Chapter 14

Simultaneous Equations Models

14.1 Introduction

- **Simultaneous equations** models differ from those we have considered in previous chapters because in each model there are *two* or more dependent variables rather than just one.
- Simultaneous equations models also differ from most of the econometric models we have considered so far because they consist of a *set of equations*
- The least squares estimation procedure *is not* appropriate in these models and we must develop new ways to obtain reliable estimates of economic parameters.

14.2 A Supply and Demand Model

- Supply and demand *jointly* determine the market price of a good and the amount of it that is sold
- An econometric model that explains market price and quantity should consist of two equations, one for supply and one for demand.

$$\text{Demand:}\quad q = \alpha_1 p + \alpha_2 y + e_d \tag{14.2.1}$$

$$\text{Supply:}\quad q = \beta_1 p + e_s \tag{14.2.2}$$

- In this model the variables p and q are called **endogenous** variables because their values are determined within the system we have created.
- The income variable y has a value that is given to us, and which is determined outside this system. It is called an **exogenous** variable.

- Random errors are added to the supply and demand equations for the usual reasons, and we assume that they have the usual properties

$$\begin{aligned} &E(e_d) = 0, \quad \mathrm{var}(e_d) = \sigma_d^2 \\ &E(e_s) = 0, \quad \mathrm{var}(e_s) = \sigma_s^2 \\ &\mathrm{cov}(e_d, e_s) = 0 \end{aligned} \tag{14.2.3}$$

- The fact that *p* is random means that on the right-hand side of the supply and demand equations we have an explanatory variable that is random.
- This is contrary to the assumption of "fixed explanatory variables" that we usually make in regression model analysis.
- Furthermore *p* and the random errors, e_d and e_s, are correlated, making the least squares estimator biased and inconsistent.

14.3 The Reduced Form Equations

- The two structural equations (14.2.1) and (14.2.2) can be solved, to express the endogenous variables *p* and *q* as functions of the exogenous variable *y*.
- This reformulation of the model is called the **reduced form** of the structural equation system.
- To find the reduced form we solve (14.2.1) and (14.2.2) simultaneously for *p* and *q*.

- To solve for *p*, set *q* in the demand and supply equations to be equal,

$$\beta_1 p + e_s = \alpha_1 p + \alpha_2 y + e_d$$

- Then solve for *p*,

$$p = \frac{\alpha_2}{(\beta_1 - \alpha_1)} y + \frac{e_d - e_s}{(\beta_1 - \alpha_1)} \qquad (14.3.1)$$
$$= \pi_1 y + v_1$$

Substitute *p* into equation 14.2.2 and simplify.

$$q = \beta_1 p + e_s = \beta_1 \left[\frac{\alpha_2}{(\beta_1 - \alpha_1)} y + \frac{e_d - e_s}{(\beta_1 - \alpha_1)} \right] + e_s$$
$$= \frac{\beta_1 \alpha_2}{(\beta_1 - \alpha_1)} y + \frac{\beta_1 e_d - \alpha_1 e_s}{(\beta_1 - \alpha_1)} \qquad (14.3.2)$$
$$= \pi_2 y + v_2$$

- The parameters π_1 and π_2 in equations 14.3.1 and 14.3.2 are called reduced form parameters.
- The error terms v_1 and v_2 are called reduced form errors, or disturbance terms.

- The reduced form equations can be estimated consistently by least squares.
- The least squares estimator is BLUE for the purposes of estimating π_1 and π_2.

- The reduced form equations are important for economic analysis.
  - These equations relate the *equilibrium* values of the endogenous variables to the exogenous variables. Thus, if there is an increase in income y, π_1 is the expected increase in price, after market adjustments lead to a new equilibrium for p and q.
  - Secondly the estimated reduced form equations can be used to *predict* values of equilibrium price and quantity for different levels of income.

14.4 The Failure of Least Squares Estimation in Simultaneous Equations Models

$$\text{Demand:} \quad q = \alpha_1 p + \alpha_2 y + e_d \tag{14.2.1}$$

$$\text{Supply:} \quad q = \beta_1 p + e_s \tag{14.2.2}$$

- In the supply equation, (14.2.2), the random explanatory variable p on the right-hand side of the equation is *correlated* with the error term e_s.

