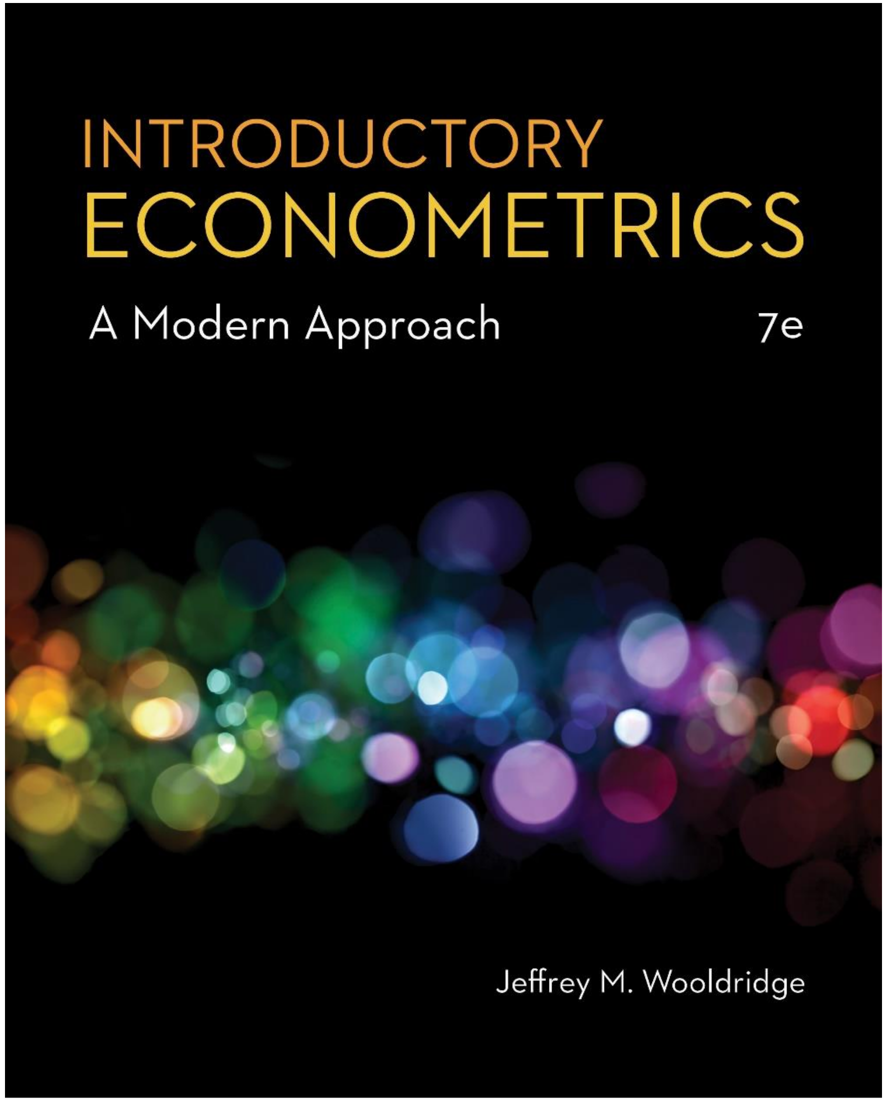
Chapter 16

Simultaneous Equations Models



Simultaneous Equations Models (1 of 14)

- **Simultaneity is another important form of endogeneity**
  - Simultaneity occurs if at least two variables are jointly determined.
  - A typical case is when observed outcomes are the result of separate behavioral mechanisms that are coordinated in an equilibrium.

  - The prototypical case is a system of demand and supply equations:
    - D(p) = how high would demand be if the price was set to p?
    - S(p) = how high would supply be if the price was set to p?
    - Both mechanisms have a ceteris paribus interpretation.
    - Observed quantity and price will be determined in equilibrium.

