LECTURE 10

Multivariate ARMA Processes

A vector sequence $y(t)$ of n elements is said to follow an n-variate ARMA process of orders p and q if it satisfies the equation

$$\begin{aligned}(1)\qquad &A_0y(t) + A_1y(t-1) + \cdots + A_py(t-p)\\ &\quad = M_0\varepsilon(t) + M_1\varepsilon(t-1) + \cdots + M_q\varepsilon(t-q),\end{aligned}$$

wherein $A_0, A_1, \ldots, A_p, M_0, M_1, \ldots, M_q$ are matrices of order $n \times n$ and $\varepsilon(t)$ is a disturbance vector of n elements determined by serially-uncorrelated white-noise processes that may have some contemporaneous correlation.

In order to signify that the ith element of the vector $y(t)$ is the dependent variable of the ith equation, for every i, it is appropriate to have units for the diagonal elements of the matrix A_0. These represent the so-called normalisation rules. Moreover, unless it is intended to explain the value of y_i in terms of the remaining contemporaneous elements of $y(t)$, then it is natural to set $A_0 = I$. It is also usual to set $M_0 = I$. Unless a contrary indication is given, it will be assumed that $A_0 = M_0 = I$.

It is assumed that the disturbance vector $\varepsilon(t)$ has $E\{\varepsilon(t)\} = 0$ for its expected value. On the assumption that $M_0 = I$, the dispersion matrix, denoted by $D\{\varepsilon(t)\} = \Sigma$, is an unrestricted positive-definite matrix of variances and of contemporaneous covariances. However, if restriction is imposed that $D\{\varepsilon(t)\} = I$, then it may be assumed that $M_0 \neq I$ and that $D(M_0\varepsilon_t) = M_0M_0' = \Sigma$.

The equations of (1) can be written in summary notation as

$$(2)\qquad A(L)y(t) = M(L)\varepsilon(t),$$

where L is the lag operator, which has the effect that $Lx(t) = x(t-1)$, and where $A(z) = A_0 + A_1z + \cdots + A_pz^p$ and $M(z) = M_0 + M_1z + \cdots + M_qz^q$ are matrix-valued polynomials assumed to be of full rank. A multivariate process of this nature is commonly described as a VARMA process—the initial letter denoting "vector".

Example. The multivariate first-order autoregressive VAR(1) process satisfies the equation

$$(3)\qquad y(t) = \Phi y(t-1) + \varepsilon(t).$$

On the assumption that $\lim(\tau \to \infty)\Phi^\tau = 0$, the equation may be expanded, by a process of back-substitution, which continues indefinitely, so as become an infinite-order moving average:

$$y(t) = \big\{\varepsilon(t) + \Phi\varepsilon(t-1) + \Phi^2\varepsilon(t-2) + \cdots\big\}. \tag{4}$$

This expansion may also be effected in terms of the algebra of the lag operator via the expression

$$(I - \Phi L)^{-1} = \big\{I + \Phi L + \Phi^2 L^2 + \cdots\big\}. \tag{5}$$

For the convergence of the sequence $\{\Phi, \Phi^2, \ldots\}$, it is necessary and sufficient that all of the eigenvalues or latent roots of Φ should be less than unity in modulus.

The conditions of stationarity and invertibility, which apply to a VARMA(p, q) model $A(L)y(t) = M(L)\varepsilon(t)$, are evident generalisations of those that apply to scalar processes. The VARMA process is stationary if and only if $\det A(z) \neq 0$ for all z such that $|z| < 1$. If this condition is fulfilled, then there exists a representation of the process in the form of

$$\begin{aligned} y(t) &= \Psi(L)\varepsilon(t) \\ &= \big\{\Psi_0\varepsilon(t) + \Psi_1\varepsilon(t-1) + \Psi_2\varepsilon(t-2) + \cdots\big\}, \end{aligned} \tag{6}$$

wherein the matrices Ψ_j are determined by the equation

$$A(z)\Psi(z) = M(z), \tag{7}$$

and where the condition $A_0 = M_0 = I$ implies that $\Psi_0 = I$. The process is invertible, on the other hand, if and only if $\det M(z) \neq 0$ for all z such that $|z| < 1$. In that case, the process can be represented by an equation in the form of

$$\begin{aligned} \varepsilon(t) &= \Pi(L)y(t) \\ &= \big\{\Pi_0 y(t) + \Pi_1 y(t-1) + \Pi_2 y(t-2) + \cdots\big\}, \end{aligned} \tag{8}$$

wherein the matrices Π_j are determined by the equation

$$M(z)\Pi(z) = A(z), \tag{9}$$

and where the condition $A_0 = M_0 = I$ implies that $\Pi_0 = I$.

Canonical Forms

There is a variety of ways in which a VARMA equation can be reduced to a state-space model incorporating a transition equation that corresponds to a first-order Markov process. One of the more common formulations is the so-called controllable canonical state-space representation.

Consider writing equation (2) as

$$\begin{aligned} y(t) &= M(L)\left\{A^{-1}(L)\varepsilon(t)\right\} \\ &= M(L)\xi(t), \end{aligned} \tag{10}$$

where $\xi(t) = A^{-1}(L)\varepsilon(t)$. This suggests that, in generating the values of $y(t)$, we may adopt a two-stage procedure, which, assuming that $A_0 = I$, begins by calculating the values of $\xi(t)$ via the equation

$$\xi(t) = \varepsilon(t) - \left\{A_1\xi(t-1) + \cdots + A_r\xi(t-r)\right\} \tag{11}$$

and then proceeds to find those of $y(t)$ via the equation

$$y(t) = M_0\xi(t) + M_1\xi(t-1) + \cdots + M_{r-1}\xi(t-r+1). \tag{12}$$

Here, $r = \max(p, q)$ and, if $p \neq q$, then either $A_i = 0$ for $i = p+1, \ldots, q$ or $M_i = 0$ for $i = q+1, \ldots, p$.

In order to implement the recursion under (11), we define a set of r state variables as follows:

$$\begin{aligned} \xi_1(t) &= \xi(t), \\ \xi_2(t) &= \xi_1(t-1) = \xi(t-1), \\ &\vdots \\ \xi_r(t) &= \xi_{r-1}(t-1) = \xi(t-r+1). \end{aligned} \tag{13}$$

Rewriting equation (11) in terms of the variables defined on the LHS gives

$$\xi_1(t) = \varepsilon(t) - \{A_1\xi_1(t-1) + \cdots + A_r\xi_r(t-1)\}. \tag{14}$$

Therefore, by defining a state vector $\xi(t) = [\xi_1(t), \xi_2(t), \ldots, \xi_r(t)]'$ and by combining (13) and (14), we can construct a system in the form of

$$\begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \vdots \\ \xi_r(t) \end{bmatrix} = \begin{bmatrix} -A_1 & \ldots & -A_{r-1} & -A_r \\ I & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & I & 0 \end{bmatrix} \begin{bmatrix} \xi_1(t-1) \\ \xi_2(t-1) \\ \vdots \\ \xi_r(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \varepsilon(t). \tag{15}$$

The sparse matrix on the RHS of this equation is an example of a so-called companion matrix. The accompanying measurement equation, which corresponds to equation (12), is given by

$$y(t) = M_0\xi_1(t) + \cdots + M_{r-1}\xi_r(t). \tag{16}$$

Even for a VARMA system of a few variables and of low AR and MA orders, a state-space system of this sort is liable to involve matrices of very large dimensions. In developing practical computer algorithms for dealing with

such systems, one is bound to pay attention to the sparseness of the companion matrix.

