11

Vector Autoregressive Moving Average Processes

11.1 Introduction

In this chapter, we extend our standard finite order VAR model,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t,$$

by allowing the error terms, here ε_t, to be autocorrelated rather than white noise. The autocorrelation structure is assumed to be of a relatively simple type so that ε_t has a finite order moving average (MA) representation,

$$\varepsilon_t = u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q},$$

where, as usual, u_t is zero mean white noise with nonsingular covariance matrix Σ_u. A finite order VAR process with finite order MA error term is called a VARMA (*vector autoregressive moving average*) process.

Before we study VARMA processes in general, we will discuss some properties of finite order MA processes in Section 11.2. In Section 11.3, we consider the more general stationary VARMA processes with stable VAR part and we will learn that generally they have infinite order pure VAR and MA representations. Their autocovariance and autocorrelation properties are treated in Section 11.4 and forecasting VARMA processes is discussed in Section 11.5. In Section 11.6, transforming and aggregating these processes is considered. In that section, we will see that a linearly transformed finite order VAR(p) process, in general, does not admit a finite order VAR representation but becomes a VARMA process. Because transformations of variables are quite common in practice, this result is a powerful argument in favor of the more general VARMA class. Finally, Section 11.7 contains discussions of causality issues and impulse response analysis in the context of VARMA systems. Throughout this chapter, we consider stationary processes only.

H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*

11.2 Finite Order Moving Average Processes

In Chapter 2, we have encountered MA processes of possibly infinite order. Specifically, we have seen that stationary, stable finite order VAR processes can be represented as MA processes. Now we deal explicitly with *finite order* MA processes. Let us begin with the simplest case of a K-dimensional MA process of order 1 (MA(1) process), $y_t = \mu + u_t + M_1 u_{t-1}$, where $y_t = (y_{1t}, \ldots, y_{Kt})', u_t$ is zero mean white noise with nonsingular covariance matrix Σ_u, and $\mu = (\mu_1, \ldots, \mu_K)'$ is the mean vector of y_t, i.e., $E(y_t) = \mu$ for all t. For notational simplicity we will assume in the following that $\mu = 0$, that is, y_t is a zero mean process. Thus, we consider

$$y_t = u_t + M_1 u_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{11.2.1}$$

which may be rewritten as

$$u_t = y_t - M_1 u_{t-1}.$$

By successive substitution we get

$$\begin{aligned} u_t &= y_t - M_1(y_{t-1} - M_1 u_{t-2}) = y_t - M_1 y_{t-1} + M_1^2 u_{t-2} \\ &= \cdots = y_t - M_1 y_{t-1} + \cdots + (-M_1)^n y_{t-n} + (-M_1)^{n+1} u_{t-n-1} \\ &= y_t + \sum_{i=1}^{\infty} (-M_1)^i y_{t-i}, \end{aligned}$$

if $M_1^i \to 0$ as $i \to \infty$. Hence,

$$y_t = -\sum_{i=1}^{\infty} (-M_1)^i y_{t-i} + u_t, \tag{11.2.2}$$

which is the potentially infinite order VAR representation of the process. Because $(-M_1)^i$ may be equal to zero for i greater than some finite number p, the process may in fact be a finite order VAR(p). For instance, we get $p = 1$ for a bivariate process with

$$M_1 = \begin{bmatrix} 0 & m \\ 0 & 0 \end{bmatrix},$$

where m is some nonzero real number.

For the representation (11.2.2) to be meaningful, M_1^i must approach zero as $i \to \infty$, which in turn requires that the eigenvalues of M_1 are all less than 1 in modulus or, equivalently,

$$\det(I_K + M_1 z) \neq 0 \qquad \text{for } z \in \mathbb{C}, |z| \leq 1.$$

This condition is analogous to the stability condition for a VAR(1) process. It guarantees that the infinite sum in (11.2.2) exists as a mean square limit.

