CHAPTER 10

Simultaneous-Equations Methods

10.1 SIMULTANEOUS-EQUATIONS MODELS

When the dependent variable in one equation is also an explanatory variable in some other equation, we have a *simultaneous-equations system* or *model*. The dependent variables in a system of simultaneous equations are called *endogenous variables*. The variables determined by factors outside the model are called *exogenous variables*. There is one *behavioral or structural equation* for each endogenous variable in the system (see Example 1). Using OLS to estimate the structural equations results in biased and inconsistent parameter estimates. This is referred to as *simultaneous-equations bias*. To obtain consistent parameter estimates, the *reduced-form equations* of the model must first be obtained. These express each endogenous variable in the system only as a function of the exogenous variable of the model (see Example 2).

EXAMPLE 1. The following two equations represent a simple macroeconomic model:

$$M_t = a_0 + a_1 Y_t + u_{1t}$$
$$Y_t = b_0 + b_1 M_t + b_2 I_t + u_{2t}$$

where M_t is money supply in time period t, Y is income, and I is investment. Since M depends on Y in the first equation and Y depends on M (and I) in the second equation, M and Y are jointly determined, so we have a simultaneous-equations model. M and Y are the endogenous variables, while I is exogenous or determined outside the model. A change in u_{1t} affects M_t in the first equation. This, in turn, affects Y_t in the second equation. As a result, Y_t and u_{1t} are correlated, leading to biased and inconsistent OLS estimates of the M (and Y) equation.

EXAMPLE 2. The first reduced-form equation can be derived by substituting the second equation into the first and rearranging:

$$M_t = a_0 + a_1(b_0 + b_1 M_t + b_2 I_t + u_{2t}) + u_{1t}$$
$$= \frac{a_0 + a_1 b_0}{1 - a_1 b_1} + \frac{a_1 b_2}{1 - a_1 b_1} I_t + \frac{u_{1t} + a_1 u_{2t}}{1 - a_1 b_1}$$

or
$$M_t = \pi_0 + \pi_1 I_t + v_{1t}$$

The second reduced-form equation can be derived by substituting the first equation into the second and rearranging:

$$Y_t = b_0 + b_1(a_0 + a_1 Y_t + u_{1t}) + b_2 I_t + u_{2t}$$
$$= \frac{a_0 b_1 + b_0}{1 - a_1 b_1} + \frac{b_2}{1 - a_1 b_1} I_t + \frac{b_1 u_{1t} + u_{2t}}{1 - a_1 b_1}$$

or

$$Y_t = \pi_2 + \pi_3 I_t + v_{2t}$$

10.2 IDENTIFICATION

Identification refers to the possibility of calculating the structural parameters of a simultaneous-equations model from the reduced-form parameters. An equation of a system is *exactly identified* if the number of excluded exogenous variables from the equation is equal to the number of endogenous variables in the equation minus 1. However, an equation of a system is *overidentified* (or *underidentified*) if the number of excluded exogenous variables from the equation exceeds (or is smaller than) the number of endogenous variables included in the equation minus 1 (see Example 3). Although this is only a necessary rather than a sufficient condition for identification, it usually gives the correct answer (see Prob. 10.5). Unique structural coefficients can be calculated from the reduced-form coefficients only for an exactly identified equation (see Example 4).

EXAMPLE 3. The money supply (M) equation of Example 1 is exactly identified because it excludes one exogenous variable (I) and includes two endogenous variables (M and Y). However, the income, Y, equation is underidentified because it excludes no exogenous variable. If this second equation had included the additional exogenous variable G (government expenditures), the first, or M, equation would have been overidentified because the number of excluded exogenous variables would then have exceeded the number of endogenous variables minus 1.

EXAMPLE 4. A unique value of the structural parameters of the exactly identified M equation of Example 1 can be calculated from the reduced-form parameters of Example 2 as follows:

$$a_1 = \frac{\pi_1}{\pi_3} = \frac{\dfrac{a_1 b_2}{1 - a_1 b_1}}{\dfrac{b_2}{1 - a_1 b_1}} \qquad \text{and} \qquad a_0 = \pi_0 - a_1 \pi_2 = \frac{a_0(1 - a_1 b_1)}{1 - a_1 b_1}$$

10.3 ESTIMATION: INDIRECT LEAST SQUARES

Indirect least squares (ILS) is a method of calculating structural-parameter values for exactly identified equations. ILS involves using OLS to estimate the reduced-form equations of the system and then using the estimated coefficients to calculate the structural parameters. However, it is not easy to calculate the standard errors of the structural parameters, nor can ILS be used in cases of overidentification.

EXAMPLE 5. Table 10.1 gives the money supply (M = currency plus demand deposits), GDP Y, gross private domestic investment I, and government purchases of goods and services G, all seasonably adjusted in billions of dollars, for the United States from 1982 to 1999 (G will be used in Example 6).

The estimated reduced-form equations of Example 2 are

$$\hat{M}_t = \underset{(2.98)}{312.0608} + \underset{(5.65)}{0.5693} I_t \qquad R^2 = 0.67$$

$$\hat{Y}_t = \underset{(2.17)}{852.3203} + \underset{(14.18)}{5.3522} I_t \qquad R^2 = 0.93$$

$$\hat{a}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_3} = \frac{0.5693}{5.3522} = 0.1064$$

and

Table 10.1 Money Supply, GDP, Investments, and Government Expenditures (Seasonably Adjusted in Billions of Dollars) in the United States, 1982–1999

