LECTURE 8

Multivariate ARMA Processes

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

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

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.

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. 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 $M_0 = I$.

It is assumed the disturbance vector $\varepsilon(t)$ has $E\{\varepsilon(t)\} = 0$ for its expected value. On the assumption that $M_0 = I$, the dispersion matrix is an unrestricted positive-definite matrix denoted by $D\{\varepsilon(t)\} = \Sigma$. However, the restriction may be imposed that $D\{\varepsilon(t)\} = I$, in which case $M_0M_0' = \Sigma$.

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

$$A(L)y(t) = M(L)\varepsilon(t), \tag{2}$$

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

$$y(t) = \Phi y(t-1) + \varepsilon(t). \tag{3}$$

On the assumption that $\lim(\tau \to \infty)\Phi^\tau = 0$, the equation may be expanded, by an 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

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

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 which 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

$$(6)\qquad y(t) = \big\{\Psi_0\varepsilon(t) + \Psi_1\varepsilon(t-1) + \cdots\big\},$$

wherein the matrices Ψ_j are determined by the equation

$$(7)\qquad A(z)\Psi(z) = M(z).$$

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

$$(8)\qquad \varepsilon(t) = \big\{y(t) + \Pi_1 y(t-1) + \Pi_2 y(t-2) + \cdots\big\},$$

wherein the matrices Π_j are determined by the equation

$$(9)\qquad M(z)\Pi(z) = A(z).$$

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 which 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

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

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 begins by calculating the values of $\xi(t)$ via the equation

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

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

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

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.

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 is quite 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 will 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

$$\begin{aligned} y(t) &= A^{-1}(L)M(L)\varepsilon(t) \\ &= \frac{1}{|A(L)|}A^{*}(L)M(L)\varepsilon(t), \end{aligned} \tag{17}$$

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

$$|A(L)|y(t) = A^{*}(L)M(L)\varepsilon(t). \tag{18}$$

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 which is no greater than the maximal order of its constituent processes. It follows that the ith equation of the system can be written as

$$\alpha(L)y_i(t) = \mu_i(L)\eta_i(t). \tag{19}$$

The interdependence of the n univariate ARMA equations is manifested by the fact (a) that they share the same autoregressive operator $\alpha(L)$ and by the fact (b) that 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.

By a certain specialisation, 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 partition the equations under (2) as

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} M_{11}(L) & M_{12}(L) \\ M_{21}(L) & M_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}. \tag{20}$$

If the restriction are imposed that

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

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

$$A_{11}(L)y_1(t) + A_{12}(L)x(t) = M_{11}\varepsilon_1(t), \tag{22}$$

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

$$A_{22}(L)x(t) = M_{22}\varepsilon_1(t). \tag{23}$$

Here, in order to emphasise the distinction between endogenous and exogenous equations, we have set $y_2(t) = x(t)$.

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 (22) 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:

$$\begin{aligned} y_1(t) = &- A_{110}^{-1}\sum_{j=1}^{r} A_{11j}L^j y(t) - A_{110}^{-1}\sum_{j=1}^{r} A_{12j}L^j x(t) \\ &+ A_{110}^{-1}\sum_{j=1}^{r} M_{11j}L^j \varepsilon(t). \end{aligned} \tag{24}$$

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

The final form of the equation of (22) is given by

$$y_1(t) = -A_{11}^{-1}(L)A_{12}(L)x(t) + A_{11}^{-1}(L)M_{11}(L)\varepsilon(t). \tag{25}$$

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.

The Rational Model and the ARMAX Model

An ARMAX model is represented by the equation

$$\alpha(L)y(t) = \beta(L)x(t) + \mu(L)\varepsilon(t), \tag{26}$$

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 (26) is

$$y(t) = \frac{\beta(L)}{\alpha(L)}x(t) + \frac{\mu(L)}{\alpha(L)}\varepsilon(t). \tag{27}$$

The Rational Transfer-Function Model or RTM is represented by

$$y(t) = \frac{\delta(L)}{\gamma(L)}x(t) + \frac{\theta(L)}{\phi(L)}\varepsilon(t), \tag{28}$$

or, alternatively, by

$$\gamma(L)\phi(L)y(t) = \phi(L)\delta(L)x(t) + \gamma(L)\theta(L)\varepsilon(t). \tag{29}$$

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 (29) can be construed as a case of (26) 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 (27) can be construed as a case of equation (28) 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. 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 (28). This can be written as

$$y(t) = \omega(L)x(t) + \eta(t), \tag{29}$$

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 $x(t)$ 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 $x(t)$ and the noise $\eta(t)$ are uncorrelated, it follows that

$$\omega_\tau = \frac{C(y_t, x_{t-\tau})}{V(x_\tau)}. \tag{30}$$

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 $C(y_t, x_{t-\tau})$ and $V(x_\tau)$ 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:

$$\rho(L)x(t) = \psi(L)\xi(t). \tag{31}$$

Then, $x(t)$ can be reduced to the white-noise sequence $\xi(t)$ by the application of the filter $\pi(L) = \rho(L)/\psi(L)$. The application of the same filter to $y(t)$ and $\eta(t)$ will generate the sequences $q(t) = \pi(L)y(t)$ and $\zeta(t) = \pi(L)\eta(t)$ respectively. Hence, if the model of equation (29) can be transformed into

$$q(t) = \omega(L)\xi(t) + \zeta(t), \tag{32}$$

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

$$\omega_\tau = \frac{C(q_t, \xi_{t-\tau})}{V(\xi_\tau)}. \tag{30}$$

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, 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.