each term of which accounts for the restrictions for normalization and exclusions of endogenous and exogenous variables.

The problems of the second-order identification of sub-structures can be formulated and considered in essentially the same way that we investigate the first-order identification here.

2.5 Identification of Sub-structures with Zero-Covariance Restrictions

Consider the model $((R, r), (\Phi, \varphi))$, where (R, r) is composed of two separable restrictions on A_1 and A_2 as explained above, and (Φ, φ) implies restrictions on

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

given by

$$\Sigma_{12} = \Sigma_{21} = 0$$

In what follows, we use the same notation as in the previous section.

Corollary 6 *Let A_2^0 be identified in the model introduced above. It is necessary and sufficient that*

$$\operatorname{rank} R_1(I_{\ell_1} \otimes A_1^0) = \ell_1^2$$

for A_1^0 to be identified.

Proof Write an $\ell \times \ell$ matrix T as

$$T = (T_1, T_2) = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$$

where the partitions are made conformably with A and Σ, and consider

$$\begin{aligned} R_1(I_{\ell_1} \otimes A_0)\,\overline{\text{vec}}\,T_1 &= r_1 \\ R_2(I_{\ell_2} \otimes A_0)\,\overline{\text{vec}}\,T_2 &= r_2 \\ T_1'\Sigma_0 T_2 &= 0 \end{aligned}$$

which any admissible transformation T must satisfy. Since A_2^0 is identified, $T_{12} = 0$ and $T_{22} = I_{\ell_2}$ must be the only solution to the second equation. However, this implies that $T_{21} = 0$ from the third equation. The first equation can therefore be rewritten as

$$R_1(I_{\ell_1} \otimes A_1^0)\,\overline{\text{vec}}\,T_{11} = r_1$$

Clearly, A_1^0 is identified if and only if $T_{11} = I_{\ell_1}$ is the only solution to the above equation, and the proof is complete. ■

The zero-covariance restrictions with an unidentified sub-structure would in general not help identify the other sub-structure, as one may easily see in the proof.

With zero-covariance restrictions with the other identified sub-structure, a sub-structure is identified whenever it is identified in the corresponding subsystem regarded as if it were the complete system. The result in the previous subsection therefore applies as well. For instance, if the first sub-subsystem of the first subsystem consisting of ℓ_{11} equations is modelled with restrictions $R_{11}\,\overline{\text{vec}}\,A_{11} = r_{11}$, then a sub-sub-structure A_{11}^0 of $A_1^0 = (A_{11}^0, A_{12}^0)$ is identified if and only if $R_{11}(I_{\ell_{11}} \otimes A_1^0)$ has rank $\ell_{11}\ell_1$.

As an application, consider the so-called *recursive* model, which is characterized as: B is upper triangular with one on the diagonal, and Σ is diagonal. We now show by induction that the model is identified. Let us call the "i-th subsystem" the subsystem consisting of the first i equations. First, the first subsystem is clearly identified with ℓ restrictions. Second, suppose the $(i-1)$-th subsystem is identified. Then the i-th equation with normalization restriction is identified in the i-th subsystem, due to the result in Corollary 7. However, the i-th equation would then be identified in the system, since adding any linear combination of the rest equations to the i-th equation would violate the restrictions imposed on the equation. Consequently, the i-th subsystem is identified.

2.6 Exercises

1. Consider the following SEM's:

(A):

$$\begin{aligned} y_{1t} &= \gamma_1 x_{1t} + \gamma_2 x_{2t} + u_{1t} \\ y_{2t} &= \beta y_{1t} + \gamma_3 x_{1t} + u_{2t} \end{aligned}$$

(B):

$$\begin{aligned} y_{1t} &= \beta y_{2t} + \gamma x_t + u_{1t} \\ y_{2t} &= (1-\beta) y_{1t} + \gamma x_t + u_{2t} \end{aligned}$$

(a) Discuss the identifiability of equations in Model (A). Consider, in particular, when the true values of the parameters are $\beta = \gamma_1 = \gamma_3 = 1$ and $\gamma_2 = 0$.

(b) Discuss the identifiability of Model (B). Consider when $\beta = \gamma = 1$.

