Simultaneous Equations Models

How is the Simultaneous Equations model different from the regular regression model with which we are all familiar?

1. Multiple equations (referred to as a *system* of equations);
2. Dependent variables in some equations are explanatory variables in other equations (this is a general form of the *random regressors problem* we encountered with lagged dependent variables models). The dependent variables (the $\mathbf{Y}$ variables) feed off of each other – shocks to one dependent variable reverberate through the model.

The generic form of a simultaneous equations model is written:

$$\begin{aligned}
\gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} &= \epsilon_{t1} \\
\gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} &= \epsilon_{t2} \\
&\vdots \\
\gamma_{1M}y_{t1} + \gamma_{2M}y_{tM} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} &= \epsilon_{tM},
\end{aligned} \tag{1}$$

where the y_{ti} $(i = 1, \ldots, M)$ are the *endogenous* variables, and the x_{tj} $(j = 1, \ldots, K)$ are the *exogenous* variables. Note that M is also the number of equations in the model. Under most circumstances, the model is "normalized" by setting the coefficient of the ith endogenous variable in the ith equation equal to -1. After normalization, equations (1) can be rewritten as,

$$\begin{aligned}
y_{t1} &= \gamma_{21}y_{t2} + \gamma_{31}y_{t3} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} + \epsilon_{t1} \\
y_{t2} &= \gamma_{12}y_{t1} + \gamma_{32}y_{t3} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} + \epsilon_{t2} \\
&\vdots \\
y_{tM} &= \gamma_{1M}y_{t1} + \cdots + \gamma_{M-1,M}y_{t,M-1} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} + \epsilon_{tM}.
\end{aligned} \tag{2}$$

We will most commonly encounter equations (1) in its matrix form,

$$\underset{(T\times M)}{\mathbf{Y}} \quad \underset{(M\times M)}{\mathbf{\Gamma}} + \underset{(T\times K)}{\mathbf{X}} \quad \underset{(K\times M)}{\mathbf{B}} = \underset{(T\times M)}{\mathbf{U}} \tag{3}$$

or

$$\mathbf{Z\Delta} = \mathbf{U},$$

where

$$\mathbf{Z} = \left[\begin{array}{cc} \mathbf{Y} & \mathbf{X} \end{array} \right] \text{ and } \mathbf{\Delta} = \left[\begin{array}{c} \mathbf{\Gamma} \\ \mathbf{B} \end{array} \right].$$

• Recall that the $\mathbf{Y}$ variables are affected by shocks to the system, while the $\mathbf{X}$ variables are not. So we say that the $\mathbf{Y}$ variables are *endogenous* while the $\mathbf{X}$ variables are *exogenous*.

• Because of the random regressors problem (which is most evident when looking at the representation of the system shown in equations (2)), we know that estimation of the system by ordinary least squares will give us suboptimal results. In particular, if we make the usual assumptions about the error structure of the model,

- $E(\epsilon_{ti}) = 0$ for $i = 1, \ldots, M$;
- $E(\epsilon_{ti}^2) = \sigma_i^2$ for $i = 1, \ldots, M$;
- $E(\epsilon_{ti}\epsilon_{tj}) = 0$ for all $i \neq j$;
- $E(\epsilon_{ti}y_{tj}) = 0$ for all $i, j = 1, \ldots, M$,

it turns out that the error term is correlated with the independent variables. Thus, the strategy of estimating each equation of the system separately by OLS yields inconsistent parameter estimates.

However, if we can somehow eliminate the endogenous variables from the set of regressors, estimation by OLS would no longer be a problem. Using familiar methods from linear algebra, we could "solve" system (1) to get:

$$\begin{array}{rcl} y_{t1} & = & \pi_{11}x_{t1} + \cdots + \pi_{K1}x_{tK} + \nu_{t1} \\ & \vdots & \\ y_{tM} & = & \pi_{1M}x_{t1} + \cdots + \pi_{KM}x_{tK} + \nu_{tM}, \end{array} \tag{4}$$

or, in matrix form,

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}$$

where $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ and $\mathbf{V} = \mathbf{U\Gamma}^{-1}$. Note that the sampling properties of $\mathbf{V}$ are nearly identical to those of $\mathbf{U}$:

- For each equation i ($i = 1, \ldots, M$), $E(\mathbf{V}_i) = E(\mathbf{U}_i\mathbf{\Gamma}^{-1}) = \mathbf{0}$, where $\mathbf{0}$ is an M-dimensional zero vector.
- $\text{Var}(\mathbf{V}_i) = (\mathbf{\Gamma}^{-1})'E(\mathbf{U}'_i\mathbf{U}_i)\mathbf{\Gamma}^{-1} = (\mathbf{\Gamma}^{-1})'\mathbf{\Sigma\Gamma}^{-1} \equiv \mathbf{\Omega}$.

Equations (4) are referred to as the "reduced form" of system (2). The π's in the reduced form system are nonlinear combinations of the γ's and β's of the original system (2).

- Note: The π's are referred to as the *reduced form parameters*, while the γ's and β's are called the *structural parameters*. Similarly, the form of the model represented in equations (1) or (2) is called the *structural form* of the model.

