BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title : Econometrics

Course Code: 23ST03DEC

Unit-IV

Econometrics

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. E. Devi
Guest Faculty
Department of Statistics

UNIT – 4

SIMULTANEOUS EQUATIONS MODELS

Introduction

In economics, many variables are not independent. Instead, they influence each other at the same time. For example, the price of a product affects the quantity demanded, but at the same time, demand and supply together also decide the market price.

If we use simple regression in such cases, the results will be misleading because one variable is both a cause and an effect. To solve this problem, Econometrics uses a special approach called the Simultaneous Equations Model, or SEM.

Definition

A Simultaneous Equations Model is a system of two or more equations where some variables are endogenous, meaning they are determined inside the system, and others are exogenous, meaning they come from outside the system.

In SEM, the endogenous variables appear on both sides of the equations, showing the interdependence between them. These models are useful to study situations where variables determine each other simultaneously.

Examples

- Demand and Supply Model (Microeconomics)
- IS–LM Model (Macroeconomics)
  - IS stands for Investment–Saving curve
  - LM stands for Liquidity preference–Money supply curve
- Wages and Employment (Micro and Macro Link)

Simultaneous Equations Model Concept, Structure, Types:

A simultaneous equations model (SEM) in econometrics is a statistical framework used to analyze systems with multiple interdependent relationships between variables. Unlike simple regression, an SEM captures bidirectional causality, where a variable can be both a cause and an effect within the same system.

Concept:

The central idea is that many economic relationships are interconnected and determined simultaneously.

Consider a basic supply and demand model:

Demand equation : $Q_d = a - bP + cI + u_d$

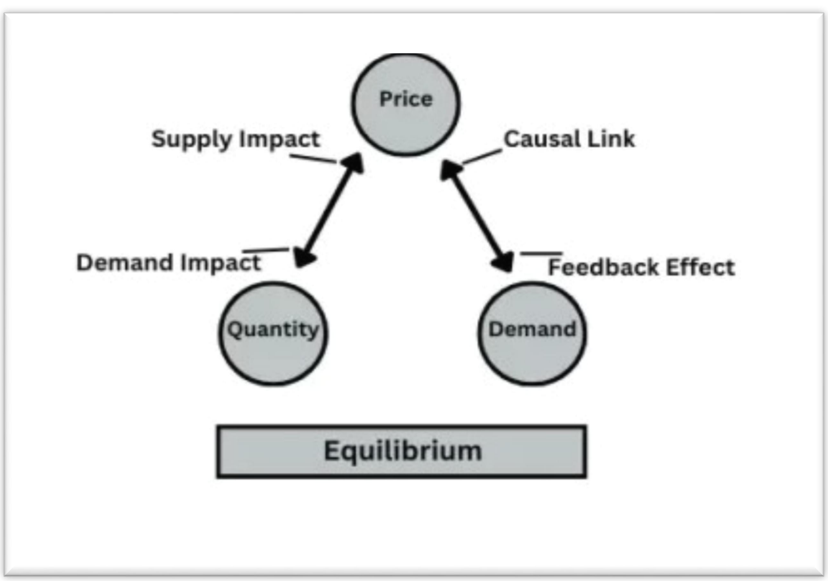
Supply equation: $Q_s = d + eP + u_s$

Equilibrium condition: $Q_d = Q_s$

Structure:

- Endogenous variables: Determined within the system by the equations themselves.
- Exogenous variables: Input variables determined outside the system.
- System of equations: Linked equations, e.g., supply and demand functions in economics or equations for speed and time in physics.

Simultaneous Equations Model



Types:

- Recursive (unidirectional): Each endogenous variable is a function of exogenous variables and possibly prior endogenous values, no feedback loops.
- Non recursive (bidirectional): Variables may influence each other simultaneously, requiring advanced estimation methods.
- Structural and reduced form: Structural form represents the original theoretical relationships, while reduced form expresses endogenous variables in terms of exogenous ones.