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990
M	474.30	520.79	551.20	619.28	724.20	749.61	786.25	792.49	824.41
Y	3315.60	3688.80	4033.50	4319.30	4537.50	4891.60	5258.30	5588.00	5847.30
I	483.50	639.50	743.60	762.30	737.10	831.60	842.00	866.70	812.80
G	710.10	742.70	829.00	905.10	963.20	1019.30	1060.70	1123.90	1213.10
Year	1991	1992	1993	1994	1995	1996	1997	1998	1999
M	896.34	1024.31	1129.69	1150.08	1126.80	1081.06	1073.94	1097.37	1122.96
Y	6080.70	6469.80	6795.50	7217.70	7529.30	7981.40	8478.60	8974.90	9559.70
I	832.10	909.80	995.80	1146.10	1155.60	1284.30	1434.50	1590.80	1723.70
G	1239.50	1281.80	1307.10	1344.00	1374.50	1438.90	1508.20	1567.20	1688.80

Source: St. Louis Federal Reserve (Bureau of Economic Analysis).

$$\hat{a}_0 = \hat{\pi}_0 - a_1\hat{\pi}_3 = 312.0608 - 0.1064(852.3203) = 221.3739$$

Thus the M equation of Example 1 estimated by ILS is

$$\hat{M}_t = 221.3739 + 0.1064Y_t$$

The same equation estimated by OLS (inappropriately) is

$$\underset{(2.13)}{\hat{M}_t = 162.7044} + \underset{(9.70)}{0.1159Y_t} \qquad R^2 = 0.85$$

10.4 ESTIMATION: TWO-STAGE LEAST SQUARES

Two-stage least-squares (2SLS) is a method of estimating consistent structural parameters for overidentified equations (for exactly identified equations, 2SLS gives the same results as ILS, but it also gives the standard errors of the estimated structural parameters). 2SLS involves regressing each endogenous variable on all the exogenous variables of the system and then using the predicted values of the endogenous variables to estimate the structural equations of the model.

EXAMPLE 6. If the second, or Y, equation of Example 1 now includes G (government expenditures) as an additonal explanatory variable, then the first, or M, equation is overidentified (see Example 3) and can be estimated by 2SLS. The first stage is

$$\hat{Y}_t = \underset{(-5.71)}{-1007.5346} + \underset{(6.10)}{1.7471I_t} + \underset{(13.57)}{4.5794G_t} \qquad R^2 = 0.99$$

The second stage is

$$\hat{M}_t = \underset{(2.07)}{166.5660} + \underset{(9.19)}{0.1153\hat{Y}_t} \qquad R^2 = 0.84$$

$\hat{a}_1 = 0.1153$ is a consistent estimate of a_1.

Solved Problems

SIMULTANEOUS-EQUATIONS MODELS

10.1 What is meant by (*a*) *Simultaneous-equations system or model*? (*b*) *Endogenous variables*? (*c*) *Exogenous variables*? (*d*) *Structural equations*? (*e*) *Simultaneous-equations bias*? (*f*) *Reduced-form equations*?

(*a*) A *simultaneous-equations system or model* refers to the case in which the dependent variable in one or more equations is also an explanatory variable in some other equation of the system. Specifically, not only are the Ys determined by the Xs, but some of the Xs are, in turn, determined by the Ys, so that the Ys and the Xs are jointly or simultaneously determined.

(*b*) The *endogenous variables* are the dependent variables in the system of simultaneous equations. These are the variables that are determined by the system, even though they also appear as explanatory variables in some other equation of the system.

(*c*) *Exogenous variables* are those variables which are determined outside of the model. These also include the lagged endogenous variables, since their values are already known in any given period. The exogenous variables and the lagged endogenous variables are sometimes called *predetermined variables*.

(*d*) *Structural* or *behavioral equations* describe the structure of an economy or the behavior of some economic agents such as consumers or producers. There is one structural equation for each endogenous variable of the system. The coefficients of the structural equations are called *structural parameters* and express the *direct effect* of each explanatory variable on the dependent variable.

(*e*) *Simultaneous-equations bias* refers to the overestimation or underestimation of the structural parameters obtained from the application of OLS to the structural equations of a simultaneous-equations model. This bias results because those endogenous variables of the system which are also explanatory variables are correlated with the error terms, thus violating the fifth assumption of OLS (see Prob. 6.4).

(*f*) *Reduced-form equations* are obtained by solving the system of structural equations so as to express each endogenous variable of the system as a function only of the exogenous or predetermined variables of the system. Since the exogenous variables of the system are uncorrelated with the error terms, OLS gives consistent reduced-form parameter estimates. These measure the total *direct and indirect effects* of a change in the exogenous variables on the endogenous variables and may be used to obtain consistent structural parameters.

10.2 The following two structural equations represent a simple demand-supply model:

$$\text{Demand:} \quad Q_t = a_0 + a_1 P_t + a_2 Y_t + u_{1t} \qquad a_1 < 0 \quad \text{and} \quad a_2 > 0$$
$$\text{Supply:} \quad Q_t = b_0 + b_1 P_t + u_{2t} \qquad b_1 > 0$$

where Q is quantity, P is price, and Y is consumers' income. It is assumed that the market is cleared in every year so that Q_t represents both quantity bought and sold in year t. (*a*) Why is this a simultaneous-equations model? (*b*) Which are the endogenous and exogenous variables of the system? (*c*) Why would the estimation of the demand and supply function by OLS give biased and inconsistent parameter estimates?

(*a*) The given demand-supply model represents a simple simultaneous-equations market system because Q and P are mutually or jointly determined. If price were below equilibrium, the quantity demanded would exceed the quantity supplied, and vice versa. At equilibrium, the (negatively sloped) demand curve crosses the (positively sloped) supply curve, jointly or simultaneously determining (the equilibrium) Q and P.

(*b*) The endogenous variables of the model are Q and P. These are the variables determined within the model. Y is the only exogenous variable of the model (i.e., determined outside the model).

