Econometrics Lecture 7: Simultaneous Equations Models: Identification, Estimation and Testing

R. G. Pierse

1 Introduction

So far this course has concentrated on models with a single dependent variable. The only exception was in lecture 3 where we looked at the seemingly unrelated regressors (*SURE*) model. However, that model was essentially a set of single equations related solely through the covariances between error terms. In this lecture we look at systems of equations determining several dependent variables jointly.

In economics we are often interested in the interaction of several equations, simultanously determining more than one variable. An example is the demand and supply model. Here we have a demand function

$$Q = \alpha_1 + \alpha_2 P + \alpha_3 Y + u_1 \tag{1.1}$$

where Q is the quantity demanded, P is the price, Y is income and u_1 is a disturbance term representing random shocks to demand. We expect that there is a negative relationship between price and quantity demanded so that α_2 should be negative. There is also a supply function

$$P = \beta_1 + \beta_2 Q + \beta_3 W + u_2 \tag{1.2}$$

relating price to the quantity supplied and an unspecified variable W with disturbance term u_2 representing shocks to supply. We expect a positive relationship between price and quantity supplied so that β_2 should be positive.

Jointly, the demand and supply equations determine price and quantity. Note that it is entirely arbitrary to write quantity as the dependent variable of the demand function and price as the dependent variable of the supply function. We could equally well have written price on the left-hand side of the demand function and quantity on the left-hand side of the supply function, or even written both

equations with the same variable on the left-hand side. The important thing is to distinguish the variables that are determined within the system (price and quantity) from the variables (Y and W) that appear in the equations but which are assumed to be determined outside the system.

However, it is very much an *assumption* as to which variables can be treated as being determined outside the system. In the context of demand and supply for a single product it might be reasonable to treat income as outside the system. In considering demand and supply at the aggregate level, this would not be reasonable.

1.1 Endogenous, Exogenous and Predetermined Variables

The variables that are determined within a simultaneous equations system are known as *endogenous* variables. Any other variables that appear in the system but are determined outside it are known as *exogenous* variables. In the example, price and quantity are both endogenous variables and Y and W are exogenous variables. Note that there are the same number of endogenous variables as equations. When this is the case we say that the system is *complete*. When there are less equations than endogenous variables then the system is *incomplete*. When there are more equations than endogenous variables, the system is said to be *overdetermined*. In this case, one or more equations is redundant and can be dropped.

The dependent variables in a system (those that appear on the left-hand side of an equation) are necessarily endogenous. Since they are functions of a random disturbance term, they are also random variables. This applies equally to *any* endogenous variable since it can always be written as the dependent variable of one of the system equations. Therefore all endogenous variables must be treated as random variables.

Furthermore, where endogenous variables appear on the right-hand side of an equation, *they will be correlated with the error term in that equation*. This follows from the simultaneity of the system which means that all equations are determined jointly. Consider a one unit positive shock to the disturbance term u_1 in the demand equation (1.1) leading to a one unit increase in quantity Q. This in turn leads to an increase of β_2 in the price variable P through the supply equation (1.2). Thus P is correlated with u_1 and this correlation is positive since β_2 is positive. Similarly, a positive unit supply shock to u_2 leads to an increase of α_2 on Q through the demand equation. Since α_2 is negative, it follows that Q is negatively correlated with u_2.

The exogenous variables in the system, by *assumption*, are independent of all current, past and future values of the error term. This assumption is known as *strict exogeneity*. It is important to stress, however, that this is just an assumption and one that it is possible to test. Consider the case of a dynamic simultaneous

system which includes lags of the endogenous variables, such as

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \alpha_4 Q_{t-1} + \alpha_5 P_{t-1} + u_{1t} \tag{1.3}$$

and

$$P_t = \beta_1 + \beta_2 Q_t + \beta_3 W_t + \beta_4 P_{t-1} + \beta_5 Q_{t-1} + u_{2t} \tag{1.4}$$

where subscript t denotes the time period. As long as the error terms u_1 and u_2 are not autocorrelated, lagged endogenous variables will be independent of all current or future values of the error terms. Variables that satisfy this condition are known as *predetermined*. Clearly, exogenous variables by assumption are also predetermined, so that the lagged endogenous variables together with all current and lagged exogenous variables form the set of *predetermined variables*.

