Our second extension of the classic regression model, to which we devote two lectures, is to a system (or model) of more than one equation. Together, the equations describe the simultaneous evolution of multiple endogenous variables and are termed simultaneous equation models.

Up to this point, we have considered structures in which only one variable, the dependent variable, is endogenous. We often think of the regressors as causing the movements in the dependent variable in that the regressors can be conditioned upon when studying the dependent variable. In many cases of interest to economists the conditional structure of a single equation breaks down, as two or more quantities are jointly determined. Perhaps the simplest such structure is to consider price and quantity for a specific market (e.g. Vans tennis shoes in Santa Barbara). One might think of the following structure

$$P_t = \beta_0 + \beta_1 Q_t + U_t,$$

where P_t is the price of a pair of shoes in period-t and Q_t is the quantity of shoes sold in the corresponding period. Without question, the price of shoes is affected by the quantity available. Surplus inventories are sold at discount and goods in short supply often sell at a premium. Yet, unless the Santa Barbara market forms only a tiny portion of total supply, the price at which the shoes are sold affects the quantity. Shoes sold at discount reflect excess inventories and indicate that production should fall, the converse holds for shoes sold at a premium. If the regressor is also a function of the dependent variable, it is clear that the regressor is correlated with the error and a classic assumption is violated. To overcome the bias (and inconsistency) of the OLS estimators, one could proceed with IV estimation. Yet not only is an adequate instrument rarely at hand, but often understanding of the economy requires that we more fully investigate the joint determination of the endogenous variables.

To do so, it is helpful to distinguish between the theoretical quantities Q_t^D, which is the quantity of shoes demanded by consumers, and Q_t^S, which is the quantity of shoes supplied by producers. We think of Q_t^D as arising from a demand function

$$Q_t^D = \alpha_1 + \alpha_2 P_t + U_{1t},$$

where we expect $\alpha_2 < 0$. We think of Q_t^S as arising from a supply function

$$Q_t^S = \beta_1 + \beta_2 P_t + U_{2t},$$

where we expect $\beta_2 > 0$. The two quantities are theoretical because they are not separately observed, the equilibrium condition that

$$Q_t^D = Q_t^S = Q_t$$

ensures that only Q_t is observed.

It is not possible to estimate the parameters of the demand and supply functions by the method of OLS. The reason is that the parameters are not identified. Recall from our earlier discussion that if a parameter is not identified, then the population value of the parameter cannot be deduced (by any method) with an infinite amount of data. For single equation models, identification fails if: the regressor does not vary over observations or if the population covariance of the regressor and the error is not zero. As P_t depends on Q_t, the population covariance between the regressor and the error is not zero and identification fails.

(Use the graph from the bottom of page 2 of handwritten notes.)
Connecting the two observations from the shifted curves, we see that we would estimate a linear combination of the coefficients from the two equations.

To understand how to restore identification, we write the system as (the equilibrium condition allows us to replace both Q_t^D and Q_t^S with Q_t, thereby reducing to a system of two equations)

$$\begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_t + U_{1t}, \\ Q_t &= \beta_1 + \beta_2 P_t + U_{2t}, \end{aligned}$$

where the first equation, again, captures demand and the second equation captures supply. Suppose that demand depends not only on the price of the shoes but also on the average household income, Y_t:

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + U_{1t}.$$

As long as the quantity of shoes supplied does not depend on average household income in Santa Barbara, the supply equation is unchanged. For such a system, if the average household income changes from one period to the next, then the demand curve shifts but the supply curve remains constant and the slope of the supply curve can be estimated.

(Use the graph from the top of page 3 of handwritten notes.)
Now suppose that supply depends not only on the price of shoes but also on the hourly wage paid to shoe makers, W_t:

$$Q_t = \beta_1 + \beta_2 P_t + \beta_3 W_t + U_{2t}.$$

As long as the quantity of shoes demanded in Santa Barbara does not depend on the wage paid to shoe makers, the demand equation is unchanged. For such a system, if the hourly wage paid to shoe makers changes from one period to the next, then the supply curve shifts but the demand curve remains constant and the slope of the demand curve can be estimated.
(Graph shifting supply, constant demand.)
With both Y_t and W_t present, one might think that we were back to the system with both demand and supply curves shifting with no identification. Yet there is an important difference. Both Y_t and W_t are observed. For all observations with a given average income, we can "trace out" the supply curve, while for all observations with a given hourly wage rate we can "trace out" the demand curve.

Of course, it is possible that only one of the equations is identified. Suppose that the demand for shoes depended on Y_t but that the supply of shoes did not depend on W_t. Because the demand function has the necessary shifter, it is possible to "trace out" the supply curve and the supply equation is identified. As there is no unique shifter of supply, it is not possible to "trace out" the demand curve, and the demand function is not identified. To make the point mathematically, consider a linear combination of the two equations

$$Q_t = \delta \left(\alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + U_{1t} \right) + (1 - \delta) \left(\beta_1 + \beta_2 P_t + U_{2t} \right),$$

where $0 < \delta < 1$. Rewriting the linear combination yields

$$Q_t = \alpha_1^* + \alpha_2^* P_t + \alpha_3^* Y_t + U_t^*,$$

which is observationally equivalent to the demand function. When estimating the demand function, we cannot distinguish between estimation of $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1^*, \alpha_2^*, \alpha_3^*)$. That is, we cannot distinguish between estimating the demand function and estimating a linear combination of the demand and supply functions and so the parameters of the demand function are not identified. As the presence of α_3^* makes the equation distinct from the supply function, the parameters of the supply function are identified.