(*c*) Since the endogenous variable P is also an explanatory variable in both the demand and supply equations, P is correlated with u_{1t} in the demand equation and with u_{2t} in the supply equation. This violates the fifth assumption of OLS, which requires that the explanatory variable be uncorrelated with the error term. As a result, estimating the demand and supply functions by OLS results in parameter estimates that are not only biased but also inconsistent (i.e., that do not converge on the true parameters even as the sample size is increased).

10.3 (*a*) Find the reduced-form equations corresponding to the structural equations of Prob. 10.2. (*b*) Why are these reduced-form equations important? What do the reduced-form coefficients measure in this market model?

(*a*) To find the reduced-form equations, the structural equations of Prob. 10.2 are solved for Q and P (the endogenous variables) as a function of only Y (the exogenous variable). Converting the supply equation into a function of P and substituting into the demand equation, we get

$$P_t = \frac{1}{b_1}(Q_t - b_0 - u_{2t})$$
$$Q_t = a_0 + \frac{a_1}{b_1}(Q_t - b_0 - u_{2t}) + a_2 Y_t + u_{1t}$$
$$Q_t\left(\frac{b_1 - a_1}{b_1}\right) = \left(\frac{a_0 b_1 - a_1 b_0}{b_1}\right) + a_2 Y_t + \left(\frac{b_1 u_{1t} - a_1 u_{2t}}{b_1}\right)$$
$$Q_t = \left(\frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}\right) + \left(\frac{b_1 a_2}{b_1 - a_1}\right) Y_t + \left(\frac{b_1 u_{1t} - a_1 u_{2t}}{b_1 - a_1}\right)$$
$$Q_t = \pi_0 + \pi_1 Y_t + v_{1t}$$

where
$$\pi_0 = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} \qquad \pi_1 = \frac{b_1 a_2}{b_1 - a_1} \qquad v_{1t} = \frac{b_1 u_{1t} - a_1 u_{2t}}{b_1 - a_1}$$

Substituting the demand equation into the supply equation as a function of P, we get

$$P_t = \frac{1}{b_1}(a_0 + a_1 P_t + a_2 Y_t + u_{1t} - b_0 - u_{2t})$$
$$P_t\left(\frac{b_1 - a_1}{b_1}\right) = \frac{1}{b_1}(a_0 + a_2 Y_t + u_{1t} - b_0 - u_{2t})$$
$$P_t = \left(\frac{a_0 - b_0}{b_1 - a_1}\right) + \left(\frac{a_2}{b_1 - a_1}\right) Y_t + \left(\frac{u_{1t} - u_{2t}}{b_1 - a_1}\right)$$
$$P_t = \pi_2 + \pi_3 Y_t + v_{2t}$$

where
$$\pi_2 = \frac{a_0 - b_0}{b_1 - a_1} \qquad \pi_3 = \frac{a_2}{b_1 - a_1} \qquad v_{2t} = \frac{u_{1t} - u_{2t}}{b_1 - a_1}$$

(*b*) Reduced-form equations

$$Q_t = \pi_0 + \pi_1 Y_t + v_{1t}$$
$$P_t = \pi_2 + \pi_3 Y_t + v_{2t}$$

are important because Y_t is uncorrelated with v_{1t} and v_{2t}, so that consistent estimates of reduced-form parameters π_0, π_1, π_2, and π_3 can be obtained by applying OLS to the reduced-form equations. π_1 and π_3 give, respectively, the total of the direct and indirect effects of a change in Y on Q and P. A change in Y causes a shift in the demand curve, which affects both the equilibrium P and Q.

10.4 Given the following three-equations system, (*a*) explain why this is not a simultaneous-equations model. (*b*) Could OLS be used to estimate each equation of this system? Why?

$$Y_{1t} = a_0 + a_1 X_t + u_{1t}$$
$$Y_{2t} = b_0 + b_1 Y_{1t} + b_2 X_t + u_{2t}$$
$$Y_{3t} = c_0 + c_1 Y_{2t} + c_2 X_t + u_{3t}$$

(*a*) The preceding system is not simultaneous because although Y_2 is a function of Y_1, Y_1 is not a function of Y_2. Similarly, although Y_3 is a function of Y_2, Y_2 is not a function of Y_3. Thus the line of causation runs only in one rather than in both directions. Once Y_1 has been estimated in the first equation, Y_1 can be used (together with X) to estimate Y_2 in the second equation. Similarly, once Y_2 has been estimated in the second equation, Y_2 can be used (together with X) to estimate Y_3 in the third equation. Models of this nature are *recursive* rather than simultaneous.

(*b*) In the first equation, exogenous variable X is uncorrelated with error term u_1, so that OLS gives unbiased parameter estimates for the first equation. In the second equation, X and Y are uncorrelated with u_2 (i.e., Y_1 is correlated with u_1 but not with u_2), so that OLS gives unbiased parameter estimates for the second equation. The same is true for the third equation. Thus recursive models can be estimated by the sequential application of OLS.

IDENTIFICATION

10.5 (*a*) What is meant by *identification*? (*b*) When is an equation of a system exactly identified? (*c*) Overidentified? (*d*) Underidentified? (*e*) Are these rules sufficient for identification?

(*a*) *Identification* refers to the possibility or impossibility of obtaining the structural parameters of a simultaneous-equations system from the reduced-form parameters. An equation of a system can be exactly identified, overidentified, or underidentified. The system as a whole is exactly identified if all its equations are exactly identified.

(*b*) An equation of a system is *just or exactly identified* if the number of excluded exogenous variables from the equation is equal to the number of endogenous variables in the equation minus 1. For an exactly identified equation, a unique value of the structural parameters can be *calculated* from the reduced-form parameters.

(*c*) An equation of a system is *overidentified* if the number of excluded exogenous variables from the equation exceeds the number of endogenous variables in the equation minus 1. For an overidentified equation, more than one numerical value can be calculated from some of the structural parameters of the equation from the reduced-form parameters.