Real Time Examples:

Economics: Market equilibrium is determined by simultaneous equations for supply and demand, both dependent on price and quantity. For example, solving for equilibrium price and quantity when both demand and supply curves are given.

Physics: Forces acting on an object may be analyzed using a system of equations, e.g., solving for tension and gravity in pulleys.

Travel & Logistics: Calculate time, speed, and distance for multiple travelers or vehicles leaving and arriving at different times.

Finance: Comparing loan plans where monthly payments, interest rates, and duration are unknowns can be solved using simultaneous equations.

Engineering: Modeling electrical circuits, where current and voltage in a network are determined together by Kirchhoff's laws.

Business Decisions: Finding the best deal among competitors (e.g., cell phones or rental cars) by equating cost equations and solving for break-even points.

Identification problem with restriction on variance and covariance

Introduction:

- In many statistical and econometric models — especially in Structural Equation Modeling (SEM) and Simultaneous Equation Models (SEM in econometrics) — we often face the identification problem, where model parameters cannot be uniquely estimated from observed data.

- ❖ To solve this problem, researchers impose restrictions on variances and covariances of variables or error terms.
- ❖ These restrictions help the model become identified, meaning all parameters can be uniquely and consistently estimated.

Variance and Covariance Restrictions

Variance Restrictions

- Fixing a variance value (often to 1) helps define the scale of a latent variable.
  Example:
  In factor analysis, we set
  Var (F) = 1 so that the factor loading values are interpretable.

Covariance Restrictions

- Setting covariances to zero means assuming no correlation between certain variables or error terms.
  Example:
  Cov (ϵ_1, ϵ_2) = 0 assumes the measurement errors are independent.

These restrictions reduce the number of free parameters → allowing unique solutions.

Advantages

1. Helps in achieving model identification — makes it possible to estimate all parameters uniquely.
2. Provides a reference scale for latent variables (e.g., fixing variance to 1 defines their measurement scale).
3. Simplifies estimation by reducing the number of free parameters.
4. Makes the model more interpretable, especially when some covariances are fixed to zero (shows independence).

5. Avoids multicollinearity among variables by reducing unnecessary parameter correlations.
6. Allows easier model comparison between different studies or models with standardized scales.

Disadvantages

1. Some restrictions may be unrealistic and not reflect real-world data relationships.
2. Restrictions are often arbitrary — chosen for convenience rather than based on theory.
3. Can lead to model misspecification if wrong variances or covariance are fixed.
4. Causes loss of flexibility — true relationships might be ignored (e.g., setting covariance = 0 when it isn't).
5. The interpretation of fixed parameters (like variance = 1) may not have real meaning.
6. Results may become sensitive to the chosen restrictions, leading to biased or inconsistent estimates.

Applications

- Psychology, sociology, education, and behavioral sciences.
- Factor analysis for customer satisfaction, product perception, and market segmentation.
- Measurement of latent constructs such as quality of life, social inclusion, or job satisfaction.
- Latent Growth Curve Modeling to track individual growth, disease progression, or behavioral trends over time.
- Econometric policy models and asset pricing models where uncorrelated shocks are assumed.

RANK AND ORDER CONDITIONS TO DETERMINE IDENTIFICATION

We use the Rank and Order Conditions to identify whether a simultaneous equation model is underidentified, exactly identified or overidentified.

If an equation in the model does not have a unique statistical form, it means that the equation is underidentified. On the other hand, an identified equation has a unique statistical form and can either be exactly identified or overidentified. Moreover, we may need to employ different estimation techniques for exactly identified and overidentified models. The method of Indirect Least Squares (ILS) can be used to estimate coefficients of exactly identified models. However, ILS cannot be applied to overidentified models, which need to be estimated using other methods like 2SLS, 3SLS or Maximum Likelihood techniques.

