Simultaneous Equation Models

- Suppose we are given the model

$$y_1 = Y_1\gamma_1 + X_1\beta_1 + u_1$$

where $E(X_1'u_1) = 0$ but $E(Y_1'u_1) \neq 0$

- We can often think of Y_1 (and more, say $Y_{\neg 1}$) as being determined as part of a system of equations along with y_1. It is convenient to summarize this system as

$$Y\Gamma \equiv XB + U$$

where $Y = (y_1 Y_{\neg 1}) \in \mathbb{R}^{nxG}$, $X = (X_1 X_{\neg 1}) \in R^{nxK}$, and $U = (U_1 U_{\neg 1}) \in R^{nxG}$ are random matrices; $\Gamma \in \mathbb{R}^{GxG}$ and $B \in \mathbb{R}^{KxG}$ are coefficient matrices. This is called the *structural form* of the model. It has coefficients and disturbances that have a structural/economic interpretation.

- Structural models can explain why some regressors are correlated with the disturbance defined by the parameters of interest. They can also suggest instruments.
- As long as Γ is invertible, we can solve for the *endogenous* variables Y in terms of the *exogenous* variables X :

$$Y \equiv X\Pi + V$$

where $\Pi = B\Gamma^{-1}$ and $V = U\Gamma^{-1}$. This is called the *reduced form* of the model because it is expressed in terms of the coefficients of the BLP that have only a statistical interpretation (and can be estimated by OLS).
- By construction, each endogenous variable is correlated at least with its own reduced form disturbance; since the latter is a linear combination of all the structural disturbances, we usually have each endogenous variable correlated with all the structural disturbances.

An Example: Labour Market Equilibrium

In what follows, I'll switch notation and denote exogenous variables by z rather than x

- Suppose we model the hours of labour supplied using

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

  where w is the wage, z_1 is an observed "supply shifter" (perhaps age or number of children), and u_1 is an unobserved "supply shifter" (call it "spunk").

- Notice α_1 and β_1 are behavioural or "structural" parameters (α_1 answers the question: "By how much would you increase your labour supply if your wage went up by one unit, holding everything else constant?")

- Suppose we model the hours of labour demanded using

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

  where z_2 is an observed "demand shifter" (perhaps relative price of output, cost of materials, or capital stock), and u_2 is an unobserved "supply shifter".

- Notice α_2 and β_2 are behavioural or "structural" parameters (α_2 answers the question: "By how much would firms increase the hours of labour demanded if the wage went up by one unit, holding everything else constant?"

- Notice that it is the difference in the variables z_1 and z_2 that allow us to distinguish between the supply and demand eqns. If they were the same (or both missing) then we couldn't distinguish demand and supply empirically.

- Equilibrium demands $h_s = h_d = h$ so, as long as $\alpha_2 \neq \alpha_1$, we can solve for the reduced form, observation by observation using

$$\begin{bmatrix} h_i \\ w_i \end{bmatrix} = \frac{1}{\alpha_2 - \alpha_1} \begin{bmatrix} \alpha_2\beta_1 & -\alpha_1\beta_2 \\ \beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} z_{1i} \\ z_{2i} \end{bmatrix} + \frac{1}{\alpha_2 - \alpha_1} \begin{bmatrix} \alpha_2 & -\alpha_1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix}$$

OR using the matrix notation above,

$$\begin{bmatrix} h & w \end{bmatrix} = \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} \frac{\alpha_2\beta_1}{\alpha_2-\alpha_1} & \frac{\beta_1}{\alpha_2-\alpha_1} \\ \frac{-\alpha_1\beta_2}{\alpha_2-\alpha_1} & \frac{-\beta_2}{\alpha_2-\alpha_1} \end{bmatrix} + \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

$$\equiv Z\Pi + V$$

- Define

$$V\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} | Z\right) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{bmatrix} \equiv \Sigma$$

- Assume that variables in Z are strictly exogenous. The implies $E(Z'U) = E(Z'V) = 0$.
- $E(w'M_{z_1}u_1) = (\alpha_2\sigma_1^2 - \alpha_1\sigma_{12})/(\alpha_2 - \alpha_1) \neq 0$ (in general), so OLS applied to the labour supply equation will not estimate the structural parameters consistently.

Inconsistency of OLS with Simulaneity Bias

- Suppose we use OLS to estimate the labour supply equation

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

- Use F-W, we get

$$\widehat{\alpha}_1 = \alpha_1 + (w' M_{z_1} w)^{-1} (w' M_{z_1} u_1)$$

- From the reduced form, we have

$$M_{z_1} w = M_{z_1} z_2 \beta_2 + M_{z_1} \frac{\alpha_2 u_1 - \alpha_1 u_2}{\alpha_2 - \alpha_1}$$

- Therefore

$$plim\left(\frac{1}{n} w' M_{z_1} w\right) = var(M_{z_1} z_2)\beta_2^2 + \frac{\alpha_2^2 \sigma_1^2 + \alpha_1^2 \sigma_2^2 - 2\alpha_1 \alpha_2 \sigma_{12}}{(\alpha_2 - \alpha_1)^2}$$

$$plim\left(\frac{1}{n} w' M_{z_1} u_1\right) = \frac{\alpha_2 \sigma_1^2 - \alpha_1 \sigma_{12}}{(\alpha_2 - \alpha_1)^2}$$

Special Cases:

1. $\alpha_2 = 0$ and $\sigma_{12} = 0$
   - If labour demand doesn't depend on wages and the unobserved shocks to labour demand and supply are uncorrelated, then wages are uncorrelated with labour supply shocks, so the usual results hold and OLS is consistent.
2. $var(M_{z_1} z_2)\beta_2^2 \uparrow \infty$
3. $\sigma_2^2 \uparrow \infty$
   - If the movement in equilibrium hours and wages is dominated by labour demand shifters, then OLS will consistently estimate the labour supply schedule.

Identification

- Identification of the supply equation requires $E(Z'(w \;\; z_1))$ to have full column rank. Clearly this has the same rank as the matrix $E(Z'(M_{z_1}w \;\; z_1))$ Expanding from the reduced form yields

$$E\begin{bmatrix} 0 & z_1'z_1 \\ z_2'M_{z_1}z_2\frac{\beta_2}{\alpha_1-\alpha_2} & z_2'z_2 \end{bmatrix}$$

  has full column rank.

- Clearly, we can't have $\alpha_1 = \alpha_2$. Otherwise, there is no equilibrium, and the model makes no sense.

- $E(Z'Z)$ must have full column rank (the instruments must have some variance and be linearly independent)

- We must also have $\beta_2 \neq 0$. This is extremely intuitive. If we want to estimate the labour supply equations we need instruments. The way to find them is to look for variables that affect labour demand but not labour supply. If all the restrictions in the structural form come from *exclusion* restrictions, we get a necessary condition called the *order condition*: There must be at least as many **excluded** exogenous variables as **included** endogenous variables for the coefficients in a structural equation to be identified.
- The n.s. condition for identification is called the *rank condition*.
- We can also get identification in other ways: other linear (or nonlinear) restrictions on the parameters of the structural equations, restrictions on the covariance matrix, restrictions coming from knowing the likelihood of the errors.