More generally, it can be shown that a (zero mean) MA(q) process (moving average process of order q),

$$y_t = u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{11.2.3}$$

has a pure VAR representation

$$y_t = \sum_{i=1}^{\infty} \Pi_i y_{t-i} + u_t, \tag{11.2.4}$$

if

$$\det(I_K + M_1 z + \cdots + M_q z^q) \neq 0 \quad \text{for } z \in \mathbb{C}, |z| \leq 1. \tag{11.2.5}$$

An MA(q) process with this property is called *invertible* in the following because we can invert from the MA to a VAR representation. Writing the process in lag operator notation as

$$y_t = (I_K + M_1 L + \cdots + M_q L^q) u_t = M(L) u_t$$

the MA operator $M(L) := I_K + M_1 L + \cdots + M_q L^q$ is invertible if it satisfies (11.2.5) and we may formally write

$$M(L)^{-1} y_t = u_t.$$

The actual computation of the coefficient matrices Π_i in

$$M(L)^{-1} = \Pi(L) = I_K - \sum_{i=1}^{\infty} \Pi_i L^i$$

can be done recursively using $\Pi_1 = M_1$ and

$$\Pi_i = M_i - \sum_{j=1}^{i-1} \Pi_{i-j} M_j, \quad i = 2, 3, \ldots, \tag{11.2.6}$$

where $M_j := 0$ for $j > q$. These recursions follow immediately from the corresponding recursions used to compute the MA coefficients of a pure VAR process (see Chapter 2, (2.1.22)).

The autocovariances of the MA(q) process (11.2.3) are particularly easy to obtain. They follow directly from those of an infinite order MA process given in Chapter 2, Section 2.1.2, (2.1.18):

$$\Gamma_y(h) = E(y_t y_{t-h}') = \begin{cases} \sum_{i=0}^{q-h} M_{i+h} \Sigma_u M_i', & h = 0, 1, \ldots, q, \\ 0, & h = q+1, q+2, \ldots, \end{cases} \tag{11.2.7}$$

with $M_0 := I_K$. As before, $\Gamma_y(-h) = \Gamma_y(h)'$. Thus, the vectors y_t and y_{t-h} are uncorrelated if $h > q$. Obviously, the process (11.2.3) is stationary because the $\Gamma_y(h)$ do not depend on t and the mean $E(y_t) = 0$ for all t.

It can be shown that a noninvertible MA(q) process violating (11.2.5) also has a pure VAR representation if the determinantal polynomial in (11.2.5) has no roots on the complex unit circle, i.e., if

$$\det(I_K + M_1 z + \cdots + M_q z^q) \neq 0 \quad \text{for } |z| = 1. \tag{11.2.8}$$

The VAR representation will, however, not be of the type (11.2.4) in that the white noise process will in general not be the one appearing in (11.2.3). The reason is that for any noninvertible MA(q) process satisfying (11.2.8), there is an equivalent invertible MA(q) satisfying (11.2.5) which has an identical autocovariance structure (see Hannan & Deistler (1988, Chapter 1, Section 3)). For instance, for the univariate MA(1) process

$$y_t = u_t + m u_{t-1}, \tag{11.2.9}$$

the invertibility condition requires that $1 + mz$ has no roots for $|z| \leq 1$ or, equivalently, $|m| < 1$. For any m, the process has autocovariances

$$E(y_t y_{t-h}) = \begin{cases} (1+m^2)\sigma_u^2 & \text{for } h = 0, \\ m\sigma_u^2 & \text{for } h = \pm 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\sigma_u^2 := \text{Var}(u_t)$. It is easy to check that the process $v_t + \frac{1}{m} v_{t-1}$, where v_t is a white noise process with $\sigma_v^2 := \text{Var}(v_t) = m^2\sigma_u^2$, has the very same autocovariance structure. Thus, if $|m| > 1$, we may choose the invertible MA(1) representation

$$y_t = v_t + \frac{1}{m} v_{t-1} \tag{11.2.10}$$

with

$$\begin{aligned} v_t &= \left(1 + \frac{1}{m} L\right)^{-1} y_t = \sum_{i=0}^{\infty} \left(\frac{-1}{m}\right)^i y_{t-i} \\ &= \left(1 + \frac{1}{m} L\right)^{-1} (1 + mL) u_t. \end{aligned}$$

The reader is invited to check that v_t is indeed a white noise process with $\sigma_v^2 = m^2 \sigma_u^2$ (see Problem 11.10). Only if $|m| = 1$ and, hence, $1 + mz = 0$ for some z on the unit circle ($z = 1$ or -1), an invertible representation does not exist.

