HANDOUT 08

SIMULTANEOUS EQUATIONS

Outline of this lecture:

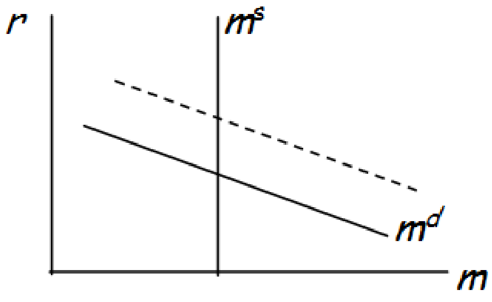
I. Simultaneity Bias: Inconsistency of OLS Estimator

So far, we only deals with single equation model

$$y = f(x) + u$$

y: endogenous, or jointly determined (effect)
x: exogenous, or predetermined (cause)

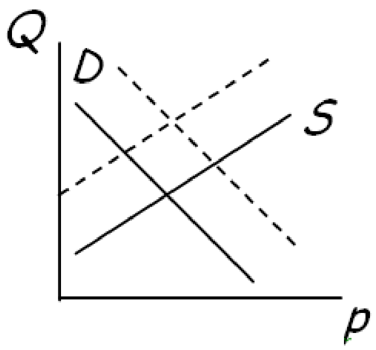
Problem There are situations where there is a two-way flow of influence among economic variables. Such as $M^d = f(r) + u$ where u is a disturbance term which denotes random shifts in M^d. r is assumed to be fixed in a single-equation model. But what happens if r depends on M^d? In a supply and demand model r depends on M^d.



If we get a positive money demand shock, $u > 0$, (or M^d increases), r will rise, thus r is positively correlated with u, which violates the classical assumption ($E(X_t u_t) = 0$), resulting in *inconsistent estimates* of parameters by OLS (so called *simultaneity bias*)

Example 1 Demand and Supply model

$$Q^d = a_1 + b_1 P + u_1 \quad b_1 < 0 \text{: demand function}$$
$$Q^s = a_2 + b_2 p + u_2 \quad b_2 > 0 \text{: supply function}$$
$$Q^d = Q^s \text{ : equilibrium condition}$$



Because of the simultaneous dependence between Q and P, u_1 and P in the demand function, and u_2 and P in the supply function cannot be independent.

Example 2 Keynesian model of income

Consumption function: $C = \beta_0 + \beta_1 Y + u$, $0 < \beta_1 < 1$
Income identity: $Y = C + I + G$

How to estimate?

Since OLS estimates are biased, we have to use some other methods, such as *indirect least squares*. The idea is to regressing each endogenous variable on all the exogenous variables by OLS (so called *reduced-form equation*), then recover the parameters in the original equation (so called *structural*, or *behavior equation*). However, this is not always possible because of what is known as *identification problem*.

II. Identification Problem

By the identification problem we mean whether numerical estimates of the parameters of a structural equation can be obtained from the estimated reduced-form coefficients. If this can be done, we say that the particular equation is identified. If this cannot be done, then we say that the equation under consideration is unidentified, or underidentified.

An identified equation may be either exactly (or fully or just) identified or overidentified. It is said to be exactly identified if unique numerical values of the structural parameters can be obtained. It is said to be overidentified if more than one numerical value can be obtained for some of the parameters of the structural equations. The circumstances under which each of these cases occurs will be shown in the following discussion.

The identification problem arises because different sets of structural coefficients may be compatible with the same set of data.

Example 1 y: income R: rainfall

$$\left.\begin{aligned} q &= a_1 + b_1 P + c_1 y + u_1 : \text{demand function} \\ q &= a_2 + b_2 p + c_2 R + u_2 : \text{supply function} \end{aligned}\right\} \text{structural equations} \qquad (1)$$

Solve these two structural equations for P and q in terms of y and R

$$\left.\begin{aligned} q &= \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y - \frac{c_2 b_1}{b_2 - b_1} R + V_1 \\ p &= \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y - \frac{c_2}{b_2 - b_1} R + V_2 \end{aligned}\right\} \text{reduced form equations } (2)$$ [1]

or

$$\left.\begin{aligned} q &= \Pi_1 + \Pi_2 y + \Pi_3 R + V_1 \\ p &= \Pi_4 + \Pi_5 y + \Pi_6 R + V_2 \end{aligned}\right\} \qquad (3)$$

where $\Pi_1 = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1}$, $\Pi_2 = \frac{c_1 b_2}{b_2 - b_1}$, etc: reduce-form parameters.

The estimation of the equations by OLS gives consistent estimates of the reduced form parameters because y and R are exogenous and are independent to the disturbances.