Numerous mathematical advantages arise from treating a pth-order autogressive system of dimension n as a first-order system of dimension $p \times n$. Thus, for example, the condition for the stability of the system can be expressed in terms of the companion matrix of equation (15), which may be denoted by Φ. The condition for stability is that the eigenvalues of Φ must lie inside the unit circle, which is the condition that $\det(I - \Phi z) \neq 0$ for all $|z| \leq 1$. It is easy to see that

$$\det(I - \Phi z) = \det(I + A_1 z + \cdots I + A_p z^p), \tag{17}$$

which indicates that the condition is precisely the one that has been stated already.

The Autocovariances of an ARMA process

The equation of an ARMA(p, q) process is

$$\sum_i^p A_i y_{t-i} = \sum_i^q M_i \varepsilon_{t-i}. \tag{18}$$

To find the autocovariances of the process, consider post-multiplying the equation by $y'_{t-\tau}$ and taking expectations. This gives

$$\sum_i A_i E(y_{t-i} y'_{t-\tau}) = \sum_i M_i E(\varepsilon_{t-i} y'_{t-\tau}). \tag{19}$$

To find $E(\varepsilon_{t-i} y'_{t-\tau})$, it is appropriate to employ the infinite-order moving-average representation of the ARMA process, which gives

$$y_{t-\tau} = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-\tau-j}, \tag{20}$$

where the coefficients Ψ_j are from the series expansion of the rational function $\Psi(z) = A^{-1}(z)M(z)$. Putting this in equation (19) gives

$$\sum_i A_i E(y_{t-i} y'_{t-\tau}) = \sum_i \sum_j M_i E(\varepsilon_{t-i} \varepsilon'_{t-\tau-j}) \Psi'_j. \tag{21}$$

Here, there are

$$E(y_{t-i} y'_{t-\tau}) = \Gamma_{\tau-i} \quad \text{and} \quad E(\varepsilon_{t-i} \varepsilon'_{t-\tau-j}) = \begin{cases} \Sigma, & \text{if } j = i - \tau, \\ 0, & \text{if } j \neq i - \tau; \end{cases} \tag{22}$$

and it should be observed that

$$\Gamma_{-i} = E(y_{t+i} y'_t) = E(y_t y'_{t-i}) = \Gamma'_i. \tag{23}$$

Therefore, equation (21) can be written as

$$\sum_{i=0}^{p} A_i\Gamma_{\tau-i} = \sum_{i=\tau}^{q} M_i\Sigma\Psi'_{i-\tau}. \tag{24}$$

Given the normalisation $A_0 = I$, equation (24) provides a means of generating Γ_τ from the known values of the parameter sequences and from the previous values $\Gamma_{\tau-1}, \ldots, \Gamma_{\tau-p}$:

$$\Gamma_\tau = \begin{cases} -\sum_{i=1}^{p} A_i\Gamma_{\tau-i} + \sum_{i=\tau}^{q} M_i\Sigma\Psi'_{i-\tau}, & \text{if } \tau \le q; \\ -\sum_{i=1}^{p} A_i\Gamma_{\tau-i} & \text{if } \tau > q. \end{cases} \tag{25}$$

In the case of a pure autoregressive process of order p, there are

$$\Gamma_0 = -\sum_{i=1}^{p} A_i\Gamma_i + \Sigma \qquad \text{and} \tag{26}$$

$$\Gamma_\tau = -\sum_{i=1}^{p} A_i\Gamma_{\tau-i} \qquad \text{if } \tau > 0. \tag{27}$$

In the case of a pure moving-average process of order q, there is $A_i = 0$ for all $i > 0$. Also, there is $\Psi_j = M_j$ for $j = 0, \ldots, q$ together with $\Psi_j = 0$ for $j > q$. It follows that the moving-average autocovariances are

$$\Gamma_\tau = \sum_{i=\tau}^{q} M_i\Sigma M'_{i-\tau} \quad \text{with} \quad \tau = 0, \ldots, q. \tag{28}$$

These values are straightforward to calculate.

In the case of the autoregressive model and of the mixed autoregressive–moving average model with autoregressive orders of p, there is a need to generate the autocovariances $\Gamma_0, \Gamma_1, \ldots, \Gamma_p$ in order to initiate a recursive process for generating subsequent autocovariances. The means of finding these initial values can be illustrated by an example.

Example. To show how the initial values are generated, we may consider equation (24), where attention can be focussed to the LHS. Consider the case of a second-order vector autoregressive model. Then, (26) and (27) yield the so-called Yule–Walker equations.

If the autocovariances Γ_0, Γ_1, Γ_2 are known, then, given that $A_0 = I$, these equations can be solved for the autoregressive parameters A_1, A_2 and for the dispersion parameters $D(\varepsilon) = \Sigma$ of the disturbances. Setting $p = 2$ and $\tau = 0, 1, 2$ in equations (26) and (27) gives

$$\begin{bmatrix} \Gamma_0 & \Gamma'_1 & \Gamma'_2 \\ \Gamma_1 & \Gamma_0 & \Gamma'_1 \\ \Gamma_2 & \Gamma_1 & \Gamma_0 \end{bmatrix} \begin{bmatrix} A'_0 \\ A'_1 \\ A'_2 \end{bmatrix} = \begin{bmatrix} \Sigma \\ 0 \\ 0 \end{bmatrix}. \tag{29}$$

Then, assuming that $A_0 = I$, the equations

$$\begin{bmatrix} A_1' \\ A_2' \end{bmatrix} = -\begin{bmatrix} \Gamma_0 & \Gamma_1' \\ \Gamma_1 & \Gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \Gamma_0 + A_1\Gamma_1 + A_2\Gamma_2 \tag{30}$$

provide the parameters if the autocovariances are known. The matrix Σ of the contemporaneous autocovariances of the disturbance vector are provided by the equation on the RHS.

Conversely, if the parameters are know, then the equations can be solved for the autocovariances. However, to begin the recursion of (26), which generates the subsequent autocovariances, it is necessary to find the values of Γ_0 and Γ_1. The solution can be found by exploiting the representation of a VARMA system that is provided by equation (15). In the case of a VAR system of order 2, this becomes

$$\begin{bmatrix} y(t) \\ y(t-1) \end{bmatrix} = \begin{bmatrix} -A_1 & -A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} y(t-1) \\ y(t-2) \end{bmatrix} + \begin{bmatrix} \varepsilon(t) \\ 0 \end{bmatrix}, \tag{31}$$

from which the following equation in the autocovariances can be derived:

$$\begin{bmatrix} \Gamma_0 & \Gamma_1 \\ \Gamma_1' & \Gamma_0 \end{bmatrix} = \begin{bmatrix} -A_1 & -A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_0 & \Gamma_1 \\ \Gamma_1' & \Gamma_0 \end{bmatrix} \begin{bmatrix} -A_1' & 0 \\ -A_2' & I \end{bmatrix} + \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}. \tag{32}$$

A summary notation for this is

$$\Gamma = \Phi\Gamma\Phi' + \Omega. \tag{33}$$

Using the rule of vectorisation, whereby $(AXB')^c = (B \otimes A)X^c$, this becomes $\Gamma^c = (\Phi \otimes \Phi)\Gamma^c + \Omega^c$, whence the vectorised matrix of the autocovariances is

$$\Gamma^c = [I - (\Phi \otimes \Phi)]^{-1}\Omega^c. \tag{34}$$

The subsequent autocovariance matrices are generated via the recursion of (27).