Although for higher order and higher-dimensional processes, where roots inside and outside the unit circle may exist, it is more complicated to find the invertible representation, it can be done whenever (11.2.8) is satisfied. In the remainder of this chapter, we will therefore assume without notice that all MA processes are invertible unless stated otherwise. It should be understood that this assumption implies a slight loss of generality because MA processes with roots on the complex unit circle are excluded.

11.3 VARMA Processes

11.3.1 The Pure MA and Pure VAR Representations of a VARMA Process

As mentioned in the introduction to this chapter, allowing finite order VAR processes to have finite order MA instead of white noise error terms, results in the broad and flexible class of vector autoregressive moving average (VARMA) processes. The general form of a process from this class with VAR order p and MA order q is

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \\ & \quad t = 0, \pm 1, \pm 2, \ldots . \end{aligned} \tag{11.3.1}$$

Such a process is briefly called a VARMA(p, q) process. As before, u_t is zero mean white noise with nonsingular covariance matrix Σ_u.

It may be worth elaborating a bit on this specification. What kind of process y_t is defined by the VARMA(p, q) model (11.3.1)? To look into this question, let us denote the MA part by ε_t, that is, $\varepsilon_t = u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}$ and

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t.$$

If this process is *stable*, that is, if

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for } |z| \leq 1, \tag{11.3.2}$$

then, by the same arguments used in Chapter 2, Section 2.1.2, and by Proposition C.9 of Appendix C.3,

$$\begin{aligned} y_t &= \mu + \sum_{i=0}^{\infty} D_i \varepsilon_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} D_i (u_{t-i} + M_1 u_{t-i-1} + \cdots + M_q u_{t-i-q}) \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \end{aligned} \tag{11.3.3}$$

is well-defined as a limit in mean square, given a well-defined white noise process u_t. Here

$$\mu := (I_K - A_1 - \cdots - A_p)^{-1}\nu,$$

the D_i are $(K \times K)$ matrices satisfying

$$\sum_{i=0}^{\infty} D_i z^i = (I_K - A_1 z - \cdots - A_p z^p)^{-1},$$

and the Φ_i are $(K \times K)$ matrices satisfying

$$\sum_{i=0}^{\infty} \Phi_i z^i = \left(\sum_{i=0}^{\infty} D_i z^i\right)(I_K + M_1 z + \cdots + M_q z^q).$$

In the following, when we call y_t a stable VARMA(p, q) process, we mean the well-defined process given in (11.3.3). For instance, if u_t is Gaussian white noise, it can be shown that y_t is a Gaussian process with all finite subcollections of vectors $y_t, \ldots, y_{t+h}$ having joint multivariate normal distributions. The representation (11.3.3) is a pure MA or simply MA representation of y_t.

To make the derivation of the MA representation more transparent, let us write the process (11.3.1) in lag operator notation,

$$A(L)y_t = \nu + M(L)u_t, \tag{11.3.4}$$

where $A(L) := I_K - A_1 L - \cdots - A_p L^p$ and $M(L) := I_K + M_1 L + \cdots + M_q L^q$. A pure MA representation of y_t is obtained by premultiplying with $A(L)^{-1}$,

$$y_t = A(1)^{-1}\nu + A(L)^{-1}M(L)u_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}.$$

