UNIT 6 INTRODUCTION TO SIMULTANEOUS EQUATIONS MODELS*

Structure

6.0 Objectives

6.1 Introduction

6.2 Simultaneous Equations Model (SEM)

6.3 Structural Form and Reduced Form

6.4 Identification Problem

6.4.1 Order Condition

6.4.2 Rank Condition

6.4 General Structure of SEM

6.5 Simultaneity Bias

6.6 Let Us Sum Up

6.7 Key Words

6.8 Some Useful Books

6.9 Answers/Hints to Check Your Progress Exercises

6.0 OBJECTIVES

After reading this unit, you will be able to:

- distinguish between a 'single equation linear regression model' and a 'simultaneous equation model' (SEM);
- specify the 'types of variables' in a SEM with illustrations;
- outline the components of the 'structural form' and the 'reduced form' of SEM, specifying the types of variables in each;
- discuss the 'identification problem' in the SEM;
- describe, with an illustration, the meaning of 'observationally equivalent structural rule';
- present the general structure of equations in a 'simultaneous equation model (SEM)' using notations from linear algebra; and
- explain the phenomenon of 'simultaneous equations bias' in a SEM.

* Prof. Poulomi Roy, Jadavpur University, Kolkata

6.1 INTRODUCTION

In the previous Unit we mentioned about regression models having multiple equations. In these types of regression models, the equations are inter-dependent. For example, in the two-sector Keynesian model, income depends upon consumption and investment ($Y = C + I$). Further, consumption depends upon income($C = \alpha + \beta Y$). Therefore, estimation of a single equation, without considering the other, will lead to inefficient estimators. We need to consider both the equations simultaneously.

Single equation linear regression models express the relationship between one endogenous (dependent) variable and one or more exogenous (independent) variables. But most economic models include certain dependent variables which interact simultaneously on other dependent variables. This means, economics is interdependent in the sense that 'everything depends on everything else'. Hence, economic data is "a product of economic system, described as a 'system of simultaneous relations' among many random variables". These relations involve both the current and past values of many interacting variables.

A single equation model considers one-way causality. The explanatory variables in a single equation model influence the dependent variable but the explanatory variables are not reciprocally influenced by the dependent variable. A restricted framework like this is inconsistent with the data generation process for the more involved interactive economic relations. For instance, in a partial equilibrium framework for a single commodity, we visualize price, quantity-demanded and quantity-supplied as determined simultaneously. We know that it is inappropriate to say that price causes consumption or consumption causes price since the values that these variables take are jointly determined. Such an interactive mechanism can be better explained by a system of relations or equations. Such a system of equations gives due importance to the underlying process of complex economic relations. Such an analytical process is known as a 'simultaneous equations system (SES)' or simultaneous equations model (SEM).

6.2 SIMULTANEOUS EQUATIONS MODEL (SEM)

A simultaneous equations model (SEM) includes the following types of variables:

(i) Endogenous Variables: Endogenous variables are 'jointly determined variables'. They are those variables whose outcome values are determined within the system, through the joint interaction with other variables. A simultaneous equations model (SEM) is formulated with the objective of analysing the behaviour of such endogenous variables. Examples of such variables are price, consumption, production, income, etc.

(ii) Exogenous Variables: These are variables which are determined entirely from outside the system but are known to affect the outcome of the endogenous variables. They are assumed to condition the outcome values

of the endogenous variables; but exogenous variables are not reciprocally affected. The reciprocal effect implies that no feedback relationship is assumed. Examples of this type of variables are weather related variables (e.g. temperature and rainfall), international price of a commodity, etc.

(iii) Lagged Endogenous Variables: These are past values of the endogenous variables. For example, if Y_t is an endogenous variable in time 't', Y_{t-1}, Y_{t-2}, etc. are past values for lagged Y_t's. It is also called lagged dependent variable.

(iv) Pre-determined Variables: These are current and past values of the exogenous variables and lagged values of the endogenous variables. You should remember that the number of pre-determined variables in a SEM is the sum of the number of exogenous variables and lagged endogenous variables.

(v) Error Variables: These are residual variables which appear in the final specification of the model to account for non-observable random errors for shocks or disturbances. We make some specifications about these unobservable random variables to imagine an ideal condition. These conditions are taken as assumptions, and they are in respect of the mean, variances and inter-correlations of the exogenous and pre-determined variables.

A system of simultaneous equations is said to be 'complete' if it has as many equations as there are endogenous variables. Let us consider two examples of simultaneous equations model (SEM).

Example 6.1: Demand-Supply Model

Let q_{dt} be the demand for rice in period t, q_{st} be the supply of rice in period t and q_t be the equilibrium quantity demanded and supplied in period t. The demand-supply model can now be explained with the help of the following system of equations:

Demand: $q_{dt} = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 p_{t-1} + u_t$ … (6.1)

Supply: $q_{st} = \beta_0 + \beta_1 p_t + \beta_2 r_t + \beta_3 p_{t-1} + e_t$ … (6.2)

Equilibrium condition is: $q_{dt} = q_{st} = q_t$

Here, p_t is the price in period t, y_t is the income in period t, r_t is the rainfall in period t, p_{t-1} is the price in period $(t-1)$, i.e., the lagged endogenous variable. Note that q_t and p_t are endogenous variables while y_t, r_t, p_{t-1} are predetermined variables. u_t and e_t are the error terms of demand and supply equations respectively.

Example 6.2: **Keynesian Model**

A simple Keynesian model comprises consumption and investment equations and is expressed as:

Consumption function: $c_t = \alpha_0 + \alpha_1 y_t + u_t$

Autonomous investment function: $i_t = \bar{i}$

Equilibrium condition: $y_t = c_t + i_t$

where c_t is consumption in period t, $y_t$is income in period $t$and i_t is autonomous investment in time t taken as equal to $\bar{i}$. In this model, c_t and y_t are endogenous variables, i_t is the exogenous variable and u_t is the random error term.You should note there are two endogenous variables (c_t and y_t) and two regression equations in this model.

6.3 STRUCTURAL FORM AND REDUCED FORM

The set of equations which actually describes the economic phenomenon is called the 'structural form'. In structural equations, the variation of a current endogenous variable is described by:

(i) current endogenous variables,

(ii) past endogenous variables,

(iii) current and past exogenous variables, and

(iv) random error term.

The demand-supply model and the simple Keynesian model discussed above are examples of structural form model. An equation of a structural model is known as the *structural equation* and its coefficients are known as *structural coefficients*. The error terms of the equations are known as *structural disturbances*.

The reduced form of the SEM is obtained from the structural form after eliminating the current endogenous variables from the right-hand side of the structural equations. It implies that in an equation, an endogenous variable is expressed as a function of pre-determined variables and error terms. Thus, in the reduced form, on the right-hand side of an equation we have

(i) past endogenous variables,

(ii) current and past exogenous variables, and

(iii) random error terms.