Comparing (2) with (3) we get $\hat{b}_1 = \frac{\hat{\Pi}_3}{\hat{\Pi}_6}$, $\hat{b}_2 = \frac{\hat{\Pi}_2}{\hat{\Pi}_5}$

$\hat{c}_1 = -\hat{\Pi}_5(\hat{b}_1 - \hat{b}_2)$, $\hat{c}_2 = \hat{\Pi}_6(\hat{b}_1 - \hat{b}_2)$, $\hat{a}_1 = \hat{\Pi}_1 - \hat{b}_1\hat{\Pi}_4$, $\hat{a}_2 = \hat{\Pi}_1 - \hat{b}_2\hat{\Pi}_4$

$\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{c}_1, \hat{c}_2$ (Single-valued function of the $\hat{\Pi}$s) are consistent estimates of the corresponding structural parameters (Indirect Least

[1] ***Reduced-form equation*** It expresses the endogenous variable solely as a function of the exogenous (or predetermined) variables and the stochastic disturbance term u. The reduced-form coefficients are also known as impact, or short-run, multipliers, because they measure the immediate impact on the endogenous variable of a unit change in the value of the exogenous variable

Squares Method). Both demand and supply functions are exactly identifies in this case.

- When unique estimates for the structural parameters of an equation are got from the reduced-form parameters, the equation is exactly identified.

- When multiple estimates for the structural parameters of an equation are got from the reduced-form parameters, the equation is over-identified.

- When no estimates for structural parameters can be got from parameters, the equation is underidentified (or not identified).

Example 2

$q = a_1 + b_1 P + c_1 y + u_1$: demand function
$q = a_2 + b_2 p + u_2$: supply function

The reduced-form is

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y + V_1$$

$$p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y + V_2$$

or $$q = \Pi_1 + \Pi_2 y + V_1$$
$$p = \Pi_3 + \Pi_4 y + V_2$$

$\hat{b}_2 = \hat{\Pi}_2 / \hat{\Pi}_4$, $\hat{a}_2 = \hat{\Pi}_1 - \hat{b}_2 \hat{\Pi}_3$, but there's no way of getting $\hat{a}_1, \hat{b}_1, \hat{c}_1$ thus the supply function is identified, but the demand function is not.

Example 3

$q = a_1 + b_1 p + u_1$ demand function (identified)
$q = a_2 + b_2 p + cR_2 + u_2$ supply function (not identified)

Example 4

D: $q = a_1 + b_1 p + c_1 y + d_1 R + u_1$ (people do less shopping if it rains)
S: $q = a_2 + b_2 p + u_2$

Reduce form equations

$$q = \frac{a_1 b_2 - a_2 b_1}{b_2 - b_1} + \frac{c_1 b_2}{b_2 - b_1} y + \frac{d_1 b_2}{b_2 - b_1} R + V_1$$
$$p = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} y + \frac{d_1}{b_2 - b_1} R + V_2$$

or
$$q = \Pi_1 + \Pi_2 y + \Pi_3 R + V_1$$
$$p = \Pi_4 + \Pi_5 y + \Pi_6 R + V_2$$

$\hat{b}_2 = \hat{\Pi}_2 / \hat{\Pi}_5$, $\hat{b}_2 = \hat{\Pi}_3 / \hat{\Pi}_6$, two different estimates of b_2, thus two different estimates of a_2 (since $\hat{a}_2 = \hat{\Pi}_1 - \hat{b}_2 \hat{\Pi}_4$). Therefore, supply function is over-identified.

No estimates for a_1, b_1, c_1 and d_1: demand function is not identified.

A. Identification Conditions

1. Order condition (*<u>necessary but not sufficient</u>*)
2. Rank condition (*necessary and sufficient*)

1. Order Condition

In a model of M simultaneous equations in order for an equation to be identified, it must exclude at least $M-1$ variables (endogenous as well as predetermined or exogenous) appearing in the model. If it excludes exactly $M-1$ variables, the equation is *just identified*. If it excludes more than $M-1$ variables, it is *overidentified*.

Let k be the number of variables (endogenous and exogenous) missing from the equation under consideration. Then

- If $k = M - 1$, the equation is exactly identified
- If $k > M - 1$, the equation is overidentified
- If $k < M - 1$, the equation is underidentified.