When the model contains a moving-average component, the column containing the matrix Σ is replaced by something more complicated that is provided by the RHS of equation (24).

The Final Form and Transfer Function Form

A VARMA model is mutable in the sense that it can be represented in many different ways. It is particularly interesting to recognise that it can be reduced in a straightforward way to a set of n interrelated ARMA models. It may be acceptable to ignore the interrelations and to concentrate on building models for the individual series. However, ignoring the wider context in which these series arise may result in a loss of statistical efficiency in the course of generating estimates of their parameters.

The assumption that $A(z)$ is of full rank allows us to write equation (2) as

$$\begin{aligned} y(t) &= A^{-1}(L)M(L)\varepsilon(t) \\ &= \frac{1}{|A(L)|}A^*(L)M(L)\varepsilon(t), \end{aligned} \tag{35}$$

where $|A(L)|$ is the scalar-valued determinant of $A(L)$ and $A^*(L)$ is the adjoint matrix. The process can be written equally as

$$|A(L)|y(t) = A^*(L)M(L)\varepsilon(t). \tag{36}$$

Here is a system of n ARMA processes that share a common autoregressive operator $\alpha(L) = |A(L)|$. The moving-average component of the ith equation is the ith row of $A^*(L)M(L)\varepsilon(t)$. This corresponds to a sum of n moving-average processes—one for each element of the vector $\varepsilon(t)$—A sum of moving-average processes is itself a moving-average process with an order that is no greater than the maximal order of its constituent processes. It follows that the ith equation of the system can be written as

$$\alpha(L)y_i(t) = \mu_i(L)\eta_i(t). \tag{37}$$

The interdependence of the n univariate ARMA equations is manifested in the facts that (a) they share the same autoregressive operator $\alpha(L)$ and that (b) their disturbance processes $\eta_i(t); i = 1, \ldots, n$ are mutually correlated.

It is possible that $\alpha(L)$ and $\mu_i(L)$ will have certain factors in common. Such common factors must be cancelled from these operators. The effect will be that the operator $\alpha(L)$ will no longer be found in its entirely in each of the n equations. Indeed, it is possible that, via such cancellations, the resulting autoregressive operators $\alpha_i(L)$ of the n equations will become completely distinct with no factors in common.

When specialised in a certain way, the VARMA model can give rise to a dynamic version of the classical simultaneous-equation model of econometrics. This specialisation requires a set of restrictions that serve to classify some of the variables in $y(t)$ as exogenous with respect to the others which, in turn, are classified as endogenous variables. In that case, the exogenous variables may be regarded as the products of processes that are independent of the endogenous processes.

To represent this situation, let us take the equations of (2), where it may be assumed that $A_0 \neq I$, albeit that the diagonal elements of A_0 are units. Let the equations be partitioned as

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} M_{11}(L) & M_{12}(L) \\ M_{21}(L) & M_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}. \tag{38}$$

Here, $y(t)$ and $\varepsilon(t)$ are denoting particular subsets of the systems variables, whereas previously they have denoted the complete sets. If $\varepsilon(t)$ and $\eta(t)$ are mutually uncorrelated vectors of white-noise disturbances and if the restrictions are imposed that

$$A_{21}(L) = 0, \qquad M_{21}(L) = 0 \qquad \text{and} \qquad M_{12}(L) = 0, \tag{39}$$

then the system will be decomposed into a set of structural equations

$$A_{11}(L)y(t) + A_{12}(L)x(t) = M_{11}\varepsilon(t), \tag{40}$$

and a set of equations that describe the processes generating the exogenous variables,

$$A_{22}(L)x(t) = M_{22}\eta(t). \tag{41}$$

The leading matrix of $A_{11}(z)$, which is associated with z^0, may be denoted by A_{110}. If this is nonsingular, then the equations of (41) may be premultiplied by A_{110}^{-1} to give the so-called reduced-form equations, which express the current values of the endogenous variables as functions of the lagged endogenous variables and of the exogenous variables:

$$\begin{aligned} y(t) = &- A_{110}^{-1}\sum_{j=1}^{r} A_{11j}L^j y(t) - A_{110}^{-1}\sum_{j=1}^{r} A_{12j}L^j x(t) \\ &+ A_{110}^{-1}\sum_{j=1}^{r} M_{11j}L^j \varepsilon(t). \end{aligned} \tag{42}$$

Each of the equations is in the form of a so-called ARMAX model.

The final form of the equation of (40) is given by

$$y(t) = -A_{11}^{-1}(L)A_{12}(L)x(t) + A_{11}^{-1}(L)M_{11}(L)\varepsilon(t). \tag{43}$$

Here, the current values of the endogenous variables are expressed as functions of only the exogenous variables and the disturbances. Each of the individual equations of the final form constitutes a so-called rational transfer-function model or RTM.

Granger Causality

Consider the system

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ 0 & A_{22}(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} M_{11}(L) & 0 \\ 0 & M_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}, \tag{44}$$

wherein $\varepsilon(t)$ and $\eta(t)$ are mutually uncorrelated vectors of white-noise processes. Here, it can be said that, whereas $x(t)$ is affecting $y(t)$, there is no effect running from $y(t)$ to $x(t)$. On referring to equation (38), it can be seen that the assertion depends, not only on the condition that $A_{21}(L) = 0$, but also on the condition that $M_{21}(L) = 0$ and $M_{12}(L) = 0$. In the absence of these two conditions, $y(t)$ and $x(t)$ would be sharing some of the same disturbances. Therefore, knowing the past and present values of the sequence $y(t)$ would help in predicting the values of $x(t)$.

This observation has led Granger to enunciate a condition of non-causality. In the sense of Granger, $y(t)$ does not cause $x(t)$ if $E\{x(t)|y(t-j); j \geq 0\}$, which is that the conditional expectation of $x(t)$ given present and past values of $y(t)$, is no different from the unconditional expectation $E\{x(t)\}$. If there are no influences affecting $y(t)$ and $x(t)$ other than those depicted in the model, then

this definition of causality accords with the commonly accepted notion of cause and effect.

A practical way of demonstrating the absence of an effect running from $y(t)$ to $x(t)$ would be to fit the model of equation (38) to the data and to demonstrate that none of the estimates of the coefficients that are deemed to be zero-valued are significantly different from zero. A simple alternative would be to regress the values of $y(t)$ on past current and future values of $x(t)$. If none of the resulting coefficients associated with the future are significantly different from zero, then it can be said that $y(t)$ does not cause $x(t)$.