Since the predetermined variables are independent of current and future values of the error term, regression of the endogenous variables on the predetermined variables alone satisfies the conditions of OLS

2 The Simultaneous Equations Model

The general linear simultaneous equations model with m equations can be written formally as

$$\mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}\mathbf{z}_t = \mathbf{u}_t, \quad t = 1, \cdots, T \tag{2.1}$$

where $\mathbf{y}_t$ is an $m \times 1$ vector of observations on the m current endogenous variables at period t, $\mathbf{z}_t$ is a $q \times 1$ vector of observations on the q predetermined variables, $\mathbf{u}_t$ is an $m \times 1$ vector of disturbances, $\mathbf{B}$ is a $m \times m$ square matrix of coefficients on the endogenous variables and $\mathbf{\Gamma}$ is an $m \times q$ matrix of coefficients on the predetermined variables. It is assumed that

$$\mathrm{E}(\mathbf{u}_t) = 0$$

and

$$\mathrm{var}(\mathbf{u}_t) = \mathbf{\Sigma}$$

where $\mathbf{\Sigma}$ is a positive definite matrix. Thus disturbances in different equations in the same time period t are allowed to be correlated. However, the disturbances are assumed *not* to be autocorrelated so that

$$\mathrm{cov}(\mathbf{u}_t, \mathbf{u}_s) = \mathrm{E}(\mathbf{u}_t \mathbf{u}_s') = \mathbf{0}, \quad \forall t \neq s.$$

The dynamic demand–supply model (1.3) and (1.4) can be rewritten in the

form of (2.1) as

$$\begin{bmatrix} 1 & -\alpha_2 \\ -\beta_2 & 1 \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} + \begin{bmatrix} -\alpha_1 & -\alpha_4 & -\alpha_5 & -\alpha_3 & 0 \\ -\beta_1 & -\beta_5 & -\beta_4 & 0 & -\beta_3 \end{bmatrix} \begin{bmatrix} 1 \\ Q_{t-1} \\ P_{t-1} \\ Y_t \\ W_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{2.2}$$

Equation (2.1) is known as the *structural form* of a simultaneous system. It corresponds to the behavioural equations of the economic model and the coefficient matrices $\mathbf{B}$ and $\mathbf{\Gamma}$ will typically contain zeros or other restrictions corresponding to assumptions in the economic model. For example, in the demand–supply model, the economic assumption is that variable W_t does not affect the demand function and that Y_t does not affect the supply function so that the matrix $\mathbf{\Gamma}$ contains two zeros.

Assuming that the matrix $\mathbf{B}$ in (2.1) is nonsingular, it is possible to pre-multiply through by $\mathbf{B}^{-1}$ giving

$$\mathbf{y}_t = -\mathbf{B}^{-1}\mathbf{\Gamma}\mathbf{z}_t + \mathbf{B}^{-1}\mathbf{u}_t \tag{2.3}$$

or

$$\mathbf{y}_t = \mathbf{\Pi}\mathbf{z}_t + \mathbf{v}_t. \tag{2.4}$$

where $\mathbf{\Pi} = -\mathbf{B}^{-1}\mathbf{\Gamma}$ and $\mathbf{v}_t = \mathbf{B}^{-1}\mathbf{u}_t$. This is known as the *reduced form* of the system and it relates the endogenous variables $\mathbf{y}$ solely to the predetermined variables $\mathbf{z}$, removing the simultaneity in the structural form. In this formulation, the economic assumptions in the model are less obvious but are embodied in the restriction that $\mathbf{\Pi} = -\mathbf{B}^{-1}\mathbf{\Gamma}$. The reduced form disturbances $\mathbf{v}_t$ no longer correspond to disturbances on particular behavioural equations.

3 Identification in Simultaneous Equations

Consider pre-multiplying the simultaneous equation system (2.1) by the $m \times m$ nonsingular matrix $\mathbf{F}$ to give

$$\mathbf{FBy}_t + \mathbf{F\Gamma z}_t = \mathbf{Fu}_t. \tag{3.1}$$

The reduced form of this transformed model is

$$\begin{aligned} \mathbf{y}_t &= -(\mathbf{FB})^{-1}\mathbf{F\Gamma z}_t + (\mathbf{FB})^{-1}\mathbf{Fu}_t \\ &= -\mathbf{B}^{-1}\mathbf{F}^{-1}\mathbf{F\Gamma}z_t + \mathbf{B}^{-1}\mathbf{F}^{-1}\mathbf{F}u_t \\ &= -\mathbf{B}^{-1}\mathbf{\Gamma z}_t + \mathbf{B}^{-1}\mathbf{u}_t \end{aligned}$$

which is identical to the reduced form (2.3) of the original model (2.1). Considering the observations $\mathbf{y}_t$ as having being generated by the equation (2.3), there is clearly a problem in determining whether the structural parameters are given by (2.1) or by (3.1).

In general it is impossible to estimate the parameters of a simultaneous equation system unless there are sufficient restrictions on the elements of $\mathbf{B}$ and $\mathbf{\Gamma}$ (or $\mathbf{\Sigma}$) to uniquely identify the parameters of the model. This is known as the *problem of identification*. Identification conditions can therefore be viewed as the conditions under which it is possible to recover the structural form parameters from the reduced form.

3.1 Rank and Order Conditions for Identification

Identification can be ensured by restrictions involving any of the structural form parameters $\mathbf{B}$, $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$. Consequently, the conditions for identification in the most general case are complicated to state. Here we consider only the most common form of restriction, namely *zero restrictions* on $\mathbf{B}$ and $\mathbf{\Gamma}$. These correspond to the exclusion of some variables from particular equations and so these restrictions are also known as *exclusion restrictions*.

