University of Illinois
Fall 1998

Department of Economics
Roger Koenker

Economics 472

Lecture 9

Introduction to Simultaneous Equation Econometric Models

Review of Linear "Seemingly Unrelated Regressions"

The simplest example of simultaneous equation models in econometrics is the model which Zellner labeled SUR and statisticians usually call just multivariate regression.

$$y_i = X_i\beta_i + u_i \qquad i = 1, \ldots, m$$

where

$y_i \sim n$-vector of observed responses
$X_i \sim n \times p_i$ matrix of exogenous variables
$u_i \sim n$-vector of "errors"

A typical example would be a *system* of m demand equations in which X_i would be composed of prices and incomes and perhaps other commodity specific exogenous influences on demands. By exogenous in this preliminary setting we will simply mean that

$$EX_i'u_j = 0 \qquad i, j = 1, \ldots, m.$$

which is the natural extension of the orthogonality condition underlying ordinary linear regression with a single response variable.

It is convenient to write the whole system of equations as

$$y = X\beta + u$$

which may be interpreted as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & \cdots & 0 \\ 0 & & & & \\ & & \ddots & & \\ 0 & & & & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}$$

in which the equations have simply been stacked one on top of another. We will suppose that the full mn-vector, u, is normal with mean 0, and covariance matrix

$$Euu' = \Omega \otimes I_n = (\omega_{ij} \;\; I_n)$$

and we may then, immediately, write the optimal (unbiased) estimator of the parameter vector β as,

$$\hat{\beta} = (X'(\Omega \otimes I)^{-1} X)^{-1} X'(\Omega \otimes I)^{-1} y$$

where we note that $(\Omega \otimes I)^{-1} = \Omega^{-1} \otimes I$. Typically, Ω is unknown, but we may estimate it by $\Omega = (\hat{\omega}_{ij})$, with

$$\hat{\omega}_{ij} = \hat{u}_i' \hat{u}_j / n$$

where $\hat{u}_i, i = 1, \ldots, m$ are the n-vectors of residuals from any initial (consistent) estimate of the model, typically from an OLS fit to the individual equations.

An important observation is that there is no efficiency gain from the reweighting by $(\Omega \otimes I)^{-1}$ if $X = (I \otimes X_0)$. That is, if $X_i = X_0$ for all i as would be the case in some demand system contexts, we gain nothing from doing the system estimate over what is accomplished in the equation-by-equation OLS case. To see this write

$$(\Omega^{-1} \otimes I)(I \otimes X_0) = \Omega^{-1} \otimes X_0.$$

We are solving the equations in the weighted case

$$X'(\Omega \otimes I)^{-1} \hat{u} = 0$$

but if $X = (I \otimes X_0)$, this is equivalent to

$$(\Omega^{-1} \otimes X_0') \hat{u}$$

but this is satisfied by assuring that

$$X_0' \hat{u}_i = 0 \qquad i = 1, \ldots, m$$

which are just the normal equations for the separate OLS regressions.

A useful introduction to maximum likelihood estimation of systems of equations may be provided by the SUR model. For this purpose it is convenient to stack the observations in "the opposite way" that is, to write

$$y_j = X_j \beta + u_j \qquad j = 1, \ldots, n$$

where

$$X_j = \begin{bmatrix} x_{j1} & 0 & \cdots & 0 \\ 0 & x_{j2} & \cdots & 0 \\ & & & \\ 0 & \cdots & & x_{jm} \end{bmatrix}$$

where x_{ji} is a p_i *row* vector. Now stacking the model we have,

$$y = X\beta + u$$

and now $u \sim \mathcal{N}(0, I \otimes \Omega)$. Note that, with this formulation

$$\begin{aligned} \hat{\beta} &= (X'(I \otimes \Omega^{-1})X)^{-1} X'(I \otimes \Omega^{-1}) y \\ &= (\sum_{j=1}^{n} X_j' \Omega^{-1} X_j)^{-1} \sum_{j=1}^{n} X_j' \Omega^{-1} y_j \end{aligned}$$

The convenient aspect of this formulation is that we can view $u_j,\ j = 1, \ldots, n$ as independent realizations of an m-variate normal vector and thus the likelihood for the model may be written as,

$$\mathcal{L}(\beta, \Omega) = (2\pi)^{\frac{mn}{2}} |\Omega|^{-\frac{n}{2}} \exp\{-\frac{1}{2}\sum_{j=1}^{n} u_j' \Omega^{-1} u_j\}$$

where implicitly we recognize that the u_j's are functions of the β vector. As usual it is more convenient to work with log likelihood,

$$\ell(\beta, \Omega) = K - \frac{n}{2}\log|\Omega| - \frac{1}{2}\sum u_j'\Omega^{-1}u_j$$

We have already seen how to estimate β in this model. We now consider two variants on estimation of Ω.