(c) The second equation in Model (A) is not identified when $\gamma_1 = \gamma_3 = 2, \beta = 1$ and $\gamma_2 = 0$. Construct an observationally equivalent structure.

2. Let a system of simultaneous equations be given by

$$\begin{aligned} y_{1t} &= \quad y_{2t} - x_t + u_{1t} \\ y_{2t} &= -y_{1t} + x_t + u_{2t} \end{aligned}$$

and denote by A_0 the corresponding structure.

(a) Define a model (R, r) in which A_0 is identified.

(b) Consider the model (R, r) with

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad r = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Show that the first equation is identified, while the second equation is not. Find an observationally equivalent system with the second equation different from the given system.

3. Consider the three-equation system

$$\begin{aligned} \alpha_{11} y_{1t} + \alpha_{21} y_{2t} + \alpha_{31} y_{3t} + \alpha_{41} x_{1t} + \alpha_{51} x_{2t} + \alpha_{61} x_{3t} &= u_{1t} \\ \alpha_{12} y_{1t} + \alpha_{22} y_{2t} + \alpha_{32} y_{3t} + \alpha_{42} x_{1t} + \alpha_{52} x_{2t} + \alpha_{62} x_{3t} &= u_{2t} \\ \alpha_{13} y_{1t} + \alpha_{23} y_{2t} + \alpha_{33} y_{3t} + \alpha_{43} x_{1t} + \alpha_{53} x_{2t} + \alpha_{63} x_{3t} &= u_{3t} \end{aligned}$$

with $\operatorname{var}(u_t) = \Sigma, u_t = (u_{1t}, u_{2t}, u_{3t})'$, where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

Suppose that the prior restrictions on coefficients are

$$\alpha_{11} = \alpha_{22} = \alpha_{33} = 1 \quad \text{and} \quad \alpha_{41} = \alpha_{61} = \alpha_{52} = \alpha_{63} = 0$$

and the covariance restrictions are

$$\sigma_{12} = \sigma_{13} = 0$$

Examine the identification for each equation.

3. Estimation by OLS and ILS

3.1 OLS and Simultaneous Equation Bias

Consider the i-the equation estimated by OLS, which yields

$$\hat{\alpha}^i_{\text{OLS}} = (Z_i'Z_i)^{-1}Z_i'y_i$$

The OLS estimate $\hat{\alpha}^i_{\text{OLS}}$ of α_i is, in general, inconsistent, since

$$\frac{Z_i'u_i}{n} = S_i'\frac{Z'u_i}{n} \xrightarrow{\mathcal{P}} G_i'\sigma_i^0 \neq 0$$

with $G_i = (B_0^{-1}, 0)S_i$, where B_0 is the true value of B and σ_i^0 is the i-th column of the true value Σ_0 of Σ. The result simply states that unless S_i selects endogenous variables uncorrelated with the errors the OLS estimator becomes inconsistent. The inconsistency is due to the presence of endogenous variables in the right-hand side of the regression, and the resulting bias is often referred to as the *simultaneous equation bias.*

3.2 Indirect Least Squares

We may obtain an estimate for the SF parameter A or its i-th column a_i from a LS estimate for the RF parameter Π using the relationship $\Pi = -CB^{-1}$, which can be more conveniently formulated in terms of

$$D = (\Pi, I_m)$$

as $DA = 0$ and $Da_i = 0$, or equivalently, as

$$\begin{aligned} (I_\ell \otimes D)(s - S\alpha) &= 0 \\ D(s_i - S_i\alpha_i) &= 0 \end{aligned}$$

The procedure is called *indirect least squares* (ILS).

Clearly, such procedure is possible only when

$$\begin{aligned} (I_\ell \otimes D)s &\in \mathcal{R}((I_\ell \otimes D)S) \\ Ds_i &\in \mathcal{R}(DS_i) \end{aligned}$$

for a given value of D. We may expect that the procedure yields a unique solution for α or α_i when the system or the equation is just identified. In such a case, the matrix $I_\ell \otimes D$ or DS_i becomes square. When underidentified, it generally yields multiple solutions. If a system or an equation is overidentified, then the solution does not exist, unless Π is estimated with the restrictions implied by the above conditions. With the restrictions, the RF model in general becomes a multivariate regression model with nonlinearity in parameter.