It is possible that some equations in a model may be identified while others are not. A model is identified only if *all* equations in the model are identified. Therefore, identification needs to be checked separately for each equation in a model .

Let the number of variables *excluded* from the jth equation by denoted by r_j. Then the *order condition* for identification of the jth equation by exclusion restrictions is that r_j is greater than or equal to $m-1$. This condition is *necessary but not sufficient*. When the number of exclusion restrictions is strictly greater than $m-1$ then the equation is said to be *over-identified* whereas when $r_j = m-1$ then the equation is said to be *exactly identified*. The order condition has the advantage of being very easy to check.

A condition which is both *necessary and sufficient* is the *rank condition* for identification of the jth equation. This considers the rank of the matrix formed from the columns of the matrices $\mathbf{B}$ and $\mathbf{\Gamma}$, corresponding to the *excluded* variables in the jth equation, but excluding the jth row. This matrix will be of dimension $(m-1) \times r_j$. The rank condition states that the rank of this matrix must be equal to $m-1$.

3.2 Some examples

Consider the two equation dynamic demand–supply model (2.2). In each equation, a single variable is excluded (W_t in the demand equation and Y_t in the second).

Thus both equations satisfy the order condition for identification.

Consider now the rank condition. for the first equation, the matrix to be considered is the 1×1 matrix

$$[-\beta_3]\,.$$

Clearly this will have rank $m - 1 = 1$ as long as $\beta_3 \neq 0$. If $\beta_3 = 0$ however, then the equation will not be identified. Similarly, the rank condition for the second equation is that the 1×1 matrix

$$[-\alpha_3]$$

has rank $m - 1 = 1$. Again, this will be the case except if $\alpha_3 = 0$.

As a second example, consider a four equation *IS-LM* model based on Stewart (1991) p 253

$$\begin{aligned} C_t &= -\gamma_{11} - \alpha_{14} Y_t + u_{1t} \\ I_t &= -\gamma_{21} - \alpha_{23} R_t - \alpha_{24} Y_t + u_{2t} \\ R_t &= -\alpha_{34} Y_t - \gamma_{32} M_t + u_{3t} \\ Y_t &= C_t + I_t + Z_t \end{aligned}$$

where C_t is consumption, I_t is investment, R_t is the rate of interest, Y_t is income, M_t is the money stock and Z_t is autonomous expenditure. This model can be rewitten in the form (2.1) as

$$\begin{bmatrix} 1 & 0 & 0 & \alpha_{14} \\ 0 & 1 & \alpha_{23} & \alpha_{24} \\ 0 & 0 & 1 & \alpha_{34} \\ -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ I_t \\ R_t \\ Y_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{21} & 0 & 0 \\ 0 & \gamma_{32} & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ M_t \\ Z_t \end{bmatrix} = \mathbf{u}_t.$$

Consider the first equation. The number of excluded regressors is $4 > m - 1 = 3$ so that the order condition is satisfied. The rank condition is based on the rank of the matrix

$$\begin{bmatrix} 1 & \alpha_{23} & 0 & 0 \\ 0 & 1 & \gamma_{32} & 0 \\ -1 & 0 & 0 & -1 \end{bmatrix}$$

which has rank 3 even if both parameters α_{23} and γ_{32} are equal to zero. This equation is over-identified.

For the second equation, the number of excluded regressors is 3 and the rank condition is based on the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \gamma_{32} & 0 \\ -1 & 0 & -1 \end{bmatrix}$$

which has rank 3 as long as the parameter γ_{32} is not equal to zero. In this case, the equation is exactly identified. If however, $\gamma_{32} = 0$ then it is clear that the rank of this matrix is only 2 in which case the equation is not identified.

The third equation has four exclusion restrictions and the rank condition depends on the matrix

$$\begin{bmatrix} 1 & 0 & \gamma_{11} & 0 \\ 0 & 1 & \gamma_{21} & 0 \\ -1 & -1 & 0 & -1 \end{bmatrix}$$

which has rank 3 so that the rank condition is satisfied. the equation is over-identified.

Finally, the fourth equation has 3 exclusion restrictions and the rank condition depends on the matrix

$$\begin{bmatrix} 0 & \gamma_{11} & 0 \\ \alpha_{23} & \gamma_{21} & 0 \\ 1 & 0 & \gamma_{32} \end{bmatrix}$$

which has rank 3 as long as all coefficients are non-zero. Then the equation is exactly identified. However, if $\gamma_{32} = 0$ then the matrix only has rank 2 so that the equation is then not identified.

In practice, it can not be known whether any of the parameters have true value zero. Thus in specifying a model, only the order condition can be guaranteed to hold by construction. It is still possible that the model may not be identified through failure of the rank condition. If this happens, then estimation will break down.

4 Estimation: Single Equation Methods

In the estimation of simultaneous equations systems there are two basic approaches. The first is to consider the estimation of each equation in isolation. This approach ignores the information about the covariances between the equations given by the covariance matrix $\mathbf{\Sigma}$ and information about the exclusion restrictions on all other equations. Consequently, this approach is called a *limited information* approach. The second approach estimates the complete system jointly, taking into account all the identifying restrictions in the model and the covariance information provided by $\mathbf{\Sigma}$. Since this approach uses all available information, it is known as the *full information* approach.