Let us consider the demand-supply model (**Example 6.1**) once again. To obtain the reduced form of this model, we first equate the demand equation (6.1) and the supply equation (6.2). We obtain

$$\alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 p_{t-1} + u_t = \beta_0 + \beta_1 p_t + \beta_2 r_t + \beta_3 p_{t-1} + e_t$$

On re-arrangement of terms, we obtain the reduced form equation of p_t as follows:

$$(\alpha_1 - \beta_1)p_t = (\beta_0 - \alpha_0) + \beta_2 r_t - \alpha_2 y_t + (\beta_3 - \alpha_3)p_{t-1} + (e_t - u_t)$$

Or, $p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} r_t - \frac{\alpha_2}{\alpha_1 - \beta_1} y_t + \frac{\beta_3 - \alpha_3}{\alpha_1 - \beta_1} p_{t-1} + \frac{e_t - u_t}{\alpha_1 - \beta_1}$ … (6.3).

We can write the above equation as

$$p_t = \pi_{10} + \pi_{11} r_t + \pi_{12} y_t + \pi_{13} p_{t-1} + v_{1t} \quad \text{… (6.4)}$$

where

$\pi_{10} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$

$\pi_{11} = \beta_2/(\alpha_1 - \beta_1)$

$\pi_{12} = -\alpha_2/(\alpha_1 - \beta_1)$

$\pi_{13} = (\beta_3 - \alpha_3)/(\alpha_1 - \beta_1)$

$v_{1t} = (e_t - u_t)/(\alpha_1 - \beta_1)$

Look into equation (6.4) carefully. It expresses the endogenous variable p_t as a function of pre-determined variables and error terms only; there is no current endogenous variable on the right-hand side of the equation. Further, note that the coefficients π_{10}, π_{11}, π_{12} and π_{13} are the reduced form coefficients.

Now, let us substitute equation (6.4) in either the demand equation to obtain the reduced form of q_t(we can substitute equation (6.4) in the supply equation also – we will obtain the same result).

From equation (6.1), we get

$$q_t = \alpha_0 + \alpha_1(\pi_{10} + \pi_{11} r_t + \pi_{12} y_t + \pi_{13} p_{t-1} + v_{1t}) + \alpha_2 y_t + \alpha_3 p_{t-1} + u_t.$$

By re-arranging terms in the above, we get

$$q_t = (\alpha_0 + \alpha_1\pi_{10}) + \alpha_1\pi_{11} r_t + (\alpha_1\pi_{12} + \alpha_2) y_t + (\alpha_1\pi_{13} + \alpha_3) p_{t-1} + (\alpha_1 v_{1t} + u_t) q_t \quad \text{… (6.5)}$$

Note that we again have expressed the endogenous variable q_t in terms of pre-determined variables and error terms only. Thus, we can write the reduced form of q_t as

$$q_t = \pi_{20} + \pi_{21} r_t + \pi_{22} y_t + \pi_{23} p_{t-1} + v_{2t} \quad \text{… (6.6)}$$

where

$\pi_{20} = \alpha_0 + \alpha_1\pi_{10}$ $\quad \pi_{21} = \alpha_1\pi_{11}$ $\quad \pi_{22} = \alpha_1\pi_{12} + \alpha_2$

$\pi_{23} = \alpha_1\pi_{13} + \alpha_3$ $\quad v_{2t} = \alpha_1 v_{1t} + u_t$

The reduced form of p_t and q_t are functions of pre-determined variables only (i.e., current and past values of exogenous variables and past values of endogenous variables). The current endogenous variables do not appear on the right-hand side

of the reduced form model. All the predetermined variables of the system appear only on the right-hand side of the reduced form equations.

Since the reduced form equations have only exogenous variables as explanatory variables, we can estimate the reduced form parameters by applying OLS method. Our objective, however, is to estimate the structural equations given at equations. Thus, we should be in a position to obtain estimators for the structural parameters (a_1, b_1, c_1, a_2, b_2 and c_2) from the reduced form parameters (π_1, π_2, π_3, π_4, π_5, and π_6). By using the estimated values of the parameters in the reduced form model, we can obtain the estimated structural form parameters in certain cases; not always. We present some examples below to bring forward this point.

Example: 6.3: Exactly identified model

Let us consider the following model:

$q_d = a_1 + b_1 p + c_1 y + u_1$ … (6.7)

$q_s = a_2 + b_2 p + c_2 r + u_2$ … (6.8)

where q is quantity, p is price, y is income, r is rainfall andu_1 and u_2 are disturbance terms. Here, p and q are jointly determined endogenous variables and y and r are exogenous variables. The exogenous variables should be uncorrelated with the random error terms. Since p is correlated with the residuals u_1 and u_2 (by assumption), we get inconsistent estimators for the parameters. To see how this is, let us now derive the reduced form of p and q as:

$a_2 + b_2 p + c_2 r + u_2 = a_1 + b_1 p + c_1 y + u_1$

Or,$(b_2 - b_1)p = (a_1 - a_2) + c_1 y - c_2 r + (u_1 - u_2)$

Or, $p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y - \frac{c_2}{b_2 - b_1} r + v_2$

We can write the above equation as

$p = \pi_4 + \pi_5 y + \pi_6 r + v_2$ … (6.9)

Substituting (6.9) in equation (6.7) we get

$q = a_1 + b_1(\pi_4 + \pi_5 y + \pi_6 r + v_2) + c_1 y + u_1$

Or, $q = a_1 + b_1\pi_4 + (b_1\pi_5 + c_1)y + b_1\pi_6 r + (u_1 + b_1 v_2)$

Or, $q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y - \frac{c_2 b_1}{b_2 - b_1} r + (u_1 + b_1 v_2)$

We can write the above equation as

$q = \pi_1 + \pi_2 y + \pi_3 r + v_1$ … (6.10)

where$\pi_1 = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1}$,$\pi_2 = \frac{c_1 b_2}{b_2 - b_1}$,$\pi_3 = \frac{-c_2 b_1}{b_2 - b_1}$, and $v_1 = (u_1 + b_1 v_2)$.

Since the reduced form equations have only exogenous variables as explanatory variables, we can estimate the reduced form parameters by applying OLS method.