(*d*) An equation of a system is *underidentified or unidentified* if the number of excluded variables from the equation is smaller than the number of endogenous variables excluded from the equation minus 1. In this case, no structural parameters can be calculated from the reduced-form parameters.

(*e*) The preceding rules for identification (called the *order condition*) are necessary but not sufficient. However, since these rules do give the correct result in most cases, they are the only ones actually used here. A sufficient condition for identification is given by the *rank condition*, which states that in a system of G equations, any particular equation is identified if and only if it is possible to obtain one nonzero determinant of order $G-1$ from the coefficients of the variables excluded from that particular equation but included in the other equations of the model. When this rank condition is satisfied, the order condition is automatically satisfied. However, the reverse is not true.

10.6 Given the following demand-supply model (*a*) determine if the demand and/or supply is exactly identified, overidentified, or underidentified.

$$\text{Demand:} \quad Q_t = a_0 + a_1 P_t + u_{1t} \qquad a_1 < 0$$
$$\text{Supply:} \quad Q_t = b_0 + b_1 P_t + u_{2t} \qquad b_1 > 0$$

(*b*) What would a regression of Q_t on P_t indicate?

(*a*) Since this demand-supply model does not include any exogenous variable, both the demand and supply functions are underidentified. In this case, there are no reduced-form equations, and no structural parameters can be calculated. Each price-quantity observation represents the equilibrium quantity bought and sold at the given price and corresponds to the interception of an (unknown) demand and supply curve.

(*b*) Regressing Q_t on P_t gives neither a demand curve nor a supply curve, but rather a hybrid of demand and supply, which should be referred to simply as a regression line.

10.7 With reference to the demand-supply model in Prob. 10.2 (*a*) determine if the demand and/or supply function is exactly identified, overidentified, or underidentified. (*b*) Give a graphical interpretation of your answer to part *a*. (*c*) Derive the formula for the structural coefficients from the reduced-form coefficients.

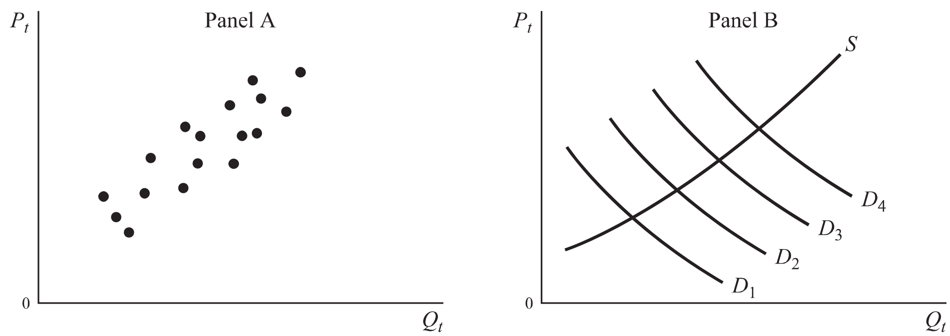

Fig. 10-1

(*a*) The demand function is underidentified because it does not exclude any exogenous variable. However, since there is one excluded exogenous variable from the supply equation (that is, Y) and two included endogenous variables (i.e., Q and P), the supply function is exactly identified.

(*b*) Changes in Y cause shifts in the demand curve, thus tracing the supply curve. Figure 10-1*a* shows a hypothetical scatter of points resulting from changes in Y and the error terms, while Fig. 10-1*b* shows the resulting supply curve that could be generated.

(*c*) Unique values of the structural coefficients of the supply equation (the exactly identified equation) can be calculated from the reduced-form coefficients in Prob. 10.3 as follows:

$$b_1 = \frac{\pi_1}{\pi_3} = \frac{\dfrac{b_1 a_2}{b_1 - a_1}}{\dfrac{a_2}{b_1 - a_1}}$$

$$b_1 = \pi_0 - b_1\pi_2 = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} - \frac{b_1 a_0 + b_0 b_1}{b_1 - a_1} = \frac{b_0(b_1 - a_1)}{b_1 - a_1}$$

The formula for the structural coefficients of the demand function cannot be derived from the reduced-form coefficients because the demand function in this model is underidentified.

10.8 With reference to the demand-supply model given below, (*a*) determine if the demand and/or supply functions are exactly identified, overidentified, or underidentified. (*b*) Find the reduced-form equations. (*c*) Derive the formula for the structural parameters.

$$\text{Demand:} \quad Q_t = a_0 + a_1 P_t + a_2 Y_t + u_{1t} \qquad a_1 < 0, \qquad a_2 > 0$$
$$\text{Supply:} \quad Q_t = b_0 + b_1 P_t + b_2 T + u_{2t} \qquad b_1 > 0, \qquad b_2 \lessgtr 0$$

where T = trend.

(*a*) The supply equation is exactly identified (as in Prob. 10.7) because it excludes one exogenous variable (Y) and includes two endogenous variables (P and Q). The demand equation is now also exactly identified because it excludes one exogenous variable (T) and includes two endogenous variables (P and Q).