The condition of non-causality can be stated in terms both of the autoregressive and of the moving average forms of the process in question. Since

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ 0 & A_{22}(L) \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1}(L) & -A_{11}^{-1}(L)A_{12}A_{22}^{-1}(L) \\ 0 & A_{22}^{-1}(L) \end{bmatrix}, \tag{45}$$

it follows that the moving-average form of the joint process is

$$\begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ 0 & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}. \tag{46}$$

It can be deduced, in a similar way, that the autoregressive form of the process is

$$\begin{bmatrix} \Pi_{11}(L) & \Pi_{12}(L) \\ 0 & \Pi_{22}(L) \end{bmatrix} \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}. \tag{47}$$

The Impulse Response Function

Within the context of the moving-average representation of the system, which is equation (6), questions can be asked concerning the effects that impulses on the input side have upon the output variables.

Given that $\varepsilon_j(t)$ is regarded as a disturbance that is specific to the jth equation, an innovation in $\varepsilon_j(t)$ is tantamount to an innovation in $y_j(t)$. Therefore, by considering an isolated impulse in $\varepsilon_j(t)$ and by holding all other disturbances at zero for all times, one can discern the peculiar effect upon the system of a change in $y_j(t)$.

If e_j denotes the vector with a unit in the jth position and with zeros elsewhere, and if

$$\delta(t) = \begin{cases} 1, & \text{when } t = 0, \\ 0, & \text{when } t \neq 0, \end{cases} \tag{48}$$

denotes the impulse sequence, then the response of the system to an impulse in the jth equation may be represented by

$$y(t) = \Psi(L)\{\delta(t)e_j\} \tag{49}$$

The system will be inactive until time $t = 0$. Thereafter, it will generate the following sequence of vectors: $\{\Psi_0 e_j, \Psi_1 e_j, \Psi_2 e_j, \ldots\}$.

The effect of the impulse will reverberate throughout the system, and it will return to $y_j(t)$ via various circuitous routes. To trace these routes, one would have to take account of the structure of the autoregressive polynomial $A(z)$ of equation (1). The net effect of the impulse to the jth equation upon the output of the ith equation is given by the sequence $\{e_i'\Psi_0 e_j, e_i'\Psi_1 e_j, e_i'\Psi_2 e_j, \ldots\}$

The Cholesky Factorisation

It must be conceded that, in practice, an isolated impulse to a single equation is an unlikely event. In fact, it has been assumed that the innovations or disturbances of the n equations of the system have a joint dispersion matrix $D\{\varepsilon(t)\} = \Sigma$, which is, in general, a non-diagonal symmetric positive-definite matrix. This matrix is amenable to a Cholesky factorisation of the form $\Sigma = T\Delta T'$, where T is a lower triangular matrix with units on the diagonal and where Δ is a diagonal matrix of positive elements. Then, $\eta_t = T^{-1}\varepsilon_t$ has a diagonal dispersion matrix $D\{\eta(t)\} = T^{-1}\Sigma T'^{-1} = \Delta$.

The effect of this transformation is to create a vector $\eta(t)$ of mutually independent disturbances, which is also described as a matter of orthogonalising the disturbances. Applying the transformation the equation $A(L)y(t) = M(L)\varepsilon(t)$ of (2) gives

$$A(L)y(t) = \{M(L)T\}\eta(t) \quad \text{with} \quad \eta(t) = T^{-1}\varepsilon(t). \tag{50}$$

The infinite-order moving-average form of this equation is

$$\begin{aligned} y(t) &= A^{-1}(L)\{M(L)T\}\eta(t) \\ &= \{\Psi(L)T\}\eta(t), \end{aligned} \tag{51}$$

whereas the infinite-order autoregressive form is

$$\begin{aligned} \eta(t) &= \{T^{-1}M^{-1}(L)\}A(L)y(t) \\ &= T^{-1}\Pi(L)y(t). \end{aligned} \tag{52}$$

The effect of the transformation T upon the structure of the system is best understood in the context of an purely autoregressive model of order p. On taking account of the normalisation $A_0 = I$ and on setting $-A_j = \Phi_j; j = 1, \ldots, p$, the transformed autoregressive equation can be represented by

$$T^{-1}y(t) = T^{-1}\Phi_1 y(t) + \cdots + T^{-1}\Phi_p y(t-p) + \eta(t). \tag{53}$$

Here, T^{-1} is a lower-triangular matrix with units on the diagonal, which are features it shares with T. Therefore, the elements of $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]'$ appear to be determined in a recursive manner.

At any given time, the leading element y_1 is determined without reference to the remaining elements. Then, y_2 is determined with reference to y_1. Next, y_3 is determined with reference to y_1 and y_2, and so on. Therefore, there is what

may be described, in the terminology of Wold, as a causal ordering amongst the variables. See, for example, Strotz and Wold (1960) or Simon (1953).

The infinite-order vector moving average corresponding to a finite-order autoregression is

$$\begin{aligned} y(t) &= \{\Psi(L)T\}\{T^{-1}\varepsilon(t)\} \\ &= \{\Psi_0 T\}\{T^{-1}\varepsilon(t)\} + \{\Psi_1 T\}\{T^{-1}\varepsilon(t-1)\} + \cdots \\ &= \Theta_0\eta(t) + \Theta_1\eta(t-1) + \Theta_2\eta(t-2) + \cdots, \end{aligned} \tag{54}$$

where $T^{-1}\varepsilon(t) = \eta(t)$ has $D\{\eta(t)\} = I$ and where $\Theta_j = \Psi_j T$. Also, if $A_0 = M_0 = I$, there will be $\Psi_0 = I$ and $\Theta_0\eta(t) = \varepsilon(t)$.

If the indexing of the variables has been arbitrary, then the causal ordering will have no substrative interpretation. However, the variables might be indexed with the intention of revealing what is thought to be a genuine causal ordering.

Variance Decompositions

One of the virtues of the Cholesky decomposition is that it assists in the decomposition of the variances of the elements of $y(t)$. In terms of the MA(∞) representation of equation (6), the dispersion matrix of $y(t)$ can be decomposed as follows:

$$D\{y(t)\} = \Psi_0\Sigma\Psi_0' + \Psi_1\Sigma\Psi_1' + \Psi_2\Sigma\Psi_2' + \cdots. \tag{55}$$

This result is in consequence of the serial independence in time of the successive vectors of disturbances. The effect of applying the Cholesky transformation T is to give

$$D\{y(t)\} = \Theta_0\Delta\Theta_0' + \Theta_1\Delta\Theta_1' + \Theta_2\Delta\Theta_2' + \cdots, \tag{56}$$

where $\Theta_j = \Psi_j T$ and where $\Delta = D\{\eta(t)\} = T^{-1}\Sigma T'^{-1}$ is a diagonal matrix. Then, in the jth equation, there is an innovation η_j that is peculiar that equation and there are the remaining innovations $\eta_k; k \neq j$ which are orthogonal to or uncorrelated with η_j and which can be attributed to the other equations. However, except in the case of the first equation, which has $\eta_1(t) = \varepsilon_1(t)$, these innovations will represent combinations of the elements of the disturbance vector $\varepsilon(t)$. For that reason, the innovations might not be amenable to an easy interpretation.