14.4.1 An Intuitive Explanation of the Failure of Least Squares

- Suppose there is a small change, or blip, in the error term e_s, say Δe_s.
- The blip Δe_s in the error term of (14.2.2) is directly transmitted to the equilibrium value of p.
- This is clear from the reduced form equation 14.3.1. Every time there is a change in the supply equation error term, e_s, it has a direct linear effect upon p.
- Since $\beta_1 > 0$ and $\alpha_1 < 0$, if $\Delta e_s > 0$, then $\Delta p < 0$.
- Thus, every time there is a change in e_s there is an associated change in p in the opposite direction. Consequently, p and e_s are negatively correlated.

- Ordinary least squares estimation of the relation between q and p gives “credit” to price for the effect of changes in the disturbances.
- In large samples, the least squares estimator will tend to be negatively biased.
- This bias persists even when the sample size is large, and thus the least squares estimator is inconsistent.

The least squares estimator of parameters in a structural simultaneous equation is biased and inconsistent because of the correlation between the random error and the endogenous variables on the right-hand side of the equation.

14.4.2 An Algebraic Explanation of the Failure of Least Squares

- First, let us obtain the covariance between p and e_s.

$$
\begin{aligned}
\operatorname{cov}(p, e_s) &= E\left[p - E(p)\right]\left[e_s - E(e_s)\right] \\
&= E(pe_s) && \text{[since } E(e_s) = 0\text{]} \\
&= E\left[\pi_1 y + v_1\right] e_s && \text{[substitute for } p\text{]} \\
&= E\left[\frac{e_d - e_s}{\beta_1 - \alpha_1}\right] e_s && \text{[since } \pi_1 y \text{ is exogenous]} \\
&= \frac{-E\left(e_s^2\right)}{\beta_1 - \alpha_1} && \text{[since } e_d, e_s \text{ assumed uncorrelated]} \\
&= \frac{-\sigma_s^2}{\beta_1 - \alpha_1} < 0
\end{aligned}
\tag{14.4.1}
$$

- The least squares estimator in equation 14.2.2 is

$$b_1 = \frac{\sum p_t q_t}{\sum p_t^2} \tag{14.4.2}$$

- Substitute for q from equation 14.3.2 and simplify,

$$b_1 = \frac{\sum p_t (\beta_1 p_t + e_{st})}{\sum p_t^2} = \beta_1 + \sum \left(\frac{p_t}{\sum p_t^2} \right) e_{st} = \beta_1 + \sum h_t e_{st} \tag{14.4.3}$$

where

$$h_t = \frac{p_t}{\sum p_t^2}$$

- The expected value of the least squares estimator is

$$E(b_1) = \beta_1 + \sum E(h_t e_{st}) \neq \beta_1 \tag{14.4.4}$$

- The expectation $E(h_t e_{st}) \neq 0$ because e_s and p are correlated.

- In large samples there is a similar failure.
- Multiply through the supply equation by price, p, take expectations and solve.

$$\begin{aligned} pq &= \beta_1 p^2 + pe_s \\ E(pq) &= \beta_1 E(p^2) + E(pe_s) \\ \beta_1 &= \frac{E(pq)}{E(p^2)} - \frac{E(pe_s)}{E(p^2)} \end{aligned} \tag{14.4.5}$$

- In large samples, as $T \to \infty$, sample analogs of expectations, which are averages, converge to the expectations. That is, $\sum q_t p_t / T \to E(pq), \quad \sum p_t^2 / T \to E(p^2)$.
- Consequently, because the covariance between p and e_s is negative, from equation 14.4.1,

$$b_1 = \frac{\sum q_t p_t / T}{\sum p_t^2 / T} \to \frac{E(pq)}{E(p^2)} = \beta_1 + \frac{E(pe_s)}{E(p^2)} = \beta_1 - \frac{\sigma_s^2 / (\beta_1 - \alpha_1)}{E(p^2)} < \beta_1 \qquad (14.4.6)$$

The least squares estimator of the slope of the supply equation, in large samples, converges to a value less than β_1.

14.5 The Identification Problem

- In the supply and demand model given by equations 14.2.1 and 14.2.2, the parameters of the demand equation, α_1 and α_2, can not be consistently estimated by *any* estimation method.
- The slope of the supply equation, β_1, can be consistently estimated.
- The problem lies with the model that we are using. There is no variable in the supply equation that will shift it relative to the demand curve.
- It is the *absence* of variables from an equation that makes it possible to estimate its parameters. A general rule, which is called a condition for *identification* of an equation, is this:

A Necessary Condition for Identification: In a system of M simultaneous equations, which jointly determine the values of M endogenous variables, at least $M-1$ variables must be absent from an equation for estimation of its parameters to be possible. When estimation of an equation's parameters is possible, then the equation is said to be *identified*, and its parameters can be estimated consistently. If less than $M-1$ variables are omitted from an equation, then it is said to be *unidentified* and its parameters can not be consistently estimated.

