Advanced Econometrics
Simultaneous Equation Methods

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute

New Delhi

April 25, 2025

Contents

1 Introduction and Motivation

1.1 Why Simultaneous Equations?

Many economic relationships involve variables that are determined together, not in isolation. For example, in a market, price and quantity are determined simultaneously by supply and demand. In macroeconomics, output, consumption, and investment are jointly determined. Modeling such systems requires simultaneous equation models (SEMs), which allow for feedback and mutual causation.

1.2 Examples

- **Market Model:** Demand and supply for a commodity.
- **Klein's Model I:** A macroeconomic model for the U.S. economy (see Section 7).
- **Recursive Systems:** Causal chains, e.g., the CAPM as a recursive system.

1.3 Limitations of Single-Equation Methods

Ordinary Least Squares (OLS) assumes regressors are exogenous. In SEMs, endogenous regressors are correlated with the error term, leading to bias and inconsistency in OLS estimates.

2 Structure of Simultaneous Equation Models

2.1 General Form

A SEM consists of G equations in M endogenous variables $(Y_1, ..., Y_M)$ and K exogenous variables $(X_1, ..., X_K)$:

$$\begin{aligned} &y_{1t} + \gamma_{12} y_{2t} + \cdots + \gamma_{1M} y_{Mt} = \beta_{10} + \beta_{11} x_{1t} + \cdots + u_{1t} \\ &\vdots \\ &y_{Gt} + \gamma_{G2} y_{2t} + \cdots + \gamma_{GM} y_{Mt} = \beta_{G0} + \beta_{G1} x_{1t} + \cdots + u_{Gt} \end{aligned}$$

where u_{gt} are structural disturbances.

2.2 Classification of Variables

- **Endogenous variables**: Determined within the system (e.g., price, quantity).
- **Exogenous (predetermined) variables**: Determined outside the system (e.g., income, weather, policy).

2.3 Structural vs. Reduced Form

- **Structural form**: Original system of equations, with economic interpretation.
- **Reduced form**: Each endogenous variable is expressed as a function of all exogenous variables and disturbances.

2.4 Example: Supply and Demand

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + u_{1t}$$
$$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 W_t + u_{2t}$$
$$Q_t^d = Q_t^s = Q_t$$

Here, P_t and Q_t are endogenous; Y_t (income), W_t (wage) are exogenous.

3 Identification in Simultaneous Equation Models

3.1 The Identification Problem

Identification addresses whether it is possible to obtain unique estimates of structural parameters from observed data. If an equation is not identified, its parameters cannot be estimated, regardless of the estimation method or sample size.

3.2 Order Condition (Necessary)

- Let M be the number of endogenous variables in the system.
- For an equation to be identified, the number of excluded variables (endogenous and exogenous) from that equation must be at least $M - 1$.
- **Just identified:** Exactly $M - 1$ excluded.
- **Overidentified:** More than $M - 1$ excluded.
- **Underidentified:** Fewer than $M - 1$ excluded.

3.3 Rank Condition (Necessary and Sufficient)

- The matrix of coefficients of the excluded variables must have rank at least $M - 1$.
- If the order condition is not met, the rank condition will not be met.
- If the order condition is met, check the rank condition for sufficiency.

3.4 Practical Example

Suppose a three-equation system ($M = 3$). For an equation to be identified, it must exclude at least 2 variables from the system.

3.5 Illustrative Example: Keynesian Model

$$C_t = \beta_{10} + \beta_{11} Y_t + u_{1t}$$
$$I_t = \beta_{20} + \beta_{21} Y_t + \beta_{22} Y_{t-1} + u_{2t}$$
$$Y_t = C_t + I_t + G_t$$

G_t and Y_{t-1} are predetermined. Which equations are identified? (See exercises.)

4 Estimation Methods for Simultaneous Equation Models

4.1 Why OLS Fails

In SEMs, endogenous regressors are correlated with the error term, violating OLS assumptions and yielding biased, inconsistent estimates.

4.2 Single-Equation Methods

1. **Indirect Least Squares (ILS):** For exactly identified equations. Estimate the reduced form by OLS, then recover structural parameters.
2. **Two-Stage Least Squares (2SLS):** For overidentified (and exactly identified) equations. Replace endogenous regressors with their predicted values from first-stage regressions using all exogenous variables as instruments.

4.3 System Methods

- **Three-Stage Least Squares (3SLS):** Combines 2SLS with Seemingly Unrelated Regressions (SUR) to account for correlation across equations.
- **Full Information Maximum Likelihood (FIML):** Estimates all equations jointly, using the full system likelihood.
- **Limited Information Maximum Likelihood (LIML):** Estimates a single equation, accounting for system structure.

5 Two-Stage Least Squares (2SLS): Detailed Derivation and Steps

5.1 Basic Idea

2SLS is used to obtain consistent estimates when endogenous regressors are present. It uses instruments (exogenous variables) to "purge" endogeneity.

5.2 Step-by-Step 2SLS Procedure

1. **First Stage:** Regress each endogenous regressor in the structural equation on all exogenous variables in the system. Obtain fitted values.
2. **Second Stage:** Regress the dependent variable on the fitted values (from stage 1) and the included exogenous variables.

5.3 Mathematical Formulation

Suppose the structural equation is:

$$y_{1t} = \gamma_{12} y_{2t} + \beta_{11} x_{1t} + u_{1t}$$

where y_{2t} is endogenous, x_{1t} is exogenous.