When the model is correctly specified, then full information estimation is more efficient than limited information estimation. However, because of its system nature, any mistakes made in the specification of one equation will affect the estimates of all the equations. Consequently, if there is uncertainty about the

specification of the system, there may be an argument in favour of using limited information methods. They also have the advantage of being computationally much cheaper to implement.

Consider the estimation of the jth equation from the system (2.1). This equation can be written as

$$y_{jt} = \mathbf{x}_j'\boldsymbol{\beta}_j + u_{jt}, \quad t = 1, \cdots, T$$

where the $k \times 1$ vector $\mathbf{x}_j$ represents all the variables in $\mathbf{y}_t$ and $\mathbf{z}_t$ that have unrestricted coefficients. Note that the number of regressors k is equal to $m-1+q-r_j$ and the order condition for identification ensures that $k \leq q$.

Stacking all the T observations together we can write

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j$$

where $\mathbf{y}_j$ is a $T \times 1$ vector, $\mathbf{X}_j$ is a $T \times k$ matrix, and $\mathbf{u}_j$ is a $T \times 1$ vector satisfying

$$\mathrm{E}(\mathbf{u}_j) = 0 \quad \text{and} \quad \mathrm{var}(\mathbf{u}_j) = \sigma_{jj}\mathbf{I}_T.$$

5 Indirect Least Squares

Consider the two equation demand and supply model

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + u_1$$

and

$$P_t = \beta_1 + \beta_2 Q_t + \beta_3 W_t + \beta_4 R_t + u_2$$

which differs from the model (1.1) and (1.2) in that the supply equation now includes an extra exogenous variable R. In matrix form the model can be written as

$$\begin{bmatrix} 1 & -\alpha_2 \\ -\beta_2 & 1 \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} + \begin{bmatrix} -\alpha_1 & -\alpha_3 & 0 & 0 \\ -\beta_1 & 0 & -\beta_3 & -\beta_4 \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ W_t \\ R_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

and by inspection it can be verified that the first equation is overidentified while the second equation is exactly identified.

The reduced form of the system is given by

$$\begin{aligned}\begin{bmatrix} Q_t \\ P_t \end{bmatrix} &= \begin{bmatrix} 1 & -\alpha_2 \\ -\beta_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 & \alpha_3 & 0 & 0 \\ \beta_1 & 0 & \beta_3 & \beta_4 \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ W_t \\ R_t \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \\ &= \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \end{bmatrix} \begin{bmatrix} 1 \\ Y_t \\ W_t \\ R_t \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}\end{aligned}$$

where π_{ij} represent the reduced form parameters.

Noting that

$$\begin{bmatrix} 1 & -\alpha_2 \\ -\beta_2 & 1 \end{bmatrix}^{-1} = \frac{1}{(1-\alpha_2\beta_2)} \begin{bmatrix} 1 & \alpha_2 \\ \beta_2 & 1 \end{bmatrix}$$

we have the parameter correspondence

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \end{bmatrix} = \frac{1}{(1-\alpha_2\beta_2)} \begin{bmatrix} \alpha_1+\alpha_2\beta_1 & \alpha_3 & \alpha_2\beta_3 & \alpha_2\beta_4 \\ \beta_1+\alpha_1\beta_2 & \alpha_3\beta_2 & \beta_3 & \beta_4 \end{bmatrix}.$$

Can the structural form parameters α_i and β_i be uniquely recovered from the reduced form parameters π_{ij}? For the parameters of the second equation, the answer is yes with the results

$$\beta_1 = \pi_{21} - \pi_{11}\pi_{22}/\pi_{12} \; ; \; \beta_2 = \pi_{22}/\pi_{12}$$

$$\beta_3 = \pi_{23} - \pi_{13}\pi_{22}/\pi_{12} \; ; \; \beta_4 = \pi_{24} - \pi_{14}\pi_{22}/\pi_{12}.$$

This unique correspondence is the result of the fact that the second equation is *exactly identified*.

For the first equation, there is no unique way to recover all the structural parameters since, for example

$$\alpha_2 = \pi_{13}/\pi_{23} = \pi_{14}/\pi_{24}$$

so that there are two alternative ways of defining α_2. This is the meaning of the fact that the first equation is *overidentified*.

