JOHANSEN'S METHOD OF ESTIMATING COINTEGRATED VECTOR AUTOREGRESSIVE SYSTEMS

Cointegrated Vector Autoregressive Systems

Consider a vector autoregressive equation in the form of

$$A(L)y(t) = y(t) - \Phi_1 y(t-1) - \cdots - \Phi_n y(t-p) = \varepsilon(t), \tag{1}$$

which purports to describe how the vector $y(t)$ of M variables is generated. In this equation, the individual processes within the disturbance vector $\varepsilon(t)$ on the RHS are presumed to be stationary and, therefore, the combination $A(L)y(t)$ of the LHS must be stationary likewise.

There are a variety of ways in which the stationarity of the LHS can arise. It may indeed be attributable to the stationarity of each of the elements of $y(t)$. Alternatively, it may be that the operator $A(L)$ is effective in taking differences of the nonstationary elements of $y(t)$. However, it is also possible for the stationarity to result, in part at least, from the combination of cointegrated nonstationary processes which follow common trends.

In order to demonstrate this third possibility, let us, for convenience, make the assumption that $p = 2$. Then equation (1) can be written as

$$y(t) - \Phi_1 y(t-1) - \Phi_2 y(t-2) = \varepsilon(t). \tag{2}$$

The equation can be transformed to give

$$\begin{aligned} &[\, I \quad -\Phi_1 \quad -\Phi_2 \,] \begin{bmatrix} I & I & I \\ 0 & I & I \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} I & -I & 0 \\ 0 & I & -I \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} y(t) \\ y(t-1) \\ y(t-2) \end{bmatrix} \\ &= [\, I \quad -\Pi \quad -\Gamma \,] \begin{bmatrix} \nabla y(t) \\ \nabla y(t-1) \\ y(t-2) \end{bmatrix} = \varepsilon(t), \end{aligned} \tag{3}$$

where

$$-\Pi = I - \Phi_1, \qquad -\Gamma = I - \Phi_1 - \Phi_2. \tag{4}$$

Thus, in place of (2), we have an equivalent equation

$$\nabla y(t) - \Pi \nabla y(t-1) - \Gamma y(t-2) = \varepsilon(t). \tag{5}$$

Equation (5) contains a mixture of differenced and undifferenced variables. We imagine that the differencing is sufficient to reduce the variables to stationarity. Therefore, if the model is to be consistent, the term $\Gamma y(t-2)$ must also be stationary. This will be impossible if $y(t)$ is nonstationary and if Γ has full rank. It will only be possible if there are one or more cointegrating relationships between the variables such that there exist linear combinations

$A'y(t-2)$, embedded within $\Gamma y(t-2) = \Delta A'y(t-2)$, which render the variables stationary. Here, $A = [\alpha_1, \alpha_2, \ldots, \alpha_S]$ is a matrix of order $M \times S$ with $\text{Rank}(A) = S \leq M$. Its columns are the so-called cointegrating vectors. Observe that $\Gamma = \Delta A' = \{\Delta Q\}\{Q^{-1}A'\}$, where Q is an arbitrary nonsingular matrix of order S. Therefore, the cointegrating vectors are not uniquely determined unless further restrictions are imposed upon A.

An individual cointegrating relationship of the form $\alpha_i' y(t-2)$ represents a restriction on the variables of the system which asserts that, in the long run, they will tend to maintain a certain proportionality. The greater the number of cointegrating relationships, the more closely are these proportions governed. In the limiting case, where the number of relationship is one less that the number of variables, every ratio amongst the variables is governed.

The number of linearly independent cointegrating relationships is equal to the rank of Γ. If the matrix Γ is of full rank, then every arbitrary combination of the sequences in the vector $y(t)$ must be stationary; which means that each of the sequences must be stationary. Then there will be no call for differencing. On the other hand, if Γ is null, with a rank of zero, then there will be no cointegrating relationships, and each sequence will be following its own independent random walk, which will be present in the equation only in its stationary differenced form.

