Simultaneous Equations Models

Michael Curran

1 Introduction

Simultaneous models describe an interrelated system of relationships. Consider the simple equation model:

$$Y = f(X, u)$$

where X is fixed or independent, which implies that

$$Cov(X_{ji}, u_i) = 0 \quad \forall i$$

i.e. X is exogenous. This describes *one way causation*; see graph in class.

One way causation will not hold in most econometric applications. Simply put, such an equation cannot exist in our interrelated system, which is dynamic – we are concerned with simultaneity. Note that zero correlation between X and u implies zero covariance between X and u.

'Strict exogeneity' refers to lack of correlation, which implies that any omitted variable (subsumed) under u cannot be correlated with X; else our assumption does not hold. With measurement errors (random), X will be stochastic.

Without lagged variables, there can be no dynamics. Lagged variables of Y on the right hand side are determined by u in the last period (Y is stochastic). As shown in the omitted variable case, we end up with a problem a lot worse than autocorrelation or heteroscedasticity, namely OLS will be biased and inconsistent (i.e. the bias does not go to zero as we increase the sample size). With endogenous X, we get *two way causation*, i.e. joint dependency so $Cov(X_{ji}, u_i) \neq 0$, which violates the Classical assumptions; see graph in class.

2 Example of Simultaneity

2.1 Competitive Market

Let us now look at an example from micro, say the competive market where $P = f(Q)$ ad $Q = g(P)$; so, Q affects P and P affects Q. Suppose

$$\begin{aligned} Q_D &= \alpha + \beta P + u \\ Q_S &= \gamma + \delta P + v \\ Q_D &= Q_S \end{aligned} \tag{1}$$

where v is the supply shock (disturbance) and u is the demand shock (disturbance). There are three endogenous variables Q_D, Q_S and P. Equation (1) is the market clearing condition, an identity. It follows that with one unknown P:

$$\begin{aligned} \alpha + \beta P + u &= \gamma + \delta P + v \\ (\beta - \delta)P &= \gamma - \alpha + v - u \\ P &= \frac{\gamma - \alpha}{\beta - \delta} + \frac{v}{\beta - \delta} - \frac{u}{\beta - \delta} \\ &= \pi + \frac{v - u}{\beta - \delta} \end{aligned}$$

where π is a constant. Note that $P = f(u)$ and $P = g(v)$ implies that $Cov(P, u) \neq 0$ and $Cov(P, v) \neq 0$, i.e. P is correlated with the disturbance term in the demand equation and also in the supply equation.

What if we were asked to calculate measures for these dependencies?

$$\begin{aligned} Cov(P, u) &= E(P - E(P))(u - E(u)) \\ &= E\left[\pi + \frac{v - u}{\beta - \delta} - \pi\right][u - 0] \\ &= E\left[\left(\frac{v - u}{\beta - \delta}\right) u\right] \\ &= E\left(\frac{vu}{\beta - \delta}\right) - E\left(\frac{u^2}{\beta - \delta}\right) \\ &= \frac{-\sigma_u^2}{\beta - \delta} \neq 0 \end{aligned}$$

Note that assuming v is independent of u, denoted $v \perp\!\!\!\perp u$ will mean that there is no correlation between the demand shock and the supply shock. With a typical product, $\beta < 0$, i.e. higher prices reduce quantity demanded and *vice-versa* and $\delta > 0$, i.e. higher prices increase the quantity supplied and *vice-versa*. So, $Cov(P, u) > 0$, i.e. a positive demand shock raises the price of a normal good; hence, price and demand shocks move in the same direction.[1]

[1] In your projects, qualify your findings by acknowledging endogeneity, simultaneity, etc.

2.2 Simple Macro Model

Looking at a simple Keynesian macro model of a more complex, interrelated system, we know that GNP Y is not exogenous, it is endogenous and so it is jointly determined with consumption C:

$$Y \equiv C + I + G + X - M$$

So $C = f(Y, \ldots)$ and $Y = g(C, \ldots) = C + I + G + X - M$. Therefore, GNP affects C and C affects GNP. Suppose that

$$\begin{aligned} C &= a_0 + a_1 Y + u \\ I &= b_0 + b_1 Y + b_2 r + v \\ Y &= C + I + G \end{aligned}$$

So C depends on u and Y depends on C so Y depends on u; also, since I depends on v and Y depends on I, Y depends on v. The explanatory variable Y is correlated with both the consumption disturbance u and the investment disturbance v, i.e. $Cov(Y, u) \neq 0$ and $Cov(Y, v) \neq 0$ and so the Classical assumptions are violated.