Our objective, however, is to estimate the structural equations of the SEM. Thus, we should be in a position to obtain estimators for the structural parameters (a_1, b_1, c_1, a_2, b_2 and c_2) from the reduced form parameters (π_1, π_2, π_3, π_4, π_5, and π_6). By using the estimated values of the parameters in the reduced form model, we can obtain the estimated structural form parameters as follows:

$\hat{b}_1 = \frac{\hat{\pi}_3}{\hat{\pi}_6}$, $\hat{b}_2 = \frac{\hat{\pi}_2}{\hat{\pi}_5}$, $\hat{c}_2 = -\hat{\pi}_6(\hat{b}_1 - \hat{b}_2)$, $\hat{c}_1 = \hat{\pi}_5(\hat{b}_1 - \hat{b}_2)$, $\hat{a}_1 = \hat{\pi}_1 - \hat{b}_1\hat{\pi}_4$ and $\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2\hat{\pi}_4$.

This method of obtaining the structural parameters from reduced form parameters is known as the 'indirect least squares(ILS) method' (we will discuss further on this estimation method in Unit 16).

In the above example we could obtain unique estimators of all the structural parameters. It is however not always possible to get the structural parameters from the reduced form parameters. There could be two different situations: (i) we do not get estimators of some of the structural parameters, and (ii) we get multiple estimators for some of the structural parameters. These two cases are illustrated in Example 6.4 and Example 6.5 respectively.

Example 6.4: Under-identified model

$q_d = a_2 + b_2p + u_2$ … (6.11)

$q_s = a_1 + b_1p + c_1y + u_1$ … (6.12)

In equilibrium, $q_d = q_s$

Or, $a_2 + b_2p + u_2 = a_1 + b_1p + c_1y + u_1$

Or, $(b_2 - b_1)p = (a_1 - a_2) + c_1y + (u_1 - u_2)$

Or, $p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1}y + \frac{1}{b_2 - b_1}(u_1 - u_2)$ ... (6.13)

The above equation can be written as

$p = \pi_3 + \pi_4y + v_2$ … (6.14)

Now, substituting (6.14) in any of the structural equations we get the reduced form of q (since, $q_s = q_d$). Let us substitute (6.14) in (6.12).

$q_s = a_2 + b_2(\pi_3 + \pi_4y + v_2) + u_2 = (a_2 + b_2\pi_3) + b_2\pi_4y + (u_2 + b_2v_2)$

If we express the reduced form parameters (π_3 and π_4) in terms of structural parameters, as in (6.14), and re-arrange terms, we obtain

$q = \frac{a_1b_2 - a_2b_1}{b_2 - b_1} + \frac{c_1b_2}{b_2 - b_1}y + v_1$ … (6.15)

The above equation can be written as

$q = \pi_1 + \pi_2 y + v_1$ ... (6.16)

Solving for the structural form parameters, in terms of the reduced form parameters, we get:

$\hat{b}_2 = \frac{\hat{\pi}_2}{\hat{\pi}_4}$, $\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2\hat{\pi}_3$.

There is now no way of getting the estimates of a_1, b_1 and c_1. We therefore have a situation where the demand function is estimable, but the supply function is not.

Example 6.5: Under-identified model

Let us consider the following two equations:

$q_d = a_1 + b_1 p + u_1$ … (6.17)

$q_s = a_2 + b_2 p + c_2 r + u_2$ … (6.18)

From equations (6.17) and (6.18) we get the reduced form of p as

$a_1 + b_1 p + u_1 = a_2 + b_2 p + c_2 r + u_2$

Or, $(a_1 - a_2) - c_2 r + (u_1 - u_2) = (b_2 - b_1)p$

Or, $p = \frac{a_1 - a_2}{b_2 - b_1} - \frac{c_2}{b_2 - b_1} r + \frac{1}{b_2 - b_1}(u_1 - u_2)$

We can write the above as

$p = \pi_3 + \pi_4 r + v_2$ … (6.19)

Now substituting (6.19) in (6.18) we get the reduced form of q as

$q = a_2 + b_2(\pi_3 + \pi_4 r + v_2) + c_2 r + u_2$

$= (a_2 + b_2\pi_3) + (b_2\pi_4 + c_2)r + (b_2 v_2 + u_2)$

We can write the above as

$q = \pi_1 + \pi_2 r + v_1$ … (6.20)

Now, we can estimate $\hat{b}_1 = \frac{\hat{\pi}_2}{\hat{\pi}_4}$, $\hat{a}_1 = \hat{\pi}_1 - \hat{b}_1\hat{\pi}_3$ but we cannot estimate a_2, b_2, c_2.

Hence, in this situation, the demand function is estimable, but the supply function is not.

Example 6.6: Over-identified Model

Now, let us consider another demand supply model:

$q_s = a_1 + b_1 p + c_1 y + d_1 r + u_1$ … (6.21)

$q_d = a_2 + b_2 p + u_2$ … (6.22)

By equating (6.21) and (6.22) we get the reduced form of p as follows:

$a_1 + b_1 p + c_1 y + d_1 r + u_1 = a_2 + b_2 p + u_2$

Or, $(a_1 - a_2) + c_1 y + d_1 r + (u_1 - u_2) = (b_2 - b_1)p$

Or, $p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y + \frac{d_1}{b_2 - b_1} r + \frac{1}{b_2 - b_1}(u_1 - u_2)$

We can write the above equation as

$$p = \pi_4 + \pi_5 y + \pi_6 r + v_2 \quad \ldots (6.23)$$

Now, substituting (6.23) in (6.22) we get

$$\begin{aligned} q &= a_2 + b_2(\pi_4 + \pi_5 y + \pi_6 r + v_2) + u_2 \\ &= (a_2 + b_2\pi_4) + b_2\pi_5 y + b_2\pi_6 r + (b_2 v_2 + u_2) \\ &= \pi_1 + \pi_2 y + \pi_3 r + v_1 \quad \ldots (6.24) \end{aligned}$$

If we try to solve the structural form parameters, in terms of the above reduced form parameters, then we get two estimates of b_2 as

$\hat{b}_2 = \frac{\hat{\pi}_2}{\hat{\pi}_5}$ and $\hat{b}_2 = \frac{\hat{\pi}_3}{\hat{\pi}_6}$.

For each of these $\hat{b}_2$, we get one estimate of $\hat{a}_2 = \hat{\pi}_1 - \hat{b}_2\hat{\pi}_4$ and we get no estimate of the parameters of the demand function, i.e.,a_1, b_1, c_1, d_1.

In Example 6.3 we could find estimators for parameters of the demand function and the supply function. In Example 6.4 and Example 6.5 we could find estimators for parameters of the demand function while estimators for the parameters of the supply function could not obtained. In these cases the demand function is identifiable but the supply function is not identifiable. In the case of Example 6.6 we could find more than one estimator for the parameters of the demand function but no estimator for the supply function. In this case the demand function is over-identified while the supply function is under-identified. This leads us to the problem of identification which we discuss in Section 6.4.

Check Your Progress 1 [answer within the space given in about 50-100 words]

1) Distinguish between a 'single equation linear regression model' and a 'simultaneous equations regression model'.

