Econometrics Lecture 8: Granger Causality and Vector Autoregressive Models

R. G. Pierse

1 Introduction

Last week we looked at simultaneous equations systems. These systems involve the imposition of two types of restriction: firstly some variables are taken as *exogenous* and are not explained within the model, secondly some parameters are restricted in order to achieve identification. Sims (1980) argues that both these steps involve arbitrary assumptions about the direction and timing of causal linkages in economics that may be rejected when tested against data. Instead Sims proposes a model with a very free dynamic specification which can be used to test causal linkages rather than impose them. The alternative he advocates is the *vector autoregressive* or *VAR* model.

The *VAR* model is a multiple variable generalisation of the autoregressive model. The regressors are lags of all the variables in the model. Because there are no current endogenous variables on the right-hand side of equations, no restrictions need to be imposed to identify the system. The set of regressors is the same for each equation so that each equation is estimated efficiently using the *OLS* estimator.

The *VAR* model ignores economic restrictions and so can be thought of as 'atheoretic'. It can also be thought of as the reduced form of a simultaneous equations system where the predetermined variables are all lagged dependent variables and there are no pure exogenous variables. *VAR* models have often been used for forecasting and have been claimed to forecast better than simultanous equations models. The implication of this is that the theory derived over-identifying restrictions imposed on simultaneous equations systems are being rejected by the data. In the past, simultanous equations systems have tended to neglect dynamics and their poor forecasting performance may be partly the result of this. *VAR* models provide a benchmark against which their forecasts can be judged.

A practical problem with *VAR* models is that the number of parameters to be estimated increases fast with the number of variables and the maximum order of lag. This acts as a constraint on the maximum size of a *VAR* model that can be built.

2 Granger Causality

The concept of Granger causality starts with the premise that the future cannot cause the past. If event A occurs after event B, then A cannot cause B. Granger (1969) applies this concept to economic time series to determine whether one time series 'causes' in the sense of precedes another. However, merely because event A occurs before B does not mean that A causes B. For example, Christmas shopping does not cause Christmas. Granger causality is therefore nothing to do with the notion of causality in the common (or philosophical) sense. It is related to the question of how useful one variable (or set of variables) y_2 is for forecasting another variable (or set of variables) y_1. If y_2 does not Granger cause y_1 then y_2 does not help to forecast y_1.

Consider the case of two variables x_t and y_t. Then x_t does not cause y_t if, in a regression of y_t on lagged x_t and lagged y_t then *all* the coefficients on the former are zero. Formally, in the regression

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + u_t$$

then x_t does not cause y_t if $\beta_i = 0$, $i = 1, \cdots, p$. A test of Granger non-causality can be based on a test of the hypothesis that

$$H_0 : \beta_i = 0, \quad i = 1, \cdots, p.$$

This test is only valid asymptotically since the regression includes lagged dependent variables, but in practice, standard F tests are often used.

2.1 Causality and Exogeneity

It is important to distinguish carefully between the concept of causality and the various definitions of exogeneity. Two concepts related to exogeneity were considered in the last lecture.

Definition 2.1. *Predeterminedness. A variable is predetermined in a particular equation if it is independent of current and future errors in that equation.*

Definition 2.2. *Strict exogeneity. A variable is strictly exogenous in a particular equation if it is independent of past, current and future errors in that equation.*

One problem with these definitions is that whether or not a variable is exogenous depends on the parameter under consideration. A much more general concept of exogeneity, *weak exogeneity*, due to Engle, Hendry and Richard (1983) makes this clear.

Definition 2.3. *Weak exogeneity. A variable x_t is weakly exogenous for estimating a set of parameters λ_1 if the joint probability density function $f(y_t, x_t; \lambda)$ can be partitioned as*

$$f(y_t, x_t; \lambda) = g(y_t|x_t; \lambda_1)h(x_t; \lambda_2)$$

so that inference on λ_1 can be conducted with no loss of information from the conditional density function $g(y_t|x_t; \lambda_1)$.

Weak exogeneity is a condition for efficient estimation of the parameters λ_1 conditioning on the variables x_t. The Wu-Hausman test considered in last week's lecture can be shown to be a test for weak exogeneity, and this is the most useful definition of exogeneity in econometrics.