First stage:

$$y_{2t} = \pi_{10} + \pi_{11}x_{1t} + \pi_{12}z_{1t} + v_t$$

where z_{1t} is another exogenous variable.

Second stage:

$$y_{1t} = \gamma_{12}\hat{y}_{2t} + \beta_{11}x_{1t} + \varepsilon_t$$

where $\hat{y}_{2t}$ is the fitted value from the first stage.

5.4 Properties and Interpretation

- 2SLS estimators are consistent and asymptotically normal.
- In exactly identified equations, 2SLS and ILS are equivalent.
- In overidentified equations, 2SLS uses all available instruments.

5.5 Instrumental Variables (IV) and 2SLS

- IV estimation is a special case of 2SLS with a single instrument.
- The relevance (correlation with endogenous regressor) and validity (uncorrelated with error) of instruments are crucial.

6 System Estimation: 3SLS, FIML, and LIML

6.1 Three-Stage Least Squares (3SLS)

- Combines 2SLS with Seemingly Unrelated Regressions (SUR).
- Accounts for correlation among error terms across equations.
- More efficient than 2SLS if errors are correlated.

6.2 Full Information Maximum Likelihood (FIML)

- Estimates all structural parameters jointly by maximizing the system likelihood.
- Requires correct specification of the entire system.
- Asymptotically efficient, but sensitive to misspecification.

6.3 Limited Information Maximum Likelihood (LIML)

- Estimates parameters of a single equation, accounting for the system structure.
- Less efficient than FIML but more robust to misspecification.

7 Detailed Example: Klein's Model I

7.1 Model Equations

Klein's Model I for the U.S. economy (1920–1941) is a classic SEM:

$$\begin{aligned} C_t &= a_0 + a_1 P_t + a_2 (W_t + W_t') + a_3 P_{t-1} + u_{1t} \\ I_t &= b_0 + b_1 P_t + b_2 P_{t-1} + b_3 K_{t-1} + u_{2t} \\ W_t &= c_0 + c_1 X_t + c_2 X_{t-1} + c_3 t + u_{3t} \end{aligned}$$

where C = consumption, I = investment, W = wage bill, P = profits, K = capital stock, X = output, t = time.

7.2 Estimation Results (from Gujarati Table 20.6)

OLS, Reduced Form, and 2SLS estimates are given for each equation. Discuss the bias in OLS, the consistency of 2SLS, and the interpretation of the reduced form.

7.3 Identification

- Apply order and rank conditions to each equation.
- Discuss which equations are exactly identified, overidentified, or underidentified.

7.4 Interpretation and Policy Implications

- Discuss economic meaning of coefficients.
- Analyze feedback between consumption, investment, and wage bill.

8 Special Topics

8.1 Recursive Systems

- Recursive systems have a unidirectional causal structure.
- OLS can be used in recursive systems, as there is no feedback.
- Example: Capital Asset Pricing Model (see Gujarati Example 20.3).

8.2 Simultaneity Bias and Endogeneity

- Simultaneity bias arises when an explanatory variable is jointly determined with the dependent variable.
- Endogeneity can also arise from omitted variables, measurement error, or reverse causality.
- IV/2SLS methods address endogeneity by using valid instruments.

8.3 Testing for Endogeneity

- Hausman test compares OLS and IV/2SLS estimates.
- Significant difference suggests endogeneity.

8.4 Weak Instruments

- Instruments must be strongly correlated with endogenous regressors.
- Weak instruments lead to biased and imprecise estimates.
- Use first-stage F-statistics to assess instrument strength.

9 Summary Table: Estimation Methods in SEMs

Method	Use Case	Identification	Notes
OLS	Recursive systems only	Any	Biased if simultaneity
ILS	Exactly identified	Exactly identified	Use reduced form
2SLS	Over/just identified	Over/just identified	Most common, uses instruments
3SLS	System estimation	Over/just identified	Efficient if errors correlated
FIML	Full system, all eqns	Any identified	Most efficient, sensitive to misspecificatio
LIML	Single equation	Over/just identified	Robust to misspecification

Glossary

- **Simultaneous Equation Model (SEM)**: A system of equations with mutually dependent endogenous variables.
- **Endogenous Variable**: Determined within the system.
- **Exogenous Variable**: Determined outside the system.
- **Identification**: Ability to uniquely estimate structural parameters.
- **Order Condition**: Necessary condition for identification based on excluded variables.
- **Rank Condition**: Necessary and sufficient condition for identification based on matrix rank.
- **Instrumental Variable (IV)**: Exogenous variable used to address endogeneity.
- **2SLS**: Two-stage least squares, a method for estimating SEMs.
- **3SLS**: Three-stage least squares, a system estimation method.
- **FIML**: Full information maximum likelihood estimation.
- **LIML**: Limited information maximum likelihood estimation.
- **Recursive System**: SEM with unidirectional causality.

Practice Questions

1. Explain the difference between structural and reduced-form equations in SEMs.
2. State and explain the order and rank conditions for identification.
3. Why is OLS generally inconsistent in simultaneous equation models?
4. Outline the steps of 2SLS and explain the role of instruments.
5. Using Klein's Model I, identify which equations are just identified, overidentified, or underidentified.
6. What are the main differences between 2SLS, 3SLS, FIML, and LIML?
7. Describe the problem of weak instruments and how to detect it.
8. Provide an example of a recursive system and explain why OLS is valid there.
9. How does the Hausman test help detect endogeneity?
10. Discuss the policy implications of simultaneous feedback in macroeconomic models.

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.