Example 1

Demand: $k = 1$ (R), $M = 2$ (q, p), $k = M - 1$, exactly identified
Supply: $k = 1$ (y), $M = 2$ (q, p), $k = M - 1$, exactly identified

Example 2

Demand: $k = 0$, $M = 2$ (q, P), $k < M - 1$, underidentified
Supply: $k = 1$ (y), $M = 2$ (q, p), $k = M - 1$, exactly identified

Example 3

Demand: $k = 1$ (R), M=2, $k = M - 1$, exactly identified
Supply: $k = 0$, M=2, $k < M - 1$, not identified

Example 4

Demand: $k = 0$, $M = 2$, $k < M - 1$, not identified
Supple: $k = 2$ (y, R), $M = 2$, $k > M - 1$, over-identified

Order Condition is a necessary but not sufficient condition for identification. Even if *order condition* holds the equation may be unidentified, because the exogenous variables excluded from an equation may not all be independent so there may not be one-to-one correspondence between the structural coefficients and the reduced-form coefficients.

2. Rank Condition

The order condition discussed previously is a necessary but not sufficient condition for identification; that is, even if it is satisfied, it may happen that an equation is not identified.

Consider a simultaneous equation system of three endogenous variables y_1, y_2, y_3 and three exogenous variables z_1, z_2, z_3:

$$y_1 = a_1 + a_2 z_1 + a_3 z_3 + u_1$$
$$y_2 = b_1 + b_2 y_3 + b_3 z_1 + b_4 z_2 + u_2$$
$$y_1 = c_1 + c_2 y_3 + c_3 z_1 + c_4 z_3 + u_3$$

Mark with 1 (or the coefficient) if a variable appears in an equation and a 0 if not

Equation	y_1	y_2	y_3	z_1	z_2	z_3
1	1	0	0	a_2	0	a_3
2	0	1	b_2	b_3	b_4	0
3	1	0	c_2	c_3	0	c_4

The rank condition for identification of an equation is as follows:

1) Delete the particular row (say the 1st row for the 1st equation)
2) Pick up the columns corresponding to the zero elements in that row (excluded variables, y_2, y_3, z_2 in the 1st equation)
3) Consider this array of columns, if there are at least (M - 1) rows *and* columns that are not all zeros (where M is the number of endogenous variables) and no column (or row) is proportional to another column (or row) for all parameter values, then the equation is identified. Otherwise not.

Hence:

- 1st Equation $\begin{pmatrix} 1 & b_2 & b_4 \\ 0 & c_2 & 0 \end{pmatrix}$, M–1=3–1=2, two rows with not all elements zero and no row is proportional to another row, the 1st equation is identified. In fact, according to order condition, k=3>M–1=2, this equation is over-identified.

- 2nd Equation $\begin{pmatrix} 1 & a_3 \\ 1 & c_4 \end{pmatrix}$, two rows with not all elements zero and no row (or column) is proportional to the other, the 2nd equation is identified; According to order condition $k = M – 1 = 2$, it is exactly identified.

- 3rd Equation $\begin{pmatrix} 0 & 0 \\ 1 & b_4 \end{pmatrix}$ only one row with not all elements zero, M-1=2>1, not identified. According to order condition k=2=M–1=2, it is identified (therefore, order condition is not sufficient).

Another definition of rank condition is as follows (you are not responsible from this):

> *In a model containing M equations in M endogenous variables, an equation is identified if and only if at least one nonzero determinant of order (M−1)(M−1) can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from that particular equation but included in the other equations of the model.*

Example 1

Demand function $q = a_1 + b_1 p + c_1 y + u_1$

Supply function $q = a_2 + b_2 p + c_2 R + u_2$

	q	P	y	R	rank condition	order condition
Demand	1	b_1	c_1	0	1 = 2 – 1: identified	identified
Supply	1	b_2	0	c_2	1 = 2 – 1: identified	identified

Example 2

Demand $q = a_1 + b_1 p + c_1 y + u_1$

Supply function $q = a_2 + b_2 p + u_2$

	q	p	y	rank condition	order condition
Demand	1	b_1	c_1	$0 \neq 2$ - 1: not identified	0<2: not identified
Supply	1	b_2	0	1 = 2 – 1: identified	1=2-1: identified

Example 3

Demand $q = a_1 + b_1 p + u_1$

Supply $q = a_2 + b_2 p + c_2 R + u_2$

	q	p	R	rank condition	order condition
Demand	1	b_1	0	identified	identified
Supply	1	1	1	not identified	not identified

Example 4

Demand $q = a_1 + b_1 p + c_1 y + d_1 R + u_1$

Supply $q = a_2 + b_2 p + u_2$

	q	p	y	R	rank condition	order condition
Demand	1	b_1	c_1	d_1	not identified	not identified
Supply	1	b_2	0	0	identified (1=2-1)	overidentified (2>2-1)

Remarks

- *Rank condition tells us whether the equation is identified or not.*
- *Order condition tells us if it is just-identified (exactly identified) or over-identified.*