  - Simultaneous equations systems can be estimated by 2SLS/IV

Simultaneous Equations Models (2 of 14)

- **Example: Labor demand and supply in agriculture**

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1$$

- h_s: Annual labor hours supplied by workers in a given county if the average hourly wage offered to such workers is w
- α_1: labor supply elasticity
- z_1: Observed supply shifters e.g. manufacturing wage
- u_1: Unobserved supply shifters e.g. immigration flows

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2$$

- h_d: Annual hours demanded by employers in a given county if the average hourly wage paid to workers is w
- α_2: labor demand elasticity
- z_2: Observed demand shifters e.g. agricultural land area
- u_2: Unobserved demand shifters e.g. food market shocks

Simultaneous Equations Models (3 of 14)

- **Example: Labor demand and supply in agriculture (cont.)**

  Competition on the labor market in each county i will lead to a county wage w_i so that the total num-ber of hours h_{is} supplied by workers in this county equals the total number of hours h_{id} demanded by agricultural employers in this county:

$$h_{is} = h_{id} \;\; \Rightarrow \;\; (h_i, w_i)$$ (= observed equilibrium outcomes in each county)

<u>Simultaneous equations model (SEM):</u>

$$h_i = \alpha_1 w_i + \beta_1 z_{i1} + u_{i1}$$

$$h_i = \alpha_2 w_i + \beta_1 z_{i2} + u_{i2}$$

Endogenous variables

Exogenous variables

Structural error terms (uncorrelated with the exogenous variables)

<u>Note</u>: Without separate exogenous variables in each equation, the two equations could never be distiguished/separately identified

Simultaneous Equations Models (4 of 14)

- **Example: Murder rates and the size of the police force**

Murders per capita → $murdpc$; Police officers per capita → $polpc$; Income per capita → $incpc$

"Behavioral equation" of murderer population →

$$murdpc = \alpha_1\, polpc + \beta_{10} + \beta_{11} incpc + u_1$$

"Behavioral equation" of city government →

$$polpc = \alpha_2 murdpc + \beta_{20} + other\ factors$$

- polpc will not be exogenous because the number of police officers will dependent on how high the murder rate is ("reverse causation").

- The interesting equation for policy purposes is the first one. City governments will want to know by how much the murder rate decreases if the number of police officers is exogenously increased. This will be hard to measure because the number of police officers is not exogenously chosen (it depends on how much crime there is in the city, see second equation).

Simultaneous Equations Models (5 of 14)

- **Simultaneity bias in OLS**

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

Variable y_2 is correlated with the error u_1 because u_1 is indirectly a part of y_2. OLS applied to this equation will be therefore be inconsistent.

Insert the first equation into the second

$$\Rightarrow \; y_2 = \left[\frac{\alpha_2\beta_1}{1-\alpha_2\alpha_1}\right] z_1 + \left[\frac{\beta_2}{1-\alpha_2\alpha_1}\right] z_2 + \left[\frac{\alpha_2 u_1 + u_2}{1-\alpha_2\alpha_1}\right]$$

$$\Leftrightarrow \; y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2$$

(reduced form equation for y_2)

Simultaneous Equations Models (6 of 14)

- **Identification in simultaneous equations systems**
- Example: Supply and demand system