In consequence of this difficulty, one is bound to think of applying the Cholesky orthgonalisation procedure successively, with each variable in turn appointed to be the leading variable. Then, in every case, there will be a disturbance innovation that is peculiar to the appointed variable, which we may describe as an *auto*-innovation, and there will be a set of innovations, originating in the other equations, which are uncorrelated with the *auto*-innovation, which we might describe as the *allo*-innovations. The matter of *auto*-innovations and *allo*-innovations will be pursued in a subsequent section.

Structural Models

A Cholesky factorisation is by no means the only way in which the dispersion matrix $D\{\varepsilon(t)\} = \Sigma$ can be diagonalised. The choice of a particular form of the diagonalising matrix is often associated with the specification of a structural model, which contains equations that purport to have a behavioural interpretation.

To reveal the range of possibilities for diagonalising a positive definite matrix, we may consider a Cholesky factorisation that reduces the matrix Σ to the identity matrix. If L is a lower-triangular Cholesky factor such that

$$L^{-1}\Sigma L'^{-1} = I \quad \text{and} \quad \Sigma = LL', \tag{57}$$

and if Q is an orthonormal matrix such that $Q'Q = QQ' = I$, then there is also

$$QL^{-1}\Sigma L'^{-1}Q' = T^{-1}\Sigma T'^{-1} = I, \quad \text{and} \quad TT' = \Sigma, \tag{58}$$

where $T = LQ'$ is a diagonalising matrix that is not triangular.

Example. A particular form of the diagonalising matrix T has been used by Blanchard and Quah (1989) to give expression to their theory concerning the impact of demand disturbances and supply disturbances upon the level Gross Domestic Product (GDP). Blanchard and Quah have described the economy in terms of a bivariate autoregressive model comprising thevector $[y_1(t), y_2(t)] = [\nabla Y(t), U(t)]$, wherein Y and U are the logarithms of GDP, i.e. aggregate output, and unemployment respectively, so that ∇Y is rate of growth of GDP.

Two mutually uncorrelated structural disturbances are identified that have different effects on the economy. These are the elements of the disturbance vector $\eta(t) = [\eta_s(t), \eta_d(t)]'$. It is proposed that neither of these disturbances has any long-run effect on employment. However, it is assumed that the supply shock $\eta_s(t)$ does have a long-run effect on output, whereas the demand shock $\eta_d(t)$, which impacts directly on unemployment, has no long-run effect on output.

The unrestricted moving-average representation of the system is the equation

$$y(t) = \varepsilon(t) + \Psi_1\varepsilon(t-1) + \Psi_2\varepsilon(t-2) + \cdots, \tag{59}$$

where $\Psi_0 = I$ and $D\{\varepsilon(t)\} = \Sigma$. This can be obtained by estimating the vector autoregressive representation of $y(t)$ and by inverting it. From the autoregressive representation, an estimate of $\Sigma = LL'$ can be obtained

The conditions affecting the disturbances provide restrictions that serve to identify the structure of the model. The structural form gives rise to the following infinite-order vector moving average:

$$y(t) = \Theta_0\eta(t) + \Theta_1\eta(t-1) + \Theta_2\eta(t-2) + \cdots, \tag{60}$$

where the condition $D\{\eta(t)\} = I$ reflects the mutually uncorrelated nature of the structural disturbances.

The transformation of equation (59) into equation (60) is by virtue of the diagonalising matrix T. By comparing equations (59) and (60), it can be seen that $\varepsilon(t) = \Theta_0\eta(t)$. But $\eta(t) = T^{-1}\varepsilon(t)$, so $\Theta_0 = T$. More generally $\{\Theta_j = \Psi_j T; j = 0, 1, 2, \ldots\}$; and these are the matrices that are entailed in the calculation of the impulse responses of the structural disturbances.

The summation of the impulse response of the structural system, i.e. its steady-state gain, is given by

$$\Theta(1) = \Psi(1)T = A^{-1}(1)T, \tag{61}$$

where

$$\begin{aligned} A^{-1}(1) &= (I - \Phi_1 - \cdots - \Phi_p)^{-1} \\ &= \{1 + \Psi_1 + \Psi_2 + \cdots\} = \Psi(1). \end{aligned} \tag{62}$$

The condition that the steady-state gain of the mapping from the demand shock η_d to the change in output ∇Y is zero is the condition that $e_s'\Theta(1)e_d = 0$, where $e_s' = [1, 0]$ and $e_d = [0, 1]'$. Using $\Theta(1) = A^{-1}(1)T$ and $T = LQ'$, the condition becomes

$$e_s'A^{-1}(1)Te_d = e_s'A^{-1}(1)LQ'e_d = 0. \tag{63}$$

Given that $A^{-1}(1)$ and L can be determined from the estimates estimates of the unrestricted model, it sufficient to find an orthonormal matrix Q that will satisfy the condition by transforming $A^{-1}(1)L$ or its inverse $L^{-1}(A(1)$ to a matrix with a zero in the top right corner. If

$$L^{-1}A(1) = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \tag{64}$$

then

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \frac{1}{\rho}\begin{bmatrix} \gamma_{22} & -\gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{bmatrix}, \tag{65}$$

where $\rho^2 = \gamma_{22}^2 + \gamma_{12}^2$. In effect, Q is the matrix that rotates the vector $[\gamma_{12}, \gamma_{22}]'$ through an angle of $\theta = \tan^{-1}(\gamma_{12}/\gamma_{22})$ degrees so as to align it with the vertical axis, i.e. to set its leading element to zero.

With Q and $T = LQ' = \Theta_0$ available, and given that the sequence $\{\Psi_j; j = 0, 1, 2, \ldots\}$ has been derived via the inversion of the estimated autoregressive model, it is straightforward to calculate the sequence $\{\Theta_j = \Psi_j T; j = 0, 1, 2, \ldots\}$ from which the impulse response functions of the structural model are derived.

In deriving the impulse respone, it is appropriate to replace the diagonal elements of Θ_0 by units, which entails premultiplying the matrices of the sequence $\{\Theta_j; j = 0, 1, 2, \ldots\}$ by the diagonal matrix $\text{diag}\{\theta_{11}^{-1}, \theta_{22}^{-1}\}$. The effect is to transfer the normalisation from the variances of the innovations to the leading coefficient matrix Θ_0.

Auto-innovations and Allo-innovations

Consider a vector autoregressive equation of the form $A(L)y(t)\varepsilon(t)$, which is inverted to give a moving-average equation $y(t) = \psi(L)\varepsilon(t)$. This can be partitioned as

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}, \tag{66}$$

to allow attention to be focussed in turn on the subvectors $y_1(t)$ and $y_2(t)$.

It may be assumed that there are no instantaneous cross-connections between the current elements of $y_1(t)$ and $y_2(t)$, which implies that $A_0 = \Psi_0 = I$. However, there may be contemporaneous correlation between the elements of $\varepsilon_1(t)$ and those of $\varepsilon_2(t)$, such that their joint dispersion matrix is

$$D\left(\begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \Sigma. \tag{67}$$

The first purpose is to determine the effects that the innovations in the sequence $\varepsilon_1(t)$ have upon the variables in $y_1(t)$ and to separate them from the effects of the disturbances in $\varepsilon_2(t)$, which impinge directly on $y_2(t)$.