For an exactly identified equation, one possible way of estimating the parameters is to estimate the unrestricted reduced form parameters by *OLS* on the reduced form equations

$$\mathbf{y}_j = \mathbf{Z}_j\pi_j + \mathbf{u}_j, \quad j = 1, \cdots, m$$

where $\mathbf{Z}_j$ is the matrix of predetermined variables appearing in the jth equation. The structural parameters can then then recovered from $\widehat{\pi}_j$ using the relationships

$$\widehat{\beta}_1 = \widehat{\pi}_{21} - \widehat{\pi}_{11}\widehat{\pi}_{22}/\widehat{\pi}_{12} \; ; \; \widehat{\beta}_2 = \widehat{\pi}_{22}/\widehat{\pi}_{12}$$

$$\widehat{\beta}_3 = \widehat{\pi}_{23} - \widehat{\pi}_{13}\widehat{\pi}_{22}/\widehat{\pi}_{12} \; ; \; \widehat{\beta}_4 = \widehat{\pi}_{24} - \widehat{\pi}_{14}\widehat{\pi}_{22}/\widehat{\pi}_{12} \; .$$

This is known as *indirect least squares*. It can be shown that the indirect least squares (ILS) estimator is an *instrumental variables* (IV) estimator.

5.1 The Method of Instrumental Variables

We have seen that the endogenous variables on the right-hand side of an equation from a simultaneous equation system are correlated with the disturbance term. This violates the fundamental assumption of the regression model that regressors and error term are uncorrelated, or formally that

$$\mathrm{E}(\mathbf{X}'\mathbf{u}) = \mathbf{0}$$

or, in large samples, that

$$\operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{n} = \mathbf{0}.$$

Consequently, since this assumption is violated, OLS estimates will be both *biased* and *inconsistent*.

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

where there are k regressors and

$$\mathrm{E}(\mathbf{X}'\mathbf{u}) \neq \mathbf{0}.$$

Suppose that we can find a set of k variables $\mathbf{W}$ satisfying the condition that

$$\mathrm{E}(\mathbf{W}'\mathbf{u}) = \mathbf{0}$$

and

$$\operatorname{plim} \frac{\mathbf{W}'\mathbf{u}}{n} = \mathbf{0}$$

where the variables $\mathbf{W}$ are correlated with $\mathbf{X}$, then the variables $\mathbf{W}$ are called *instruments* for $\mathbf{X}$ and the estimator

$$\widetilde{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{y} \tag{5.1}$$

is called an *instrumental variables* or IV estimator.

Note that

$$\widetilde{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{W}'\mathbf{X})^{-1}\mathbf{W}'\mathbf{u}$$

so that, as long as

$$\text{plim}\left(\frac{\mathbf{W}'\mathbf{X}}{n}\right) = \mathbf{Q}, \quad \mathbf{0} < \mathbf{Q} < \infty.$$

then

$$\text{plim}(\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \text{plim}\left(\frac{\mathbf{W}'\mathbf{X}}{n}\right)^{-1} \text{plim}\left(\frac{\mathbf{W}'\mathbf{u}}{n}\right) = \boldsymbol{\beta}$$

so that the *IV* estimator is *consistent*.

Sometimes there may be more instruments $\mathbf{W}$ available than there are regressors $\mathbf{X}$. In this case we can take a subset of any k columns from $\mathbf{W}$. This will be a valid set of instruments. However, the higher the correlation between the instrument set and the regressors $\mathbf{X}$ the better, and so a better strategy is to use the $T \times k$ linear combination of the instruments

$$\widehat{\mathbf{X}} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X}.$$

Note that $\widehat{\mathbf{X}}$ can be interpreted as the fitted values from a regression of $\mathbf{X}$ on the set of instruments $\mathbf{W}$. This is the linear combination that maximises the correlation with $\mathbf{X}$ and is hence the best combination of instruments.

The *IV* estimator is then given by

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} = \left(\widehat{\mathbf{X}}'\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \end{aligned} \tag{5.2}$$

This estimator is known as the *Two Stage Least Squares* or *2SLS* estimator. This can be regarded as a *generalised IV estimator* or *GIVE* estimator since, when the number of instruments and regressors is the same, it collapses to the standard form (5.1).

In general, the problem with *IV* estimation is to find a set of variables that satisfy the conditions for being valid instruments.

5.2 Two Stage Least Squares Estimator

In the simultaneous equations context of estimating the equation

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j$$

there is a ready set of valid instruments available. This is the $T \times q$ matrix of predetermined variables $\mathbf{Z}$ from the system (2.1) defined by

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_T' \end{bmatrix}$$

These variables comprise lagged dependent variables plus exogenous variables that by assumption satisfy the condition that

$$\mathrm{E}(\mathbf{Z}'\mathbf{u}_j) = \mathbf{0}, \quad j = 1, \cdots, m$$

and

$$\operatorname{plim} \frac{\mathbf{Z}'\mathbf{u}_j}{T} = \mathbf{0}.$$

In general, $q > k_j$ so that there are more instruments than regressors.