……
……
……

2) Specify the 'types of variables' in a 'simultaneous equations system'.

……
……
……

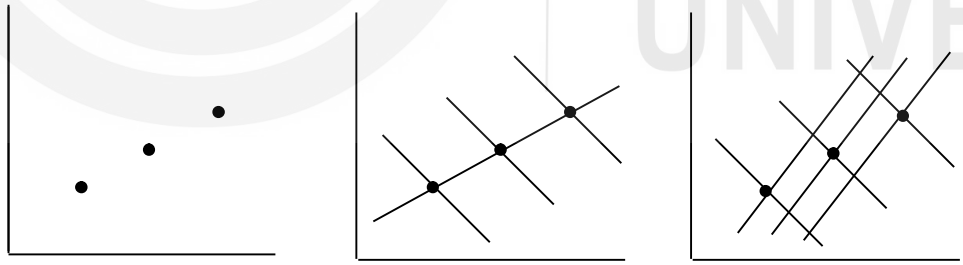
3) Consider the following simultaneous equations system, where c_t and y_t are endogenous variables while i_t is exogenous.
Consumption function: $c_t = \alpha_0 + \alpha_1 y_t + u_t$

Autonomous investment function: $i_t = \bar{i}$
Equilibrium condition: $y_t = c_t + i_t$
Find out the reduced form of the model.

……
……
……
……
……

6.4 IDENTIFICATION PROBLEM

Let us consider the demand function$q = \alpha + \beta p + u$where we assume that p and u are correlated. To measure the coefficients of the demand equation, suppose we use time series data on the quantity bought of the commodity at different prices. Quantity bought, however, is identical with the quantity sold at the corresponding price. Thus, market data register points of interaction of equilibrium supply and demand at prevailing prices at certain points of time. That is, it only shows the points of interactions of demand and supply. If we use these data for estimation, we actually measure the coefficients of a function of the form Q=(p). This equation may be either the demand function or the supply function. How can we be sure that this equation represents demand function or supply function? If one is interested in measurement of the demand function, then (s)he can use the data. Similarly, a person interested in measurement of the supply equation will also be using the same data. Therefore, there is a need for some criteria, which will enable us to verify whether the estimated coefficients belong to a particular relationship or another. Such criteria are known as 'identification conditions' of an equation in a simultaneous equations model.



(a) (b) (c)

Fig. 6.1: Supply and Demand Functions

We begin with the concept of identification using Fig. 6.1. The observed data consist of the market outcomes as shown in Fig. 6.1. Fig. 6.1(a) shows three data points. From the Fig. 6.1(b) and 14.1(c) we can see that we can get the same data for two different situations. In Fig. 6.1(b) the supply curve is stable, and the demand curve is shifting. The data represents points on the supply curve under the assumption that the supply function is fairly stable and the demand curve shifts. In Fig. 14(c) both demand curve and supply curve are shifting randomly. If

we wish to estimate $Q = f(p)$ from the same data set then we estimate neither the demand nor the supply curve; rather we end up estimating a mongrel equation (i.e., a combination of two functions).

A second inference that we draw from Fig. 15(b) is very important: the shift in the demand curve helps us in identification of the supply curve. There is a certain variable in the demand equation (that is excluded from the supply equation), which causes a shift in the demand curve. The result is paradoxical – the exclusion of certain variables from the supply curve helps us in identification of the supply curve. This is often termed as the 'paradox of identification'. To identify an equation, we have to look at the variables that are absent from the equation (while being present in other equations of the model). In identification the relevant variables are those which are absent from the equation!

We can make three different statements on the status of identification of equations in a SEM.

(i) If we do not get estimator of the structural form parameters in terms of the reduced form parameters, then the equation is **under-identified**.

(ii) If we get unique estimator of the structural form parameters in terms of the reduced form parameters of an equation, then the equation is **exactly identified**.

(iii) If we get multiple estimators of the structural form parameters in terms of the reduced form parameters, then the equation is **over-identified.**

It is quite tedious and time consuming to transform the structural form of the SEM to its reduced form and then examine whether structural parameters can be expressed in terms of reduced form parameters. There is a counting rule, which tells us the identification status of an equation on the basis of variables excluded from an equation. There are two conditions which should be checked: (i) order condition, and (ii) rank condition.

6.4.1 Order Condition for Identification

For an equation to be identified the total number of variables (endogenous and exogenous) excluded from the equation must be equal to or greater than the number of endogenous variables in the model less one. Since the number of endogenous variables is equal to the number of equations in a SEM, the order condition for identification can be stated as follows:

For an equation to be identified the total number of variables excluded from it but included in other equations should be equal to more than the number of equations of the system less one.

Let G =total number of equations (= total number of endogenous variables)

K =total number of variables in the model (endogenous and predetermined)

M = the number of variables (endogenous and exogenous) included in a particular equation.

Then the order condition for identification is

$$(K - M) \geq (G - 1) \quad \text{... (6.25)}$$

If there is equality (i.e., $(K - M) = (G - 1)$) the equation is exactly identified. If there is inequality (i.e., $(K - M) > (G - 1)$) the equation is over identified.

Example 6.8

There are10 equations in a simultaneous equations model with 15 variables (10 endogenous and 5 exogenous).

(a) Check the identification status of the first equation in the model, if it contains 11 variables.

For the first equation, G = 12, K = 15 and M = 11.

The **order condition** is given by$(K - M) \geq (G - 1)$. By substituting the values in the equation, we obtain (15–11) < (10–1). Thus the first equation in the model is not identified.

(b) Check the identification status of the second equation if it includes 5 variables.

G **= 10** K **= 15** M **= 5**

Order condition: $(K - M) \geq (G - 1)$

(15–5) > (10–1)

Thus, the second equation is over-identified.

The order condition for identification is a necessary condition for an equation to be identified; but it is not a sufficient condition. It means that the order condition may be fulfilled in an equation; yet the relation may not be identified.

6.4.2 The Rank Condition for Identification

According to the rank condition, in a system of G equations a particular equation is identified if and only if the rank of the matrix of parameters of the excluded variables (endogenous and predetermined) from that equation is equal to $(G - 1)$. In other words, an equation is identified if it is possible to construct at least one non-zero determinant of order $(G - 1)$ from the coefficients of the variables excluded from that particular equation but contained in the other equations of the model. The rank condition is a necessary and sufficient condition for identification.

When you are required to check the identification status of an equation in a SEM, there are certain steps you should follow. For illustration purpose we take the following SEM.

$$y_1 = 3y_2 + 2x_1 + x_2 + u_1$$

$$y_2 = y_3 + x_3 + u_2$$

$$y_3 = y_1 - y_2 - 2x_3 + u_3 \quad \text{... (6.26)}$$

There are 3 endogenous variables in the model (G=3); total number of variables is 6 (implies, K=6).