Case 1. Suppose that Ω is known up to a scalar, i.e., $\Omega = \omega\Omega_0$ with the matrix Ω_0 known.

Recall that $|\omega\Omega_0| = \omega^m|\Omega_0|$ so

$$\ell(\beta, \omega) = K - \frac{n}{2}(m\log\omega + \log|\Omega_0|) \quad - \frac{1}{2\omega}\sum u_j'\Omega_0^{-1}u_j$$

so

$$\frac{\partial l}{\partial \omega} = -\frac{nm}{2\omega} + \frac{1}{2\omega^2}\sum j_j'\Omega_0^{-1}u_j = 0$$

implies

$$\hat{\omega} = (mn)^{-1}\sum u_j'\Omega_0^{-1}u_j.$$

Case 2. If Ω is completely unknown, we simply differentiate with respect to Ω. Recall that

$$(i) \qquad \frac{\partial \log|A|}{\partial A} = (A')^{-1} \qquad \text{if } |A| > 0$$

$$(ii) \qquad \frac{\partial x'Ax}{\partial A} = xx'$$

$$(iii) \qquad \frac{\partial y'A^{-1}y}{\partial A} = \frac{\partial x'Ax}{\partial A} = xx' = -A^{-1}yy'A^{-1}$$

Verify that these formulae work in the "A scalar" case. Now,

$$\nabla_\Omega \ell = -\frac{n}{2}\Omega^{-1} + \frac{1}{2}\sum \Omega^{-1}u_ju_j'\Omega^{-1}$$

so

$$\hat{\Omega} = n^{-1}\sum u_ju_j'$$

which is the same formula suggested earlier in the lecture.

Now, concentrating the log likelihood as in the single equation case we may simplify the last term,

$$\sum \text{ tr } (u_j'(\sum u_ju_j')^{-1}u_j) = \sum \text{ tr } u_ju_j'(\sum u_ju_j')^{-1} = mn$$

so for purposes of computing likelihood ratios or SIC numbers we have

$$\ell(\hat{\beta}, \hat{\Omega}) = K^* - \frac{n}{2} \log |\hat{\Omega}|$$

where K^* is a constant in dependent of the data.

Introduction to Vector Autoregressive Models

An important class of models in time-series which draw upon the ideas of SUR models are the so called VAR models. Consider an m-vector y_t observed at time t and a model

$$y_t = \mu + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + u_t$$

Again exploiting the lag operator notation we may write this as

$$A(L) y_t = \mu + u_t$$

where

$$A(L) = I - A_1 L - A_2 L^2 - \ldots - A_p L^p.$$

Again, stability is crucial determined by the characteristic equation*,

$$|A(z)| = 0.$$

If the roots of this equation lie outside the unit circle, then all is well, if some roots lie on the unit circle, then it is useful to reformulate the model in the error correction form

$$\Delta y_t = \mu + B_1 \Delta y_{t-1} + B_2 \Delta y_{t-2} + B_{p-1} \Delta y_{t-p+1} - \Pi y_{t-1} + u_t$$

where we have

$$\Pi = A(1) = I - A_1 - \ldots - A_p$$

and has rank less than m. We then factor Π into pieces that have rank $m - r$ and this leads to the theory of cointegrated time series, a topic which is dealt with in some depth in 473. Rather than delve into this any further I will briefly mention one related topic.

Impulse Response Functions, Again

Since we have a somewhat different setting than our single equation demand model, it is worth revisiting the question "what is an IRF for a VAR?" In the VAR context we have no exogenous variables which might be regarded as candidates for a permanent policy shock of the type we have already discussed.

However, we can still ask what would be the path of the system if it were in equilibrium and was then "shocked" by a permanent increase in one of the error realizations. So we are really asking what happens to the whole system of equations, how does it evolve after encountering a once and for all increase in one element of the error vector u_t. Formally, we have the same problem except that now we have matrices everywhere we used to have scalars.

*Note that this characteristic equation now yields roots of an eigenvalue problem, not simply an ordinary polynomial but the principle is the same as before.

If the model is stable in the sense we have already described, we can "invert" the VAR representation and put the model in the MA form,

$$y_t = m + A(L)^{-1}u_t$$

where $A(L)^{-1}u_t$ is interpretable in much the same way that we interpreted

$$D(L)x_t = A(L)^{-1}B(L)x_t$$

in the earlier, simpler, models. To illustrate, it may be helpful to consider the example,

$$A(L) = I - AL$$

In this case the invertible MA representation would have

$$(I - AL)^{-1} = I + A + A^2 + \dots$$

Note that as in the simple case we can verify this directly. Obviously we require that the right hand side converge, for this to make any sense. The MA or impulse response formulation of the model has some inherent ambiguity in the typical case of correlated errors. The underlying thought experiment is rather implausible in this case and there has been considerable discussion about various schemes to orthogonalize the errors, but these "solutions" introduce new problems having to do with the nonuniqueness of the orthogonalization.