To achieve this separation, a transformation T^{-1} may be applied to the vector $\varepsilon(t)$ to remove the correlation between the two sets of innovations. The compensating transformation T must be applied to the operator $\psi(L)$. Thus, in summary notation, the transformed system in denoted by

$$y(t) = \{\Psi(L)T\}\{T^{-1}\varepsilon(t)\} = \Theta(L)\eta(t). \tag{68}$$

The partitioned form of the transformation that is applied to the disturbance vector is

$$T^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix}. \tag{69}$$

This block-Cholesky transformation creates a block-diagonal dispersion matrix, of which the partitioned form is

$$\begin{aligned} T^{-1}\Sigma T'^{-1} &= \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix}. \end{aligned} \tag{70}$$

In the context of a bivariate normal distribution of ε_{1t} and ε_{2t}, the transformation will be effective in factorising of the joint distribution $N(\varepsilon_{1t}, \varepsilon_{2t}) = N(\varepsilon_{1t})N(\varepsilon_{2t}|\varepsilon_{1t})$ into a marginal distribution $N(\varepsilon_{1t})$ and a conditional distribution $N(\varepsilon_{2t}|\varepsilon_{1t}) = N(\eta_{12t})$. By definition, the marginal and the conditional distributions are statistically independent, in consequence of the orthogonality of their elements.

The partitioned form of the equation $T^{-1}\varepsilon(t) = \eta(t)$ is

$$\begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) - \Sigma_{21}\Sigma_{11}^{-1}\varepsilon_1(t) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(t) \\ \eta_{21}(t) \end{bmatrix}. \tag{71}$$

The disturbances within $\varepsilon_1(t) = \eta_{11}(t)$ are the *auto*-innovations of the first block and the innovations within $\eta_{21}(t) = \varepsilon_2(t) - \Sigma_{21}\Sigma_{11}^{-1}\varepsilon_1(t)$ are the *allo*-innovations. The factorisation has ensured the mutual orthogonality of these two sets of innovations. A corresponding factorisation is available that will generate the *auto*-innovations for the second block together with the orthogonal *allo*-innovations,

The partitioned form of the equation $\Psi(L)T = \Theta(L)$ is

$$\begin{aligned} &\begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} I & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} \Psi_{11}(L) + \Psi_{12}(L)\Sigma_{21}\Sigma_{11}^{-1} & \Psi_{12}(L) \\ \Psi_{21}(L) + \Psi_{22}(L)\Sigma_{21}\Sigma_{11}^{-1} & \Psi_{22}(L) \end{bmatrix} = \begin{bmatrix} \Theta_{11}(L) & \Theta_{12}(L) \\ \Theta_{21}(L) & \Theta_{22}(L) \end{bmatrix}. \end{aligned} \tag{72}$$

By gathering these results, it is found that

$$\begin{aligned} y_1(t) &= \Theta_{11}(L)\varepsilon_1(t) + \Theta_{12}(L)\eta_{21}(t) \\ &= \{\Psi_{11}(L) + \Psi_{12}(L)\Sigma_{21}\Sigma_{11}^{-1}\}\varepsilon_1(t) + \Psi_{12}(L)\eta_{21}(t), \end{aligned} \tag{73}$$

and, on substituting for $\eta_{21}(t) = \varepsilon_2(t) - \Sigma_{21}\Sigma_{11}^{-1}\varepsilon_1(t)$, it can be seen that this is just a reparametrisation of the equation $y_1(t) = \Psi_{11}\varepsilon_1(t) + \Psi_{12}\varepsilon_2(t)$. On applying the corresponding analysis to the second block of the partitioned system of (66), it will be found that

$$y_2(t) = \{\Psi_{22}(L) + \Psi_{21}(L)\Sigma_{12}\Sigma_{22}^{-1}\}\varepsilon_2(t) + \Psi_{21}(L)\eta_{12}(t), \tag{74}$$

where $\eta_{12}(t) = \varepsilon_1(t) - \Sigma_{12}\Sigma_{22}^{-1}\varepsilon_2(t)$.

Since the additive components in the LHS of equations (73) and (74) are mutually orthogonal, the equations are amenable to a simple analysis of variance. By these means, it is possible to measure the relative influences of the *auto*-innovations and the *allo*-innovations in both equations.

Example. The foregoing results can used in an attempt to determine whether it is supply side or the demand side of the economy that drives the business cycle. The relevant indices are the growth rates of gross domestic production and aggregate consumption, which are defined as $y_p(t) = \ln\{P(t) - P(t-4)\}$ and $y_c(t) = \ln\{C(t) - C(t-4)\}$, where $P(t)$ and $C(t)$ are sequences of quarterly observations on production and consumption, respectively.

A bivariate autoregressive VAR(2) model comprising lagged values of the growth rates can be represented by the following equations:

$$\alpha_{cc}(L)y_c(t) + \alpha_{cp}(L)y_p(t) = \varepsilon_c(t), \tag{75}$$

$$\alpha_{pc}(L)y_c(t) + \alpha_{pp}(L)y_p(t) = \varepsilon_p(t). \tag{76}$$

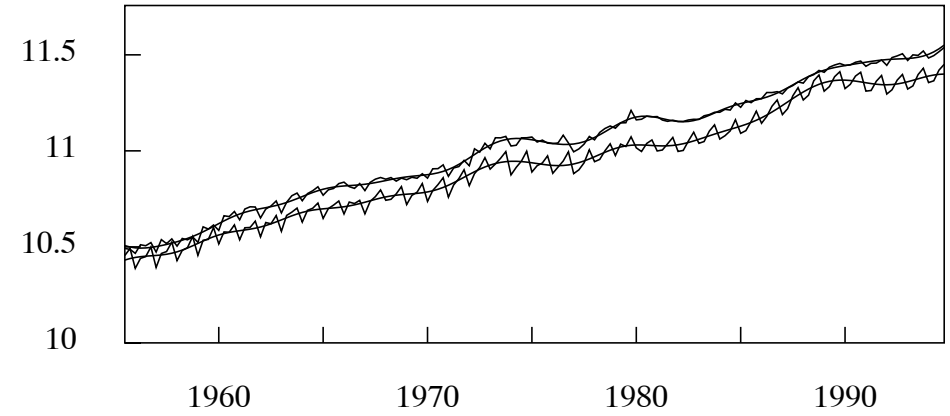

Figure 1. The quarterly series of the logarithms of income (upper) and consumption (lower) in the U.K., for the years 1955 to 1994, together with their interpolated trends.