If we were to estimate the reduced form equations (4) by OLS, we would get parameter estimates

$$\hat{\mathbf{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and

$$\hat{\mathbf{\Omega}} = \frac{1}{T-K}\hat{\mathbf{V}}'\hat{\mathbf{V}}$$

with the classical sampling properties (provided the error term obeyed the classical assumptions).

However, except for the case in which we would like a forecast of $\mathbf{Y}$ given $\mathbf{X}$, we're generally not interested in the reduced form parameters. We'd like estimates of the structural parameters – the γ's and the β's. However, this may be easier said than done. Why? Note the following:

1. The number of reduced form parameters is $M \times K$;
2. The number of γ's is $M \times (M-1)$ (assuming we normalize the system), while the number of β's is $M \times K$.

Adding the dimensions of the structural form, we get $M \times (K+M-1)$ structural parameters. This means that if we can get the number of estimable structural parameters down to a mere $M \times K$, we can estimate the reduced form of the system, and reconstruct the structural parameters from the reduced form parameters (this is known as *indirect least squares*). The process of placing restrictions on the structural parameters so as to get the estimable number of them down to $M \times K$ is known as the *identification* issue, and it is one of two major issues in simultaenous equations modelling (the other being the estimation itself).

A simple example will illustrate the importance of identification. Suppose we construct the following model of the market for some good:

$$\begin{aligned} Q^d &= \alpha_0 + \alpha_1 P + e^d \\ Q^s &= \beta_0 + \beta_1 P + e^s \\ Q^s &= Q^d \end{aligned} \tag{5}$$

where P is the price of some good, Q^s is the supply of the good, Q^d is the demand for the good, and the subscripts on P, e^s and e^d have been suppressed to avoid clutter. Equating supply and demand yields:

$$\alpha_0 + \alpha_1 P + e^d = \beta_0 + \beta_1 P + e^s$$

$$\longrightarrow P = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{e^s - e^d}{\alpha_1 - \beta_1} \equiv \pi_1 + \nu_1$$

and

$$Q = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} - \frac{\alpha_1 e^s - \beta_1 e^d}{\alpha_1 - \beta_1} \equiv \pi_2 + \nu_2.$$

Since P affects both Q^s and Q^d, there is no way that we can reconstruct the structural parameters from the reduced form parameters.

But now, add an income component to the demand equation:

$$\begin{aligned} Q^d &= \alpha_0 + \alpha_1 P + \alpha_2 Y + e^d \\ Q^s &= \beta_0 + \beta_1 P + e^s \\ Q^s &= Q^d \end{aligned} \tag{6}$$

• Note that Y is assumed to affect demand, but not supply. In other words, if we were to write the supply equation as $Q^s = \beta_0 + \beta_1 P + \beta_2 Y + e^s$, we would be assuming that $\beta_2 = 0$ in this example. This is known as an *identifying restriction.*

Equating supply and demand in the second model yields:

$$\alpha_0 + \alpha_1 P + \alpha_2 Y + e^d = \beta_0 + \beta_1 P + e^s$$

$$\longrightarrow P = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2 Y}{\alpha_1 - \beta_1} + \frac{e^s - e^d}{\alpha_1 - \beta_1} \equiv \pi_{10} + \pi_{11} Y + \nu_1$$

and

$$Q = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} - \frac{\alpha_2\beta_1 Y}{\alpha_1 - \beta_1} - \frac{\alpha_1 e^s - \beta_1 e^d}{\alpha_1 - \beta_1} \equiv \pi_{20} + \pi_{21} Y + \nu_2.$$

At this point, we can be very clever and notice that

$$\frac{\pi_{21}}{\pi_{11}} = \left(\frac{\alpha_2\beta_1 Y}{\alpha_1 - \beta_1}\right) \Big/ \frac{\alpha_2 Y}{\alpha_1 - \beta_1} = \beta_1.$$

In other words, we have just "identified" β_1. With a little more work, we can deduce that $\beta_0 = \pi_{11} - \beta_1\pi_{10}$. Thus, the entire structural supply equation has been identified. However, none of the structural parameters of the demand equation can be identified. If we were to add a third exogenous variable that was assumed to affect supply but not demand, we could, by a similar process, identify the entire demand equation.

• **Conditions for Identification**

Recall equations (6):

$$\begin{aligned} Q^d &= \alpha_0 + \alpha_1 P + \alpha_2 Y + e^d \\ Q^s &= \beta_0 + \beta_1 P + \beta_2 Y + e^s \\ Q^s &= Q^d. \end{aligned}$$

Our identifying restriction was setting $\beta_2 = 0$; this allowed us to fully identify the supply equation. Our restriction basically took the form of placing a zero in the matrix of slope parameters where we might otherwise have had an estimable parameter. In other words, we have *excluded* the income variable Y from the supply equation.

Exclusion of variables from certain equations (along with normalization) comprises one of the primary techniques used to identify simultaneous equations models. Consequently, we will be interested in just how many identifying restrictions we will need to make in order to be able to recover the model's structural parameters from its reduced-form parameters. There are two sets of rules that we use to determine whether or not we have made enough identifying restrictions, called the *order* and *rank* conditions.