(*b*) The reduced-form equations can be obtained as in Prob. 10.3(*a*):

$$Q_t = \left(\frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}\right) + \left(\frac{a_2 b_1}{b_1 - a_1}\right) Y_t + \left(\frac{-a_1 b_2}{b_1 - a_1}\right) T + \left(\frac{b_1 u_{1t} - a_1 u_{2t}}{b_1 - a_1}\right)$$

$$P_t = \left(\frac{a_0 - b_0}{b_1 - a_1}\right) + \left(\frac{a_2}{b_1 - a_1}\right) Y_t + \left(\frac{-b_2}{b_1 - a_1}\right) T + \left(\frac{u_{1t} - u_{2t}}{b_1 - a_1}\right)$$

or,

$$Q_t = \pi_0 + \pi_1 Y_t + \pi_2 T + v_{1t}$$
$$P_t = \pi_3 + \pi_4 Y_t + \pi_5 T + v_{2t}$$

where

$$\pi_0 = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} \qquad \pi_1 = \frac{a_2 b_1}{b_1 - a_1} \qquad \pi_2 = \frac{-a_1 b_2}{b_1 - a_1} \qquad v_{1t} = \frac{b_1 u_{1t} - a_1 u_{2t}}{b_1 - a_1}$$

$$\pi_3 = \frac{a_0 - b_0}{b_1 - a_1} \qquad \pi_4 = \frac{a_2}{b_1 - a_1} \qquad \pi_5 = \frac{-b_2}{b_1 - a_1} \qquad v_{2t} = \frac{u_{1t} - u_{2t}}{b_1 - a_1}$$

(*c*) $a_1 = \dfrac{\pi_2}{\pi_5}$ and $b_1 = \dfrac{\pi_1}{\pi_4}$

$$a_2 = \pi_4(b_1 - a_1) = \pi_4\left(\frac{\pi_1}{\pi_4} - \frac{\pi_2}{\pi_5}\right) \quad \text{and} \quad b_2 = -\pi_5(b_1 - a_1) = \pi_5\left(\frac{\pi_2}{\pi_5} - \frac{\pi_1}{\pi_4}\right)$$

$$a_0 = \pi_3(b_1 - a_1) + b_0 = \pi_3\left(\frac{\pi_0}{\pi_3} - \frac{\pi_2}{\pi_5}\right) \quad \text{and} \quad b_0 = -\pi_3(b_1 - a_1) + a_0 = \pi_3\left(\frac{\pi_0}{\pi_3} - \frac{\pi_1}{\pi_4}\right)$$

10.9 With reference to the demand-supply model given below, (*a*) determine if the demand and/or supply equation is exactly identified, overidentified, or underidentified. (*b*) Calculate the structural slope parameters.

$$\text{Demand:} \quad Q_t = a_0 + a_1 P_t + a_2 Y_t + a_3 W_t + u_{1t}$$
$$\text{Supply:} \quad Q_t = b_0 + b_1 P_t + u_{2t}$$

where W_t is wealth and the expectation is that $a_3 > 0$.

(*a*) The demand equation is underidentified because it does not exclude any exogenous variable. However, since there are two excluded exogenous variables from the supply equation (i.e., Y and W) and two included endogenous variables (i.e., Q and P), the supply function is overidentified.

(*b*) In order to calculate the structural slope parameters, the reduced-form equations must be found. They are obtained as in Prob. 10.7(*c*) and are

$$Q_t = \pi_0 + \pi_1 Y_t + \pi_2 W_t + v_{1t}$$
$$P_t = \pi_3 + \pi_4 Y_t + \pi_5 W_t + v_{2t}$$

where

$$\pi_0 = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} \qquad \pi_1 = \frac{a_2 b_1}{b_1 - a_1} \qquad \pi_2 = \frac{a_3 b_1}{b_1 - a_1}$$

$$\pi_3 = \frac{a_0 - b_0}{b_1 - a_1} \qquad \pi_4 = \frac{a_2}{b_1 - a_1} \qquad \pi_5 = \frac{a_3}{b_1 - a_1}$$

The value of b_1 can be calculated from

$$\frac{\pi_1}{\pi_4} = b_1 \quad \text{or} \quad \frac{\pi_2}{\pi_5} = b_1$$

These two estimates of b_1 will generally be different, reflecting the fact that the supply equation is now overidentified. As in Prob. 10.7(*c*), the structural coefficients of the demand function cannot be calculated from the reduced-form coefficients because the demand function in this model is underidentified.

ESTIMATION: INDIRECT LEAST SQUARES

10.10 (*a*) When can indirect least squares be used? (*b*) What does it involve? (*c*) What are some of the shortcomings of using indirect least squares?

(*a*) *Indirect least squares* (ILS) is a method of calculating consistent structural parameter values for the exactly identified equations in a system of simultaneous equations.

(*b*) ILS involves using OLS to estimate the reduced-form equations of the system and then using the estimated reduced-form parameters to calculate unique and consistent structural parameter estimates, as indicated in Probs. 10.7(*c*), 10.8(*c*), and 10.9(*b*).

(*c*) One disadvantage of using ILS is that it does not give the standard error of the calculated structural parameters, and it is rather complicated (and beyond the scope of this book) to calculate them. Another disadvantage of ILS is that it cannot be used to calculate unique and consistent structural-parameter estimates from the reduced-form coefficients for the overidentified equations of a simultaneous-equations model.

10.11 Table 10.2 gives the index of crop output Q (indexed to 1992), crop prices P (indexed to 1991–1992), and disposable income per capita Y (in 1996 dollars), in the United States from 1975 to 1996. Assume that the market is cleared in every year so that Q_t represents both the quantity bought and sold in year t. (*a*) Estimate by OLS the reduced-form equations given in Prob. 10.3(*a*). (*b*) Calculate the supply structural parameters from the reduced-form coefficients. (*c*) How do these compare with the structural parameters obtained by regressing Q_t on P_t directly?

Table 10.2 Index of Crop Output, Prices, and Disposable Income per Capita in 1996 Dollars: United States, 1975–1996

Year	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Q	68	68	74	76	83	75	87	87	68	85	89
P	88	87	83	89	98	107	111	98	108	111	98
Y	14,236	14,653	15,010	15,627	15,942	15,944	16,154	16,250	16,564	17,687	18,120
Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Q	84	86	75	86	92	92	100	90	106	96	103
P	87	86	104	109	103	101	101	102	105	112	127
Y	18,536	18,790	19,448	19,746	19,967	19,892	20,359	20,354	20,675	21,032	21,385

Source: *Economic Report of the President, 2000*.