Step1: Write down the parameters of all the equations of the model in a separate table, noting that the parameter of a variable excluded from an equation is equal to zero.

Thus, the SEM given at (6.26) can be written as

$$-y_1 + 3y_2 + 0y_3 - 2x_1 + x_2 + 0x_3 + u_1 = 0$$

$$0y_1 + y_2 + y_3 - 0x_1 + 0x_2 + x_3 + u_2 = 0$$

$$y_1 + y_2 + y_3 - 0x_1 + 0x_2 + 2x_3 + u_3 = 0 \quad \text{… (6.27)}$$

Table 6.1: Parameters of the Equation System

Equations	Variables					
	y_1	y_2	y_3	x_1	x_2	x_3
1st equation	–1	3	0	–2	1	0
2nd equation	0	–1	1	0	0	1
3rd equation	1	–1	–1	0	0	–2

You should prepare a table similar to Table 6.1. Note that we ignore the error terms while checking the identification status of an equation. Let us assume that we have to check the identification of the second equation in the above SEM.

Step2: Eliminate the row of coefficients of the equation, which is being examined for identification.

For example, for checking the identification status of the second equation, we strike out the second row of the coefficients in Table 6.1.

Step3: Eliminate the columns in which a non-zero coefficient of the equation being examined appears. By deleting the relevant row and columns we are left with the coefficients of variables not included in the particular equation, but contained in the other equations of the model.

For example, for identification the second equation of the system, we will eliminate the second, third and the sixth columns of Table 6.1. After carrying out Steps 2 and 3, we obtain Table 6.2.

Table 6.2: Strike-out the Rows and Columns

Equations	Variables					
	y_1	y_2	y_3	x_1	x_2	x_3
1st equation	–1	3	0	–2	1	0
2nd equation	0	–1	1	0	0	1
3rd equation	1	–1	–1	0	0	–2

Step4: Form the determinant(s) of order $(G-1)$ and examine their value. If at least one of these determinants is non-zero, the equation is identified. If all the determinants of order $(G-1)$ are zero, the equation is under-identified.

Table 6.3: Remaining Parameters

y_1	x_1	x_2
−1	−2	1
1	0	0

We have three determinants of order $(G-1) = 3-1=2$ on the basis of Table 6.3.

They are

$$\Delta_1 = \begin{vmatrix} -1 & -2 \\ 1 & 0 \end{vmatrix} \neq 0$$

$$\Delta_2 = \begin{vmatrix} -2 & 1 \\ 0 & 0 \end{vmatrix} = 0$$

$$\Delta_3 = \begin{vmatrix} -1 & 1 \\ 1 & 0 \end{vmatrix} \neq 0$$

(the symbol 'Δ' stands for determinant). We see that we can form two non-zero determinants of order $G-1 = 3-1 = 2$; hence the second equation of our system is identified.

Step5: To see whether the equation is exactly identified or over-identified we use the order condition $(K-M) > (G-1)$.With this criterion, if the equality sign is satisfied, that is, if $(K-M) = (G-1)$, the equation is exactly identified. If the inequality sign holds, that is, if $(K-M) > (G-1)$, then the equation is over-identified.

In the case of the second equation we have

$G = 3$ $K = 6$ $M = 3$

and the counting rule $(K-M) \geq (G-1)$ gives

$(6-3) > (3-1)$.

Therefore the second equation of the model is over-identified.

Check Your Progress 2[answer within the space given in about 50-100 words]

1) What is meant by the 'identification problem' in a SES?

……
……
……
……
……

2) What is a 'necessary condition' for identification in a SEM?

……
……
……
……

……

3) What is a 'sufficient condition' for an equation of SES to be identified?

.....
.....
.....
.....

4) Consider the three-equation model

$$Y_1 = \beta_{13}Y_3 + \gamma_{12}X_2 + u_1$$
$$Y_2 = \beta_{21}Y_1 + \beta_{23}Y_3 + \gamma_{21}X_1 + \gamma_{22}X_2 + u_2$$
$$Y_3 = \gamma_{33}X_3 + u_3$$

where Y_1, Y_2 and Y_3 are endogenous, and X_1, X_2 and X_3 are exogenous variables. Discuss the identification status of each equation of the model, based on the order and the rank condition.

.....
.....
.....
.....

4) The structure of a model with four endogenous and three exogenous variables is as follows (1 indicates presence and 0 absence of the variable in the equation):

1 0 1 1 1 0 0

1 1 1 0 0 1 1

0 0 1 0 1 0 0

1 0 1 1 0 1 0

Which of the above four equations is/are identified?

.....
.....
.....
.....

6.4 GENERAL STRUCTURE OF SEM

Suppose there are G number of endogenous variables, Y_1, Y_2,….,Y_G and K number of exogenous variables, X_1,X_2,….,X_K in a simultaneous equations system.

Thus, the total number of equations in the system are G. Let $\beta_1, \beta_2, \ldots\ldots, \beta_G$ be the coefficients associated with the endogenous variables and $\gamma_1, \gamma_2, \ldots\ldots, \gamma_k$ be the coefficients associated with the exogenous variables. The system of SEM at any time point t can be written as:

$$\begin{aligned} Y_1\beta_{11} + Y_2\beta_{21} + \ldots + Y_G\beta_{G1} + X_1\gamma_{11} + X_2\gamma_{21} + \ldots + X_K\gamma_{K1} + e_1 &= 0 \\ Y_1\beta_{12} + Y_2\beta_{22} + \ldots + Y_G\beta_{G2} + X_1\gamma_{12} + X_2\gamma_{22} + \ldots + X_K\gamma_{K2} + e_2 &= 0 \\ \vdots & \\ Y_1\beta_{1G} + Y_2\beta_{2G} + \ldots + Y_G\beta_{GG} + X_1\gamma_{1G} + X_2\gamma_{2G} + \ldots + X_K\gamma_{KG} + e_G &= 0 \end{aligned}$$

... (6.28)

We assume that the structural disturbances are generated by a stationary multivariate process such that:

(i) $E(e_i) = 0$, $i = 1,2,3,\ldots,G$

(ii) $E(e_i e_i') = \sigma_{ii} I_T$ for $i = 1,2,3,\ldots,G$

(iii) $E(e_i e_j') = \sigma_{ij} I_T$, for $i \neq j$, I,j=1,2,3,..,G

Thus, if $e' = \left(e_1' e_2' e_3' \ldots .. e_G'\right)$

Then $E(ee') = E\begin{bmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_G \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ \vdots \\ e_G \end{pmatrix}' \end{bmatrix} = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \cdots & \sigma_{1G}I \\ \sigma_{21}I & \sigma_{22}I & \cdots & \cdots & \sigma_{2G}I \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \sigma_{G1}I & \sigma_{G2}I & \cdots & \cdots & \sigma_{GG}I \end{bmatrix} = \sum \otimes I_T$