Consider a system and an equation that are just identified. The ILS estimators are then easily obtained from the above relationships with

$$\hat{D} = (\hat{\Pi}, I_m) = (X'X)^{-1}X'Z$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$ is the OLS estimate for Π in the RF model. Since $(I_\ell \otimes \hat{D})S$ is a square matrix that is nonsingular a.s., and $I_\ell \otimes \hat{D} = (I_\ell \otimes (X'X)^{-1})(I_\ell \otimes X')(I_\ell \otimes Z)$, the ILS estimator $\hat{\alpha}_{\mathrm{ILS}}$ for α is given by

$$\hat{\alpha}_{\mathrm{ILS}} = ((I_\ell \otimes X')Z_*)^{-1}(I_\ell \otimes X')y_*$$

Moreover, $\hat{D}S_i = (X'X)^{-1}X'ZS_i = (X'X)^{-1}X'Z_i$ is a square matrix and nonsingular a.s., and the ILS estimator $\hat{\alpha}^i_{\mathrm{ILS}}$ for α_i becomes

$$\hat{\alpha}^i_{\mathrm{ILS}} = (X'Z_i)^{-1}X'y_i$$

for the i-th equation.

4. IV Estimation

4.1 Instrumental Variables

As we have seen above, the simultaneous equation bias is due to the correlation between the errors and the endogenous variables included in the right-hand side of the equation. One obvious solution to correct the bias is to use the IV method. In a SEM, system exogenous variables are the most natural candidate for the instrumental variables. More precisely, X for a single equation and $I_\ell \otimes X$ for the entire system trivially satisfy the IV conditions (a), (c) and (d) by the assumptions on X. To see when the IV condition (b) holds, define for the true parameter value $A_0 = (B_0', C_0')'$

$$D_0 = (\Pi_0, I_m) \quad \text{with} \quad \Pi_0 = -C_0 B_0^{-1}$$

conformably with the previous definition of D, and let a_i^0 be the i-th column of A_0. Notice that

$$\frac{X'Z_i}{n} = \frac{X'Z}{n} S_i \xrightarrow{\mathcal{P}} MD_0S_i$$

and

$$\frac{(I_\ell \otimes X)Z_*}{n} = \left(I_\ell \otimes \frac{XZ}{n}\right) S \xrightarrow{\mathcal{P}} (I_\ell \otimes MD_0)S = (I_\ell \otimes M)(I_\ell \otimes D_0)S$$

The following lemma tells us exactly when the set of exogenous variables becomes a valid instrument.

Lemma 7 *The matrix $(I_\ell \otimes D_0)S$ is of full column rank if and only if A_0 is identified. Similarly, the matrix D_0S_i has full column rank if and only if a_i^0 is identified.*

Proof Notice that

$$\mathcal{N}(R) = \mathcal{R}(S) \quad \text{and} \quad \mathcal{R}(I_\ell \otimes A_0) = \mathcal{N}(I_\ell \otimes D_0)$$

since $RS = 0$ and $(I_\ell \otimes D_0)(I_\ell \otimes A_0) = I_\ell \otimes D_0A_0 = 0$, and therefore

$$\mathcal{N}(R) \cap \mathcal{R}(I_\ell \otimes A_0) = \mathcal{N}(I_\ell \otimes D_0) \cap \mathcal{R}(S)$$

which consists only of 0 if and only if $R(I_\ell \otimes A_0)$ and $(I_\ell \otimes D_0)S$ are full column rank. The proof for $D_0 S_i$ is completely analogous. ■

In sum, X for a single equation and $I_\ell \otimes X$ for the system become valid instruments when and only when the true structure A_0 and the sub-structure a_i^0 are identified.