• **The Order Condition** The order condition is a *necessary* condition for identification – if the order condition is satisfied, we know that we can recover the structural parameters from the reduced form paramters (though the solution may not be unique). In the discussion that follows, any parameter or variable matrix marked with an asterisk (*) has been *excluded* from the equation at hand.

Note that the jth equation of a simultanous equations model can be written as

$$\mathbf{y}_j = \mathbf{Y}_j \boldsymbol{\gamma}_{\ j} + \mathbf{Y}_j^* \boldsymbol{\gamma}_{\ j}^* + \mathbf{x}_j \boldsymbol{\beta}_{\ j} + \mathbf{x}_j^* \boldsymbol{\beta}_{\ j}^* + \boldsymbol{\epsilon}_{\ j}, \tag{7}$$

with the implication that $\boldsymbol{\gamma}_{\ j}^*$ and $\boldsymbol{\beta}_{\ j}^*$ are both equal to zero. We say that there are M_j^* excluded endogenous variables and K_j^* excluded exogenous variables in the jth equation, and M_j included endogenous variables and K_j included exogenous variables in the jth equation. This means that $M_j^* + M_j + 1 = M$ (the total number of endogenous variables in the model) and $K_j^* + K_j = K$ (the total number of exogenous variables in the model).

Recall that the reduced form coefficient matrix is

$$\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1},$$

which can be rewritten as

$$\boldsymbol{\Pi}\boldsymbol{\Gamma} = -\mathbf{B},$$

of which the jth equation is

$$\boldsymbol{\Pi}\boldsymbol{\Gamma}_j = -\mathbf{B}_j. \tag{8}$$

We can then separate equation (8) into two parts – those which contain *included* variables and those which contain *excluded* variables:

$$\begin{bmatrix} \boldsymbol{\pi}_{\ j} & \boldsymbol{\Pi}_j^1 & \boldsymbol{\Pi}_j^2 \\ \boldsymbol{\pi}_{\ j}^{\ *} & \boldsymbol{\Pi}_j^{*1} & \boldsymbol{\Pi}_j^{*2} \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\gamma}_{\ j} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{\ j} \\ \mathbf{0} \end{bmatrix}.$$

The K_j equations with included variables are:

$$\boldsymbol{\pi}_{\ j} - \boldsymbol{\Pi}_j^1 \boldsymbol{\gamma}_{\ j} = \boldsymbol{\beta}_{\ j} \tag{9}$$

(which is also the solution to $\mathbf{B}$ in terms of $\boldsymbol{\Gamma}$, since these are where the nonzero parameters are located). The K_j^* equations with excluded variables are:

$$\boldsymbol{\pi}_{\ j}^{\ *} - \boldsymbol{\Pi}_j^{*1} \boldsymbol{\gamma}_{\ j} = \mathbf{0},$$

or

$$\boldsymbol{\Pi}_j^{*1} \boldsymbol{\gamma}_{\ j} = \boldsymbol{\pi}_{\ j}^{\ *}. \tag{10}$$

If we can solve equation (10) for $\boldsymbol{\gamma}_{\ j}$, then we can plug the solution into (9) and obtain an estimate of $\boldsymbol{\beta}_{\ j}$. If we can do this for all j equations in the model, we can fully reconstruct $\mathbf{B}$ and $\boldsymbol{\Gamma}$, the matrices of structural parameters. This then leads us to the Order Condition – namely, that we need to be able to solve equations (9) and (10). In equation (10), there are K_j^* equations in M_j unknowns. Rules from linear algebra require that there be at least as many equations as unknowns. Consequently, we have the following:

The Order Condition for Identification: $K_j^* \geq M_j$. In other words, for the jth equation to be identified, the number of exogenous variables excluded from the jth equation must be at least as large as the number of endogenous variables included in the jth equation.

• **The Rank Condition** While the order condition for identification needed to be satisfied in order to reconstruct the structural parameters from the reduced form parameters, such a reconstruction may not be unique. In fact, if the strict inequality in the Order Condition holds, and $K_j^* > M_j$, then the jth equation is identified, but it is *overidentified* – there will be more than one way to reconstruct the structural parameters from the reduced form parameters. Consider again equations (6),

$$\begin{aligned} Q^d &= \alpha_0 + \alpha_1 P + \alpha_2 Y + e^d \\ Q^s &= \beta_0 + \beta_1 P + e^s \\ Q^s &= Q^d, \end{aligned}$$

where we have imposed the identifying restriction that Y has no effect on Q^s. Suppose that there were a third variable which was thought to influence demand, such as the price of a substitute good (call this variable P^a). Further, suppose that the price of this substitute good was thought to have

no influence on supply (this is probably unrealistic, but this is just an example). We would then write equations (6) as

$$\begin{aligned} Q^d &= \alpha_0 + \alpha_1 P + \alpha_2 Y + \alpha_3 P^a + e^d \\ Q^s &= \beta_0 + \beta_1 P + \beta_2 Y + \beta_3 P^a + e^s \\ Q^s &= Q^d, \end{aligned}$$

and impose the identifying restriction that $\beta_2 = \beta_3 = 0$. From our earlier example we have seen that β_1 can be derived from the reduced-form parameters as

$$\beta_1 = \frac{\pi_{21}}{\pi_{11}} = \left(\frac{\alpha_2\beta_1 Y}{\alpha_1 - \beta_1}\right) \Big/ \frac{\alpha_2 Y}{\alpha_1 - \beta_1}.$$