(*a*) The estimated reduced-form equations [from Prob. 10.3(a)] are

$$\hat{Q}_t = 14.2802 + 0.0039Y_t \qquad R^2 = 0.67$$
$$(1.26) \qquad (6.31)$$

$$\hat{P}_t = 54.1671 + 0.0026Y_t \qquad R^2 = 0.30$$
$$(3.36) \qquad (2.91)$$

(*b*)
$$\hat{b}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_3} = \frac{0.0039}{0.0026} = 1.5000 \quad \text{[see Prob. 10.7(c)]}$$

$$\hat{b}_0 = \hat{\pi}_0 - b_1\hat{\pi}_3 = 14.2802 - 1.5000(54.1671) = -66.9705$$

where $\hat{b}_0$ and $\hat{b}_1$ are consistent estimators of b_0 and b_1, respectively, and the structural supply equation (estimated by ILS) is

$$\hat{Q}_t = -66.9705 + 1.5000P_t$$

(*c*) Regressing Q_t on P_t directly, we get

$$\hat{Q}_t = 33.1984 + 0.5145P_t \qquad R^2 = 0.26$$
$$(1.67) \qquad (2.63)$$

The values of $\hat{b}_0$ and $\hat{b}_1$ obtained by regressing Q_t on P_t are biased and inconsistent estimates of the supply parameters.

10.12 With reference to the demand-supply model of Prob. 10.8 and using the data in Table 10.2 and trend values $T = 1, 2, 3, \ldots, 30$, (*a*) calculate consistent structural parameters for the demand equation. (*b*) How do these compare with the structural parameters obtained by estimating the demand equation directly by OLS?

(*a*) Since the demand equation is exactly identified [see Prob. 10.8(*a*)], we can use ILS to obtain consistent demand structural-parameter values. The estimated reduced-form equations [from Prob. 10.8(*b*)] are

$$\hat{Q}_t = \underset{(1.73)}{102.6080} - \underset{(-0.57)}{0.0024Y_t} + \underset{(1.51)}{2.2520T} \qquad R^2 = 0.70$$

$$\hat{P}_t = \underset{(2.58)}{211.3674} - \underset{(-1.49)}{0.0087Y_t} + \underset{(1.95)}{4.0079T} \qquad R^2 = 0.41$$

where $\hat{\pi}_0 = 102.6080, \hat{\pi}_1 = -0.0024, \hat{\pi}_2 = 2.2520$

$$\hat{\pi}_3 = 211.3674, \hat{\pi}_4 = -0.0087, \hat{\pi}_5 = 4.0079$$

Using the formulas given in Prob. 10.8(c), we get

$$\hat{a}_1 = \frac{\hat{\pi}_2}{\hat{\pi}_5} = \frac{2.2520}{4.0079} = 0.5619$$

$$\hat{a}_2 = \hat{\pi}_4\left(\frac{\hat{\pi}_1}{\hat{\pi}_4} - \frac{\hat{\pi}_2}{\hat{\pi}_5}\right) = (-0.0087)\left(\frac{(-0.0024)}{(-0.0087)} - \frac{2.2520}{4.0079}\right) = 0.0025$$

$$\hat{a}_0 = \hat{\pi}_3\left(\frac{\hat{\pi}_0}{\hat{\pi}_3} - \frac{\hat{\pi}_2}{\hat{\pi}_5}\right) = 211.3674\left(\frac{102.6080}{211.3674} - \frac{2.2520}{4.0079}\right) = -16.1573$$

Thus the demand equation estimated by ILS (and showing consistent parameter estimates) is

$$\hat{Q}_t = -16.1573 + 0.5619P_t + 0.0025T$$

(*b*) The OLS estimation of the demand function is

$$\hat{Q}_t = \underset{(0.66)}{9.4529} + \underset{(0.56)}{0.0891P_t} + \underset{(4.89)}{0.0037T} \qquad R^2 = 0.67$$

The values of $\hat{a}_0$, $\hat{a}_1$, and $\hat{a}_2$ estimated by OLS are biased and inconsistent. Indeed, $\hat{a}_1$ is less than 20% of the ILS estimate, and $\hat{a}_0$ even has the wrong sign (but is not statistically significant).

ESTIMATION: TWO-STAGE LEAST SQUARES

10.13 When can 2SLS be used? (*b*) What does it involve? (*c*) What are the advantages of 2SLS with respect to ILS?

(*a*) *Two-stage least squares* (2SLS) is a method of estimating consistent structural-parameter values for the exactly identified or overidentified equations of a simultaneous-equations system. For exactly identified equations, 2SLS gives the same result as ILS.

(*b*) 2SLS estimation involves the application of OLS in two stages. In the first stage, each endogenous variable is regressed on all the predetermined variables of the system. These are now the reduced-form equations. In the second stage, the predicted rather than the actual values of the endogenous variables are used to estimate the structural equations of the model. The predicted values of the endogenous variables are obtained by substituting the observed values of the exogenous variables into the reduced-form equations. The predicted values of the endogenous variables are uncorrelated with the error terms, leading to consistent 2SLS structural-parameter estimates.

(*c*) One advantage of 2SLS over ILS is that 2SLS can be used to obtain consistent structural-parameter estimates for the overidentified as well as for the exactly identified equations in a system of simultaneous equations. Another important advantage is that 2SLS (but not ILS) gives the standard error of the estimated structural parameters directly. Since most identified models are in fact overidentified, 2SLS is very useful. Indeed, 2SLS is the simplest and one of the best and most common of all simultaneous-equations estimators.