- In our supply and demand model there are M=2 equations and there are a total of three variables: p, q and y.
- In the demand equation none of the variables are omitted; thus it is unidentified and its parameters can not be estimated consistently.
- In the supply equation M–1=1 and one variable, income, is omitted; the supply curve is identified and its parameter can be estimated.
- The identification condition must be checked <u>before</u> trying to estimate an equation.

Remark: The two-stage least squares estimation procedure is developed in Chapter 13 and shown to be an instrumental variables estimator. The number of instrumental variables required for estimation of an equation within a simultaneous equations model is equal to the number of right-hand-side endogenous variables. In a typical equation within a simultaneous equations model several exogenous variables appear on the right-hand-side. Thus instruments must come from those exogenous variables omitted from the equation in question. Consequently, identification requires that the number of omitted exogenous variables in an equation be at least as large as the number of right-hand-side endogenous variables. This ensures an adequate number of instrumental variables.

14.6 The Two-Stage Least Squares Estimation Procedure

- In this section we briefly describe *two-stage least squares* (2*SLS*) estimation

$$q = \beta_1 p + e_s \tag{14.2.2}$$

- The variable p is composed of a systematic part, which is its expected value $E(p)$, and a random part, which is the reduced form random error v_1.

$$p = E(p) + v_1 = \pi_1 y + v_1 \tag{14.6.1}$$

- In the supply equation (14.2.2) the portion of p that causes problems for the least squares estimator is v_1, the random part.
- Suppose we *knew* the value of π_1. Then we could replace p in (14.2.2) by (14.6.1) to obtain

$$
\begin{aligned}
q &= \beta_1[E(p) + v_1] + e_s \\
&= \beta_1 E(p) + (\beta_1 v_1 + e_s) \qquad (14.6.2) \\
&= \beta_1 E(p) + e_*
\end{aligned}
$$

- We could apply least squares to equation 14.6.2 to consistently estimate β_1.
- We can *estimate* π_1 using $\hat{\pi}_1$ from the reduced form equation for p.
- A consistent estimator for $E(p)$ is

$$\hat{p} = \pi_1 y$$

- Using $\hat{p}$ as a replacement for $E(p)$ in (14.6.2) we obtain

$$q = \beta_1 \hat{p} + e_* \qquad (14.6.3)$$

- In large samples, $\hat{p}$ and the random error $\hat{e}_*$ are uncorrelated, and consequently the parameter β_1 can be consistently estimated by applying least squares to (14.6.3).

- Estimating the equation 14.6.3 by least squares generates the so-called **two-stage least squares** estimator of β_1, which is consistent and asymptotically normal.

13.4 An Example of Two Stage Least Squares Estimation

- Consider a supply and demand model for truffles:

$$\text{demand:} \quad q_t = \alpha_1 + \alpha_2 p_t + \alpha_3 ps_t + \alpha_4 di_t + e_t^d \qquad (13.4.1)$$

$$\text{supply:} \quad q_t = \beta_1 + \beta_2 p_t + \beta_3 pf_t + e_t^s \qquad (13.4.2)$$

- In the demand equation *q* is the quantity of truffles traded in a particular French market at time *t*, *p* is the market price of truffles, *ps* is the market price of a substitute for real truffles (another fungus much less highly prized), and *di* is per capita disposable income.

- The supply equation contains the market price and quantity supplied. Also it includes *pf*, the price of a factor of production, which in this case is the hourly rental price of truffle-pigs used in the search process.
- In this model we assume that *p* and *q* are endogenous variables.
- The exogenous variables are *ps*, *di*, *pf* and the intercept variable.

13.4.1 Identification

- The rule for identifying an equation is that in a system of M equations at least $M-1$ variables must be omitted from each equation in order for it to be identified.
- In the demand equation the variable *pf* is not included and thus the necessary $M-1=1$ variable is omitted.
- In the supply equation both *ps* and *di* are absent; more than enough to satisfy the identification condition.
- We conclude that each equation in this system is identified and can thus be estimated by two-stage least squares.