Hence the application of *IV* leads to the *2SLS* estimator

$$\widetilde{\boldsymbol{\beta}}_j = (\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_j \tag{5.3}$$

This estimator can be interpreted as a two stage estimation procedure. In the first stage, the regressors $\mathbf{X}_j$ are regressed on the set of instruments $\mathbf{Z}$. This is an estimation of the parameters of the jth equation of the reduced form (2.3). Then $\mathbf{y}_j$ is regressed on the fitted values from this regression

$$\widehat{\mathbf{X}}_j = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_j = \mathbf{P}_z\mathbf{X}_j$$

where $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is an idempotent projection matrix, to give the second stage estimates

$$\widetilde{\boldsymbol{\beta}}_j = (\widehat{\mathbf{X}}_j'\widehat{\mathbf{X}}_j)^{-1}\widehat{\mathbf{X}}_j'\mathbf{y}_j$$

which is formally identical to (5.3). The $\widehat{\mathbf{X}}_j$ are sometimes called *constructed regressors*. They are defined so that, by construction, they are not correlated with the error term $\mathbf{u}_j$. In this interpretation of *2SLS*, the original regressors are *replaced* by the constructed regressors in the second stage of estimation. Although this was the original rationale for the *2SLS* estimator that was invented by Theil (1958), it is generally more helpful to think of *2SLS* as the *instrumental variables estimator*

$$\widetilde{\boldsymbol{\beta}}_j = (\widehat{\mathbf{X}}_j'\mathbf{X}_j)^{-1}\widehat{\mathbf{X}}_j'\mathbf{y}_j$$

where the original regressors are not *replaced* but are *instrumented* by $\widehat{\mathbf{X}}_j$, even though formally, the two expressions are identical in this case.

The variance of the 2SLS estimator is given by

$$\text{var}(\widetilde{\boldsymbol{\beta}}_j) = \sigma_{jj}(\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_j)^{-1}$$

and a consistent estimator of σ_{jj} can be obtained from the *2SLS* residuals

$$\mathbf{e} = \mathbf{y}_j - \mathbf{X}_j\widetilde{\boldsymbol{\beta}}_j \tag{5.4}$$

using the expression

$$\widetilde{\sigma}_{jj} = \frac{\mathbf{e}'\mathbf{e}}{T-k}.$$

The degrees of freedom correction in the denominator of this expression does not affect the consistency. Note that the *2SLS* residuals (5.4) are *not the same* as the residuals obtained from the second stage of the two stage estimation procedure, which would be given by

$$\widehat{\mathbf{e}} = \mathbf{y}_j - \widehat{\mathbf{X}}_j\widetilde{\boldsymbol{\beta}}_j.$$

This is one reason why it can be unhelpful to think of *2SLS* as a two stage estimation procedure, since it would lead to an incorrect expression for the estimated error variance $\widetilde{\sigma}_{jj}$.

In the special case of an exactly identified equation where $q = k_j$, then the estimator (5.3) collapses to

$$\widetilde{\boldsymbol{\beta}}_j = (\mathbf{Z}'\mathbf{X}_j^{-1}\mathbf{Z}'\mathbf{y}_j.$$

This is equivalent to the *indirect least squares* (*ILS*) estimator.

On the assumption that the plims of the moment matrices

$$\text{plim}\,\frac{\mathbf{Z}'\mathbf{Z}}{T} = \mathbf{Q}_{ZZ} \quad \text{and} \quad \text{plim}\,\frac{\mathbf{X}_j'\mathbf{Z}}{T} = \mathbf{Q}_{XZ}$$

exist and are finite, it can be shown that the *2SLS* estimator is consistent since

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_j &= (\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j)\\ &= \boldsymbol{\beta}_j + (\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}_j\end{aligned}$$

and

$$\begin{aligned}\text{plim}\,\widetilde{\boldsymbol{\beta}}_j &= \boldsymbol{\beta}_j + \left(\text{plim}\,\frac{\mathbf{X}_j'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\frac{\mathbf{Z}'\mathbf{X}_j}{T}\right)^{-1}\text{plim}\left(\frac{\mathbf{X}_j'\mathbf{Z}}{T}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\frac{\mathbf{Z}'\mathbf{u}_j}{T}\right)\\ &= \boldsymbol{\beta}_j.\end{aligned}$$

Finally, we can apply a central limit theorem

$$T^{-\frac{1}{2}}\mathbf{Z}'\mathbf{u}_j \sim_a N(0, \sigma_{jj}\mathbf{Q}_{ZZ})$$

to show that $\widetilde{\boldsymbol{\beta}}_j$ is asymptotically normally distributed with

$$\sqrt{T}(\widetilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) \sim_a N(0, \sigma_{jj}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX})^{-1}). \tag{5.5}$$

This result can be used as the basis for hypothesis testing in the simultaneous equations model, using the same large sample principles as those considered for the classical model.

5.3 Limited Information Maximum Likelihood

The maximum likelihood principle can also be used to construct an estimator in the simultaneous equations system using limited information. This estimator is known as the *LIML* estimator. The details will not be presented here. However, it can be shown that the *LIML* estimator is asymptotically equivalent to the *2SLS* estimator and so has the same asymptotic distribution (5.5).