4.2 Single Equation IV Estimation

Consider the IV estimation of the i-th equation on the assumption that the equation is identified. The IV estimator for the i-th equation is given explicitly by

$$\hat{\alpha}_{\mathrm{IV}}^i = (Z_i' P_X Z_i)^{-1} Z_i' P_X y_i$$

if $\operatorname{var}(u_i) = \sigma_i^2 I$ as we assume here. Since it is just an IV estimator, all the results for the IV estimator clearly apply. For instance, it is consistent, and its asymptotic distribution is given by

$$\sqrt{n}\,(\hat{\alpha}_{\mathrm{IV}}^i - \alpha_i^0) \xrightarrow{\mathcal{D}} \mathcal{N}\,(0, V_i)$$

where $V_i = \left(0, \sigma_i^2 (S_i D_0' M D_0 S_i)^{-1}\right)$, which can be consistently estimated by

$$\hat{\sigma}_i^2 \left(\frac{Z_i' P_X Z_i}{n}\right)^{-1}$$

Notice also that the ILS estimator is just the IV estimator for a just identified equation.

The IV estimator is called the *two stage least squares* (2SLS) estimator, if applied to an equation with only zero-one restrictions. In this case, $Z_i = (Y_i, X_i)$ and therefore $\hat{\alpha}_{\mathrm{IV}}^i$ can be viewed as the "two stage" LS estimator: Regress Y_i on X to get the fitted value $\hat{Y}_i$ in the first stage, and in the second stage regress y_i on $\hat{Z}_i = (\hat{Y}_i, X_i)$. Using $\hat{Y}_i$ instead of Y_i can be motivated as to "purge" endogeneity of Y_i.

The IV estimator can be regarded as a member of a broader class of estimators defined by

$$\hat{\alpha}_\kappa^i = (Z_i' Q_\kappa Z_i)^{-1} Z_i' Q_\kappa y_i$$

where

$$Q_\kappa = \kappa P_X + (1 - \kappa) I$$

called the *k-class estimator* for α_i. It becomes the IV estimator for $\kappa = 1$, while $\kappa = 0$ yields the OLS estimator. It is quite clear that any $\hat{\alpha}_\kappa^i$ with $\kappa = 1 + o_p(1)$ is consistent. Moreover, if $\kappa = 1 + o_p(\frac{1}{\sqrt{n}})$, then $\hat{\alpha}_\kappa^i$ has the same asymptotic distribution as the IV estimator.

4.3 System IV Estimation

Assume the system is identified. The instruments $I_\ell \otimes X$ and $(\Sigma^{-1} \otimes I_n)(I_\ell \otimes X) = \Sigma^{-1} \otimes X$ are valid, and both IV-OLS and IV-GLS analogue estimations are applicable. However, the two IV approaches yield the identical estimator, which is given by

$$\hat{\alpha}_{\mathrm{IV}} = \left(Z_*' (\tilde{\Sigma}^{-1} \otimes P_X) Z_*\right)^{-1} Z_*' (\tilde{\Sigma}^{-1} \otimes P_X) y_*$$

where $\tilde{\Sigma}$ is a consistent estimator for Σ, which can be obtained by applying single equation IV method equation by equation and using their residuals. If the system IV method is applied to a model with zero-one restrictions, then the resulting estimator is often called the *three stage least squares* (3SLS) estimator. Since it involves one additional step for the covariance matrix estimation using the 2SLS residuals, comes the name "three stage" LS.

As for the single equation IV, the general results for the IV estimator applies to the system IV estimator considered here. It is consistent, and its asymptotic distribution is given by

$$\sqrt{n}\,(\hat{\alpha}_{\mathrm{IV}} - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}\,(0, V)$$

with $V = \left(S'(\Sigma_0^{-1} \otimes D_0' M D_0) S\right)^{-1}$, which can be consistently estimated by

$$\left(\frac{Z_*'(\tilde{\Sigma}^{-1} \otimes P_X) Z_*}{n}\right)^{-1}$$

We also define *system k-class estimator* given by

$$\hat{\alpha}_\kappa = \left(Z_*'(\Sigma^{-1} \otimes Q_\kappa) Z_*\right)^{-1} Z_*'(\Sigma^{-1} \otimes Q_\kappa) y_*$$

where Q_κ is as defined in the previous subsection. When $\kappa = 0$, it is simply the SUR estimator, while if $\kappa = 1$ then the resulting estimator is the system IV estimator that we consider here. Clearly, any $\hat{\alpha}_\kappa$ with $\kappa = 1 + o_p(1)$ is consistent, and with $\kappa = 1 + o_p(\frac{1}{\sqrt{n}})$ it has the same asymptotic distribution as the system IV estimator.

