THE STARTING-VALUE PROBLEM ASSOCIATED WITH ARMA PROCESSES

The ARMA(p, q) model can be represented by the equation

$$\sum_{i=0}^{p} \alpha_i y(t-i) = \sum_{i=0}^{q} \mu_i \varepsilon(t-i), \quad \text{with} \quad \alpha_0 = 1, \tag{1}$$

where $\varepsilon(t)$ is a white-noise process with a variance of σ^2. In matrix terms, the system that generates a sample $y = [y_0, y_1, \ldots, y_{T-1}]'$ of T observations, can be written as

$$Ay + A_* y_* = M\varepsilon + M_* \varepsilon_*. \tag{2}$$

Here, $\varepsilon = [\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}]'$ is a vector of independently and identically distributed elements, whilst $y_* = [y_{-p}, \ldots, y_{-2}, y_{-1}]'$ and $\varepsilon_* = [\varepsilon_{-q}, \ldots, \varepsilon_{-2}, \varepsilon_{-1}]'$ are vectors of presample elements.

The matrices A and M, which are of a lower-triangular Toeplitz form, are completely characterised by their leading columns, which are the vectors $[\alpha_0, \alpha_1, \ldots \alpha_p, 0, \ldots, 0]'$ and $[\mu_0, \mu_1, \ldots, \mu_q, 0, \ldots 0]'$, respectively.

The matrices $A_* = [A'_{**}, 0]'$ and $M_* = [M'_{**}, 0]'$ contain the parameters associated with the presample elements. The principal minor of $A_* = [A'_{**}, 0]'$ is a nonsingular upper-triangular matrix A_{**} of order p. Likewise, M_{**}, which is the leading minor of M_*, is a nonsingular matrix of order q.

An example is provided by the following display that relates to the case where the autoregressive order is $p = 3$ and the size of the sample is $T = 6$:

$$A_* = \begin{bmatrix} \alpha_3 & \alpha_2 & \alpha_1 \\ 0 & \alpha_3 & \alpha_2 \\ 0 & 0 & \alpha_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} \alpha_0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & \alpha_0 & 0 & 0 & 0 & 0 \\ \alpha_2 & \alpha_1 & \alpha_0 & 0 & 0 & 0 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & 0 & 0 \\ 0 & \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & 0 \\ 0 & 0 & \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 \end{bmatrix}. \tag{3}$$

From (2), we obtain the following expressions for y and ε:

$$y = A^{-1} M\varepsilon + A^{-1}(M_* \varepsilon_* - A_* y_*), \tag{4}$$

$$\varepsilon = M^{-1} Ay - M^{-1}(A_* y_* - M_* \varepsilon_*). \tag{5}$$

We wish to generate the vector ε of the within-sample disturbances knowing the parameter matrices and the data. Equation (5) indicates a way of generating ε that would be available if the presmaple values of y_* and ε_* were known. In the absence of known values, we might think of using the conditional expectations $E(y_*|y)$ and $E(\varepsilon_*|y)$. Whereas $E(\varepsilon_*|y) = 0$, we could obtain the elements of $E(y_*|y)$ by a process of "backwards forecasting" that applies the conventional ARMA forecasting technique in reverse time. However, this requires a knowledge of elements of the vector ε that are, as yet, unknown to us.

Using $y_0, \ldots, y_{p-1}$ as starting values and setting $\varepsilon_0 = \cdots = \varepsilon_{q-1} = 0$, we can generate initial estimates of $\varepsilon_p, \ldots, \varepsilon_{T-1}$ via a recursion based on (1). Then, presample values of $y(t)$ can be estimated that can be used in generating fresh estimates of the values of $\varepsilon(t)$ within the sample period. Thereafter, new presample values of $y(t)$ can be estimated, and so on, back and forth. This kind of procedure was used by Box and Jenkins (1976).

A more fruitful approach is to concentrate on finding the conditional expectation $E(\varepsilon|y)$ directly. The formula is

$$E(\varepsilon|y) = E(\varepsilon) + C(\varepsilon, y)D^{-1}(y)\{y - E(y)\}, \tag{6}$$

where $E(\varepsilon) = 0$ and $E(y) = 0$ are unconditional expectations, $D(y) = \sigma^2 Q$ is the dispersion matrix of y and $C(\varepsilon, y)$ is the matrix of the covariances of ε and y. The essential result is that

$$\begin{aligned} C(\varepsilon, y) &= E(\varepsilon\varepsilon')M'A'^{-1} \\ &= \sigma^2 M'A'^{-1}. \end{aligned} \tag{7}$$