However, we could also derive β_1 a second way, namely

$$\beta_1 = \left(\frac{\alpha_3\beta_1 P^a}{\alpha_1 - \beta_1}\right) \Big/ \frac{\alpha_3 P^a}{\alpha_1 - \beta_1}.$$

(In practice we would hope that β_1 would be the same no matter which formula we used!) The Rank Condition for identification tells us when there is only one way to reconstruct the structural parameters from the reduced form parameters. The formal statement is the following:

Rank Condition for Identification $\text{rank}[\boldsymbol{\pi}_{j}^{*} \; \boldsymbol{\Pi}_j^*] = \text{rank}\,[\boldsymbol{\Pi}_j^*] = M_j$. This is just another way of saying that equation (10) has a unique solution.

Note that, in practice, the Order Condition becomes more useful than the Rank Condition, since the Rank Condition requires that $\boldsymbol{\Pi}_j^*$ be known, which may not always be true, especially with a large number of equations and/or variables. The Order Condition, on the other hand, simply amounts to the following collection of counting rules:

For the jth equation, if:

1. $M_j + K_j < K$, the equation is overidentified;
2. $M_j + K_j = K$, the equation is just identified (i.e., there is only one combination of the reduced form parameters that will reconstruct the structural parameters in equation j);
3. $M_j + K_j > K$, the equation is not identified, and we cannot reconstruct the equation's structural parameters.

• **Estimation of Simultaneous Equations Models**

1. OLS and Instrumental Variables

As we will see, because we generally will have endogenous variables as regressors when estimating the model, OLS parameter estimates are biased and inconsistent.

Write the jth equation in the system as

$$\begin{aligned} \mathbf{y}_j &= \mathbf{Y}_j\boldsymbol{\gamma}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{e}_j \\ &\equiv \mathbf{Z}_j\boldsymbol{\delta}_j + \mathbf{e}_j \end{aligned} \tag{11}$$

The OLS estimate of $\boldsymbol{\delta}_j$ is

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{j,\text{OLS}} \equiv \mathbf{d}_j &= (\mathbf{Z}'_j\mathbf{Z})^{-1}\mathbf{Z}'_j\mathbf{y}_j. \\ &= \boldsymbol{\delta}_j + \begin{bmatrix} \mathbf{Y}'_j\mathbf{Y}_j & \mathbf{Y}'_j\mathbf{X}_j \\ \mathbf{X}'_j\mathbf{Y}_j & \mathbf{X}'_j\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_j\mathbf{e}_j \\ \mathbf{X}'_j\mathbf{e}_j \end{bmatrix}. \end{aligned} \tag{12}$$

Unfortunately, we see that $\mathbf{d}_j$ is biased because no term in the inverse matrix above converges to zero in the probability limit. While it is true that $\text{plim}(1/T)\mathbf{X}'_j\mathbf{e}_j = 0$, it is unfortunately also true that

$$\begin{aligned} \text{plim}(1/T)\mathbf{Y}'_j\mathbf{e}_j &= \text{plim}(1/T)\mathbf{Y}'_j\mathbf{V}\boldsymbol{\Gamma}_j \text{ (since } \mathbf{V} = \mathbf{U}\boldsymbol{\Gamma}^{-1}\text{, so } \mathbf{U} = \mathbf{V}\boldsymbol{\Gamma}) \\ &= \boldsymbol{\Gamma}_j\text{plim}(1/T)\mathbf{V}'_j\mathbf{V} \neq 0. \end{aligned}$$

Since OLS will not work well under the model's assumptions, we turn instead to an instrumental variables (IV) estimator. Again considering the jth equation of the system,

$$\mathbf{Z}_j\boldsymbol{\delta}_j + \mathbf{e}_j,$$

Let $\mathbf{W}_j$ be a $T \times (M_j + K_j)$ matrix satisfying the usual IV assumptions:

$$\begin{cases} \text{plim}(1/T)\mathbf{W}'_j\mathbf{Z}_j = \boldsymbol{\Sigma}_{wz} \text{ ($\boldsymbol{\Sigma}_{wz}$ a finite nonsingular matrix)} \\ \text{plim}(1/T)\mathbf{W}'_j\mathbf{e}_j = \mathbf{0} \\ \text{plim}(1/T)\mathbf{W}'_j\mathbf{W}_j = \boldsymbol{\Sigma}_{ww} \text{ ($\boldsymbol{\Sigma}_{ww}$ a positive matrix).} \end{cases}$$

Then, the IV estimator

$$\hat{\boldsymbol{\delta}}_{j,\text{IV}} = (\mathbf{W}'_j\mathbf{Z}_j)^{-1}\mathbf{W}'_j\mathbf{y}_j \tag{13}$$

is consistent and has

$$\text{asym. var.}(\hat{\boldsymbol{\delta}}_{j,\text{IV}}) = (\sigma_{jj}/T)\boldsymbol{\Sigma}_{wz}^{-1}\boldsymbol{\Sigma}_{ww}\boldsymbol{\Sigma}_{zw}^{-1} \tag{14}$$

σ_{jj} is consistently estimated as

$$\hat{\sigma}_{jj} = (1/T)(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_{\ j,\mathrm{IV}})'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_{\ j,\mathrm{IV}}).$$

For small samples, a degrees of freedom correction of $T - M_j - K_j$ is sometimes suggested.