10.14 For the demand-supply model in Prob. 10.8 and using the data in Table 10.2 to estimate the demand equation, (*a*) show the first-stage result of 2SLS estimation. (*b*) Show the second-stage result of 2SLS estimation. (*c*) How do these results compare with the ILS estimation of the demand equation found in Prob. 10.12(*a*)?

(*a*) The first-stage result of the 2SLS estimation of the demand equation is

$$\hat{P}_t = \underset{(2.58)}{211.3674} - \underset{(-1.49)}{0.0087Y_t} + \underset{(1.95)}{4.0079T} \qquad R^2 = 0.41$$

(*b*) The second-stage result of 2SLS estimation of the demand equation is

$$\hat{Q}_t = \underset{(-0.70)}{-16.16} + \underset{(2.18)}{0.56\hat{P}_t} - \underset{(1.51)}{0.0025Y_t} \qquad R^2 = 0.70$$

(*c*) Since the demand equation in Prob. 10.8 is exactly identified, 2SLS estimation gives identical results to ILS estimation [see Prob. 10.12(*a*)]. However, with 2SLS estimation (as opposed to ILS), we also get the standard errors of the estimated structural parameters directly.

10.15 Table 10.3 includes the additional variable wealth W, measured here by total liquid assets, in billions of dollars, to the data in Table 10.2 for the United States for the years 1975 to 1996. For the demand-supply model in Prob. 10.9, estimate the supply equation by (*a*) 2SLS and (*b*) OLS.

Table 10.3 Index of Crop Output, Prices, Disposable Income per Capita, and Total Liquid Assets in Billions of Dollars in the United States, 1975–1996

Year	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
Q	68	68	74	76	83	75	87	87	68	85	89
P	88	87	83	89	98	107	111	98	108	111	98
Y	14,236	14,653	15,010	15,627	15,942	15,944	16,154	16,250	16,564	17,687	18,120
W	1366.5	1516.7	1705.4	1911.3	2121.2	2330.0	2601.8	2846.0	3150.7	3518.7	3827.1
Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Q	84	86	75	86	92	92	100	90	106	96	103
P	87	86	104	109	103	101	101	102	105	112	127
Y	18,536	18,790	19,448	19,746	19,967	19,892	20,359	20,354	20,675	21,032	21,385
W	4122.4	4340.0	4663.7	4893.2	4977.5	5008.0	5081.4	5173.3	5315.8	5702.3	6083.6

Source: *Economic Report of the President, 2000*.

(*a*) Since the supply equation in Prob. 10.9 is overidentified, 2SLS is an appropriate estimating technique to obtain consistent structural parameters. The first stage is

$$\hat{P}_t = \underset{(1.84)}{197.51} - \underset{(-1.05)}{0.01Y_t} + \underset{(1.35)}{0.02W_t} \qquad R^2 = 0.36$$

The second stage is

$$\hat{Q}_t = \underset{(-1.15)}{-32.11} + \underset{(4.21)}{1.16\hat{P}_t} \qquad R^2 = 0.47$$

(*b*) The (inappropriate) OLS estimation of the supply equation is

$$\hat{Q}_t = \underset{(-1.67)}{-33.20} + \underset{(2.63)}{0.51\hat{P}_t} \qquad R^2 = 0.26$$

Supplementary Problems

SIMULTANEOUS-EQUATIONS MODELS

10.16 The following two equations represent a simple wage-price model:

$$W_t = a_0 + a_1 P_t + a_2 Q_t + u_{1t}$$
$$P_t = b_0 + b_1 W_t + u_{2t}$$

where W_t is the wage in time period t, P represents prices, and Q is productivity. (*a*) Why is this a simultaneous-equations model? (*b*) Which are the endogenous and exogenous variables? (*c*) Why would the estimation of W and P equations by OLS give biased and inconsistent parameter estimates?
Ans. (*a*) This two-equations model is simultaneous in nature because $W = f(P)$ and $P = f(W)$; thus W and P are jointly determined. (*b*) The endogenous variables are W and P. The exogenous variable is Q. (*c*) The estimation of the W function by OLS gives biased and inconsistent parameter estimates because P is correlated with u_1. Similarly, estimating the second, or P, equation by OLS also gives biased and inconsistent parameter estimates because W and u_2 are correlated.

10.17 (*a*) Find the reduced-form equations for the model in Prob. 10.16. (*b*) Why are they important? (*c*) What do the reduced-form coefficients measure in this macro model?

Ans. (*a*)

$$W_t = \frac{a_0 + a_1 b_0}{1 - a_1 b_1} + \frac{a_2}{1 - a_1 b_1} Q_t + \frac{u_{1t} + a_1 u_{2t}}{1 - a_1 b_1} \quad \text{or} \quad W_t = \pi_0 + \pi_1 Q_t + v_{1t}$$

$$P_t = \frac{b_0 + a_0 b_1}{1 - a_1 b_1} + \frac{a_2 b_1}{1 - a_1 b_1} Q_t + \frac{b_1 u_{1t} + u_{2t}}{1 - a_1 b_1} \quad \text{or} \quad P_t = \pi_2 + \pi_3 Q_t + v_{2t}$$

(*b*) The reduced-form equations are important because they express each endogenous variable of the model as a function of the exogenous variable(s) only, so that OLS gives consistent parameter estimates. (*c*) The reduced-form parameters give the total direct and indirect effects of a change in any exogenous variable of the model on each endogenous variable of the model.

10.18 (*a*) What type of model is the following? (*b*) How can the equations of this model be estimated?

$$Y_{1t} = a_0 + a_1 X_{1t} + u_{1t}$$
$$Y_{2t} = b_0 + b_1 Y_{1t} + b_2 X_{2t} + u_{2t}$$
$$Y_{3t} = c_0 + c_1 Y_{1t} + c_2 Y_{2t} + c_3 X_{3t} + u_{3t}$$

Ans. (*a*) The model is recursive. (*b*) The equations of the model can be estimated by applying OLS sequentially, starting with the first equation.