3 Structural Form of SEM

The previous examples provided an economic description of systems. We can label the variables according to the following types:

- *Endogenous*: jointly determined by the operation of the system (e.g. C and Y).
- *Exogenous*: not jointly determined by the operation of the system: forced into the system from outside / elsewhere.
- *Predetermined*: both exogenous (time / weather) and lagged endogenous variables (determined in previous periods).

We can label the following types of relationships:

- *Behavioural*: (most) e.g. supposedly explain the behaviour of economic agents (e.g. demand equations / production functions / consumption function / investment function / export function / wage equation).
- *Technological*: behavioural – production function.
- *Identities*: definition of national income – nothing to estimate (equilibrium condition).
- *Equilibrium condition*: as above.

Note that *completeness* means that the number of equations is equal to the number of endogenous variables, so that we can solve the SEM and the solution is unique.

3.1 Example

Recall the competitive market example:

$$Q_D = \alpha + \beta P + u \tag{2}$$
$$Q_S = \gamma + \delta P + v \tag{3}$$
$$Q_D = Q_S \tag{4}$$

Equations (2) & (3) are examples of *behavioural* equations while (4) is an equilibrium condition.

There are four structural parameters: $\alpha, \beta, \gamma, \delta$. Two particular ones are of interest, β and δ.

There is one *exogenous variable*: 1 (the intercept).

There are three *endogenous variables*: Q_D, Q_S and P.
Is the system complete? Yes, the system is complete because there are three equations and three endogenous variables.[2]

4 Simultaneity Bias

The central bank model of the Irish economy used to use OLS, which of course was biased.[3] Single equation methods to estimate SEM include OLS, Indirect Least Squares (ILS – in the case of just identification, defined shortly) and 2SLS (when just identified or over-identified). System methods include 3SLS and the very powerful Full Information Maximum Likelihood Estimation procedure.

5 Reduced Form of SEM

The reduced form of SEM is arrived by solving for and expressing endogenous variables in terms of predetermined variables. Endogenous variables are on the left-hand side and only predetermined are allowed on the right hand side.
In the micro example[4]

$$\begin{aligned} P &= \frac{\gamma - \alpha}{\beta - \delta} + \frac{v}{\beta - \delta} - \frac{u}{\beta - \delta} \\ &= \pi + \frac{v - u}{\beta - \delta} \\ &= \pi_1 + w_1 \end{aligned}$$

[2] Problem set 2 contains a further example – simple macro model.
[3] Problem set 2 asks you to derive simultaneity bias.
[4] Problem set 2 provides an example of derivation of the reduced form for a very simple macro model.

which is the reduced form price equation. We can similarly get

$$\begin{aligned} Q &= \frac{\beta\gamma - \alpha\delta}{\beta - \delta} + \frac{\beta v - \delta u}{\beta - \delta} \\ &= \pi_2 + w_2 \end{aligned}$$

Obviously, this would be rather tricky for instance with a macro model of the Irish economy in the central bank having more than 100 equations where we must solve the structural model to write endogenous variables in terms of predetermined variables. With all the effort, why would we need reduced form models? After all, reduced form parameters π's are complicated functions of structural parameters α, β, etc. and reduced form disturbances w's are complicated functions of structural parameters α, β, etc. and structural disturbances u, v. Sometimes we can interpret reduced form parameters as multipliers and sometimes we cannot. Since exogenous variables are usually under government control, we can use reduced form equations for forecasting too.

The final form of SEM for dynamic models expresses endogenous variables purely in terms of exogenous variables through a process of continuous subsitution. There is no difference in static models between reduced form and final form models. With dynamic models (reduced form could include lagged endogenous variables), there is a difference in terms of multipliers and parameters.