The IV method of estimation is the basic building block for nearly all of the methods of estimation we will consider.

2. Indirect Least Squares

If a system of equations is just identified, an IV-based estimator called *indirect least squares* (which we have briefly met beforehand) works well. Indirect Least Squares (ILS) uses the $\mathbf{W}$ matrix

$$\mathbf{W}_j = [\mathbf{X}_j^* \ \mathbf{X}_j] = \mathbf{X}. \tag{15}$$

A consistent estimator of $\boldsymbol{\delta}_{\ j}$ is:

$$\hat{\boldsymbol{\delta}}_{\ j,\mathrm{ILS}} = (\mathbf{X}'\mathbf{Z}_j)^{-1}\mathbf{X}'\mathbf{y}_j. \tag{16}$$

Note that the ILS estimator only works because there is one excluded exogenous variable serving as an instrument for each included endogenous variable.

3. Two-Stage Least Squares (2SLS)

Strikingly similar to the regular IV estimate of equation (13), Two- Stage Least Squares (2SLS) is an option when faced with an overidentified simultaneous equations model.

The 2SLS estimator uses predicted $\hat{\mathbf{Y}}_j$ values (from a regression of $\mathbf{Y}_j$ on the entire $\mathbf{X}$ matrix) as the instruments:

$$\hat{\mathbf{Y}}_j = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j \equiv \mathbf{X}\mathbf{P}_j.$$

Thus,

$$E(\hat{\mathbf{Y}}_j) = E(\mathbf{X}\mathbf{P}_j) = \mathbf{X}E(\mathbf{P}_j) = \mathbf{X}\boldsymbol{\Pi}_j.$$

(Note that $\hat{\mathbf{Y}}_j$ is built from the reduced form parameter matrix)

Now, in true IV fashion, let

$$\begin{aligned} \mathbf{W}_j &= [\hat{\mathbf{Y}}_j \ \mathbf{X}_j] \\ &= [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j \ \mathbf{X}_j] \end{aligned} \tag{17}$$

Writing the jth equation as

$$\mathbf{y}_j = \mathbf{W}_j\hat{\boldsymbol{\delta}}_{\ j},$$

the 2SLS estimator is

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{j,2\text{SLS}} &= (\mathbf{W}'_j\mathbf{W}_j)^{-1}\mathbf{W}'_j\mathbf{y}_j \\ &= \begin{bmatrix} \hat{\mathbf{Y}}'_j\hat{\mathbf{Y}}_j & \hat{\mathbf{Y}}'_j\mathbf{X}_j \\ \mathbf{X}'_j\hat{\mathbf{Y}}_j & \mathbf{X}'_j\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}'_j\mathbf{y}_j \\ \mathbf{X}'_j\mathbf{y}_j \end{bmatrix}. \end{aligned} \tag{18}$$

Computationally, we can make our lives simpler by noting that $\mathbf{W}'_j\mathbf{W}_j = \mathbf{W}'_j\mathbf{Z}_j$, so that equation (18) becomes:

$$\hat{\boldsymbol{\delta}}_{j,2\text{SLS}} = \begin{bmatrix} \hat{\mathbf{Y}}'_j\mathbf{Y}_j & \hat{\mathbf{Y}}'_j\mathbf{X}_j \\ \mathbf{X}'_j\mathbf{Y}_j & \mathbf{X}'_j\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}'_j\mathbf{y}_j \\ \mathbf{X}'_j\mathbf{y}_j \end{bmatrix}.$$

This turns out to be equivalent to

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{j,2\text{SLS}} &= (\hat{\mathbf{Z}}'_j\hat{\mathbf{Z}}_j)^{-1}\hat{\mathbf{Z}}'_j\mathbf{y}_j \\ &= (\mathbf{Z}_j[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}_j)^{-1}(\mathbf{Z}_j[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}_j) \end{aligned} \tag{19}$$

which is how we are used to seeing the 2SLS estimator. The sampling properties of $\hat{\boldsymbol{\delta}}_j$ are:

$$\widehat{\text{asym.}}\,\text{var.}(\hat{\boldsymbol{\delta}}_{j,2\text{SLS}}) = \hat{\sigma}_{jj}(\hat{\mathbf{Z}}'_j\hat{\mathbf{Z}}_j)^{-1}$$

where

$$\hat{\sigma}_{jj} = (1/T)(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j).$$

2SLS, then, may be summed up as the following (unsurprisingly) two-step procedure:

1. 1. Obtain predicted values $\hat{\mathbf{Y}}_j$ of $\mathbf{Y}_j$ via an ordinary least squares regression of $\mathbf{Y}_j$ on $\mathbf{X}$.
2. 2. Obtain the 2SLS estimate of $\boldsymbol{\delta}_j$ by ordinary least squares regression of $\mathbf{y}_j$ on $\hat{\mathbf{Y}}_j$ and $\mathbf{X}_j$.