Supply of milk → $q = \alpha_1 p + \beta_1 z_1 + u_1$

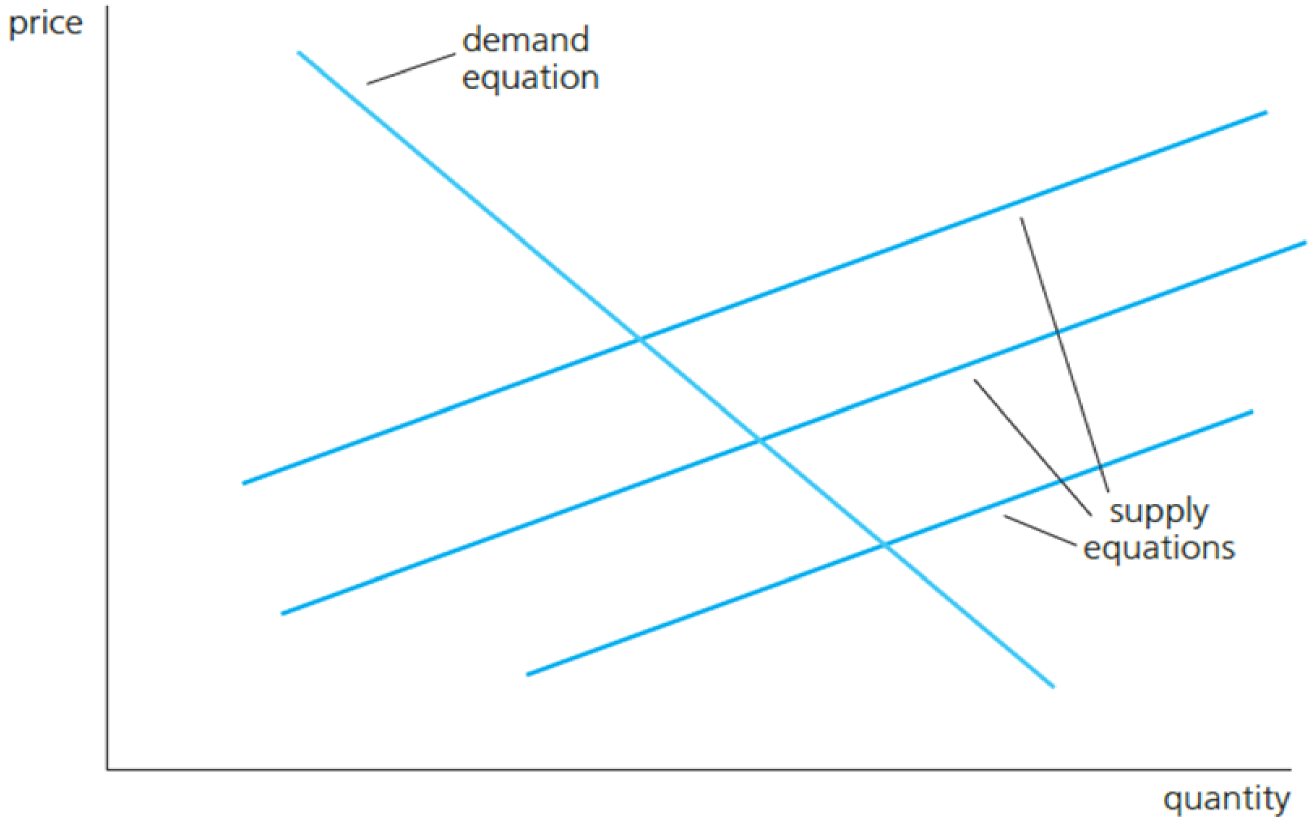
For example, price of cattle feed (→ z_1)

Demand for milk → $q = \alpha_2 p + u_2$

- Which of the two equations is identified?
  - The supply function cannot be consistently estimated because one of the regressors is endogenous and we do not have an instrument.
  - The demand function can be consistently estimated because we can take z_1 as an instrument for the endogenous price variable.

Simultaneous Equations Models (7 of 14)

- **Graphical illustration of identification problem**



- Intuitively, it is clear why the demand equation can be identified:
- We have an observed variable z_1 that shifts the supply equation while not affecting the demand equation.
- In this way the demand equation can be traced out.

Simultaneous Equations Models (8 of 14)

- **General rules for identification in simultaneous equation systems**

$$y_1 = \beta_{10} + \alpha_1 y_2 + \beta_{11} z_{11} + \beta_{12} z_{12} + \ldots + \beta_{1k_1} z_{1k_1} + u_1$$

$$y_2 = \beta_{20} + \alpha_2 y_1 + \beta_{21} z_{21} + \beta_{22} z_{22} + \ldots + \beta_{2k_2} z_{2k_2} + u_2$$

- Order condition
  - A necessary condition for the first equation to be identified is that at least one of all exogenous variables is excluded from this equation.
- Rank condition
  - The first equation is identified if, and only if, the second equation contains at least one exogenous variable that is excluded from the first equation.