6 Identification

6.1 Introduction

Let us introduce this section with an example:

$$\begin{aligned} Q_D &= \alpha + \beta P + u \\ Q_S &= \gamma + \delta P + v \\ Q_D &= Q_S \end{aligned}$$

Identification concerns whether we can meaningfully estimate the structural parameters. If we can, then the equation is identifiable; otherwise, the equation is unidentifiable.

There are two ways of viewing the problem of identification of SEMs. One is the mathematical view and the other is the statistical view. We will look at each in turn.

With the mathematical view, consider the OLS estimation on the reduced form where we are interested in α, β, γ and δ. The reduced form is

$$\begin{aligned} P &= \pi_1 + w_1 \\ Q_D = Q_S &= \pi_2 + w_2 \end{aligned}$$

where π_1 and π_2 are means of P and Q, respectively:

$$\pi_2 = \frac{\gamma - \alpha}{\beta - \delta}$$
$$\pi_1 = \frac{\beta\gamma - \alpha\delta}{\beta - \delta}$$

We have 2 equations in four unknowns, so it is not possible to solve uniquely for four unknown structural parameters given 2 reduced form parameters, i.e. structural parameters are not identified. Therefore the supply and demand functions are unidentifiable. Generally, for a complete SEM, we use the following notation:

M: endogenous variables and equations.

K: predetermined variables.

So the model has up to $M^2 - M$ structural parameters on endogenous variables and up to MK structural parameters on predetermined variables but there can be only MK reduced form parameters (endogenous in tems of only predetermined variables), so MK equations relating structural and reduced form parameters. So, it is not possible in general to get back from a knowledge of reduced form to a knowledge of structural form parameters. If $M^2 - M > MK$ and we have that the number of structural parameters exceeds the number of reduced form parameters, then no meaningful estimation can be carried out since we cannot estimate more than MK parameters in structural form with only MK equations. We can solve the model, but the solution will not be unique.

With the statistical view, see the graphs in class. The problem is that statistically, the demand and supply functions look the same – both are linear functions of price and quantities and require the same data for estimation. The functions are indistinguishable statistically from any arbitrary function of P and Q, of which there is an infinite number. Let $Q_D = Q_S = Q$ and k, c be arbitrary reals:

$$\begin{aligned}
kQ &= k\alpha + k\beta P + ku \\
cQ &= c\gamma + c\delta P + cv \\
\therefore (k+c)Q &= k\alpha + c\gamma + (k\beta + c\delta)P + ku + cv \\
Q &= \frac{k\alpha + c\gamma}{k+c} + \left(\frac{k\beta + c\delta}{k+c}\right) P + \frac{ku + cv}{k+c} \\
Q &= A + BP + W
\end{aligned}$$

where A is the intercept, B is the slope and W is a disturbance term in the last equation, which is devoid of meaning and statistically the same form as the supply and demand equation, i.e. another linear function in P and Q. So, using data on P and Q, we are unsure if we are estimating demand function parameters α, β, supply function parameters γ, δ or some 'mongrel' equation's parameters A, B devoid of economic meaning. This is known as *observational equivalence.* There are an infinite number of functions that are observationally equivalent to the demand and supply function in our model (depending on the choice of values for c and k) and which have the same reduced form. We can estimate the reduced form, but when trying to get back to the structural form, we cannot know that there is a unique structural form. So, the problem of identification is the problem of ascertaining that we can, from a given sample, estimate properly any behavioural equations of interest.

6.2 Solution

(*Intuition.*)
The mathematical approach would suggest that we place restrictions on the structural form so that the number of structural form parameters is reduced to be equal to the number of reduced form parameters. The statistical approach would suggest that we need an equation to be stable while others vary so that we can identify the stable equation, e.g. stable demand, variable supply could identify the demand function.
(*Formal rules.*)
M: number of endogenous variables / equations in the model (complete model).
m: number of endogenous variables in equation of interest.
K: number of predetermined variables in model.
k: number of predetermined variables in equation of interest.

Order condition (necessary):

$$K - k \geq m - 1$$

i.e. the number of excluded predetermined variables is at least as many as the number of endoengous variables on the right-hand side of the structural form equation minus 1. Equivalently

$$M + K - (m + k) \geq M - 1$$

so the total number of excluded variables (endogenous and predetermined) is no less than the number of equations minus one.