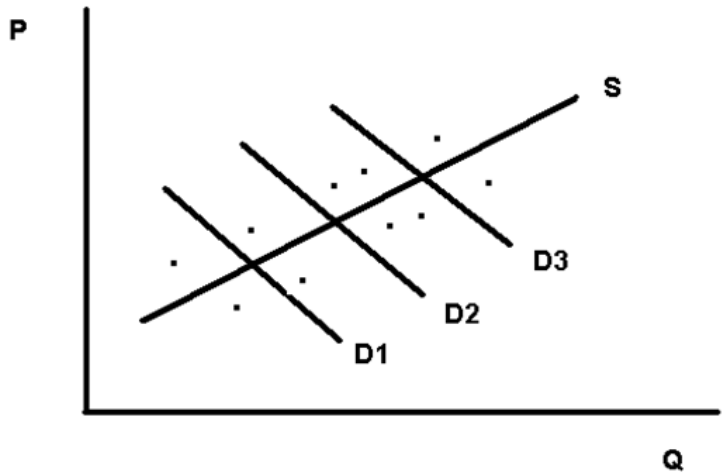
III. More on Identification Problem

The identification problem was first brought to the attention of economists in an article by the agricultural economist E.J. Working, which was published in 1927 and is still worth reading today. Using simple supply and demand, Working was able to show that statistical estimates which were claimed to be of the parameters of demand functions may not be so after all. This difficulty is not specific to supply and demand models of markets and can be found in all areas of economics such as the IS-LM model, the Phillips curve and etc.

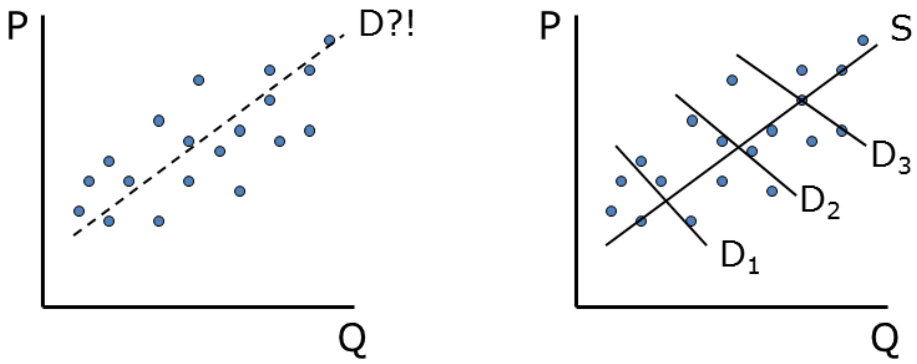
The crux of the identification problem is as follows: Recall the demand-and-supply model. Suppose that we have time series data on Q and P only and no additional information (such as income of the consumer, price prevailing in the previous period, and weather condition). The identification problem then consists in seeking an answer to this question: Given only the data on P and Q, how do we know whether we are estimating the demand function or the supply function? Alternatively, if we think we are fitting a demand function, how do we guarantee that it is, in fact, the demand function that we are estimating and not something else?

As an example assume that you want to estimate the electricity demand curve for your ECON302 term project. You collect the monthly data on the electricity prices and electricity consumption for, say, 5 years. You expect to find a negative relationship—if prices are high, the consumptions should be low. Unfortunately, the world tends to be more complicated than that: you estimate the demand curve and you find a positive relationship between prices and consumption levels! What is happening? You may start to insult econometrics. However, do not be impatient and hasty!

In order to understand the point think of the standard market supply and demand curves: price on the vertical axis, quantity on the horizontal axis, demand slopes down, supply slopes up; market price and quantity is where the two curves intersect. Suppose that during the period you are studying, due to the nationwide increase in the per capita GDP, the demand for the electricity has been increased: the demand curve has shifted right several times.