One further definition of exogeneity will be briefly mentioned: that is the concept of strong exogeneity. If x_t is weakly exogenous and x_t is not Granger caused by y_t then x_t is said to be *strongly exogenous*. Strong exogeneity is not necessary for efficient estimation.

The various definitions can be compared by considering the simple bivariate dynamic model

$$\begin{aligned} y_t &= \alpha_1 x_t + \beta_{11} y_{t-1} + \beta_{12} x_{t-1} + u_{1t} \\ x_t &= \alpha_2 y_t + \beta_{21} y_{t-1} + \beta_{22} x_{t-1} + u_{2t} \end{aligned}$$

where the error terms u_{1t} and u_{2t} are assumed to be independent. If $\alpha_2 = 0$ then x_t is predetermined for y_t in the first equation. If $\alpha_2 = 0$ and $\beta_{21} = 0$ then x_t is strictly exogenous for y_t in the first equation. In this simple model, the condition $\alpha_2 = 0$ is sufficient for x_t to be weakly exogenous for estimation of the parameters α_1, β_{11} and β_{12} in the first equation. The condition $\alpha_2 = 0$ is thus the condition under which the parameters α_1, β_{11} and β_{12} can be efficiently estimated by *OLS* on the first equation. In this model, weak exogeneity coincides with predeterminedness. However, in more general models, the two definitions will result in different conditions.

If $\alpha_2 = 0$ and $\beta_{21} = 0$ then y_t does not Granger cause x_t. Thus in this model Granger non-causality happens to coincide with strict exogeneity. In general however, Granger causality is neither necessary nor sufficient for (weak) exogeneity.

3 Vector Autoregressive Models

The pth order vector autoregressive model or *VAR* model can be written as

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

with

$$\boldsymbol{\varepsilon}_t \sim iid\ N(\mathbf{0}, \mathbf{\Sigma}).$$

where y_t is a $n \times 1$ vector of variables at time t and $\mathbf{c}$ is an intercept. Other deterministic components can be added to the model without affecting the analysis. There are pn^2 parameters in the $\mathbf{\Phi}$ matrices. Making use of the lag operator L, defined by $L^k x_t = x_{t-k}$, the equation can be rewritten as

$$\mathbf{\Phi}(L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \tag{3.1}$$

where

$$\mathbf{\Phi}(L) = \mathbf{\Phi}_0 L^0 - \mathbf{\Phi}_1 L^1 - \cdots - \mathbf{\Phi}_p L^p,$$

$\mathbf{\Phi}_0 = \mathbf{I}$, and, to ensure stationarity, the roots of $|\mathbf{\Phi}(L)|$ lie *outside the unit circle*.

In 'pure' *VAR* models no *a priori* economic restrictions are imposed on $\mathbf{\Phi}$.

3.1 Estimating *VAR* Models

The restriction that $\mathbf{\Phi}_0 = \mathbf{I}$ implies that there are no current endogenous variables in the model. Equations are related solely through the off-diagonal elements in the covariance matrix $\mathbf{\Sigma}$. This is the special case of a *SURE* model where the regressors are the same for each equation. In this case the maximum likelihood estimator for $\mathbf{\Phi}$ is simply the *OLS* estimator. The error covariance matrix $\mathbf{\Sigma}$ is consistently estimated by

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t'$$

where $\widehat{\boldsymbol{\varepsilon}}_t$ is the $n \times 1$ vector of *OLS* residuals.

3.2 Granger causality and *VARs*

Granger causality can be tested easily in a *VAR* model. Consider the two sets of variables $\mathbf{y}_1$ and $\mathbf{y}_2$.

Definition 3.1. $\mathbf{y}_2$ *does not Granger cause* $\mathbf{y}_1$ *if for all* $s > 0$

$$
\begin{aligned}
& \text{MSE}\ [\widehat{E}(\mathbf{y}_{1,t+s} \mid \mathbf{y}_{1,t}, \mathbf{y}_{1,t-1}, \cdots)] \\
= \ & \text{MSE}\ [\widehat{E}(\mathbf{y}_{1,t+s} \mid \mathbf{y}_{1,t}, \mathbf{y}_{1,t-1}, \cdots, \mathbf{y}_{2,t}, \mathbf{y}_{2,t-1}, \cdots)]
\end{aligned}
$$

