Econometrics, Lecture 15A2, Exogeneity, simultaneity and identification

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- We looked at VARs where all the variables were endogenous, but determined by their predetermined lagged values.
- Now look at systems where there is feedback between the variables. We started this in the whiteboard session. Will repeat some of that material in case you couldn't read my writing.
- Originally a variable was endogenous if simultaneously determined within a system. But endogeneity has widened to any failure of exogeneity including omitted variables and measurement error.
- This raises two issues (a) identification can we estimate the parameters and (b) how do we estimate them since OLS will not work.
- Identification is the prior issue and we looked at that through demand and supply in the whiteboard session.

Exogeneity

- One definition of exogeneity is where a variable is independent or uncorrelated with the unobserved errors.
- The other definition is in terms of observables. The joint distribution iss the product of the conditional distribution and the marginal:

  $$D_j(y_t, x_t; \theta_j) = D_c(y_t \mid x_t; \theta_c) D_m(x_t; \theta_m)$$

  Weak exogeneity is when the parameters of interest are functions only of, $\theta_c = \theta = (\beta, \sigma^2)$, and the parameters of the conditional and marginal distributions should be 'variation free': no restrictions linking them. There is no information in θ_m about θ_c. If it fails we need to know how x_t is generated.
- Exogeneity is defined relative to the parameters of interest. x may be weakly exogenous for some parameters and not others.

The Simultaneous Equations Model

- A vector of m endogenous variables, $\mathbf{y}_t$, are jointly determined by k exogenous variables $\mathbf{x}_t$. The structural form is

$$\mathbf{B}\mathbf{y}_t = \boldsymbol{\Gamma}\mathbf{x}_t + \mathbf{u}_t; \quad \mathbf{E}(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Omega} \tag{1}$$

  B is an $m \times m$ matrix, that describes how the endogenous variables interact Γ is a $m \times k$ matrix.

- The reduced form is

$$\mathbf{y}_t = \mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{x}_\mathbf{t} + \mathbf{B}^{-1}\mathbf{u}_t \tag{2}$$

$$\mathbf{y}_t = \boldsymbol{\Pi}\mathbf{x}_t + \mathbf{v}_t \tag{3}$$

$$\mathbf{E}(\mathbf{v}_t\mathbf{v}_t') = \boldsymbol{\Sigma} = \mathbf{B}^{-1}\boldsymbol{\Omega}\mathbf{B}^{-1\prime}$$

- Equation (2) is called the restricted reduced form, RRF, since it reflects any restrictions on $\mathbf{B}$ and $\boldsymbol{\Gamma}$, while (3) is called the unrestricted reduced form, URF.

Identification

- We can estimate the $m \times k$ matrix $\mathbf{\Pi}$, since the $\mathbf{x}_t$ are exogenous, this is just m equations estimated by OLS.
- However we want to estimate the structural parameters, since these are often the objects of economic interest, like price elasticity of demand.
- Thus we need to estimate both B an $m \times m$ matrix and Γ a $m \times k$ matrix. But we only have the $m \times k$ elements of the matrix $\mathbf{\Pi}$, so we are m^2 elements short.
- This is the "identification problem", to find m^2 prior restrictions to obtain all the parameters of the structural form.

Observational Equivalence

- Since, $\mathbf{\Pi} = \mathbf{B}^{-1}\mathbf{\Gamma} = (\mathbf{BP})^{-1}\mathbf{P\Gamma}$ for any $m \times m$ non-singular matrix $\mathbf{P}$, then $\mathbf{B}$ and $\mathbf{B}^* = (\mathbf{BP})$ and $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^* = \mathbf{P\Gamma}$ are observationally equivalent.
- This is the same problem we had with identifying cointegrating vectors and impulse response functions. We need to specify $\mathbf{P}$ *a priori by the* m^2 "just identifying" restrictions.
- For m^2 restrictions on the system we need at least m restrictions on each of the m equations.
- If there are d restrictions available for an equation,
  - when $d < m$, the equation is said to be underidentified or not identified and cannot be estimated;
  - when $d = m$ it is said to be exactly identified or just identified;
  - when $d > m$ it is said to be overidentified. Over identifying restrictions can be tested.
- You can have a system with some equations identified and others not identified.

Restrictions

- One restriction on each equation will come from normalisation: the coefficient of a dependent variable equals unity.
- The requirement that $d \geqslant m$ for each equation is called the order condition, a necessary but not sufficient condition for identification.
- The sufficient condition is the rank condition. Typically that a variable excluded from an equation must have an effect on the system in another equation.
- The order condition can be written in lots of different but equivalent ways. One way of expressing it for a particular equation is that the number of excluded exogenous variables (not appearing in that equation) must be greater or equal to the number of included right hand side endogenous variables.

Demand and supply

- The model in structural form is

$$\begin{aligned} q_t^d &= \gamma_{10} + \beta_{12} p_t + \gamma_{11} y_t + u_{1t}, && (4) \\ q_t^s &= \gamma_{20} + \beta_{22} p_t + \gamma_{22} w_t + u_{2t}. && (5) \end{aligned}$$

- Price adjusts so that quantity demanded equals that supplied $q_t^d = q_t^s = q_t$.
- The two endogenous variables p_t and q_t are determined by the exogenous income y_t, and the weather $w_t$and the errors.
- With $\Phi = [\beta_{12} - \beta_{22}]$, $v_{1t} = \Phi^{-1}(u_{2t} - u_{1t})$, $v_{2t} = \Phi^{-1}(\beta_{12} u_{2t} - \beta_{22} u_{1t})$ The RRF and URF are:

$$\begin{aligned} p_t &= \Phi^{-1}\{(\gamma_{20} - \gamma_{10}) - \gamma_{11} y_t + \gamma_{22} w_t\} + v_{1t} \\ q_t &= \Phi^{-1}\{(\beta_{12}\gamma_{20} - \beta_{22}\gamma_{10}\} - \beta_{22}\gamma_{11} y_t + \beta_{12}\gamma_{22} w_t\} + v_{2t}, \end{aligned}$$

$$\begin{aligned} p_t &= \pi_{10} + \pi_{11} y_t + \pi_{12} w_t + v_{1t} \\ q_t &= \pi_{20} + \pi_{21} y_t + \pi_{22} w_t + v_{2t} \end{aligned}$$