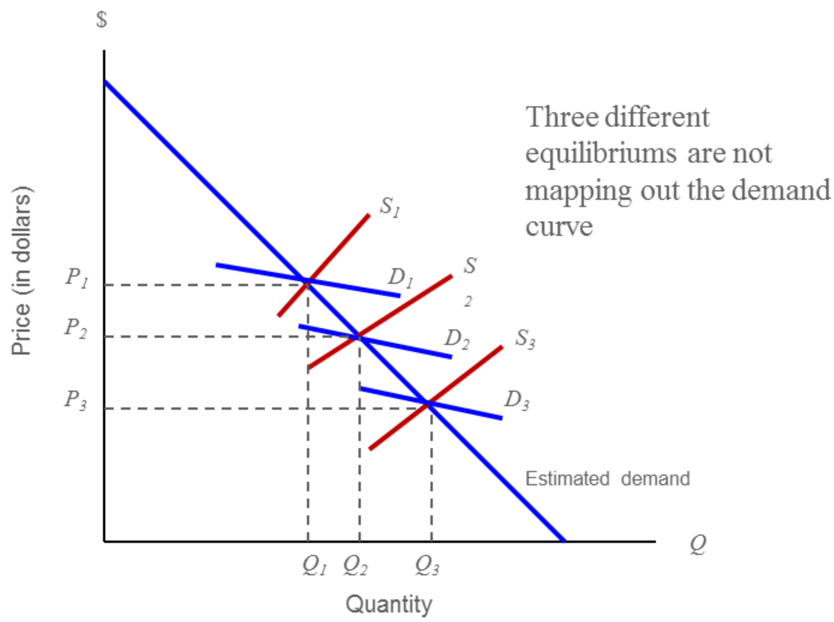


As a result of this development the higher prices are associated with higher electricity consumptions. As can be seen from the figure above that in this example since the demand curve shifts right several times, the data that you have for the last years traces the supply curve not the demand curve! That is why you obtain a positive relationship between prices and consumption levels: in fact you are estimating the supply curve not the demand curve!



Can demand be upward sloping?!

OR...?

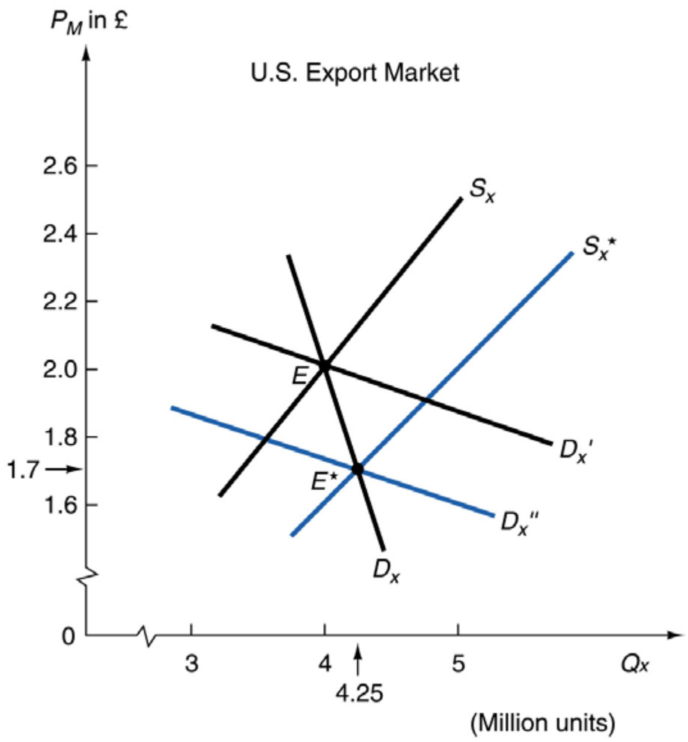


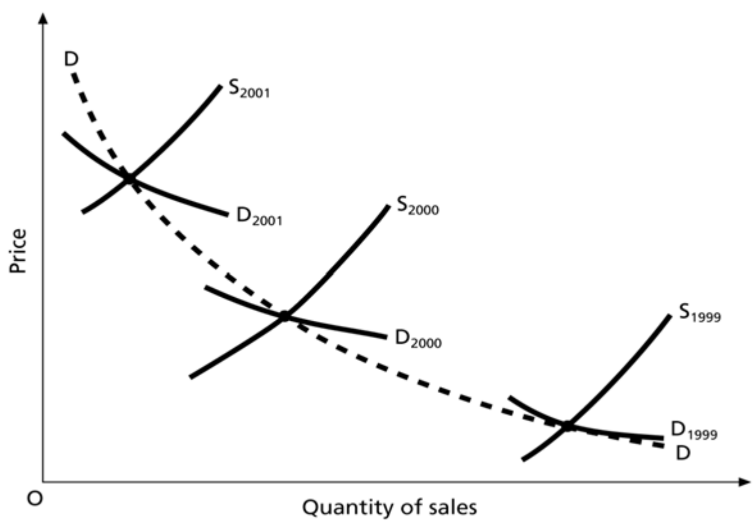
The problem is that if the demand curve is shifting, points on the supply curve are traced out, but if the supply curve is shifting, points on the demand curve are traced out. In the first case, the supply curve is "identified;" in the second, you can "identify" the demand curve. In general, though, both supply and demand curves can be shifting at the same time, and it can be difficult to separate the two effects. That is known as the "*identification problem*" in economics.

If we simply put: the difficulty of deriving the demand curve for a commodity from observed priced-quantity points that results from the intersection of different and unobserved demand and supply curves for the commodity is referred to as the identification problem.