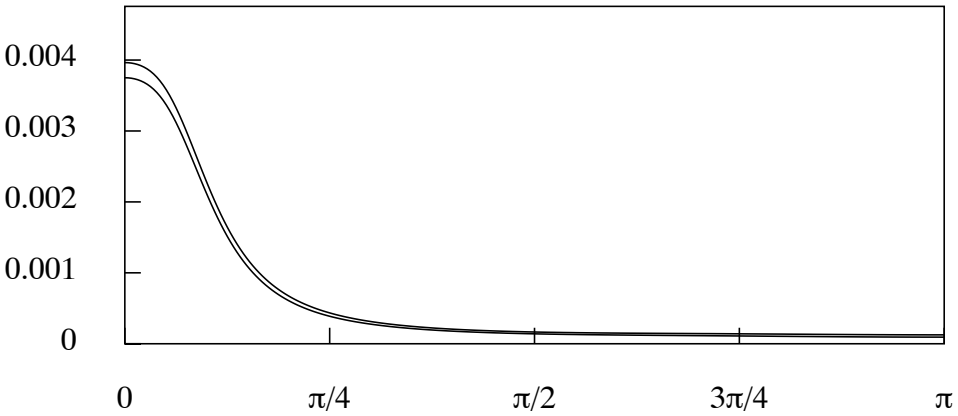

Figure 2. The spectrum of the consumption growth sequence $y_c(t)$ (the outer envelope) and that of its auto-innovation component $\{\pi_{cc}(L)/\delta(L)\}\varepsilon_c(t)$ (the inner envelope).

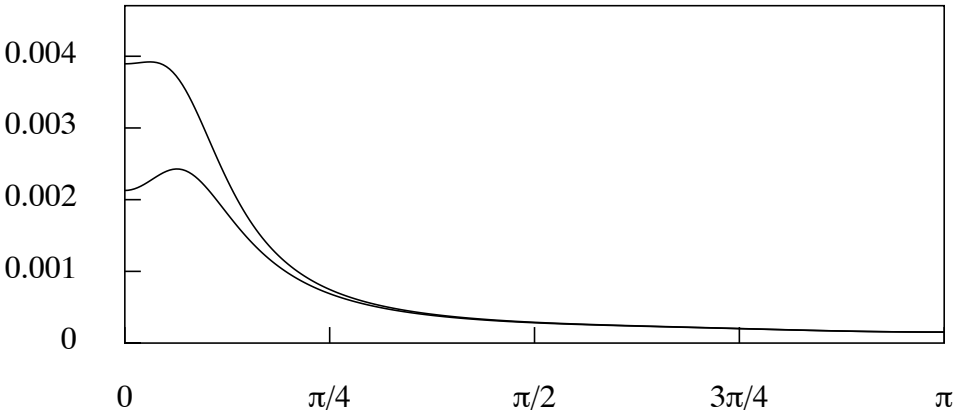

Figure 3. The spectrum of the income growth sequence $y_p(t)$ (the outer envelope) and that of its auto-innovation component $\{\pi_{pp}(L)/\delta(L)\}\varepsilon_p(t)$ (the inner envelope).

The notion that the sequence $y_c(t)$ is driving the sequence $y_p(t)$ would be substantiated if the influence of the innovations sequence $\varepsilon_c(t)$ upon $y_c(t)$ were found to be stronger than the influence of $\varepsilon_p(t)$ upon the corresponding sequence $y_p(t)$.

The matter can be investigated via the moving-average forms of the equations, which express $y_p(t)$ and $y_c(t)$ as functions only of the innovations sequences $\varepsilon_c(t)$ and $\varepsilon_p(t)$. The moving-average equations, which are obtained by inverting equations (75) and (76) jointly, are

$$y_c(t) = \frac{\alpha_{pp}(L)}{\delta(L)}\varepsilon_c(t) - \frac{\alpha_{cp}(L)}{\delta(L)}\varepsilon_p(t), \tag{77}$$

$$y_p(t) = -\frac{\alpha_{pc}(L)}{\delta(L)}\varepsilon_c(t) + \frac{\alpha_{cc}(L)}{\delta(L)}\varepsilon_p(t), \tag{78}$$

where $\delta(L) = \alpha_{cc}(L)\alpha_{pp}(L) - \alpha_{cp}(L)\alpha_{pc}(L)$.

These two equations can be reparametrised so that each is expressed in terms of a pair of uncorrelated innovations. The reparametrised version of the consumption equation is

$$y_c(t) = \frac{\pi_{cc}(L)}{\delta(L)}\varepsilon_c(t) - \frac{\alpha_{cp}(L)}{\delta(L)}\eta_p(t), \tag{79}$$

where

$$\begin{aligned} \pi_{cc}(L) &= \alpha_{pp}(L) + \frac{\sigma_{pc}}{\sigma_c^2}\alpha_{cp}(L), \\ \eta_p(t) &= \varepsilon_p(t) - \frac{\sigma_{pc}}{\sigma_c^2}\varepsilon_c(t). \end{aligned} \tag{80}$$

Here, $\sigma_c^2 = V\{\varepsilon_c(t)\}$ is the variance of the consumption innovations and $\sigma_{pc}^2 = C\{\varepsilon_p(t), \varepsilon_c(t)\}$ is the covariance of the income and consumption innovations. We may describe the sequence $\varepsilon_c(t)$ as the *auto*-innovations of $y_c(t)$ and $\eta_p(t)$ as the *allo*-innovations.

By a similar reparametrisation, the equation (78) in $y_p(t)$ becomes

$$y_p(t) = \frac{\pi_{pp}(L)}{\delta(L)}\varepsilon_p(t) - \frac{\alpha_{pc}(L)}{\delta(L)}\eta_c(t), \tag{81}$$

where

$$\begin{aligned} \pi_{pp}(L) &= \alpha_{cc}(L) + \frac{\sigma_{cp}}{\sigma_p^2}\alpha_{pc}(L), \\ \eta_c(t) &= \varepsilon_c(t) - \frac{\sigma_{cp}}{\sigma_p^2}\varepsilon_p(t), \end{aligned} \tag{82}$$

and where $\sigma_c^2 = V\{\varepsilon_c(t)\}$.

The relative influences of $\varepsilon_c(t)$ on $y_c(t)$, the growth rate of consumption, and of $\varepsilon_p(t)$ on $y_p(t)$, the growth rate of income, can now be assessed by an analysis of the corresponding spectral density functions.

The quarterly data, from which the autoregressive parameters are estimated after taking four-period differences, are from the United Kingdom over the period 1955 to 1994; and Figure 1 shows the logarithms of these data.

Figure 2 shows the spectrum of $y_c(t)$ together with that of its *auto* -innovation component $\{\pi_{cc}(L)/\delta(L)\}\varepsilon_c(t)$, which is the lower envelope. Figure 3 shows the spectrum of $y_p(t)$ together with that of its *auto*-innovation component $\{\pi_{pp}(L)/\delta(L)\}\varepsilon_p(t)$.

From a comparison of the figures, it is clear that the innovation sequence $\varepsilon_c(t)$ accounts for a much larger proportion of $y_c(t)$ than $\varepsilon_p(t)$ does of $y_p(t)$. Thus, the consumption growth series appears to be driven largely by its *auto* -innovations. These innovations also enter the income growth series to the extent that the latter is not accounted for by its *auto*-innovations. Figure 3 shows that this is a considerable extent.