From this, we get

$$\begin{aligned} E(\varepsilon|y) &= C(\varepsilon, y)D^{-1}(y)y \\ &= M'A'^{-1}Q^{-1}y. \end{aligned} \tag{8}$$

The dispersion matrix of the estimate is

$$\begin{aligned} D\{E(\varepsilon|y)\} &= C(\varepsilon, y)D^{-1}(y)C(y, \varepsilon) \\ &= \sigma^2 M'A'^{-1}Q^{-1}A^{-1}M. \end{aligned} \tag{9}$$

By combining (8) and (9), we find that

$$E'(\varepsilon|y)D^{-1}\{E(\varepsilon|y)\}E(\varepsilon|y) = \frac{1}{\sigma^2}y'Q^{-1}y, \tag{10}$$

which is as we might expect.

Next, there is the question of how to represent the dispersion matrix $D(y) = \sigma^2 Q$, how it might be approximated and how to deal with it in the process of computing the disturbance vector ε. Pollock (1999) has provided the expressions

$$D(y) = \sigma^2 A^{-1}(V\Omega V' + MM')A'^{-1} \quad \text{and} \tag{11}$$

$$\begin{aligned} D^{-1}(y) = \frac{1}{\sigma^2}A'M'^{-1}[I_T - M^{-1}V\{\Omega^{-1} \\ + V'(MM')^{-1}V\}V'M'^{-1}]M^{-1}A. \end{aligned} \tag{12}$$

where

$$V = [\,-A_* \quad M_*\,] \quad \text{and} \quad \Omega = D(u_*) \quad \text{with} \quad u_* = \begin{bmatrix} y_* \\ \varepsilon_* \end{bmatrix}. \tag{13}$$

The purpose of these expressions was to provide a portmanteau which contains the expression of the pure autoregressive (AR) process and a pure moving average (MA) process as special cases.

For the AR case, there are

$$D(y) = \sigma^2 A^{-1}\{I - A_*(A_*'A_* + I)A_*'\}A'^{-1} \quad \text{and} \tag{14}$$

$$D^{-1}(y) = \frac{1}{\sigma^2}\{A'A - A_*A_*'\}. \tag{15}$$

whereas, for the MA case there, are

$$D(y) = \sigma^2(MM' + M_*M_*') \quad \text{and} \tag{16}$$

$$\begin{aligned} D^{-1}(y) = \frac{1}{\sigma^2}M'^{-1}[I - M^{-1}M_*\{I_q \\ + M_*'(MM')^{-1}M_*\}^{-1}M_*'M'^{-1}]M^{-1}. \end{aligned} \tag{17}$$

The commonest form of approximations arise from ignoring the matrices A_* and M_* within (14) and (16) respectively by setting them to zero. Within (11), it is a matter of ignoring the matrix $V\Omega V'$. The effect of these simplifications within (8) is to give the approximation

$$\begin{aligned} E(\varepsilon|y) &\simeq M'A'^{-1}\{A'M'^{-1}M^{-1}A\}y \\ &\simeq M^{-1}Ay, \end{aligned} \tag{18}$$

which is what we would expect on the basis of equation (5). However, it is the very inadequacy of this approximation that has motivated us to find a exact expression for $E(\varepsilon|y)$.

A computational rendering of the algebra of (11), which is sparing of computer storage, has been provided in Pollock (1999). The algorithm is relatively complex; and its complexities can be avoided at the cost of using more storage space.

The essential computational task is to evaluate the expression $p = Q^{-1}y$, which is found in equation (8). To begin with, it is straightforward to find the autocovariances of an ARMA process which constitute the elements of the dispersion matrix $D(y) = \sigma^2 Q$. In the case of an MA process or an ARMA process, we need to avoid the direct inversion of Q, which is a large matrix full of nonzero elements.

To these ends, we may compute the Cholesky factorisation $Q = LL'$, in which L stands for a lower-triangular matrix. Then we may write the equation $Qp = y$ as $LL'p = Ld = y$. We proceed to find d from $Ld = y$ by a simple process of forwards recursions. Thereafter, we find p from $L'p = d$ by a backwards recursion. The remaining steps in finding $E(\varepsilon|y)$ consist of a backward recursion, for finding $g = A'^{-1}p$ from $A'g = p$, and a direct multiplication, for finding $E(\varepsilon|y) = M'g$.

References

Box, G.E.P., and G.M. Jenkins, (1976), *Time Series Analysis: Forecasting and Control,* Revised Edition, Holden-Day, San Francisco.

Pollock, (1999), *Time-Series Analysis, Signal Processing and Dynamics,* The Academic Press, London.