5. ML Estimation

5.1 Full Information Maximum Likelihood

Consider the ML estimation of the entire system, called the *full information maximum likelihood* (FIML) procedure. Upon noticing that the Jacobian of the transformation $U = YB + XC$ is $|\det B|^n$, the density of Y can easily be deduced from that of U given in Section 3.1. Ignoring the constant term, the loglikelihood function of A and Σ is given by

$$\ell(A, \Sigma) = n\,\log|\det B| - \frac{n}{2}\log\det\Sigma - \frac{1}{2}\mathrm{tr}\,\Sigma^{-1} A'Z'ZA$$

Moreover, since for any given A

$$\Sigma = \frac{1}{n}\,A'Z'ZA$$

maximizes the likelihood function, we get the following concentrated loglikelihood function of A:

$$\ell(A) = n\,\log|\det B| - \frac{n}{2}\log\det\frac{1}{n}A'Z'ZA$$

ignoring the constant term again.

We have

$$\begin{aligned} d\ell(A) &= n\,\mathrm{tr}B^{-1}dB - \mathrm{tr}\Sigma^{-1}A'Z'Z\,dA \\ &= -\mathrm{tr}\Sigma^{-1}A'Z'XD\,dA \end{aligned}$$

where $D = (\Pi, I_m)$ with $\Pi = -CB^{-1}$ as before, since

$$nB^{-1}dB = \Sigma^{-1}A'Z'ZAB^{-1}(I_\ell, 0)dA \quad \text{and} \quad ZAB^{-1}(I_\ell, 0) = (Y - X\Pi, 0)$$

from which we may readily get the FOC's for the maximization of $\ell(A)$.

Now let $\hat{\alpha}_{\mathrm{FIML}}$ be the FIML estimator for α, and let $\hat{A}$, $\hat{D}$ and $\hat{\Sigma}$ be the corresponding parameters evaluated with $\hat{\alpha}_{\mathrm{FIML}}$. Furthermore, we define

$$\hat{Z} = X\hat{D} = (X\hat{\Pi}, X)$$

Writing $d\ell(A) = -(\overline{\mathrm{vec}}\,A)'(\Sigma^{-1} \otimes Z'XD)(\overline{\mathrm{vec}}\,dA)$ and replacing $\overline{\mathrm{vec}}\,A = s - S\alpha$ and $\overline{\mathrm{vec}}\,dA = -S\,d\alpha$, the FOC for $\hat{\alpha}_{\mathrm{FIML}}$ becomes

$$S'(\hat{\Sigma}^{-1} \otimes \hat{Z}'Z)(s - S\hat{\alpha}_{\mathrm{FIML}}) = 0$$

We therefore have

$$\hat{\alpha}_{\mathrm{FIML}} = \left(\hat{Z}_*'(\hat{\Sigma}^{-1} \otimes I_n)Z_*\right)^{-1} \hat{Z}_*'(\hat{\Sigma}^{-1} \otimes I_n)y_*$$

where $\hat{Z}_* = (I_\ell \otimes \hat{Z})S$.

To compare with $\hat{\alpha}_{\mathrm{FIML}}$, we write the system IV estimator $\hat{\alpha}_{\mathrm{IV}}$ as

$$\hat{\alpha}_{\mathrm{IV}} = \left(\tilde{Z}_*'(\tilde{\Sigma}^{-1} \otimes I_n)Z_*\right)^{-1} \tilde{Z}_*'(\tilde{\Sigma}^{-1} \otimes I_n)y_*$$

where $\tilde{Z}_* = (I_\ell \otimes \tilde{Z})S$ with

$$\tilde{Z} = P_X Z = (X\tilde{\Pi}, X)$$

$\tilde{\Pi} = (X'X)^{-1}X'Y$, and $\tilde{\Sigma}$ is the estimate of Σ based on single equation IV residuals. It is now straightforward to see that $\hat{\alpha}_{\mathrm{FIML}}$ and $\hat{\alpha}_{\mathrm{IV}}$ are asymptotically equivalent. In finite samples, the two may diverge by the amount that $\hat{Z}$ and $\tilde{Z}$, and $\hat{\Sigma}$ and $\tilde{\Sigma}$ differ.