Order Condition

The Order Condition is necessary but not a sufficient condition for the identification of the chosen equation. The formula for Order Condition is as follows:

$$K - M \geq E - 1$$

Where;

K is total number of variables in the model,

M is number of variables in equation under consideration and

E is number of equations in the model (which is equal to number of endogenous variables in model).

If this condition is not satisfied, we conclude that the equation is underidentified. If this condition is fulfilled, we move on to test the Rank Condition. Hence, the Order condition alone is not sufficient for identification.

Rank Condition

The Rank Condition states that an equation is identified if it has at least one determinant that is non-zero, from the matrix constructed by excluding coefficients from the given equation, but including coefficients in other equations of the model.

Hence, we need to construct a matrix of coefficients that contains only the coefficients of variables that are not present in the equation under consideration. Then, we calculate the determinant of that matrix which will be of the order $(E - 1)$If any of the determinants is non-zero, we conclude that the equation is identified. Conversely, if every possible determinant is zero, then the equation is underidentified.

Implementation Of Rank And Order Conditions Of Identification

To illustrate the implementation of these conditions, let us take the example of the following model:

$$D = \beta_1 + \beta_2 + \beta_3 Y + \beta_4 P_S + \mu$$

$$S = \alpha_1 + \alpha_2 P + \alpha_3 W + \vartheta$$

$$D = S$$

Where;

D is quantity demanded,

S is quantity supplied

P is price

P_S is price of substitute

Y is income and

W is labour wages.

This model is complete because it contains 3 equations and 3 endogenous variables (D,S and P). Furthermore, the P_S, Y and W variables are exogenous.

Demand Equation: Order Condition

Let us examine the identification of the demand equation. First, we examine the order condition:

$$K - M \geq E - 1$$

$$6 - 4 \geq 3 - 1$$

$$2 \geq 2$$

Therefore, the Order Condition is satisfied. Now, we move on to the Rank Condition of the demand equation.

Demand Equation: Rank Condition

To examine the rank condition, we need to form a matrix that includes coefficients of variables that are excluded from the demand equation. We rearrange the equations in the model to have all variables on the same side of the equation:

$$-D + \beta_1 + \beta_2 + \beta_3 Y + \beta_4 P_S + \mu = 0$$

$$-S + \alpha_1 + \alpha_2 P + \alpha_3 W + \vartheta = 0$$

$$-D + S = 0$$

First, let us express all the variables in every equation in the form of a table:

	Variables					
Equations	**D**	**P**	**Y**	P_s	**S**	**W**
Demand Equation	-1	0	β_3	β_4	0	0
Supply equation	0	a_2	0	0	-1	a_3
Equilibrium Equation	-1	0	0	0	1	0

In this table ,the Zero Values imply that the variables is not present in the equation . The values of other variables correspond to the coefficient of those variables.

From this table, we can construct the matrix required for the Rank Condition:

$$\begin{bmatrix} -1 & 0 & \beta_3 & \beta_4 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 & -1 & \alpha_3 \\ -1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

As we discussed, we need to form a matrix with coefficient of variables excluded from the equation i.e., demand equation in this case. To do this, we strike out the entire demand equation from the table because those variables are included in the equation. And, we remover all the variables or coefficients that are present in the demand equation .

Hence, we remove the first row and first to the fourth column to obtain the matrix of coefficient that are excluded from the demand equation and we end up with

$$\begin{bmatrix} -1 & \alpha_3 \\ 1 & 0 \end{bmatrix}$$

From this matrix, we can calculate one determinant of order 2, i.e. of order (E-1).

$$\Delta = \begin{bmatrix} -1 & \alpha_3 \\ 1 & 0 \end{bmatrix}$$

$$\Delta = (0 * -1) - (1 * \alpha_3)$$

Here;

$\Delta \neq \alpha_3$, Since, $\alpha_3 \neq 0$ we have a non-zero determinant and we conclude that the demand equation is identified.