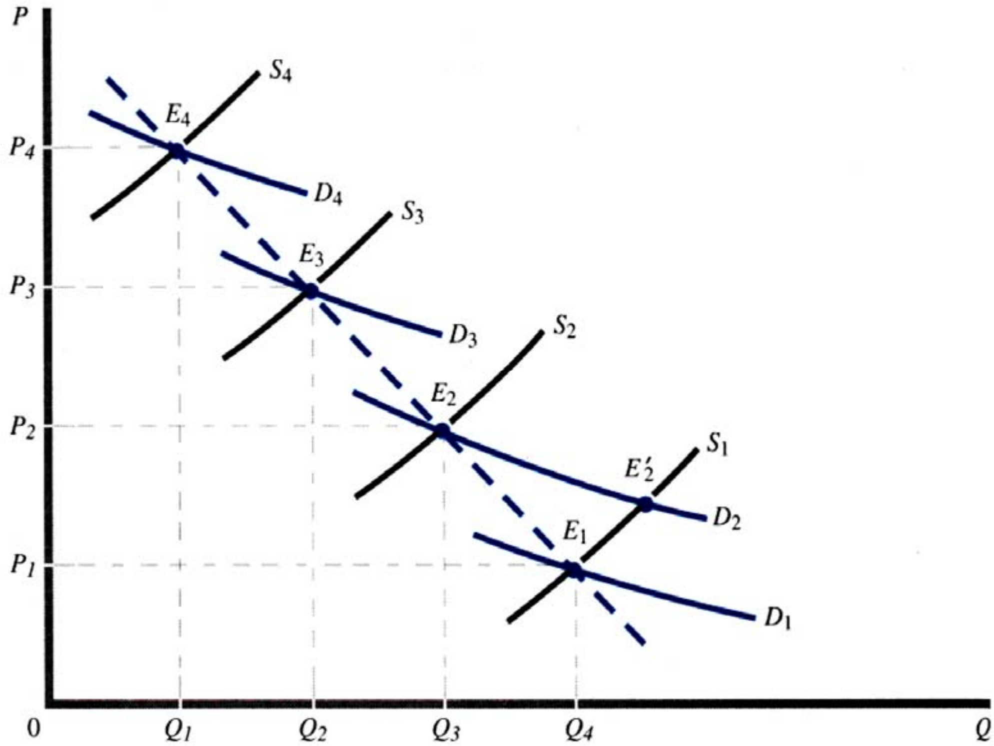
The demand curve for a commodity is generally estimated from market data on the quantity purchased of the commodity at various price over time (i.e. Time-series data) or various consuming units at one point in time (i.e. Cross-sectional data). Simply joining priced-quantity observations on a graph does not generate the demand curve for a commodity. The reason is that each priced-quantity observation is given by the intersection of a different and unobserved demand and supply curve of commodity. In other words, the difficulty of deriving the demand curve for a commodity from observed priced-quantity points that results from the intersection of different and unobserved demand and supply curves for the commodity is referred to as the identification problem.

It can be said that supply is identified when demand equation includes at least one exogenous variable that is not also in the supply equation (such as "per capita GDP" or simply "income" in our example). Likewise, the demand is identified when supply equation includes at least one exogenous variable that is not also in the demand equation.





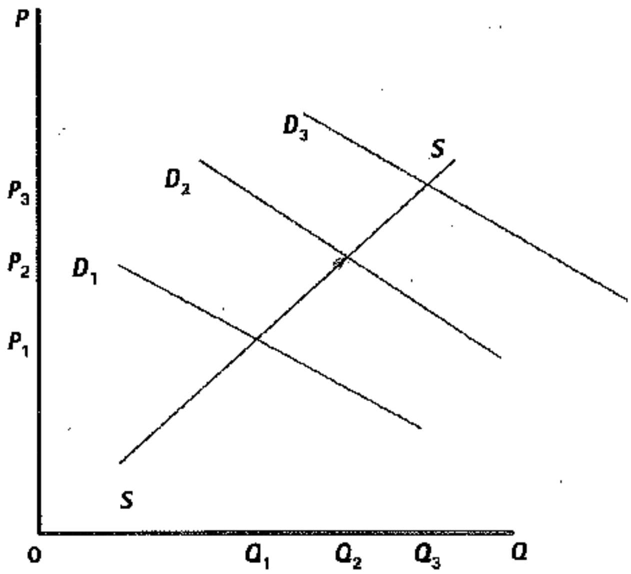
In the following demand curve, observed price-quantity data points E_1, E_2, E_3, and E_4, result respectively from the intersection of unobserved demand and supply curves D_1 and S_1, D_2 and S_2, D_3 and S_3, and D_4 and S_4. Therefore, the dashed line connecting observed points E_1, E_2, E_3, and E_4 is not the demanded curve for the commodity. To derive a demand curve for the commodity, say, D_2, we allow the supply to shift or to be different and correct, through regression analysis, for the forces that cause demand curve D_2 to shift or to be different as can be seen at points E_2, E'_2. This is done by regression analysis.