5.2 Limited Information Maximum Likelihood

We now consider the ML procedure for a subsystem on the presumption that there are no overidentifying restrictions elsewhere. The procedure is called the *limited information maximum likelihood* (LIML). For a given structure (A, Σ), partition A and Σ as in Sections 2.4 and 2.5. Then

Proposition 8 *Suppose there are no overidentifying restrictions in the second subsystem. Then the concentrated likelihood function of* (A_1, Σ_{11}) *for the first subsystem is given by*

$$\ell(A_1, \Sigma_{11}) = \frac{n}{2}\log\det\frac{1}{n}A_1'Z'(I - P_X)ZA_1 - \frac{n}{2}\log\det\Sigma_{11} - \frac{1}{2}\mathrm{tr}\,\Sigma_{11}^{-1}A_1'Z'ZA_1$$

ignoring the constant term.

Proof We assume that Σ is of the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \mathrm{I} \end{pmatrix}$$

This causes no loss in generality because for any structure (A, Σ), we may consider (A_*, Σ_*) given by $A_* = AT$ and $\Sigma_* = T'\Sigma T$ with

$$T = \begin{pmatrix} \mathrm{I} & -\Sigma_{22\cdot 1}^{-1/2}\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \Sigma_{22\cdot 1}^{-1/2} \end{pmatrix}$$

where $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. The parameter in the first subsystem of (A_*, Σ_*) is (A_1, Σ_{11}), as in that of (A, Σ), and Σ_* is of the form given above. Moreover, (A, Σ) and (A_*, Σ_*) are observationally equivalent, and therefore yield the same density for Y.

The likelihood function of A and Σ is then given by

$$n \log|\det B| - \frac{n}{2}\log\det\Sigma_{11} - \frac{1}{2}\mathrm{tr}\,\Sigma_{11}^{-1}A_1'Z'ZA_1 - \frac{1}{2}\mathrm{tr}\,A_2'Z'ZA_2$$

To get the likelihood function concentrated on (A_1, Σ_{11}), we maximize for given A_1

$$n\log|\det B| - \frac{1}{2}\mathrm{tr}\,A_2'Z'ZA_2$$

with respect to A_2, or maximize

$$n\log|\det B| - \frac{1}{2}\mathrm{tr}\,B_2'Y'(I - P_X)YB_2$$

with respect to B_2, since for a given B_2 the maximizer of C_2 is given by $C_2 = (X'X)^{-1}X'YB_2$. The FOC for this maximization problem is

$$\mathrm{tr}\,\left(nJB^{-1} - B_2'Y'(I - P_X)Y\right)dB_2$$

where $J = (0, I)$. The FOC's are

$$\frac{1}{n}B_2'Y'(I - P_X)YB_1 = 0 \quad \text{and} \quad \frac{1}{n}B_2'Y'(I - P_X)YB_2 = I$$

and notice that

$$\begin{aligned} n\log|\det B| &= \frac{n}{2}\log\det\frac{1}{n}B'Y'(I - P_X)YB + \text{const} \\ &= \frac{n}{2}\log\det\frac{1}{n}B_1'Y'(I - P_X)YB_1 + \text{const} \end{aligned}$$

and $B_1'Y'(I - P_X)YB_1 = A_1'Z'(I - P_X)ZA_1$ to get the stated result. ■

The loglikelihood function further concentrated to A_1 is then given by

$$\ell(A_1) = \frac{n}{2}\log\det\frac{1}{n}A_1'Z'(I - P_X)ZA_1 - \frac{n}{2}\log\det\frac{1}{n}A_1'Z'ZA_1$$

since $\Sigma_{11} = A_1' Z' Z A_1 / n$ maximizes the likelihood for any given A_1.