Exactly-Identified or Overidentified

To check whether the demand equation is exactly identified or overidentified, we have to revisit the Order Condition:

If K – M = E – 1 ➔ exactly identified

If K – M > E – 1 ➔ overidentified

In our example, the demand equation is exactly identified because "2 = 2" (K−M=E−1).

We can follow this same procedure to check the identification of the supply function as well, where we apply the Order Condition followed by the Rank Condition.

Note: the problem of identification may arise for equations that have coefficients to be estimated. We do not need to check the identification of equilibrium equations. For instance, the 3rd equation in the above example is an equilibrium equation (D=S). Therefore, we do not need to check the identification of such an equation.

Indirect Least Squares Method

1. Definition

- **Indirect Least Squares (ILS)** is a method used to estimate parameters in a system of simultaneous equations.
- It reduces the system of equations to a single equation by solving one endogenous variable in terms of others, and then estimates the parameters using Ordinary Least Squares (OLS).
- ILS is useful when dealing with **endogeneity** (when variables are correlated with the error term) in simultaneous equations models.

2. Steps in Indirect Least Squares

1. **Rearrange the System of Equations**
   - Express one of the endogenous variables in terms of the exogenous variables and parameters.
2. **Eliminate Endogenous Variables**
   - Solve for endogenous variables in terms of exogenous variables (substitute endogenous variables in the system).
   - This will give the reduced form of the system.
3. **Apply OLS**
   - After substitution, the model should resemble a linear regression model with only exogenous variables.
   - Apply Ordinary Least Squares (OLS) to estimate the parameters of the reduced-form equation.

4. **Example of Application**

Consider the following simultaneous equations:

1. Equation 1: $Y = \alpha_1 + \beta_1 X + \epsilon_1$
2. Equation 2: $X = \alpha_2 + \gamma Y + \epsilon_2$

Step-by-Step Process

1. **Solve Equation 2 for X:**

$$X = \alpha_2 + \gamma Y + \epsilon_2$$

2. **Substitute X into Equation 1:**

$$Y = \alpha_1 + \beta_1(\alpha_2 + \gamma Y + \epsilon_2) + \epsilon_1$$

Simplify:

$$Y = \alpha_1 + \beta_1\alpha_2 + \beta_1\gamma Y + \beta_1\epsilon_2 + \epsilon_1$$

Group terms involving Y:

$$(1 - \beta_1\gamma)Y = \alpha_1 + \beta_1\alpha_2 + \beta_1\epsilon_2 + \epsilon_1$$

3. **Solve for Y:**

$$Y = \frac{\alpha_1 + \beta_1\alpha_2 + \beta_1\epsilon_2 + \epsilon_1}{1 - \beta_1\gamma}$$

4. **Estimate using OLS:**
   - Now Y is expressed in terms of exogenous variables.
   - Apply OLS to estimate the parameters of the reduced-form equation.

.

4. Advantages of Indirect Least Squares

- **Simplicity**: Useful for simple systems where the equations can be reduced to a single equation.
- **Alternative to 2SLS**: ILS can be a more straightforward alternative when the system is over-identified and can be reduced.
- **Applicable in Over-Identified Systems**: Works when you have more instruments than endogenous variables.

5. Limitations of Indirect Least Squares

- **Requires Identified System**: The system must be identified (i.e., it must be possible to solve for one endogenous variable and substitute it into the other equations).
- **Endogeneity**: If the instruments are weak or invalid (correlated with the error terms), estimates will be biased.
- **Complexity with More Equations**: As the number of equations increases, it becomes more difficult to apply ILS effectively.

6. When to Use ILS

- When dealing with over-identified systems where there are more instruments than endogenous variables.
- When the system of equations can be reduced to a form that allows for the estimation of parameters using OLS.

TWO STAGE LEAST SQUARES METHOD OF ESTIMATION

The two-stage least squares (2SLS) method is an essential technique in econometrics for dealing with endogeneity—when one or more explanatory variables in a regression model are correlated with the error term, leading to biased and inconsistent ordinary least squares (OLS) estimates.