The Rational Model and the ARMAX Model

An ARMAX model is represented by the equation

$$\alpha(L)y(t) = \beta(L)x(t) + \mu(L)\varepsilon(t), \tag{83}$$

where $y(t)$ and $x(t)$ are observable sequences and $\varepsilon(t)$ is a white-noise disturbance sequence, which is statistically independent of $x(t)$. For the model to be viable, the roots of the polynomial equations $\alpha(z) = 0$ and $\mu(z) = 0$ must lie outside the unit circle. This is to ensure that the coefficients of the series expansions of $\alpha^{-1}(z)$ and $\mu^{-1}(z)$ form convergent sequences. The rational form of equation (83) is

$$y(t) = \frac{\beta(L)}{\alpha(L)}x(t) + \frac{\mu(L)}{\alpha(L)}\varepsilon(t). \tag{84}$$

The Rational Transfer-Function Model or RTM is represented by

$$y(t) = \frac{\delta(L)}{\gamma(L)}x(t) + \frac{\theta(L)}{\phi(L)}\varepsilon(t), \tag{85}$$

or, alternatively, by

$$\gamma(L)\phi(L)y(t) = \phi(L)\delta(L)x(t) + \gamma(L)\theta(L)\varepsilon(t). \tag{86}$$

The leading coefficients of $\gamma(L), \phi(L)$ and $\theta(L)$ are set to unity. The roots of the equations $\gamma(z) = 0, \phi(z) = 0$ and $\theta(z) = 0$ must lie outside the unit circle.

Notice that equation (86) can be construed as a case of (83) in which the factors of $\alpha(L) = \gamma(L)\phi(L)$ are shared between $\beta(L) = \phi(L)\delta(L)$ and $\mu(L) = \gamma(L)\theta(L)$. Likewise, equation (84) can be construed as a case of equation (85) in which the two transfer functions have the same denominator $\alpha(L)$. There may be scope for cancellation between the numerators and denominators of these transfer functions; after which they might have nothing in common. However, if none of the factors of $\alpha(L)$ are to be found in either $\beta(L)$ or $\mu(L)$, then

$\alpha(L)$ is present in its entirety in the denominators of both the systematic and disturbance parts of the rational form of the model. In that case, the two parts of the model have dynamic properties which are essentially the same.

An advantage of the ARMAX model when $\mu(L) = 1$ is that it may be estimated simply by ordinary least-squares regression. The advantage disappears when $\mu(L) \neq 1$. Its disadvantage lies in the fact that the two transfer functions of the rational form of the model are constrained to have the same denominators.

Unless the relationship which is being modelled does have similar dynamic properties in its systematic and disturbance parts, then the restriction is liable to induce severe biases in the estimates. See, for example, Pollock and Pitta, (1996). This is a serious problem if the estimated relationship is to be used as part of a control mechanism. If the only concern is to forecast the values of $y(t)$, then there is less harm in such biases.

The advantages of a rational transfer-function model with distinct parameters in both transfer functions is an added flexibility in modelling complex dynamic relationships as well as a degree of robustness that allows the systematic part of the model to be estimated consistently even when the stochastic part is misspecified. A disadvantage of the rational model may be the complexity of the processes of identification and estimation. In certain circumstances, the pre-whitening technique may help in overcoming such difficulties.

Fitting the Rational Model by Pre-whitening

Consider, once more, the rational transfer-function model of (85). This can be written as

$$y(t) = \omega(L)x(t) + \eta(t) \tag{87}$$

where $\omega(z) = \{\omega_0 + \omega_1 z + \cdots\}$ stands for the expansion of the rational function $\delta(z)/\gamma(z)$ and where $\eta(t) = \{\theta(L)/\phi(L)\}\varepsilon(t)$ is a disturbance generated by an ARMA process.

If the input signal $x(t) = \xi(t)$ happens to be white noise, which it might be by some contrivance, then the estimation of the coefficients of $\omega(z)$ is straightforward. For, given that the signal $\xi(t)$ and the noise $\eta(t)$ are uncorrelated, it follows that

$$\omega_\tau = \frac{C(y_t, \xi_{t-\tau})}{V(\xi_\tau)}. \tag{88}$$

The principal of estimation known as the method of moments suggest that, in order to estimate ω_τ consistently, we need only replace the theoretical moments $C(y_t, \xi_{t-\tau})$ and and $V(\xi_t)$ within this expression by their empirical counterparts.

Imagine that, instead of being a white-noise sequence, $x(t)$ is a stochastic sequence that can be represented by an ARMA process, which is stationary and invertible:

$$\rho(L)x(t) = \mu(L)\xi(t). \tag{89}$$

Then, $x(t)$ can be reduced to the white-noise sequence $\xi(t)$ by the application of the filter $\pi(L) = \rho(L)/\mu(L)$. The application of the same filter to $y(t)$ and $\eta(t)$ will generate the sequences $q(t) = \pi(L)y(t)$ and $\zeta(t) = \pi(L)y(t)$ respectively. Hence, if the model of equation (86) can be transformed into

$$q(t) = \omega(L)\xi(t) + \zeta(t), \tag{90}$$

then the parameters of $\omega(L)$ will, once more, become accessible in the form of

$$\omega_\tau = \frac{C(q_t, \xi_{t-\tau})}{V(\xi_\tau)}. \tag{91}$$

Given the coefficients of $\omega(z)$, it is straightforward to recover the parameters of $\delta(z)$ and $\gamma(z)$, when the degrees of these polynomials are known.

In practice, the filter $\pi(L)$, which would serve to reduce $x(t)$ to white noise, requires to be identified and estimated via the processes of ARMA model building, which we have described already. The estimated filter may then be applied to the sequences of observations on $x(t)$ and $y(t)$ in order to construct the empirical counterparts of the white-noise sequence $\xi(t)$ and the filtered sequence $q(t)$. From the empirical moments of the latter, the estimates of the coefficients of $\omega(z)$ may be obtained.

These estimates of $\omega(z)$, which are obtained via the technique of pre-whitening, are liable to be statistically inefficient. The reasons for their inefficiency lie partly in the use of an estimated whitening filter—in place of a known one—and partly in the fact that no attention is paid, in forming the estimates, of the nature of the disturbance processes $\eta(t)$ and $\zeta(t)$. However, the "pre-whitening" estimates of $\omega(z)$, together with the corresponding estimates of $\delta(z)$ and $\gamma(z)$, may serve as useful starting values for an iterative procedure aimed at finding the efficient maximum-likelihood estimates.

References

Blanchard, O.J., and D. Quah, (1989), The Dynamic Effects of Aggregate Demand and Supply Disturbances, *The American Economic Review,* 79, 655–673.

Enders, W., (2004), *Applied Econometric Time Series,* 2nd Edition, John Wiley and Sons, Hoboken, NJ.

Pollock, D.S.G. and Evangelia Pitta, (1996), The Misspecification of Dynamic Regression Models, *Journal of Statistical Inference and Planning,* 49, 223–229.

Simon, H., (1953), Causal Ordering and Identifiability, pps. 49–74 in W,C, Hood and T.C. Koopmans (eds.), *Studies in Econometric Methods,* John Wiley and Sons, New York

Strotz, R., and Wold, H. (1960), Recursive versus Nonrecursive Systems: An Attempt at Synthesis, *Econometrica,* 28 417-427.