Maximum Likelihood Estimation

Consider the system of vector autoregressive equations which has been rewritten in the form of

$$\nabla y(t) = \Gamma y(t-p) + \sum_{j=1}^{p-1} \Pi_j \nabla y(t-j) + \varepsilon(t). \tag{6}$$

We imagine that there is a set of T observations running from $t = 0$ to $t = T-1$. Let the observations on $\nabla y(t)$ for $t = 0, \ldots, T-1$ be gathered in a succession of *row* vectors which together form a matrix Y of order $T \times M$. Let the observations on $\nabla y(t-1), \ldots, \nabla y(t-p+1)$ be gathered likewise in a matrix X_2 of order $T \times M(p-1)$, and let X_1 be the matrix of successive observations on $y(t-p)$. Then our system of equations is

$$\begin{aligned} Y &= X_1 B_1 + X_2 B_2 + \mathcal{E} \\ &= XB + \mathcal{E}, \end{aligned} \tag{7}$$

where $B_1 = \Gamma'$ and $B_2 = [\Pi_1', \ldots, \Pi_{p-1}']$. The hypothesis of cointegration is that the $M \times M$ matrix $B_1 = \Gamma'$ is of rank $S < M$. This is equivalent to the proposition that $B_1 = A\Delta'$, where A has order $M \times S$ and Δ' has order $S \times M$ and both matrices are of rank S.

The log-likelihood function of the model is given by

$$L^*(B, \Sigma) = -\frac{MT}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}\text{Trace}\{\mathcal{E}'\mathcal{E}\Sigma^{-1}\}, \tag{8}$$

where $B = [B_1, B_2]$. We may eliminate Σ from this expression by observing that the maximum-likelihood estimate of Σ conditional upon the value of B is

$$\Sigma(B) = \frac{1}{T}(Y - XB)'(Y - XB). \tag{9}$$

Substituting this in (8) gives

$$L^*(B, \Sigma) = -\frac{MT}{2}\log(2\pi) - \frac{T}{2}\log\left|\frac{(Y - XB)'(Y - XB)}{T}\right| - \frac{MT}{2}. \tag{10}$$

For a given value of B_1, the maximum-likelihood estimate of B_2 is

$$B_2 = (X_2'X_2)^{-1}X_2'(Y - X_1B_1) \tag{11}$$

Substituting this into (7) and rearranging gives

$$\begin{aligned}\mathcal{E} &= Y - X_1B_1 - X_2(X_2'X_2)^{-1}X_2'(Y - X_1B_1)\\ &= (I - P_2)Y + (I - P_2)X_1B_1,\end{aligned} \tag{12}$$

where $P_2 = X_2(X_2'X_2)^{-1}X_2'$. On defining $V = (I - P_2)Y$ and $(I - P_2)X_1B_1 = WB_1 = WA\Delta'$, we can write the criterion function as

$$|\Sigma(A, \Delta)| = |T^{-1}(V - WA\Delta')'(V - WA\Delta')|. \tag{13}$$

This can be concentrated in respect of Δ holding A constant. The estimating equation for Δ is

$$\Delta'(A) = (A'W'WA)^{-1}A'W'V; \tag{14}$$

and the concentrated function, which is the generalised variance of the residuals from this regression, is

$$\begin{aligned}|\Sigma(A)| &= |T^{-1}(V'V - V'WA(A'W'WA)^{-1}A'W'V)|\\ &= |S_{VV} - S_{VW}A(A'S_{WW}A)^{-1}A'S_{WV}|.\end{aligned} \tag{15}$$

According to a basic matrix identity we have

$$\begin{aligned}&|A'S_{WW}A||S_{VV} - S_{VW}A(A'S_{WW}A)^{-1}A'S_{WV}|\\ &\quad = |S_{VV}||A'S_{WW}A - A'S_{WV}S_{VV}^{-1}S_{VW}A|,\end{aligned} \tag{16}$$

and this shows that we should find the estimate of A by minimising the function

$$\frac{|A'S_{WW}A - A'S_{WV}S_{VV}^{-1}S_{VW}A|}{|A'S_{WW}A|}, \tag{17}$$

which is a matter of minimising the value of the numerator subject to an arbitrary normalisation of the value of the denominator. It can be shown that the columns of the resulting matrix A are the eigenvectors of the matrix $\lambda S_{WW} - S_{WV}S_{VV}^{-1}S_{VW}$ corresponding to the S eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_S$ of greatest magnitude. This is the problem of canonical correlations.

Reference

Johansen, S., (1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control,* **12,** 231–254.