... (6.29)

The contemporaneous covariance matrix $\sum$ is a $G \times G$ symmetric positive semi-definite matrix. It may have less than full rank as some of the equations may appear in the form of identities with null error vectors. In a matrix form, the system of simultaneous equations model, for all time points, can be written as:

$YB + X\Gamma + E = 0$

where 0 is a TxG matrix of zeros,

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & \cdots & y_{1G} \\ y_{21} & y_{22} & \cdots & \cdots & y_{2G} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ y_{T1} & y_{T2} & \cdots & \cdots & y_{TG} \end{bmatrix} = (y_1 y_2 \ldots\ldots y_G)$$

and

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2K} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ x_{T1} & x_{T2} & \cdots & \cdots & x_{TK} \end{bmatrix} = (x_1 x_2 \ldots\ldots x_K)$$

Y and X are sample values taken by the jointly determined variables and the pre-determined variables, respectively, and

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & \cdots & e_{1G} \\ e_{21} & e_{22} & \cdots & \cdots & e_{2G} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ e_{T1} & e_{T2} & \cdots & \cdots & e_{TG} \end{bmatrix} = (e_1 e_2 \ldots . e_G)$$

is the matrix of unobservable values taken by the random error vectors. B is a $G \times G$ matrix of coefficients of the current endogenous variables, where each column refers to the coefficients of a particular equation. Specifically:

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2G} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GG} \end{bmatrix} = [\beta_1 \beta_2 \cdots \cdots \beta_G] \quad \text{... (6.30)}$$

is the $G \times G$ matrix of structural coefficients associated with the endogenous variables and each column contains coefficients of a particular equation. Further,

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1G} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2G} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{K1} & \gamma_{K2} & \cdots & \gamma_{KG} \end{bmatrix} = [\gamma_1 \gamma_2 \cdots \cdots \gamma_K] \quad \text{... (6.31)}$$

is the $K \times G$ matrix of structural coefficients associated with the exogenous-predetermined variables where each column contains the coefficients of a particular equation. For the system to be complete, B must be of order G. In addition, if it is non-singular, then we may rewrite the system of equations as:

$$YB + X\Gamma + E = 0 \quad \text{... (6.32)}$$

If we post-multiply the above equation by B^{-1}, we obtain

$$YBB^{-1} + X\Gamma B^{-1} + EB^{-1} = 0$$

Or, $Y + X\Gamma B^{-1} + EB^{-1} = 0$

Or, $Y = -X\Gamma B^{-1} - EB^{-1}$

We can write the above equation as

$$Y = X\pi + V \quad \text{… (6.33)}$$

Equation (6.33) is the reduced form of the simultaneous equations model

where

$$\pi = -\Gamma B^{-1} = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1G} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{KG} \end{bmatrix} = (\pi_1 \pi_2 \cdots \pi_G)$$

is a $K \times G$ matrix of reduced form coefficients.

$V = -EB^{-1}$ is a $T \times G$ matrix of error terms. That is:

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1G} \\ v_{21} & v_{22} & \cdots & v_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ v_{T1} & v_{T2} & \cdots & v_{TG} \end{bmatrix} = (v_1 v_2 \cdots v_G)$$

The G equations in (6.33) are called the *reduced form equations.* The matrix $\pi = -\Gamma B^{-1}$is the matrix of reduced form parameters. The matrix V represents the reduced form disturbances. With this transformation, we can express each endogenous variable as a linear function of the predetermined variables in the system.

The system of reduced form equations may be rewritten as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix} = \begin{bmatrix} X & & & \\ & X & & \\ & & \ddots & \\ & & & X \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_G \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_G \end{bmatrix}$$

Or, $y = (I \otimes X)\pi + v$ ... (6.34)

6.5 SIMULTANEITY BIAS

In the reduced form we keep one of the endogenous variables on the left-hand side while the predetermined variables (exogenous variables and lagged dependent variables) are kept on the right-hand side in an equation. The coefficient of the endogenous variables on the left-hand side is 1. To keep the dependent variable of the i^{th} structural equation on the left-hand side and all other variables on the right-hand side, we set the coefficient of the i^{th} endogenous variables y_i to be equal to –1. This is called the 'normalization rule'.

The i^{th} equation can therefore be written as:

$YB_i + X\Gamma_i + e_i = 0$ ... (6.35)

$$y_i = Y_i\beta_i + Y_i^*\beta_i^* + X_i\gamma_i + X_i^*\gamma_i^* + e_i$$

$$y_i = Y_i\gamma_i + X_i\beta_i + e_i = (Y_i \quad X_i)\begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} + e_i$$

Or, $y_i = Z_i\delta_i + e_i$ ... (6.36)

Here, $B_i = \begin{bmatrix} -1 \\ \beta_i \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} -1 \\ \beta_i \\ 0 \end{bmatrix}$, $\Gamma_i = \begin{bmatrix} \gamma_i \\ \gamma_i^* \end{bmatrix} = \begin{bmatrix} \gamma_i \\ 0 \end{bmatrix}$

$Y = \begin{bmatrix} y_i & Y_i & Y_i^* \end{bmatrix}$ and $X = \begin{bmatrix} y_i & Y_i & Y_i^* \end{bmatrix}$. $G = G_i^{\Delta} + G_i^{\Delta\Delta}, K = K_i^* + K_i^{**}$, Y_i^* are the $G_i^{\Delta\Delta}$ endogenous variables that appear in the i^{th} equation with zero coefficients, i.e.,they do not appear in the i^{th} equation. X_i^{**} are those exogenous or predetermined variables that appear in the i^{th} equation with zero coefficients. $y_i \quad Y_i \quad X_i$ are the endogenous variables that appear in the i^{th} equation with non zero coefficients. Consequently, y_i is a $T \times 1$ vector, Y_i is a $T \times (G_i^{\Delta} - 1)$ matrix,

β_i is a $(G_i^{\Delta} \times 1)$ vector, e_i is an unobservable $(T \times 1)$ random vector with mean vector zero and covariance $\sigma_i^2 I_T$, $Z_i = (Y_i \quad X_i)$ and $\delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix}$.