For the i-th equation, we may simply consider

$$\kappa = \frac{a_i' Z' Z a_i}{a_i' Z' (I - P_X) Z a_i}$$

which, if *minimized* with respect to α_i for $a_i = s_i - S_i \alpha_i$, yields the LIML estimator $\hat{\alpha}^i_{\text{LIML}}$ for α_i. The ratio consists of two estimates for the variance of the i-th equation error, and $\hat{\alpha}^i_{\text{LIML}}$ may be viewed as the value of α_i which minimizes the ratio of two variance estimates. For this reason, the LIML is often referred to the method of *least variance ratio*.

Redefine $Z := (y_i, Z_i)$ and $a_i := (1, -\alpha_i')'$. Under this formulation, the minimum κ_* of κ and the minimizer a_i^* of a_i are the smallest eigenvalue and the corresponding eigenvector (normalized to have the unit first entry), respectively, of $Z'Z$ with respect to $Z'(I - P_X)Z$. For given κ_*, a_i^* is defined by

$$(Z'Z - \kappa_* Z'(I - P_X)Z) a_i^* = (Z' Q_{\kappa_*} Z) a_i^* = 0$$

where Q_κ is as defined before, and $\hat{\alpha}^i_{\text{LIML}}$ is obtained from a_i^* by $Z_i' Q_{\kappa_*} (y_i - Z_i \hat{\alpha}^i_{\text{LIML}}) = 0$. We have

$$\hat{\alpha}^i_{\text{LIML}} = (Z_i' Q_{\kappa_*} Z_i)^{-1} Z_i' Q_{\kappa_*} y_i$$

which defines the LIML estimator as a k-class estimator.

The interpretation of $\hat{\alpha}^i_{\text{LIML}}$ as a k-class estimator is very useful to derive its asymptotic properties. It is indeed not difficult to show that

$$\kappa_* = 1 + O_p\left(\frac{1}{n}\right)$$

since

$$\kappa_* - 1 \leq \frac{a_i^{0\prime} Z' P_X Z a_i^0}{a_i^{0\prime} Z' (I - P_X) Z a_i^0} \leq \frac{u_i' X (X'X)^{-1} X' u_i}{u_i' u_i - u_i' X (X'X)^{-1} X' u_i}$$

The LIML estimator is therefore asymptotically equivalent to the single equation IV estimator.

Moreover, recall that we have

$$P_X (y_i - Z_i \hat{\alpha}^i_{\text{IV}}) = 0$$

for the just-identified equation. Noticing that $P_X = Q_\kappa$ with $\kappa = 1$ and $\kappa \geq 1$, we may conclude that $\kappa_* = 1$ in this case. This, however, implies that

$$\hat{\alpha}^i_{\text{LIML}} = \hat{\alpha}^i_{\text{IV}}$$

for the just-identified equation.

5.3 Exercises

1. Let a SEM be given by

$$\begin{aligned} y_{1t} &= \alpha_{11} y_{2t} + \alpha_{12} x_{1t} + u_{1t} \\ y_{2t} &= \alpha_{21} x_{1t} + \alpha_{22} x_{2t} + u_{2t} \end{aligned}$$

and answer the following questions.
(a) Compare 2SLS, 3SLS, LIML and FIML estimators for the parameters α_{11} and α_{12} in the first equation.
(b) Now suppose $\alpha_{21} = 0$. How would your answer to (a) change?

2. Consider a SEM given by

$$\begin{aligned} y_{1t} &= \beta y_{2t} + u_{1t} \\ y_{2t} &= \gamma_1 x_{1t} + \gamma_2 x_{2t} + u_{2t} \end{aligned}$$

with the sample moment matrix

	y_1	y_2	x_1	x_2
y_1	10	2	3	1
y_2	2	5	1	1
x_1	3	1	2	1
x_2	1	1	1	1

Answer the following questions.
(a) Compute 2SLS, 3SLS, LIML and FIML estimates for β.
(b) Suppose the covariance matrix of u_{1t} and u_{2t} is

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix}$$

Find the FIML estimates of β, γ_1 and γ_2.