The 2SLS and other Instrumental Variable estimators presented so far are examples of *limited information* estimators, so-called because the estimators consider only one equation at a time from the system, and do not take into account the effects of any cross-equation error correlation. Before moving on to *full information* estimators, which estimate all M equations of the system concurrently, and do allow for cross-equation error correlation (i.e., non-diagonal $\boldsymbol{\Sigma}$), we discuss one more limited information estimator based on the maximum likelihood method.

4. Limited Information Maximum Likelihood (LIML)

In the end, the LIML estimator is similar to the 2SLS estimator with the added assumption that the reduced form error matrix of each equation is normally distributed. As a matter of fact, LIML shows that 2SLS is an asymptotically efficient limited information estimator if the reduced form errors do indeed follow the normal distribution.

Write the jth reduced form equation as

$$\mathbf{Y}_j^0 = \mathbf{X}\mathbf{\Pi}_j^0 + \mathbf{V}_j^0, \tag{20}$$

where $\mathbf{Y}_j^0 = [\mathbf{y}_j \ \mathbf{Y}_j]$. If we assume that $\mathbf{V}_j^0 \sim N(0, \mathbf{\Omega}_j^0)$, the LIML estimator $\hat{\boldsymbol{\delta}}_{j,\text{LIML}}$ is obtained by maximizing the log-likelihood function

$$\mathcal{L}(\mathbf{\Pi}_j^0) = -(1/2)N\{(M_j+1)\ln(2\pi) + \ln(|\mathbf{\Omega}_j^0|) + \text{tr}[(\mathbf{Y}_j^0 - \mathbf{X}\mathbf{\Pi}_j^0)'\mathbf{\Omega}_j^{0^{-1}}(\mathbf{Y}_j^0 - \mathbf{X}\mathbf{\Pi}_j^0)]\} \tag{21}$$

with the restriction that $\mathbf{\Pi}\mathbf{B}_j = -\mathbf{\Gamma}_j$.

Full-Information Methods

Recall that with limited information methods of estimation, each equation in the system in considered separately, and the estimator does not account for any cross-equation error correlation that may exist (hence the name "limited information"). Full information estimators, on the other hand, estimate all M equations simultaneously, and explicitly accounts for contemporaneous correlation that might exist between the error terms of separate equations.

To begin with, we must stack the M equations thusly to create one big matrix equation:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_M \end{bmatrix}, \tag{22}$$

or $\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$.

We generally assume that

$$E(\mathbf{e}) = \mathbf{0} \tag{23}$$

and

$$\begin{aligned} E(\mathbf{e}\mathbf{e}') &= \mathbf{\Sigma} \otimes \mathbf{I} \\ &= \begin{bmatrix} \sigma_{11}\mathbf{I} & \cdots & \sigma_{1M}\mathbf{I} \\ \vdots & \ddots & \vdots \\ \sigma_{M1}\mathbf{I} & \cdots & \sigma_{MM}\mathbf{I} \end{bmatrix}. \end{aligned} \tag{24}$$

As with limited information methods, the OLS estimator $\mathbf{d} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ estimates the system equation-by- equation and is therefore an inconsistent estimator of $\boldsymbol{\delta}$.

1. Seemingly Unrelated Regressions

As a starting point to full information methods of estimation, we consider the Seemingly Unrelated Regressions (SUR) estimator. The SUR estimator is only useful in the special case where

the set of regressors does not contain any of the endogenous variables. In other words, SUR can be used when $\mathbf{\Gamma} = \mathbf{0}$. The model then becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \ + \mathbf{e}, \tag{25}$$

where we assume that

$$E(\mathbf{e}) = \mathbf{0} \tag{26}$$

and

$$E(\mathbf{e}\mathbf{e}') = \mathbf{V} = \mathbf{\Sigma} \otimes \mathbf{I}. \tag{27}$$

Efficient estimation of $\hat{\boldsymbol{\beta}}$ requires generalized least squares, with which we are already familiar:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GLS} &= [(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}] \\ &= [\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})^{-1}\mathbf{X}^{-1}\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})^{-1}\mathbf{y}] \\ &= \begin{bmatrix} \sigma^{11}\mathbf{X}'_1\mathbf{X}_1 & \cdots & \sigma^{1M}\mathbf{X}'_1\mathbf{X}_M \\ \vdots & \ddots & \vdots \\ \sigma^{M1}\mathbf{X}'_M\mathbf{X}_1 & \cdots & \sigma^{MM}\mathbf{X}'_M\mathbf{X}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M} \sigma^{1j}\mathbf{X}'_1\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}\mathbf{X}'_M\mathbf{y}_j \end{bmatrix}. \end{aligned} \tag{28}$$

Note that SUR assumes that $\mathbf{\Sigma}$ is known. Since this is not usually the case, the SUR estimator uses $\hat{\mathbf{\Sigma}}$, an estimate of $\mathbf{\Sigma}$. $\hat{\mathbf{\Sigma}}$ is computed from the least squares residuals, with elements

$$\hat{\sigma}_{ij} = (1/T)\mathbf{e}'_i\mathbf{e}_j.$$

For small samples, the degrees of freedom corrections

$$\hat{\sigma}^*_{ij} = \frac{\mathbf{e}'_i\mathbf{e}_j}{\sqrt{(T-K_i)(T-K_j)}} \text{ or } \hat{\sigma}^{**}_{ij} = \frac{\mathbf{e}'_i\mathbf{e}_j}{T - \max(K_i, K_j)}$$

are sometimes suggested.