Demand and supply in matrix form

- Odd to have quantity the dependent variable in both equations. Usually each endogenous variable is the dependent variable in one equation, B has ones on the diagonal.
- In form (1), with $m = 2$, $k = 3$, D&S system is

$$\begin{bmatrix} 1 & -\beta_{12} \\ 1 & -\beta_{22} \end{bmatrix} \begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} \gamma_{10} & \gamma_{11} & 0 \\ \gamma_{20} & 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} 1 \\ y_t \\ w_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

- OLS estimates of the structural form will be inconsistent since in the demand equation u_{1t} will be correlated with p_t (which is a function of u_{1t} as the reduced form equations show), so the exogeneity assumption fails.

Demand and supply identification and estimation

- In the demand and supply example, both equations are exactly identified, $d = 2, m = 2$. In demand we have $\beta_{11} = 1; \gamma_{12} = 0$. In supply $\beta_{21} = 1; \gamma_{21} = 0$.
- The rank condition is that $\gamma_{11} \neq 0$ and $\gamma_{22} \neq 0$, so that income and the weather do influence the system and are correlated with the endogenous variables.
- Estimation can be done by Two stage Least Squares, which is the same as Instrumental Variables, IV, discussed later.
- First stage: estimate the reduced form and obtain the predicted values for p_t :

$$\widehat{p}_t = \widehat{\pi}_{10} + \widehat{\pi}_{11} y_t + \widehat{\pi}_{12} w_t$$

Estimation

- $\widehat{p}_t$ is just a function of exogenous variables and so not correlated with u_{1t} (or u_{2t}) and can be used to replace p_t in the second stage estimates of (4) and (5). In (4) with $p_t = \widehat{p}_t + \widehat{v}_{1t}$

$$
\begin{aligned}
q_t &= \gamma_{10} + \beta_{12}(\widehat{p}_t + \widehat{v}_{1t}) + \gamma_{11} y_t + u_{1t} \\
q_t &= \gamma_{10} + \beta_{12}\widehat{p}_t + \gamma_{11} y_t + e_{1t} \qquad (6)
\end{aligned}
$$

- OLS on (6) is fine since $e_{1t} = u_{1t} + \beta_{12}\widehat{v}_{1t}$ and neither are correlated with $\widehat{p}_t$. If the model is not identified (no excluded w_t) $\widehat{p}_t$ will be perfectly correlated with y_t. Estimate errors as

$$\widetilde{u}_{1t} = q_t - \widetilde{\gamma}_{10} - \widetilde{\beta}_{12} p_t - \widetilde{\gamma}_{11} y_t$$

where $\widetilde{\gamma}_{10}$ etc. are the estimates got from the second stage, OLS on (6).

A Keynesian system

- Consider a national income identity and a consumption function:

$$
\begin{aligned}
Y_t &= C_t + I_t, \\
C_t &= \alpha + \beta Y_t + u_t.
\end{aligned}
$$

- Identification is not an issue for the identity, there are no coefficients to estimate. The RRF is

$$
\begin{aligned}
Y_t &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{u_t}{1-\beta} \\
C_t &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + \frac{u_t}{1-\beta}.
\end{aligned}
$$

- The coefficient of investment in the income equation is the standard Keynesian multiplier.
- The URR which we can estimate (with $v_{1t} = v_{2t},$) is

$$
\begin{aligned}
Y_t &= \pi_{10} + \pi_{11} I_t + v_{1t} \\
C_t &= \pi_{20} + \pi_{21} I_t + v_{2t},
\end{aligned}
$$

A Keynesian system, estimation

- Clearly Y_t is correlated with u_t in the consumption function, since as the reduced form shows u_t determines Y_t through consumption. What is the covariance between Y_t and u_t?
- However we can estimate β by "indirect least squares", ILS, from the reduced form coefficients of investment:

$$\widehat{\beta}^{ILS} = \frac{\widehat{\pi}_{21}}{\widehat{\pi}_{11}} = \frac{\sum c_t i_t / \sum i_t^2}{\sum y_t i_t / \sum i_t^2} = \frac{\sum c_t i_t}{\sum y_t i_t}.$$

- Only unique in exactly identified cases, where all the estimators (2SLS, IV, ILS and others) give the same estimates.
- The fourth term in the equation is of the same form as the IV estimator, coming soon.

An overidentified model

- Suppose that we had two exogenous variables, investment and government expenditure

$$
\begin{aligned}
Y_t &= C_t + I_t + G_t, \\
C_t &= \alpha + \beta Y_t + u_t.
\end{aligned}
$$

- The restricted and unrestricted reduced form equations for Y_t are

$$
\begin{aligned}
Y_t &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{1}{1-\beta} G_t + \frac{u_t}{1-\beta} \\
Y_t &= \pi_{10} + \pi_{11} I_t + \pi_{12} G_t + v_{1t}.
\end{aligned}
$$

- The testable over-identifying restriction is that $\pi_{11} = \pi_{12}$.

Next time

- Return to the $y = X\beta + u$, but with $E(X'u) \neq 0$.
- Also have some exogenous "Instrumental Variables" W comparable to income and the weather in our demand and supply example.
- The instrumental variable, IV, estimator is exactly the same as two stage least squares.