IDENTIFICATION

10.19 If the simple macroeconomic model in Prob. 10.16 did not include the variable Q_t, (*a*) would the first equation be exactly identified, overidentified, or underidentified? (*b*) What about the second equation?
Ans. (*a*) The first equation would be underidentified. (*b*) The second equation also would be underidentified.

10.20 For the macro model in Prob. 10.16, determine (*a*) if the first equation is exactly identified, overidentified, or underidentified. (*b*) What about the second equation? (*c*) What are the values of the structural parameters?
Ans. (*a*) The first equation is underidentified. (*b*) The second equation is exactly identified. (*c*) $b_1 = \pi_3/\pi_1$; $b_0 = \pi_2 - b_1\pi_0$; a_1 and a_2 cannot be calculated from the reduced-form coefficients because the W equation is underidentified.

10.21 If the second equation of the macro model in Prob. 10.16 included the additional variable Y (GNP), (*a*) determine if the W and/or P equations are exactly identified, overidentified, or underidentified. (*b*) Find the reduced-form equations. (*c*) Derive the formula for the structural parameters.

Ans. (*a*) Both the first, or W, equation and the second, or P, equation are now exactly identified.

(*b*)

$$W_t = \frac{a_0 + a_1 b_0}{1 - a_1 b_1} + \frac{a_2}{1 - a_1 b_1} Q_t + \frac{a_1 b_2}{1 - a_1 b_1} Y_t + \frac{u_{1t} + a_1 u_{2t}}{1 - a_1 b_1}$$

$$P_t = \frac{a_0 b_1 + b_0}{1 - a_1 b_1} + \frac{a_2 b_1}{1 - a_1 b_1} Q_t + \frac{b_2}{1 - a_1 b_1} Y_t + \frac{b_1 u_{1t} + u_{2t}}{1 - a_1 b_1}$$

or

$$W_t = \pi_0 + \pi_1 Q_t + \pi_2 Y_t + v_{1t}$$

$$P_t = \pi_3 + \pi_4 Q_t + \pi_5 Y_t + v_{2t}$$

(*c*)

$$a_1 = \frac{\pi_2}{\pi_5} \quad \text{and} \quad b_1 = \frac{\pi_4}{\pi_1}$$

$$a_2 = \pi_2\left(\frac{\pi_1}{\pi_2} - \frac{\pi_4}{\pi_5}\right) \quad \text{and} \quad b_2 = \pi_2\left(\frac{\pi_5}{\pi_2} - \frac{\pi_4}{\pi_1}\right)$$

$$a_0 = \pi_3\left(\frac{\pi_0}{\pi_3} - \frac{\pi_2}{\pi_5}\right) \quad \text{and} \quad b_0 = \pi_0\left(\frac{\pi_3}{\pi_0} - \frac{\pi_4}{\pi_1}\right)$$

10.22 If the first equation in Prob. 10.16 included the additional variable P_{t-1} (price lagged 1 year), (*a*) would the equations be exactly identified, overidentified, or underidentified? (*b*) What is the value of the structural slope parameters?

Ans. (*a*) The first, or W, equation is underidentified, while the second, or P, equation is overidentified. (*b*) $b_1 = \pi_4/\pi_1$ or π_5/π_2, reflecting the fact that the P equation is now overidentified; a_1, a_2, and a_3 cannot be calculated because the W equation is underidentified.

ESTIMATION: INDIRECT LEAST SQUARES

10.23 Table 10.4 gives an index of hourly earnings W, consumer prices P, output per hour in nonfarm businesses, Q, and GDP in billions of dollars Y in the United States from 1980 to 1999. (*a*) Estimate the reduced-form equations of Prob. 10.17(*a*). (*b*) Calculate the structural coefficients of the P equation from the reduced-form coefficients. (*c*) How do these compare with the structural parameters obtained by regressing P on W directly?

Table 10.4 Earnings, Price Index, Productivity, and GDP: United States, 1980–1999

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
W	56.6	61.5	65.7	68.0	71.1	74.8	78.5	81.4	84.8	87.2
P	86.4	94.1	97.7	101.4	105.5	109.5	110.8	115.7	120.8	126.4
Q	82.4	82.5	83.3	87.3	88.4	90.2	92.2	93.1	94.1	94.6
Y	2918.8	3203.1	3315.6	3688.8	4033.5	4319.3	4537.5	4891.6	5258.3	5588.0
Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
W	92.3	96.7	101.4	102.7	105.0	107.7	111.1	114.7	120.8	126.5
P	134.3	138.3	142.4	146.4	150.2	154.1	159.1	161.8	164.4	168.8
Q	94.4	94.4	101.5	101.3	102.4	103.6	105.9	108.1	111.2	115.8
Y	5847.3	6080.7	6469.8	6795.5	7217.7	7529.3	7981.4	8478.6	8974.9	9559.7

Source: St. Louis Federal Reserve (Bureau of Labor Statistics (W, P, Q values), Bureau of Economic Analysis (Y values)).

Ans. (*a*)

$$\hat{W}_t = -114.8528 + 2.1270 Q_t \qquad R^2 = 0.98$$
$$(-17.61) \qquad (31.62)$$

$$\hat{P}_t = -126.0632 + 2.6471 Q_t \qquad R^2 = 0.96$$
$$(-9.89) \qquad (20.14)$$

(*b*) $\hat{b}_1 = 1.2445$; $\hat{b}_0 = 16.8711$ (*c*) By OLS $\hat{b}_1 = 1.2550$ and $\hat{b}_0 = 15.9256$