5.4 Testing Overidentifying Restrictions

When an equation is overidentified, there are more restrictions on its parameters than are necessary to identify it. Therefore it is possible to construct a test of these extra restrictions. Such a test was developed by Sargan (1964). It is based on the quantity

$$\mathbf{e}'\mathbf{P}_z\mathbf{e}$$

where $\mathbf{e} = \mathbf{y}_j - \mathbf{X}_j\widetilde{\boldsymbol{\beta}}_j$ is the vector of equation residuals (*2SLS* or *LIML*) and $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z}^{-1})\mathbf{Z}'$ is the instrument projection matrix. The test for overidentifying restrictions is given by

$$\frac{\mathbf{e}'\mathbf{P}_z\mathbf{e}}{\widetilde{\sigma}_{jj}} \sim_a \chi^2_{q-k_j}$$

and is asymptotically distributed as a chi-squared variate with degrees of freedom $q - k_j$ which is the degree of overidentification of the equation. This test is sometimes called the Sargan *validity of instruments* test.

5.5 Testing Exogeneity

Consider the equation

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j = \mathbf{Y}_1\alpha_j + \mathbf{Z}_1\gamma_j + \mathbf{u}_j \tag{5.6}$$

where the regressors $\mathbf{X}_j$ have been partitioned into the set of p current endogenous variables $\mathbf{Y}_1$ and the $k_j - p$ predetermined variables $\mathbf{Z}_1$ appearing in the jth

equation. The special estimation methods for simultaneous equations are needed because the assumption of exogeneity that

$$\mathrm{E}(\mathbf{Y}_1'\mathbf{u}_j) = \mathbf{0}$$

will not be expected to hold. If however, this condition did hold, then *OLS* estimation would be appropriate. It is possible to devise a test of the exogeneity of the regressors $\mathbf{Y}_1$. This is known as the Wu-Hausman test for exogeneity and was developed independently by Wu (1973) and Hausman (1978). The test is computed by first estimating the equation (5.6) by *OLS* and computing the residual sum of squares S_0. Then the equation is re-estimated by *OLS* including as additional regressors the instrumental variables

$$\widehat{\mathbf{Y}}_1 = \mathbf{Z}(\mathbf{Z}'\mathbf{Z}^{-1})\mathbf{Z}'\mathbf{Y}_1$$

to give residual sum of squares S_1. then the test statistic is given by

$$\frac{S_0 - S_1}{\widehat{\sigma}^2} \sim_a \chi^2_p$$

where $\widehat{\sigma}^2$ is the estimated error variance from the first estimation. The test statistic is asymptotically distributed as chi-squared with degrees of freedom equal to p, the number of columns of $\mathbf{Y}_1$ and so the number of potentially endogenous regressors. Rejection of the null hypothesis of exogeneity would show that *IV* estimation is needed.

6 Estimation: System Methods

System methods of estimation in simultaneous equation systems use all the information in the model to estimate the parameters of all equations jointly. They will be more efficient than single equation methods but are liable to the problem that misspecification of any one equation will affect the estimates in all the equations. These methods are more costly in computational terms than single equation methods and may not be feasible when the instrument set is large. However, many econometric packages such as *TSP*, *E-Views* and *Pc-FIML* offer these estimation techniques.

6.1 Three Stage Least Squares

Consider stacking the all the equations of the model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{u}_j, \quad j = 1, \cdots, m$$

to form the stacked equation

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{bmatrix} \tag{6.1}$$

or

$$\overline{\mathbf{y}} = \overline{\mathbf{X}}\boldsymbol{\beta} + \overline{\mathbf{u}}$$

where

$$\text{var}(\overline{\mathbf{u}}) = \begin{bmatrix} \sigma_{11}\mathbf{I}_T & \cdots & \sigma_{1m}\mathbf{I}_T \\ \vdots & \ddots & \vdots \\ \sigma_{m1}\mathbf{I}_T & & \sigma_{mm}\mathbf{I}_T \end{bmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T \tag{6.2}$$

The stacked system (6.1) has a non-constant variance covariance matrix (6.2). It also has the problem that the regressors $\overline{\mathbf{X}}$ are correlated with the error term $\overline{\mathbf{u}}$.

The solution is to apply a combination of instrumental variables estimation and generalised least squares to correct these two problems. The instrument set is the matrix

$$\begin{bmatrix} \widehat{\mathbf{X}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\mathbf{X}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{P}_z\mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_z\mathbf{X}_m \end{bmatrix} = (\mathbf{I}_m \otimes \mathbf{P}_z)\overline{\mathbf{X}} \tag{6.3}$$

where $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the instrument projection matrix.

Applying both *GLS* using (6.2) and *IV* using instrument set (6.3) results in the *Three Stage Least Squares (3SLS) Estimator* of Zellner and Theil (1962)

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= \left(\overline{\mathbf{X}}'(\mathbf{I}_m \otimes \mathbf{P}_z)'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}(\mathbf{I}_m \otimes \mathbf{P}_z)\overline{\mathbf{X}}\right)^{-1} \overline{\mathbf{X}}'(\mathbf{I}_m \otimes \mathbf{P}_z)'(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)^{-1}\overline{\mathbf{y}} \\ &= \left(\overline{\mathbf{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{P}_z)\overline{\mathbf{X}}\right)^{-1} \overline{\mathbf{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{P}_z)\overline{\mathbf{y}}. \end{aligned} \tag{6.4}$$

In practice the unknown covariance matrix $\boldsymbol{\Sigma}$ needs to be replaced by a consistent estimator $\widehat{\boldsymbol{\Sigma}}$. Such an estimator can be based on the expression

$$\widehat{\sigma}_{ij} = \frac{\mathbf{e}_i'\mathbf{e}_j}{T}$$

where $\mathbf{e}_j$ is the vector of residuals

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{X}_j\widetilde{\boldsymbol{\beta}}_j$$

from *2SLS* regression on the jth equation.