Note that SUR is exactly equivalent to OLS if:

1. $\text{cov}(e_{it}, e_{js}) = 0 \ \ \forall t, s, (i \neq j)$
2. $\mathbf{X}_i = \mathbf{X}_j$ (i.e., all equations have identical regressors)

It is easy to see whether the second item is satisfied. To determine the degree of cross-equation contemporaneous error correlation, the following test is suggested. The lagrange multiplier test statistic is:

$$\lambda = T \sum_{i=2}^{M} \sum_{j=1}^{i-1} r_{ij}^2$$

where $r_{ij}^2 = \hat{\sigma}_{ij}/(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{1/2}$. λ is distributed as a chi-square with $M(M-1)$ degrees of freedom, and tests the null hypothesis that there is no cross-equation error correlation ($H_0 : \sigma_{ij} = 0 \ \forall i \neq j$).

2. Three-Stage Least Squares

As might be expected, Three-Stage Least Squares (3SLS) is in some sense an adaptation of the 2SLS method to the problem of full information estimation (and not just because the estimator has multiple stages!). In fact, 3SLS is a combination of Instrumental Variables and Generalized Least Squares.

We start with Instrumental Variables here. Let the IV matrix $\mathbf{W}$ be given by

$$\begin{aligned}\mathbf{W} = \hat{\mathbf{Z}} &= \begin{bmatrix} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_M \end{bmatrix} \\ &\equiv \begin{bmatrix} \hat{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{Z}}_M \end{bmatrix}. \end{aligned} \tag{29}$$

Use of the IV procedure gives us a consistent estimator (whereas OLS does not). However, simply using IV is not enough, because in this case the IV estimator $\hat{\boldsymbol{\delta}}_{\text{IV}} = (\hat{\mathbf{Z}}'\mathbf{Z})^{-1}\hat{\mathbf{Z}}'\mathbf{y}$ simply runs 2SLS separately on each equation. This turn out to be less efficient that Generalized Least Squares. If we combine our IV estimator from equation (29) with the GLS method, we obtain the 3SLS estimator:

$$\hat{\boldsymbol{\delta}}_{3SLS} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{Z}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{y}. \tag{30}$$

This is equivalent to the estimator (which is more commonly used)

$$\hat{\hat{\boldsymbol{\delta}}}_{3SLS} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\hat{\mathbf{Z}}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{y}. \tag{31}$$

The GLS portion of the 3SLS estimator accounts for the cross-equation error correlation. The sampling properties of $\hat{\hat{\boldsymbol{\delta}}}_{3SLS}$ are

$$\widehat{\text{asym. var.}}(\hat{\hat{\boldsymbol{\delta}}}_{3SLS}) = [\bar{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\bar{\mathbf{Z}}]^{-1}. \tag{32}$$

The asymptotic variance of $\hat{\hat{\boldsymbol{\delta}}}_{3SLS}$ is usually estimated using the inverse matrix $[\hat{\mathbf{Z}}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\hat{\mathbf{Z}}]^{-1}$ from equation (31). $\bar{\mathbf{Z}}$ is a matrix equal to the diagonal elements of $[\mathbf{X}\boldsymbol{\Pi}_j \;\; \mathbf{X}_j]$, which can be estimated using $\hat{\mathbf{Z}}$.

The last remaining difficulty is that the 3SLS estimator assumes that Σ is known. We usually get around this by estimating Σ just as we did in 2SLS, with

$$\hat{\sigma}_{ij} = (1/T)(\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i)'(\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_j). \tag{33}$$

As with 2SLS, we may sum up the 3SLS procedure in the following (three) steps:

1. Estimate the reduced form parameter matrix $\mathbf{\Pi}$ by OLS, and use $\hat{\mathbf{\Pi}}$ to yield $\hat{\mathbf{Y}}$, the predicted values of $\mathbf{Y}$.

2. For each equation, compute $\hat{\boldsymbol{\delta}}_{j,\text{2SLS}}$, and then compute $\hat{\mathbf{\Sigma}}$ as in equation (33). This corrects for the inconsistency of OLS.

3. Compute the final 3SLS estimator using the GLS procedure in equation (31) and the asymptotic covariance matrix as per step (2). This corrects for any cross-equation error correlation that might be present.