Under this specification, the least square estimator of δ_i is

$$\hat{\delta}_i = (Z_i Z_i')^{-1} Z_i' y_i \text{m} \qquad \text{... (6.37)}$$

The mean of random vector $\hat{\delta}_i$ is

$$E(\hat{\delta}_i) = E\left[(Z_i Z_i')^{-1} Z_i' y_i\right] = E\left[(Z_i Z_i')^{-1} Z_i' (Z_i \delta_i + e_i)\right] = \delta_i + E\left[(Z_i Z_i')^{-1} Z_i' e_i\right]$$

The last term does not vanish because Z_i contains endogenous variables that are jointly determined with y_i. They are thus not independent of e_i. Consequently, the estimator $\hat{\delta}_i$ is biased, i.e., $E(\hat{\delta}_i) \neq \delta_i$. Further, as the sample size increases, $\hat{\delta}_i$ does not converge to δ_i even in probability, since:

$$\text{plim}(\hat{\delta}_i) = \delta_i + p\text{lim} T(Z_i Z_i')^{-1} \text{ plim} T^{-1} Z_i' e_i \neq \delta_i$$

Specifically, plim $T^{-1} Z_i' e_i$ does not converge to zero vector because $Z_i = (Y_i \quad X_i)$ contains Y_i that is not independent of the error vector e_i. Therefore, the least squares estimator applied directly to a structural equation containing two or more endogenous variables will be inconsistent. More specifically, the statistical property of unbiasedness and consistency are not fulfilled.

Example 6.9

To illustrate the above, let us consider a simple macro model of income determination consisting solely of a consumption function and national income identity.

$C_t = \alpha + \beta Y_t + u_t$

$Y_t \equiv C_t + I_t$

Here, C_t= aggregate consumption expenditure at time t,

Y_t=national income at time point t,

I_t=investment expenditure at time point t and

u_t=error term at time point t.

We classify C_t and Y_t as endogenous variables and I_t as exogenous variable. We make two assumptions, namely: $u \sim N(0, \sigma_u^2 I)$ and I and u are independent. The first assumption implies that:

(i) $E(u_t) = 0$

(ii) $V(u_t) = \sigma_u^2$

(iii) $\text{cov}(u_t, u_s) = 0, t \neq s$

The second assumption will be satisfied if either z is a set of fixed numbers, or I is a random variable distributed independently of u. Hence, $E(u_t, I_t) = 0$. We

shall now derive the reduced form of this SEM for which we will substitute the consumption function in the income identity.

$Y_t = C_t + I_t = \alpha + \beta Y_t + u_t + I_t$

Or, $(1-\beta)Y_t = \alpha + u_t + I_t$

Or, $Y_t = \frac{\alpha}{1-\beta} + \frac{I_t}{1-\beta} + \frac{u_t}{1-\beta}$ …(6.38)

Substituting (6.38) in the 'consumption function' we get:

$C_t = \alpha + \beta(\frac{\alpha}{1-\beta} + \frac{I_t}{1-\beta} + \frac{u_t}{1-\beta}) + u_t$

Or, $C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + \frac{1}{1-\beta} u_t$

We can write the above equation as

$C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + v_t$ … (6.39)

where $v_t = \frac{1}{1-\beta} u_t$. Now from reduced form of Y_t is given by:

$Y_t = \frac{\alpha}{1-\beta} + \frac{I_t}{1-\beta} + \frac{u_t}{1-\beta}$ ... (6.40)

If we take expected value of equation (6.40) we obtain

$E(Y_t) = \frac{\alpha}{1-\beta} + \frac{I_t}{1-\beta}$

Therefore, $Y_t - E(Y_t) = \frac{u_t}{1-\beta}$

$cov(Y_t, u_t) = E\left(\left(\frac{1}{1-\beta} u_t\right) u_t\right) = \frac{1}{1-\beta} E(u_t^2) = \frac{\sigma_u^2}{1-\beta} \neq 0$ ... (6.41)

Let us now consider the consumption function in its structural form, i.e.,

$C_t = \alpha + \beta Y_t + u_t.$

On applying OLS to this equation, we have:

$\widehat{\beta} = \frac{\sum(c_t - \overline{c})(Y_t - \overline{Y})}{\sum(Y_t - \overline{Y})^2} = \frac{\sum(c_t - \overline{c}) y_t}{\sum y_t^2}$ ... (6.42)

where

$y_t = Y_t - \overline{Y} = \frac{\sum c_t y_t}{\sum y_t^2} - \overline{c} \frac{\sum y_t}{\sum y_t^2} = \frac{\sum c_t y_t}{\sum y_t^2}$

Since $\sum y_t = 0$, we have

$\frac{\sum c_t y_t}{\sum y_t^2} = \frac{\sum(\alpha + \beta Y_t + u_t) y_t}{\sum y_t^2}$

$= \frac{\alpha \sum y_t + \beta \sum Y_t y_t + \sum u_t y_t}{\sum y_t^2}$

$= \frac{\alpha \sum y_t + \beta \sum (y_t + \overline{Y}) y_t + \sum u_t y_t}{\sum y_t^2} = \frac{\alpha \sum y_t + \beta \sum y_t^2 + \beta \overline{Y} \sum y_t + \sum u_t y_t}{\sum y_t^2}$

$\widehat{\beta} = \beta . 1 + \frac{\sum u_t y_t}{\sum y_t^2}$ ... (6.43)

Let us take *plim* on both sides of equation (6.43) to check the consistency property of the estimators. We get

$$\text{plim}\widehat{\beta} = \text{plim}\beta + \text{plim}\left(\frac{\sum\frac{y_t u_t}{n}}{\sum\frac{y_t^2}{n}}\right)$$

$$=\beta + \frac{plim\sum\frac{u_t y_t}{n}}{plim\sum\frac{y_t^2}{n}} = \beta + \frac{\frac{\sigma_u^2}{1-\beta}}{\sigma_y^2} \quad \text{... (6.44)}$$

Now, $\frac{1}{1-\beta}\frac{\sigma_u^2}{\sigma_y^2} \neq 0$.

Therefore, $\left(\text{plim}\hat{\beta} - \beta\right) \neq 0$ even if $n \to \infty$.Thus, the bias will not vanish even if we increase the sample size to ∞.This bias is known as the 'simultaneous equation bias' or simultaneity bias.

Check Your Progress 2 [answer within the space given in about 50-100 words]

1) If the structural model in matrix form for a system of simultaneous equations is given by $\boldsymbol{YB + X\Gamma + E = 0}$, write down the reduced form equations.

.....
.....
.....
.....
.....

2) What is meant by 'simultaneity bias'?

.....
.....
.....
.....
.....

6.6 LETUSSUMUP

In estimating the parameters of a system of simultaneous equations, several problems arise. First, the direct application of the least squares rule to a structural equation may result in an estimator that is biased. Further, the ordinary least square estimator will not converge to the true parameter even when the sample size is increased indefinitely. In searching for consistent estimator, one possibility is to: (i) find out the reduced form model and then used the estimated values of the reduced form parameters (obtained by the least square method) and then (ii) use the reduced form parameters to solve the structural form parameters. But this also may not always work as we may not get unique solution of the structural form parameters. This raises the problem of identification.