Note that in the case where $\mathbf{\Sigma}$ is a diagonal matrix, *3SLS* estimation is identical to *2SLS* estimation on each equation. This is also the case if every equation is exactly identified. The reason is that in these cases there is no informational gain in considering all the equations together.

The *3SLS* estimator is *consistent* and *asymptotically efficient* in the class of full information models with

$$\sqrt{T}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim_a N\left(0, \left(\text{plim}\,\frac{1}{T}\overline{\mathbf{X}}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{P}_z)\overline{\mathbf{X}}\right)^{-1}\right). \tag{6.5}$$

6.2 Full Information Maximum Likelihood

The maximum likelihood principle can also be used to construct an estimator in the simultaneous equations system using full information. This estimator is known as the *Full Information Maximum Likelihood* or *FIML* estimator.

Consider again the general linear simultaneous equations model with m equations

$$\mathbf{B}\mathbf{y}_t + \mathbf{\Gamma}\mathbf{z}_t = \mathbf{u}_t, \quad t = 1, \cdots, T$$

where it is now assumed that $\mathbf{u}_t$ is distributed independently normally as

$$\mathbf{u}_t \sim IN(\mathbf{0}, \mathbf{\Sigma}).$$

The probability distribution function for $\mathbf{u}_t$ is given by

$$f(\mathbf{u}_t) = (2\pi)^{-m/2} |\mathbf{\Sigma}|^{-1/2} \exp(-\frac{1}{2}\mathbf{u}_t'\mathbf{\Sigma}^{-1}\mathbf{u}_t)$$

and the likelihood function for $\mathbf{y}_t$ is given by

$$\begin{aligned} f(\mathbf{y}_t) &= \left\|\frac{\partial \mathbf{u}_t}{\partial \mathbf{y}_t}\right\| f(\mathbf{u}_t) = \|\mathbf{B}\| f(\mathbf{u}_t) \\ &= (2\pi)^{-m/2} |\mathbf{\Sigma}|^{-1/2} \|\mathbf{B}\| \exp(-\frac{1}{2}\mathbf{u}_t'\mathbf{\Sigma}^{-1}\mathbf{u}_t). \end{aligned}$$

where $\|\|$ denotes the absolute value of the determinant. $\|\partial \mathbf{u}_t/\partial \mathbf{y}_t\|$ is called the *Jacobian* of the transformation from $\mathbf{u}_t$ to $\mathbf{y}_t$.

The likelihood of the whole sample is therefore given by

$$\begin{aligned} L(\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Sigma}; \mathbf{y}, \mathbf{z}) &= \prod_{t=1}^{T} f(\mathbf{y}_t) \\ &= (2\pi)^{-mT/2} |\mathbf{\Sigma}|^{-T/2} \|\mathbf{B}\|^T \exp(-\frac{1}{2}\sum_{t=1}^{T} \mathbf{u}_t'\mathbf{\Sigma}^{-1}\mathbf{u}_t) \end{aligned}$$

The *FIML* estimator is derived by maximising this likelihood function numerically with respect to the unknown parameters $\mathbf{B}$, $\mathbf{\Gamma}$, and $\mathbf{\Sigma}$, taking into account all the identifying restrictions imposed on these matrices.

It can be shown that the *FIML* estimator is asymptotically equivalent to the *3SLS* estimator and so has the same asymptotic distribution (6.5). *FIML* has the advantage over *3SLS* that all parameters are estimated jointly whereas in *3SLS*, $\mathbf{\Sigma}$ is pre-estimated from *2SLS* residuals. On the other hand, it requires iterative numerical optimisation and so is computationally more costly than *3SLS*.

References

[1] Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica*, 46, 1251–71.

[2] Sargan, J.D. (1964), 'Wages and prices in the United Kingdom: a study in econometric methodology', in D.F. Hendry and K.F. Wallis (eds.), *Econometrics and Quantitative Economics*, Basil Blackwell, Oxford, 1984.

[3] Stewart, J. (1991), *Econometrics*, Philip Allan, Hemel Hempstead.

[4] Theil, H. (1958), *Economic Forecasts and Policy*, North-Holland, Amsterdam.

[5] Wu, D-M. (1973), 'Alternative tests of independence between stochastic regressors and disturbances', *Econometrica*, 41, 733–750.

[6] Zellner, A. and H. Theil (1962), 'Three-stage least squares: simultaneous estimation of simultaneous equations', *Econometrica*, 30, 54–78.