Simultaneous Equations Models (9 of 14)

- **Example: Labor supply of married, working women**

Supply equation:

Hours and wages are endogenous (market equilibrium)

$$hours = \alpha_1 \log(wage) + \beta_{10} + \beta_{11} educ + \beta_{12} age$$
$$+ \beta_{13} kidslt6 + \beta_{14} nwifeinc + u_1$$

The supply equation is identified because it does not contain exper and $exper^2$

Wage offer equation:

$$\log(wage) = \alpha_2 hours + \beta_{20} + \beta_{21} educ$$
$$+ \beta_{22} exper + \beta_{23} exper^2 + u_2$$

The wage offer function is identified because it does not contain age, kidslt6, and nwifeinc

Simultaneous Equations Models (10 of 14)

- **Example: Labor supply of married, working women (cont.)**

  The rank condition, i.e. the condition that exogenous variables excluded from the equation are included in the other equation can be tested using reduced form equations.

$$\log(wage) = \pi_{20} + \pi_{21} educ + \pi_{22} age + \pi_{23} kidslt6 + \pi_{24} nwifeinc + \pi_{25} exper + \pi_{26} exper^2 + v_2$$

  Note that this equation can be consistently estimated by OLS as it contains only exogenous variables.

  The labor supply function is identified if the hypothesis $\pi_{25} = \pi_{26} = 0$ can be rejected. This is equivalent to rejecting $\beta_{22} = \beta_{23} = 0$ in the system of equations.

  The same argument applies to the identification of the wage offer function.

Simultaneous Equations Models (11 of 14)

- **Estimation of simultaneous equation systems by 2SLS**
  - Given the identification condition holds, the parameters of a simultaneous equations system can be consistenly estimated by 2SLS.
  - For this, in a first stage, each endogenous variable is regressed on the full list of exogenous variables (reduced form regressions).
  - In a second stage, the system equations are estimated by OLS but with the endogenous regressors being replaced by predictions from stage one.
  - If not all equations are identified, one can estimate only the identified ones.
  - If certain additional conditions hold, one can also use more efficient system estimation methods (Three Stage Least Squares, 3SLS).

Simultaneous Equations Models (12 of 14)

MROZ.dta

- **Example: Labor supply of married, working women using 2SLS**

$$\widehat{hours} = \underset{(574.56)}{2{,}225.66} + \underset{(470.58)}{1{,}639.56}\ \log(wage) - \underset{(59.10)}{183.75}\ educ$$

$$- \underset{(9.38)}{7.81}\ age - \underset{(182.93)}{198.15}\ kidslt6 - \underset{(6.61)}{10.17}\ nwifeinc,\ n = 428$$

$$\widehat{\log}(wage) = -\underset{(.338)}{.656} + \underset{(.00025)}{.00013}\ hours + \underset{(.016)}{.110}\ educ$$

$$+ \underset{(.019)}{.035}\ exper - \underset{(.00045)}{.00071}\ exper^2,\ n = 428$$

Simultaneous Equations Models (13 of 14)

- **Systems with more than two equations**
  - A necessary condition for identification of an equation is that there are more excluded exog. var. than endog. regressors (= order condition).
  - There is also a rank condition (but it is much more complicated).

- **Simultaneous equations models with time series**
  - Among the earliest applications of SEMs was the estimation of large systems of simultaneous equations for macroeconomic time series.
  - For a number of reasons, such systems are seldom estimated now.
  - The main problem is that most time series are not weakly dependent.
  - Another problem is the lack of enough exogenous variables.

Simultaneous Equations Models (14 of 14)

- **Example: A simple Keynesian model of aggregate demand**

Consumption → $C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + u_{t1}$

↑ Taxes (T_t) ↑ Interest rate (r_t)

Investment → $I_t = \gamma_0 + \gamma_1 r_t + u_{t2}$

Income → $Y_t \equiv C_t + I_t + G_t$ ← Government spending

Endogenous: C_t, I_t, Y_t Exogenous: T_t, G_t, r_t ← The exogeneity of these variables is very questionable