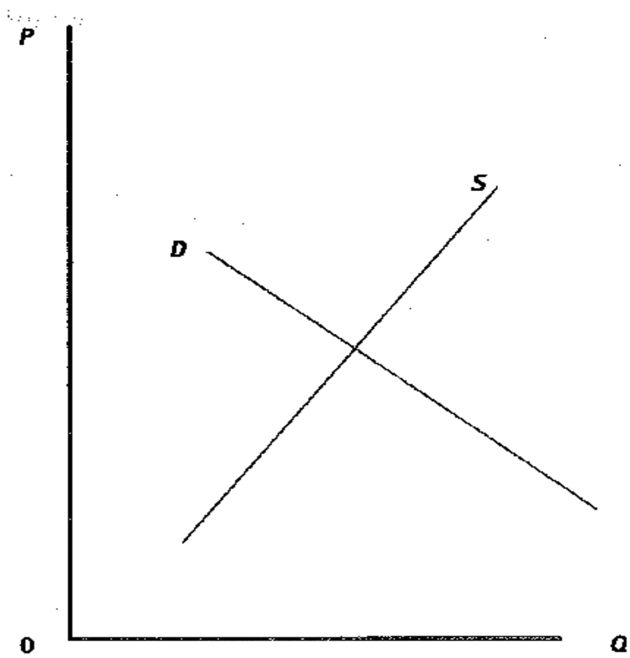
A. Diagrammatic Illustrations

1. Supply identified but demand not identified

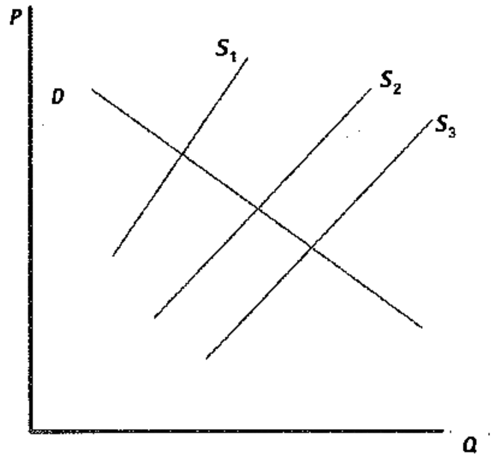
It seems that the supply relationship can be identified because the movements in the demand schedule trace out a path representing the supply schedule. Conversely, the demand schedule cannot be identified because the price-quantity relationship for demand cannot be observed.



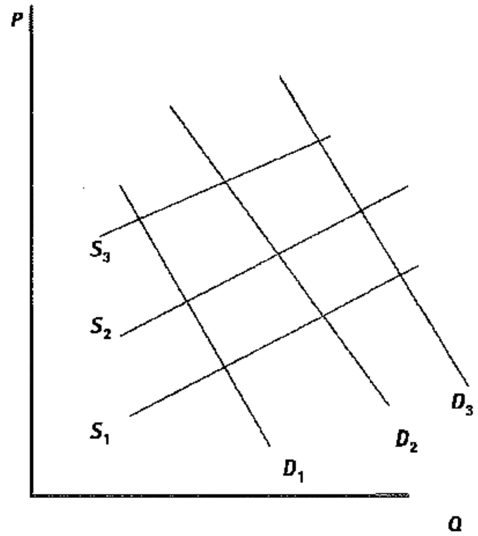
2. Neither equation identified



3. Demand identified but supply function not identified



4. Both equations identified



IV. Methods of Estimation

*Single Equation (*or *limited information*) *methods*: estimate each equation separately using only the information about the restrictions on the coefficients of that particular equation, not other equations.

System (or *full information*) *methods*: use information about the restrictions on all equations.

System methods are not commonly used for a variety of reasons:

1) *Enormous computational burden*
    - The comparatively small (20 equations) 1955 Klein-Goldberger model of the US economy had 151 nonzero coefficients, of which the authors estimated only 51 coefficients using the time-series data. The Brookings-Social Science Research Council (SSRC) econometric model of the US economy published in 1965 initially had 150 equations.

2) *Solutions are highly nonlinear in the parameters and often difficult to determine.*

3) *A specification error in one or more equations of the system is transmitted to the rest of the system, thus the system methods are very sensitive to specification errors.*

In practice, single-equation methods are often used:

1) *Indirect Least Squares Method [Estimation of a just-identified equation]* cumbersome, not often used (it works for just identified equation)

2) *Instrumental Variable Method* (We will not go into details of this method here):

3) *Two-stage Least Squares Method* (2SLS) *[Estimation of an over-identified equation]*

A. Indirect Least Squares Method (ILS): Estimation of a Just-Identified Equation

For a just or exactly identified structural equation, the method of obtaining the estimates of the structural coefficients from the OLS estimates of the reduced-form coefficients is known as the method of indirect least squares (ILS), and the estimates thus obtained are known as the indirect least squares estimates.