Definition and Purpose

2SLS is an instrumental variable estimation method used to address the endogeneity problem in regression models. It provides consistent estimates of the model parameters, even when standard OLS fails due to endogeneity

When to Use 2SLS

2SLS is used when:

- There is suspected or known correlation between explanatory variables and the regression error.
- The model suffers from omitted variable bias, measurement error, or simultaneity bias (for example, in supply and demand models where variables like price and quantity are jointly determined).

The Two Stages Explained

First Stage: Instrumental Variable Regression

- Regress the endogenous explanatory variable(s) on the instrumental variable(s) (IVs) and all exogenous control variables.
- The instruments must be correlated with the endogenous regressor(s) and uncorrelated with the error term.
- Obtain predicted values of the endogenous variable(s), which represent only the variation in the endogenous variable that is due to the IVs.

Second Stage: Structural Equation Estimation

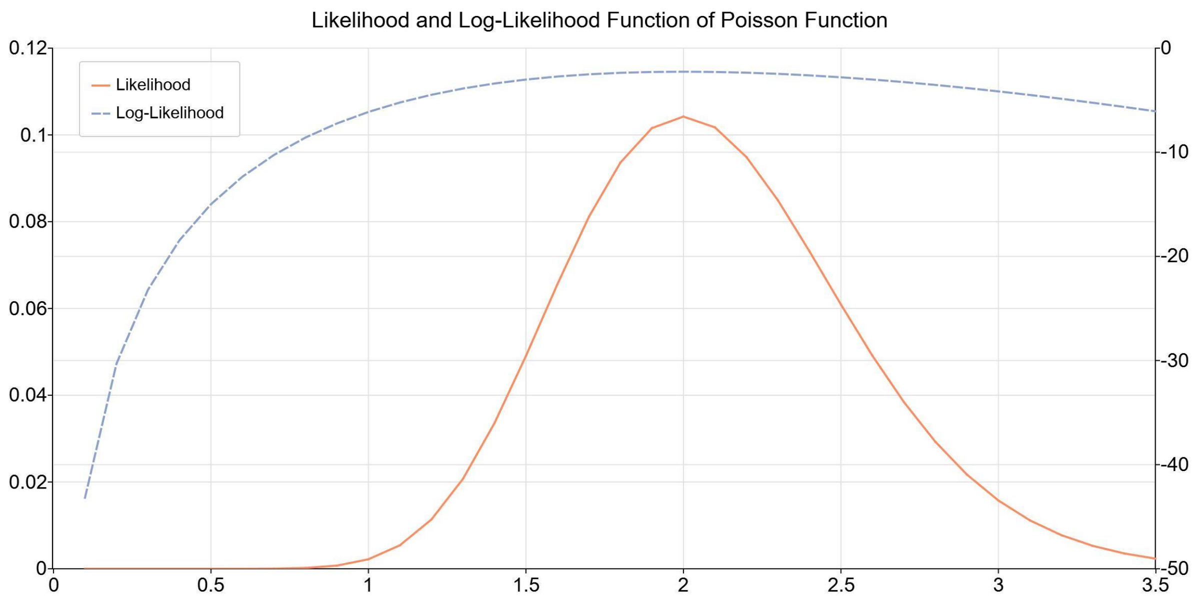
- Replace the original endogenous variable(s) in the main regression equation with the predicted values obtained from the first stage.
- Regress the dependent variable on these predicted values and the exogenous variables to estimate the structural parameters of interest.
- The coefficients from the second-stages regression are consistent estimates, even if OLS would be biased.