3

The need for an exogenous (or predetermined) shifter in each equation points toward the necessary, but not sufficient, condition for identification. This condition, termed the order condition for each equation, is

Order Condition: *The number of predetermined variables in the system is greater than or equal to the number of slope coefficients in the equation.*

To see that the order condition is only necessary, consider the case in which $Z_t = \eta Y_t$, rather than W_t, appeared in the supply function. As we have two predetermined variables, Y_t and Z_t, and two slope coefficients in each equation, the order condition is satisfied. Yet, because the supply function contains δY_t, the linear combination is not distinct from the supply function, and identification of the supply equation fails. It is necessary that the shifters to the two equations be distinct. Linearly distinct predetermined variables is also not sufficient. We typically assume the predetermined variables are distinct and then ask if it is possible to distinguish each equation from all (nontrivial) linear combinations of the other equations in the system. The sufficient condition for identification is the

Rank Condition. *For each equation: Each of the variables excluded from the equation must appear in at least one of the other equations (no zero columns). Also, at least one of the variables excluded from the equation must appear in each of the other equations (no zero rows).*

The condition is more commonly stated as

Rank Condition. *For each equation: Consider the set of variables excluded from the equation. The matrix of coefficients for these variables in the other equations must have full row rank.*

A structural equation is identified only when the predetermined variables are arranged within the system so as to use the observed equilibrium points to distinguish the shape of the equation under study.

Consider applying the conditions to a general example of such a system

$$
\begin{aligned}
Y_{1t} &= \alpha_1 + \alpha_2 X_{1t} + \alpha_3 X_{3t} + U_{1t}, \\
Y_{2t} &= \beta_1 + \beta_2 Y_{3t} + \beta_3 X_{1t} + \beta_4 X_{2t} + U_{2t}, \\
Y_{3t} &= \gamma_1 + \gamma_2 Y_{1t} + \gamma_3 X_{1t} + \gamma_4 X_{3t} + U_{3t},
\end{aligned}
$$

for which Y_{1t}, Y_{2t} and Y_{3t} are the jointly endogenous variables. The equations in the system are termed structural equations, as they characterize the economic theory underpinning the determination of each endogenous variable. A variable is endogenous because it is jointly determined (a change in Y_{1t} leads to a change in Y_{3t}, which in turn leads to a change in Y_{2t}). Exogenous variables may appear

in all equations, witness X_{1t}. As to what is endogenous and what is exogenous, why that depends on the scope of the partial equilibrium model under study.

To verify the order condition, note that there are 3 predetermined variables in the system (X_{1t}, X_{2t}, X_{3t}) and no more than 3 slope coefficients in any one equation. To verify the rank condition, we use the following table, in which $\times$ indicates a variable appears in the given equation and 0 indicates a variable does not appear in the given equation:

Equation	Y_{1t}	Y_{2t}	Y_{3t}	X_{1t}	X_{2t}	X_{3t}
1	$\times$	0	0	$\times$	0	$\times$
2	0	$\times$	$\times$	$\times$	$\times$	0
3	$\times$	0	$\times$	$\times$	0	$\times$

We study the 3×6 matrix of 0's and $\times$'s. For equation i, first select the columns corresponding to the variables that do not appear in equation i. From this submatrix, delete row i. If the remaining submatrix has rank greater than or equal to $m-1$, where m is the number of equations in the system, then the rank condition is satisfied for the equation and the parameters of the equation are identified.

Consider equation 1. The variables (Y_{2t}, Y_{3t}, X_{2t}) are all excluded, so the relevant submatrix is

$$
\begin{array}{ccc}
\times & \times & \times \\
0 & \times & 0
\end{array}.
$$

The two rows are linearly distinct, as there is no possible multiple of the first row that equals the second. The rank is 2 and the equation is identified. For equation 2, the excluded variables are (Y_{1t}, X_{3t}), so the relevant submatrix is

$$
\begin{array}{cc}
\times & \times \\
\times & \times
\end{array}.
$$

The two rows are linearly distinct as long as the coefficients of one row are not proportional to the coefficients of the other row. Absent proportionality, the rank is 2 and the equation is identified. For equation 3, the excluded variables are (Y_{2t}, X_{2t}), so the relevant submatrix is

$$
\begin{array}{cc}
0 & 0 \\
\times & \times
\end{array}.
$$

The two rows are not linearly distinct, as multiplying the first column by a constant (the proportional factor relating the two elements of the second row) produces the second column. The rank is 1 and equation 3 is not identified. An alternative way to see this is to return to our discussion of linear combinations of the equations. A linear combination of any two equations is distinct from the first and second equations. A linear combination of the first and third equations, however, is not distinct from the third equation.

What would restore identification? We need to ensure that equation 3 is distinct from a linear combination of equations 1 and 3. One way is to replace X_{1t} with X_{2t}. The solution seems natural as now each of the 3 equations has a distinct shifter: (X_1, X_3) in equation 1, (X_1, X_2) in equation 2 and (X_2, X_3) in equation 3. Before both equation 1 and equation 3 shared the same shifter (X_1, X_3). Surprisingly, this is not the only way to restore identification. Thus it is possible that both equation 1 and equation 3 share the same shifter (X_1, X_3) and yet are both identified. To do so, add Y_{2t} to the first equation. Clearly, the third equation is now distinct from a linear combination of equations 1 and 3 and the equation is identified. (By adding Y_{2t} we have made Y_{1t} a function of X_{2t}.) From the example it is clear that the distribution of the predetermined variables depends crucially on the distribution of the endogenous variables.