13.4.2 The Reduced Form Equations

- The reduced form equations express each endogenous variable, *p* and *q*, in terms of the exogenous variables *ps*, *di*, *pf* and the intercept variable, plus an error term.

$$q_t = \pi_{11} + \pi_{21} ps_t + \pi_{31} di_t + \pi_{41} pf_t + v_{t1}$$

$$p_t = \pi_{12} + \pi_{22} ps_t + \pi_{32} di_t + \pi_{42} pf_t + v_{t2}$$

- Data for each of the endogenous and exogenous variables are given in Table 13.1.
- The price *p* is measured in \$ per ounce, *q* is measured in ounces, *ps* is measured in \$ per pound, *di* is in \$1000 and *pf* is the hourly rental rate for a truffle-finding pig.

Table 13.1 Sample Truffle Supply and Demand Data

OBS	*p*	*q*	*ps*	*di*	*pf*
1	**9.88**	**19.89**	**19.97**	**21.03**	**10.52**
2	**13.41**	**13.04**	**18.04**	**20.43**	**19.67**
27	**27.90**	**20.81**	**28.98**	**46.32**	**27.80**
28	**27.00**	**14.95**	**18.52**	**48.94**	**30.34**
29	**29.48**	**26.27**	**28.16**	**51.25**	**24.12**
30	**35.15**	**20.65**	**28.43**	**48.36**	**34.01**

Table 13.2a **Reduced Form Equation for Quantity of Truffles (*q*)**

Variable	*Estimate*	*Std. Error*	*t-value*	*p-value*
Const	**7.895**	**3.019**	**2.615**	**0.0147**
PS	**0.656**	**0.133**	**4.947**	**0.0001**
DI	**0.217**	**0.065**	**3.323**	**0.0026**
PF	**–0.507**	**0.113**	**–4.491**	**0.0001**

Table 13.2b **Reduced Form Equation for Price of Truffles (*p*)**

Variable	*Estimate*	*Std. Error*	*t-value*	*p-value*
Const	**–10.837**	**2.478**	**–4.374**	**0.0002**
PS	**0.569**	**0.109**	**5.229**	**0.0001**
DI	**0.253**	**0.054**	**4.736**	**0.0001**
PF	**0.451**	**0.093**	**4.872**	**0.0001**

- The reduced form equations are used to obtain $\hat{p}_t$ which will be used in place of p_t on the right-hand side of the supply and demand equations in the second stage of two-stage least squares.

$$\begin{aligned}\hat{p}_t &= \pi_{12} + \hat{\pi}_{22}\, ps_t + \pi_{32} di_t + \hat{\pi}_{42}\, pf_t \\ &= -10.837 + .569\, ps_t + .253 di_t + .451 pf_t\end{aligned}$$

- The *2SLS* results are given in Tables 13.3a and 13.3b.

Table 13.3a *2SLS* Estimates for Truffle Demand (q_d)

Variable	*Estimate*	*Std. Error*	*t-value*	*p-value*
Const	–4.279	5.161	–0.829	0.4145
P	–1.123	0.460	–2.441	0.0217
PS	1.296	0.331	3.919	0.0006
DI	0.501	0.213	2.359	0.0261

Table 13.3b *2SLS* Estimates for Truffle Supply (q_s)

Variable	*Estimate*	*Std. Error*	*t-value*	*p-value*
Const	20.033	1.160	17.264	0.0001
P	1.014	0.071	14.297	0.0001
PF	–1.001	0.078	–12.784	0.0001

- The estimated demand curve results are in Table 13.3a.
  - Note that the coefficient of price is negative, indicating that as the market price rises the quantity demanded of truffles declines, as predicted by the law of demand.
  - The standard errors that are reported are obtained from *2SLS* software.
  - They and the *t*-values are valid in large samples.
  - The *p*-value indicates that the estimated slope of the demand curve is significantly different from zero.
  - Increases in the price of the substitutes for truffles increase the demand for truffles, which is a characteristic of substitute goods.
  - Finally the effect of income is positive, indicating that truffles are a normal good.

- The supply equation results, appear in Table 13.3b.
  - As anticipated, increases in the price of truffles increase the quantity supplied,
  - and increases in the rental rate for truffle-seeking pigs, which is an increase in the cost of a factor of production, reduces supply.
  - Both of these variables have statistically significant coefficient estimates.