Under-identification: inequality not satistifed (not identifiable) – no meaningful estimation is possible.
Just / exact identification: equality achieved – identification is possible (still need to check sufficient condition).
Over-identification: strictly greater than – over-identification is possible (still need to check sufficient condition).

6.2.1 Example

Let us consider the example

$$\begin{aligned}
&\text{Demand:} && Q_D = \alpha + \beta_1 P + \beta_2 Y + u \\
&\text{Supply:} && Q_S = \gamma + \delta P + v \\
&\text{Equilibrium:} && Q_D = Q_S
\end{aligned}$$

Rewrite this as

$$\begin{array}{l} Q_D \ +0 \ -\alpha - \beta_1 P - \beta_2 Y = u \\ \ 0 \ + Q_S - \gamma \ - \delta P \quad +0 \ = v \\ Q_D - Q_S + 0 \ \ +0 \quad\ \ +0 \ = 0 \end{array}$$

Note that $M = 3$ (endogenous variables are Q_D, Q_S and P) and $K = 2$ (exogenous variables are 1 and Y).
With the demand equation, $m = 2$ (Q_D, P) and $k = 2$ $(1, Y)$ so $K - k = 2 - 2 = 0$ and $m = 1 = 2 - 1 = 1$ so $K - k < m - 1$ and therefore the demand function is *not* identified.
With the supply equation, $m = 2$ (Q_S, P) and $k = 1$ (constant) so $K - k = 2 - 1 = 1$ and $m - 1 = 2 - 1 = 1$ so $K - k \geq m - 1$ and therefore the supply function *may* be identified and if it is then it will be *just* identified.

Rank condition (necessary and sufficient):

The *rank* condition is necessary and sufficient to be sure about identification (it is a stronger condition):

$$\rho(\Lambda) = M - 1$$

A verbal definition is tricky so it is easier to look at an example.

6.2.2 Example

For the demand function,

$$\Lambda_D = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and $\rho(\Lambda_D) = 1 \neq M - 1 = 2$ so the demand function is *not* identified, confirming the results from the order condition.

For the supply function,

$$\Lambda_S = \begin{bmatrix} 1 & -\beta_2 \\ 1 & 0 \end{bmatrix}$$

where $-\beta_2$ corresponds to the second parameter in Q_D that is missing from Q_S, the top left entry corresponds to the coefficient on the first variable in Q_D that is missing from Q_S and the bottom left entry corresponds to the coefficient on this variable in the identity. So the rank $\rho(\Lambda_S) = 2 = M - 1$. Therefore, the supply function is identified and given the order condition result, the supply function is 'just identified'.[5]

7 Estimation

7.1 Indirect Least Squares

Indirect Least Squares (ILS) is appropriate in the case of *exact* identification. ILS estimates structural form parameters from reduced form parameters and

[5] Further applications of identification rules are given in problem set 2.

works because the number of reduced form parameters, which we can estimate by OLS is the same as the number of structural form parameters, since we are in the exact identified case. Let us first look at the underidentified case. The structural form model is given by

$$\begin{aligned} &\textit{Demand:} && Q_D = \alpha + \beta_1 P + \beta_2 Y + u \\ &\textit{Supply:} && Q_S = \gamma + \delta P + v \\ &\textit{Equilibrium:} && Q_D = Q_S = Q \end{aligned}$$

There are two predetermined variables, 1 (dummy variable on intercept) and Y. The reduced form is given by

$$\begin{aligned} P &= \pi_{11} + \pi_{12} Y + w_1 \\ Q &= \pi_{21} + \pi_{22} Y + w_2 \end{aligned}$$

Each π is related to structural form parameters and only the supply function is (just) identified. Note that π_{ij} depends on $\alpha, \beta_1, \beta_2, \gamma, \delta$ for $i = 1, 2$ and $j = 1, 2$. There are four reduced form parameters but 5 structural form parameters to estimate. The problem is that the solution is non-unique!