3. Full Information Maximum Likelihood

Full Information Maximum Likelihood (FIML) is the full information analogue of Limited Information Maximum Likelihood (LIML). The same relationship that exists between LIML and 2SLS also exists between FIML and 3SLS. Namely, if the error structure is normally distributed, 3SLS is asymptotically equivalent to FIML. Because of this (and because the 3SLS estimator is much less computationally expensive than FIML), 3SLS is more commonly used than FIML.

FIML begins with the full set of reduced form equations

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V}$$

where each row of $\mathbf{V}$ is assumed to be normally distributed with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Omega}$. The log-likelihood function to be maximized is

$$\ln(\mathcal{L}) = -(T/2)\{M\ln(2\pi) + \ln|\mathbf{\Omega}| + tr[(1/T)\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_i^0)'(\mathbf{y} - \mathbf{X}\boldsymbol{\pi}_i^0)]\} \tag{34}$$

where $\boldsymbol{\pi}_i^0$ is the ith column of $\mathbf{\Pi}$. With a little effort, equation (34) can be reduced to

$$\ln(\mathcal{L}) = -(1/2)MT\ln(2\pi) + T\ln|\mathbf{\Gamma}| - (T/2)tr(\mathbf{\Sigma}^{-1}\mathbf{S}) - (T/2)\ln|\mathbf{\Sigma}|, \tag{35}$$

where

$$s_{ij} = (1/T)(\mathbf{Y\Gamma}_j + \mathbf{XB}_j)'(\mathbf{Y\Gamma}_j + \mathbf{XB}_j).$$

Equation (35) has the advantage over equation (34) in that includes the structural parameters in the log-likelihood function, rather than the reduced form parameters. The disadvantage is that equation involves $\mathbf{\Sigma}$, which is unknown. However, we can easily substitute $\hat{\mathbf{\Sigma}}$ for $\mathbf{\Sigma}$, where $\hat{\mathbf{\Sigma}}$ is calculated using equation (33).

One alternative to FIML which may have advantages using small samples (when the asymptotic efficiency of 3SLS may not hold) is the *linearized ML* method. Using linearized ML, we can obtain a FIML parameter estimate as follows:

The function to be maximized is

$$\mathcal{L}_2(\mathbf{B},\mathbf{\Gamma}) = \ln(|\mathbf{\Gamma}'\mathbf{Y}'\mathbf{Y\Gamma}|/T) - \ln(|[\mathbf{Y\Gamma} + \mathbf{XB}]'[\mathbf{Y\Gamma} + \mathbf{XB}]|/T). \tag{36}$$

Making the substutuions $Q = |\mathbf{\Gamma}'\mathbf{Y}'\mathbf{Y}\mathbf{\Gamma}|/T$ and $S = |[\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B}]'[\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B}]|/T$, the normal equations for maximizing $\mathcal{L}_2$ are

$$\frac{\partial \mathcal{L}_2}{\partial \mathbf{B}} = \mathbf{Y}'\mathbf{Y}\mathbf{\Gamma}Q^{-1} - \mathbf{Y}'(\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B})S^{-1} = 0 \tag{37}$$

$$\frac{\partial \mathcal{L}_2}{\partial \mathbf{\Gamma}} = -\mathbf{X}'(\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B})S^{-1} = 0 \tag{38}$$

The FIML estimator of $\boldsymbol{\delta}$ is then derived from

$$\hat{\boldsymbol{\delta}} = [\mathbf{Z}'(\hat{\mathbf{\Sigma}} \otimes \mathbf{I})\mathbf{Z} - \mathbf{Z}'^{0}(\hat{\mathbf{\Phi}}^{-1} \otimes \mathbf{I})\mathbf{Z}^{0}]^{-1}[\mathbf{Z}'(\hat{\mathbf{\Sigma}} \otimes \mathbf{I})\mathbf{y} - \mathbf{Z}'^{0}(\hat{\mathbf{\Phi}}^{-1} \otimes \mathbf{I})\mathbf{y}^{0}] \tag{39}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} = \begin{bmatrix} [\mathbf{Y}_1 \ \mathbf{X}_1] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & [\mathbf{Y}_2 \ \mathbf{X}_2] & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & [\mathbf{Y}_M \ \mathbf{X}_M] \end{bmatrix};$$

$$\mathbf{Z}^0 = \begin{bmatrix} \mathbf{Z}_1^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2^0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M^0 \end{bmatrix} = \begin{bmatrix} [\mathbf{Y}_1 \ \mathbf{0}] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & [\mathbf{Y}_2 \ \mathbf{0}] & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & [\mathbf{Y}_M \ \mathbf{0}] \end{bmatrix};$$

$\hat{\mathbf{\Sigma}} = \mathbf{e}'\mathbf{e}/T$ for $\mathbf{e} = (e_1, \ldots, e_M)$ and $e_j = \mathbf{y}_j - \mathbf{Z}_j\mathbf{d}_j$, $j = 1, \ldots, M$; and
$\hat{\mathbf{\Phi}} = \mathbf{e}'^{0}\mathbf{e}^{0}/T$ for $\mathbf{e}^0 = (e_1^0, \ldots, e_M^0)$ and $e_j^0 = \mathbf{y}_j^0 - \mathbf{Z}_j^0\mathbf{d}_j$, $j = 1, \ldots, M$.