Consider the partitioned *VAR* model

$$\mathbf{\Phi}(L)\mathbf{y}_t = \begin{bmatrix} \mathbf{\Phi}_{11}(L) & \mathbf{\Phi}_{12}(L) \\ \mathbf{\Phi}_{21}(L) & \mathbf{\Phi}_{22}(L) \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \end{bmatrix} = \mathbf{c} + \boldsymbol{\varepsilon}_t$$

where

$$\boldsymbol{\varepsilon}_t \sim iid\ N(\mathbf{0}, \mathbf{I}).$$

If $\mathbf{\Phi}_{12}(L) = 0$ in this model *then* $\mathbf{y}_2$ does not Granger cause $\mathbf{y}_1$.

3.3 Impulse Response Functions

The *VAR* model can always be transformed into the infinite Moving Average (*Wold*) representation:

$$\mathbf{y}_t = \mu + \boldsymbol{\varepsilon}_t + \mathbf{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Psi}_2\boldsymbol{\varepsilon}_{t-2} + \cdots +$$

where

$$\frac{\partial\ \mathbf{y}_{t+s}}{\partial\ \boldsymbol{\varepsilon}_t} = \mathbf{\Psi}_s$$

is the response of $\mathbf{y}$ in period $t+s$ to a shock in period t. Considered as a function of s, $\mathbf{\Psi}_s$ is known as the *impulse response function.*

The matrices $\mathbf{\Psi}_s$ can be computed *recursively* using the formula

$$\mathbf{\Psi}_s = \sum_{j=1}^{\min(p,s)} \mathbf{\Psi}_{s-j}\mathbf{\Phi}_j, \quad s = 1, 2, 3, \cdots$$

where $\mathbf{\Psi}_0 = \mathbf{I}$. Alternatively, they can be generated by simulation of the *VAR* model.

3.4 Orthogonal shocks

Note that the off-diagonal terms in $\mathbf{\Sigma}$ imply that the shocks $\boldsymbol{\varepsilon}_t$ are *correlated* between equations. However, $\mathbf{\Sigma}^{-1}$ can always be factorised as

$$\mathbf{\Sigma}^{-1} = \mathbf{H}\mathbf{H}'$$

where $\mathbf{H}$ is a lower triangular matrix having zeros above the diagonal (This is known as a *Cholesky decomposition*). It then follows that

$$\boldsymbol{\varepsilon}_t^* = \mathbf{H}'\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{H}'\boldsymbol{\Sigma}\mathbf{H}) = N(\mathbf{0}, \mathbf{I})$$

and the transformed shocks $\boldsymbol{\varepsilon}_t^*$ are *orthogonal.* This orthogonalisation *is not unique* and depends on the ordering of the variables in the *VAR*. Sometimes there may be a natural *recursive* ordering of the equations that justifies the Cholesky factorisation.

Pesaran (1997) proposes a generalised orthogonalisation that does not depend on the ordering of variables in the *VAR*. This is implemented in version 4 of the *MicroFit* computer package.

4 *VAR* models and Structural models

4.1 Simultaneous Equations Models

There is a close relationship between *VAR* models and simultaneous equations models. Consider the standard simultaneous equations model

$$\mathbf{B}_0\mathbf{y}_t = \boldsymbol{\Gamma}\,\mathbf{z}_t + \mathbf{u}_t, \quad \mathbf{B}_0 \neq \mathbf{I} \tag{4.1}$$

$$\mathbf{u}_t \sim iid\ N(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\mathbf{z}_t$ is a set of *predetermined* variables comprising:

a) lags of endogenous variables $\mathbf{y}_t$

b) current and lagged *exogenous* variables $\mathbf{x}_t$ that are *not* explained within the system

Let the two components of $\mathbf{z}_t$ be separated out into

$$\boldsymbol{\Gamma}\mathbf{z}_t = \mathbf{B}^*(L)\mathbf{y}_{t-1} + \mathbf{C}(L)\mathbf{x}_t.$$