ILS involves the following three steps:

Step 1 We first obtain the reduced-form equations. As noted before, these reduced-form equations are obtained from the structural equations in such a manner that the dependent variable in each equation is the only endogenous variable and is a function solely of the predetermined (exogenous or lagged endogenous) variables and the stochastic error term(s).

Step 2 We apply OLS to the reduced-form equations individually. This operation is permissible since the explanatory variables in these equations are predetermined and hence uncorrelated with the stochastic disturbances. The estimates thus obtained are consistent.

Step 3 We obtain estimates of the original structural coefficients from the estimated reduced-form coefficients obtained in Step 2. As noted before, if an equation is exactly identified, there is a one-to-one correspondence between the structural and reduced-form coefficients; that is, one can derive unique estimates of the former from the latter.

B. Two-stage Least Squares Method (2SLS or TSLS): Estimation of an Over-Identified Equation

Stage 1 we regress the endogenous variables on all the predetermined (or exogenous) variables in the system and obtain the predicted (or fitted) $\hat{y}'s$.

Stage 2 Replace the right-hand-side endogenous variables by $\hat{y}'s$ and estimate the structural equation by OLS.

Example

To illustrate 2SLS (or TSLS) further, consider the income–money supply model as follows:

$$Y_{1t} = \beta_{10} + \beta_{12}Y_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t}$$
$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + \gamma_{23}X_{3t} + \gamma_{24}X_{4t} + u_{2t}$$

It can be readily verified that both equations are overidentified. To apply 2SLS, we proceed as follows: In Stage 1 we regress the endogenous variables on all the predetermined variables in the system. Thus,

$$Y_{1t} = \hat{\Pi}_{10} + \hat{\Pi}_{11}X_{1t} + \hat{\Pi}_{12}X_{2t} + \hat{\Pi}_{13}X_{3t} + \hat{\Pi}_{14}X_{4t} + \hat{u}_{1t}$$
$$Y_{2t} = \hat{\Pi}_{20} + \hat{\Pi}_{21}X_{1t} + \hat{\Pi}_{22}X_{2t} + \hat{\Pi}_{23}X_{3t} + \hat{\Pi}_{24}X_{4t} + \hat{u}_{2t}$$

In Stage 2 we replace Y_1 and Y_2 in the original (structural) equations by their estimated values from the preceding two regressions and then run the OLS regressions as follows:

$$Y_{1t} = \beta_{10} + \beta_{12}\hat{Y}_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u^*_{1t}$$
$$Y_{2t} = \beta_{20} + \beta_{21}\hat{Y}_{1t} + \gamma_{23}X_{3t} + \gamma_{24}X_{4t} + u^*_{2t}$$

where $u^*_{1t} = u_{1t} + \beta_{12}\hat{u}_{2t}$ and $\hat{u}^*_{2t} = u_{2t} + \beta_{21}\hat{u}_{1t}$.

The estimates thus obtained will be consistent.

1. Remarks for 2SLS

a) It can be applied to an individual equation in the system without directly taking into account any other equation(s) in the system. Hence, for econometric models involving a large number of equations, 2SLS is economical.

b) Unlike ILS, which provides multiple estimates of parameters in the over-identified equations, 2SLS provides only one estimate per parameter.

c) For exactly identified equations, ILS and 2SLS give identical estimates.

d) If the R^2 in the reduced-form regressions are very high, (say, > 0.8) the classical OLS estimates and 2SLS estimates will be very close.

e) ILS, structural parameters are generally nonlinear functions of the reduced-form parameters, and there is no simple method of estimating their standard errors from the standard errors of the reduced-form coefficients.

f) The statistical justification of 2SLS is of the large-sample type. When there are no lagged endogenous variables, the 2SLS estimators are consistent if the exogenous variables are constant in repeated samples and if the disturbance are iid$(0, \sigma^2)$. If these two conditions are satisfied, the sampling distribution of 2SLS estimators becomes approximately normal for large samples.

When the system contains lagged endogenous variables, the consistency and large-sample normality of the 2SLS estimators require that as T increases, the mean square of the values taken by each lagged endogenous variable converges in probability to a positive limit.

If u_i are not independent, lagged endogenous variables are not independent of the current operation of the equation system, which means these variables are not really predetermined, the resulting estimators are not consistent.

References

• Gujarati