Identification is a problem of model formulation, rather than of model estimation or evaluation. An econometric model is identified if it is in a unique statistical

form, enabling unique estimates of its parameters to be subsequently made from sample data. If a model is not identified then we cannot say what relationship we are estimating.

The counting rule (necessary condition) for an equation with a system of G equations to be just identified is that there should be exactly G–1 variables not appearing in the equation in order to differentiate that equation from the remaining G–1 equations. The rank condition requires that the excluded variables of the i^{th} equation should appear in the remaining G–1 equations in such a way that no other equation (or linear combination of equations) appears in any other equation. Finally, the unit explains how in the presence of simultaneity, the LSEs produce biased estimates of parameter which even with $n \rightarrow \infty$, continues to remain. This is termed as the 'simultaneous equation bias' or simply 'simultaneity bias'.

6.7 KEYWORDS

Simultaneous Equations Model (SEM) : This is a model which involves different type of variables like endogenous variables, lagged endogenous variables, exogeneous variables, pre-determined variables which could be both the exogeneous variables and the endogenous variables with lags, along with the error term.

Structural and Reduced Form Models : Structural form model refers to the system or the set of equations which describes the actual economic phenomena. An equation of the structural model is known as the structural equation; its coefficients are known as the structural coefficients and its error term called the structural disturbance. The reduced form of the model is obtained from the structural form model by eliminating the current endogenous variables from the structural equations.

Identification Problem : We consider the reduced form models for arriving at the estimates of the structural parameters. This method, being indirect, is called the method of 'indirect least squares'. However, getting unique estimates of the structural parameters is possible only when the system of simultaneous equations is 'complete'. This is the case when the number of equations in the system are exactly as many as there are endogenous variables. When this requirement is not fulfilled, we get situations in which either there are no estimates of structural parameters or there are multiple estimates of the

same. Both these situations are referred to as the problem of identification. A 'no estimate' situation is termed as the 'under identification problem' and 'multiple estimates' situation termed as the problem of 'over identification'. We get unique estimates of the structural parameters only when the equations are exactly identified or the when the system of simultaneous equation is 'complete'.

Simultaneity Bias or Simultaneous Equation Bias : Refers to a situation where, in the presence of simultaneity, for the OLS estimate of β, $\hat{\beta}$, the $\left(plim\hat{\beta} - \beta\right) \neq 0$ even if $n \rightarrow \infty$. Thus, if the LSE bias refers to the violation of the property of unbiasedness of the estimate, the simultaneity bias refers to the violation of the property of the consistency of the estimated value of β.

Exogenous Variables : These are variables determined outside the model. The values of the exogenous variables are assumed to be given.

Endogenous Variables : The values of the endogenous variables are determined within the model.

6.8 SOME USEFUL BOOKS

Wooldridge, Jaffery (2010) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

A. Collin Cameron and P. K. Trivedi (2005). *Micro Econometric Methods and Applications*, Cambridge University Press.

W. H. Greene (2018). *Econometric Analysis*, Eighth Edition, Pearson Education.

6.9 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) The former expresses the relationship between one endogenous variable and one or more exogenous variables. It considers one-way causality. This means explanatory variables exert influence on the dependent variable, but they are not reciprocally influenced by it. The latter, i.e., the SES or the SEM considers the reciprocity. SEM allows for an interactive mechanism by taking into account the feedback among interdependent variables. In other words, SEM allows for two-way causality.

2) Endogenous variables (or jointly determined variables), exogenous variables (which are from outside the system without any reciprocal effect on the endogenous variable), lagged endogenous variables, pre-determined variables and an error variable to account for all residual effect.

3) A SES is said to be complete if it has as many equations as there are endogenous variables. Such a situation also helps us in getting unique estimates. Identification problem relates to getting unique estimates of structural parameters from the reduced equations parameters. This is possible only when the equations are exactly identified. In cases of over identification and under identification, we either get multiple estimates or no estimates respectively. In both these cases, we say that there is an identification problem.

Check Your Progress 2

1) Substitute the consumption function $\boldsymbol{c_t = \alpha_0 + \alpha_1 y_t + u_t}$ in $\boldsymbol{y_t = c_t + i_t}$. The reduced form equations are $\boldsymbol{y_t = \frac{a_0}{1-a_1} + \frac{1}{1-a_1} i_t + \frac{1}{1-a_1} u_t}$ and $\boldsymbol{c_t = \frac{a_0}{1-a_1} + \frac{a_1}{1-a_1} i_t + \frac{1}{1-a_1} u_t}$.

2) To form a matrix of rank *G–1*, we must have at least *G–1* number of columns in the Δ matrix. Thus, there should be at least (*G–1*) missing variables from the equation under consideration. This is known as the 'order condition' which is a necessary condition for identification.

3) A sufficient condition for an equation in the system to be identified is that the matrix Δ obtained after applying the rule for checking is of rank (G–1), where G is the number of endogenous variables in the equation system or the number of equations. This is known as the rank condition.

4) Arrange the equations as in equation (6.27). Prepare a table similar to Table 6.1. Strike out the relevant rows and columns. Compare the relationship between (K–M) and (G–1). This will give you the order condition. Prepare the determinants for the rank condition. The first and third equations are over-identified while the second equation is under-identified.

5) There are four equations. Hence, (G–1) = 3. Thus, at least 3 variables should be excluded from an equation for it to be identified. The first and fourth equations are exactly identified; the second equation is under-identified; and the third equation is over-identified.

Check Your Progress 3

1) Assuming that *B* is a non-singular matrix, the reduced form equations for the structural model indicated are: $\boldsymbol{YB + X\Gamma + E = 0}$, $\boldsymbol{YBB^{-1} + X\Gamma B^{-1} + EB^{-1} = 0}$, $\boldsymbol{Y + X\Gamma B^{-1} + EB^{-1} = 0}$, $\boldsymbol{Y = -X\Gamma B^{-1} - EB^{-1}}$ or $\boldsymbol{Y = X\pi + V}$ where *V* is the matrix of disturbance terms.

2) In the presence of a structural equation containing two or more endogenous variables, applying the least squares approach yields biased estimates. More specifically, to take an illustration, if we apply the OLS to the consumption

function [$\boldsymbol{C_t = \alpha + \beta Y_t + u_t}$], if we consider the *plim* for the estimate of $\hat{\boldsymbol{\beta}}$ obtained, even as n→∞ we find that $\left(\boldsymbol{plim\hat{\beta} - \beta}\right) \neq \boldsymbol{0}$. This violates the consistency property. This is known as the 'simultaneous equation bias' or the 'simultaneity bias'.

ignou
THE PEOPLE'S
UNIVERSITY