Now let us look at the just identified case and consider the structural form

$$\begin{aligned} &\text{Demand:} && Q_D = \alpha + \beta_1 P + \beta_2 Y + u \\ &\text{Supply:} && Q_S = \gamma + \delta_1 P + \delta_2 W + v \\ &\text{Equilibrium:} && Q_D = Q_S = Q \end{aligned}$$

Here $K - k = m - 1$ so that the order condition implies that both demand and supply equations are identified; assume that rank conditions holds (they do). Note that the reduced form will be

$$\begin{aligned} P &= \pi_{11} + \pi_{12} Y + \pi_{13} W + w_1 \\ Q &= \pi_{21} + \pi_{22} Y + \pi_{23} W + w_2 \end{aligned}$$

so there are six reduced form parameters. Supply and demand will be just identified, which implies that the entire system will be just identified. The π_{ij} depend on $\alpha, \beta_1, \beta_2, \gamma, \delta_1, \delta_2$ for $i = 1, 2$ and $j = 1, 2, 3$. There are six reduced form parameters and six structural parameters to estimate so that a unique solution is possible.

Finally, let us look at how ILS breaks down under over-identification. Consider the following structural form

$$\begin{aligned} &\text{Demand:} && Q_D = \alpha + \beta_1 P + \beta_2 Y + u \\ &\text{Supply:} && Q_S = \gamma + \delta_1 P + \delta_2 W + \delta_3 C + v \\ &\text{Equilibrium:} && Q_D = Q_S = Q \end{aligned}$$

Letting W denote weather and C denote cost of fertiliser, observe that the demand equation is overidentified (omits W and C not relevant to demand).

The reduced form is given by

$$P = \pi_{11} + \pi_{12}Y + \pi_{13}W + \pi_{14}C + w_1$$
$$Q = \pi_{21} + \pi_{22}Y + \pi_{23}W + \pi_{24}C + w_2$$

Supply is 'just' identified while demand is 'over' identified. Note that π_{ij} depends on $\alpha, \beta_1, \beta_2, \gamma, \delta_1, \delta_2, \delta_3$ for $i = 1, 2$ and $j = 1, 2, 3, 4$. There are eight reduced form parameters and 7 structural parameters to estimate, so the problem is non-uniqueness.

Summarising, ILS is only applicable for the *exact* identified case.

7.2 2SLS

2SLS introduced by Henry Theil in 1953 allows estimation of just or over-identified SEMs.

Stage 1: Estimate reduced form equations for endogenous variables on the right hand side of structural form equations of interest using OLS and obtain reduced form estimates of these 'endos'.

Stage 2: Substitute for right hand side endogenous variables in the structural equation using reduced form estimates from stage 1 as proxies (instruments) and estimate structural form equation by OLS.

For example, consider the over-identified SEM used earlier:

$$\begin{aligned} &\text{Demand:} && Q_D = \alpha + \beta_1 P + \beta_2 Y + u \\ &\text{Supply:} && Q_S = \gamma + \delta_1 P + \gamma_2 W + \delta_3 C + v \\ &\text{Equilibrium:} && Q_D = Q_S = Q \end{aligned}$$

Estimate over-identified demand equations as follows:

1. Estimate reduced form equation: $P = \pi_{11} + \pi_{12}Y + \pi_{13}W + \pi_{14}C + w_1$ via OLS and obtain $\hat{P} = \hat{\pi}_{11} + \hat{\pi}_{12}Y + \hat{\pi}_{13}W + \hat{\pi}_{14}C$ where $\hat{P}$ is a function of only predetermined variables.

2. Estimate modified demand equations $Q_D = \alpha + \beta_1 \hat{P}_1 + \beta_2 Y + u$ via OLS. This produces 2SLS estimates of structural form parameters.

7.2.1 Statistical Properties of 2SLS

2SLS will be biased but consistent (sampling distribution collapses on true parameter). Estimates will be asymptotically normal and asymptotically efficient, which makes the procedure preferable to OLS since t tests and F tests will be valid asymptotically; by asymptotically efficient, we mean in terms of minimising the variance of the estimator so inference will be more precise since confidence intervals will be smaller – with inefficient estimators, confidence intervals are wider. Efficient single stage algorithms are used in practice, yet it is important to understand the logic of the pedagogy of 2SLS. Problem for larger systems: number of observations must exceed number of predetermined variables in models; this is an issue with large models such as the Brookings model (over 1000 equations) and models of the Irish economy (over 100 equations) especially since in macro, T is small. Note that 2SLS is the same as ILS in the just identified case and 2SLS can be implemented in Stata.