Estimation by Two-stage Least Square (2SLS):

It can be considered as a special case of the instrumental variables approach. In the instrumental approach one of the exogenous variables is used as an appropriate instrument. 2SLS essentially combines all the exogenous variable to create a combined variable. There are two major problems with the ILS technique. First, it can be applied to just identified, but not to over identified, equations. Second, the derivation of manipulation of the underlying structural equations can be difficult, since considerable algebraic to both of these problems is to apply two-stage least square (2SLS), instead of ILS, to a system of simultaneous equations. Two stage least square (2SLS), is useful in the case of an over-identified equation, and as a further alternative to an exactly identified equation. One important advantage of 2SLS over ILS is that 2SLS can be used to obtain consistent structural parameter estimates for over identified as well as for the exactly identified equations in a system of simultaneous equations. When the model is exactly identified, the two stage least square and the ILS give identical estimates. Another advantage is that 2SLS (but not ILS) gives the standard error if the estimated structural parameters directly. But 2SLS ignores information contained in correlation between errors. So, 2SLS is not efficient.

DEFINITION:

Limited Information Maximum Likelihood (LIML) is an estimation method used in simultaneous equation models that estimates the parameters of a single structural equation using only the information relevant to that equation, based on the principle of maximum likelihood.



- **Limited Information:**

  Refers to the estimation process focusing on a single equation or a small subset of equations within a larger system, rather than the entire system of equations simultaneously.

- **Maximum Likelihood (ML):**

  A general statistical method for estimating model parameters by finding the parameter values that maximize the likelihood function, which represents the probability of observing the data given the parameters.

- **LIML Estimator:**
  Combines these concepts, specifically for estimating a single equation in a linear simultaneous equation model. It finds the parameter values for that specific equation that maximize the likelihood function, while also accounting for the restrictions imposed by the structure of the overall model.

Principles of Limited Information Maximum Likelihood (LIML) –

1. **Limited Information**
   - Estimates one equation at a time.
   - Does not use the full system of equations.

2. **Maximum Likelihood**
    - Finds parameter values that make the observed data most likely.
    - Ensures efficient and reliable estimates.
3. **Instrumental Variables**
    - Uses instruments for variables that affect each other (endogenous).
    - Prevents biased or inconsistent results.
4. **Small-Sample Adjustment**
    - Includes a correction factor k to improve accuracy in small datasets.
    - Adjusts estimates slightly compared to 2SLS.
5. **Single-Equation Focus**
    - Each equation in the system can be estimated separately.
    - Saves time and simplifies analysis.
6. **Consistency and Efficiency**
    - Produces trustworthy estimates even with endogenous variables.
    - Often more efficient than 2SLS in small samples.
    - In large samples, LIML and 2SLS give similar result.

ADVANTAGES:

- Computationally simpler and faster, making it more practical for complex models.
- More robust to model misspecification in some cases.

DISADVANTAGES:

- Less efficient than FIML because it does not use all available information.
- May result in less accurate parameter estimates compared to FIML when full data is available.

Limited Information Maximum Likelihood (LIML)

Formula

The likelihood function is based on the conditional distribution of the endogenous variables given the instruments and exogenous variables:

$$L(\theta) = \prod_{i=1}^{n} f(y_i \mid z_i, x_i;\ \theta)$$

Where:

- y_i : vector of endogenous variables (dependent variables affected by error term)
- z_i : vector of instrumental variables (used to correct endogeneity)
- x_i : vector of exogenous explanatory variables (independent variables not correlated with error)
- θ : parameters to be estimated

Log-Likelihood Function

To make estimation easier, we take the natural log of the likelihood function:

$$\ell(\theta)=\sum_{i=1}^{n} \log f(y_i \mid z_i, x_i;\ \theta)$$

The LIML estimator, denoted as $\hat{\theta}$, maximizes this log-likelihood function $\ell(\theta)$.

Difference from Full Information Maximum Likelihood (FIML)

Feature	LIML	FIML
Information Used	Uses information from only one equation	Uses information from the entire system of equations
Complexity	Simpler	More complex
When Used	When limited info about the system is available	When full model structure is known