Then the system can be completed by adding a *VAR* model for $\mathbf{x}_t$:

$$\mathbf{D}(L)\mathbf{x}_t = \mathbf{v}_t, \quad \mathbf{D}_0 = \mathbf{I}$$

so that, stacking,

$$\begin{bmatrix} \mathbf{B}(L) & -\mathbf{C}(L) \\ \mathbf{0} & \mathbf{D}(L) \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix}$$

where $\mathbf{B}(L) = \mathbf{B}_0 - \mathbf{B}^*(L)L$. This can be re-expressed as

$$\mathbf{\Phi}(L)\mathbf{w}_t = \boldsymbol{\eta}_t \tag{4.2}$$

where $\mathbf{w}_t = (\mathbf{y}_t' : \mathbf{x}_t')'$ and $\boldsymbol{\eta}_t = (\mathbf{u}_t' : \mathbf{v}_t')'$ and the first term in $\mathbf{\Phi}(L)$ is given by

$$\mathbf{\Phi}_0 = \begin{bmatrix} \mathbf{B}_0 & -\mathbf{C}_0 \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

The matrices $\mathbf{B}(L)$ and $\mathbf{C}(L)$ will typically have *overidentifying* restrictions (usually zero restrictions) deriving from economic theory.

4.2 The Reduced and Final Forms

Now define

$$\mathbf{\Phi}(L) = \mathbf{\Phi}_0\mathbf{\Phi}^*(L)$$

where $\mathbf{\Phi}_0^* = \mathbf{I}$. Then

$$\mathbf{\Phi}^*(L)\ \mathbf{w}_t = \mathbf{\Phi}_0^{-1}\boldsymbol{\eta}_t = \boldsymbol{\eta}_t^* \tag{4.3}$$

is the *reduced form* of the system. In form this is a *VAR* model but the structure imposes *restrictions* on $\mathbf{\Phi}^*(L)$.

Inverting the lag polynomial $\mathbf{\Phi}^*(L)$ gives the *final form* representation

$$\mathbf{w}_t = \mathbf{\Phi}^*(L)^{-1}\mathbf{\Phi}_0^{-1}\boldsymbol{\eta}_t = \mathbf{\Phi}(L)^{-1}\boldsymbol{\eta}_t$$

which is an infinite moving average representation in terms of the *structural* errors $\boldsymbol{\eta}_t$.

Thus the simultaneous equations system (4.1) can be seen to be a *VAR* system like (4.3) that is *subject to a set of (over)identifying restrictions.* The real difference between structural *VARs* and conventional structural equation systems is the nature of the identification restrictions that are imposed.

5 Identification restrictions

Identification in the model (4.2) requires the imposition of n^2 restrictions. In conventional simultaneous equations models, identification is generally achieved by imposing zero restrictions on the coefficients on the predetermined variables in the matrices $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_p$. Sims (1980) argues against this type of identifying restriction on the dynamics. Instead, structural *VAR* modellers have sought to impose identifying restrictions either on the matrix of contemporaneous coefficients $\mathbf{\Phi}_0$, on the covariance matrix $\mathbf{\Sigma}$, or on the *long run* coefficients

$$\mathbf{\Phi}(1) = \mathbf{\Phi}_0 - \mathbf{\Phi}_1 \cdots - \mathbf{\Phi}_p = \mathbf{\Psi}(1)^{-1}.$$

For example Sims (1980) suggests imposing the $n \times n - 1$ restrictions that $\mathbf{\Sigma}$ is diagonal, plus the $n \times n + 1$ restrictions that $\mathbf{\Phi}_0$ is lower triangular with ones on the diagonal. These restrictions together make the system *recursive* and exactly identified. By contrast, Blanchard and Quah (1989) and King *et al.* (1991) identify a structural *VAR* through restrictions on the long run coefficients $\mathbf{\Phi}(1)$.

References

[1] Blanchard, O.J. and D. Quah (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review*, 79, 655–73.

[2] Engle, R.F., D.F. Hendry and J-F. Richard (1983), 'Exogeneity', *Econometrica*, 51, 277–304.

[3] Granger, C.W.J. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica*, 37, 24–36.

[4] King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991), 'Stochastic trends and economic fluctuations', *American Economic Review*, 81, 819–840.

[5] Pesaran, M.H. (1997), 'MicroFit Version 4 Manual', *mimeo.*

[6] Sims, C.A. (1980), 'Macroeconomics and reality', *Econometrica*, 48, 1–48.