Hence, multiplying from the left by $A(L)$ gives

$$\begin{aligned}(I_K - A_1 L - \cdots - A_p L^p)&\left(\sum_{i=0}^{\infty} \Phi_i L^i\right) \\ &= I_K + \sum_{i=1}^{\infty}\left(\Phi_i - \sum_{j=1}^{i} A_j \Phi_{i-j}\right) L^i \\ &= I_K + M_1 L + \cdots + M_q L^q\end{aligned}$$

and, thus, comparing coefficients results in

$$M_i = \Phi_i - \sum_{j=1}^{i} A_j \Phi_{i-j}, \quad i = 1, 2, \ldots,$$

with $\Phi_0 := I_K$, $A_j := 0$ for $j > p$, and $M_i := 0$ for $i > q$. Rearranging terms gives

$$\Phi_i = M_i + \sum_{j=1}^{i} A_j \Phi_{i-j}, \quad i = 1, 2, \dots. \tag{11.3.5}$$

If the MA operator $M(L)$ satisfies the invertibility condition (11.2.5), then the VARMA process (11.3.4) is called *invertible*. In that case, it has a pure VAR representation,

$$y_t - \sum_{i=1}^{\infty} \Pi_i y_{t-i} = M(L)^{-1} A(L) y_t = M(1)^{-1}\nu + u_t,$$

and the Π_i matrices are obtained by comparing coefficients in

$$I_K - \sum_{i=1}^{\infty} \Pi_i L^i = M(L)^{-1} A(L).$$

Alternatively, multiplying this expression from the left by $M(L)$ gives

$$\begin{aligned}
(I_K + M_1 L + \dots + M_q L^q) & \left(I_K - \sum_{i=1}^{\infty} \Pi_i L^i \right) \\
&= I_K + \sum_{i=1}^{\infty} \left(M_i - \sum_{j=1}^{i} M_{i-j} \Pi_j \right) L^i \\
&= I_K - A_1 L - \dots - A_p L^p,
\end{aligned}$$

where $M_0 := I_K$ and $M_i := 0$ for $i > q$. Setting $A_i := 0$ for $i > p$ and comparing coefficients gives

$$-A_i = M_i - \sum_{j=1}^{i-1} M_{i-j} \Pi_j - \Pi_i$$

or

$$\Pi_i = A_i + M_i - \sum_{j=1}^{i-1} M_{i-j} \Pi_j \quad \text{for } i = 1, 2, \dots. \tag{11.3.6}$$

As usual, the sum is defined to be zero if the lower bound for the summation index exceeds its upper bound.

For instance, for the zero mean VARMA$(1,1)$ process

$$y_t = A_1 y_{t-1} + u_t + M_1 u_{t-1}, \tag{11.3.7}$$

we get

$$\begin{aligned}
\Pi_1 &= A_1 + M_1 \\
\Pi_2 &= A_2 + M_2 - M_1 \Pi_1 = -M_1 A_1 - M_1^2 \\
&\vdots \\
\Pi_i &= (-1)^{i-1} (M_1^i + M_1^{i-1} A_1), \quad i = 1, 2, \dots,
\end{aligned}$$

and the coefficients of the pure MA representation are

$$\begin{aligned}
\Phi_0 &= I_K \\
\Phi_1 &= M_1 + A_1 \\
\Phi_2 &= M_2 + A_1\Phi_1 + A_2\Phi_0 = A_1(M_1 + A_1) \\
&\vdots \\
\Phi_i &= A_1^{i-1}M_1 + A_1^i, \quad i = 1, 2, \ldots .
\end{aligned}$$

If y_t is a stable and invertible VARMA process, then the pure MA representation (11.3.3) is called the *canonical* or *prediction error MA representation*, in accordance with the terminology used in the finite order VAR case. In addition to the pure MA and VAR representations considered in this section, a VARMA process also has VAR(1) representations. One such representation is introduced next.

11.3.2 A VAR(1) Representation of a VARMA Process

Suppose y_t has the VARMA(p, q) representation (11.3.1). For simplicity, we assume that its mean is zero and, hence, $\nu = 0$. Let

$$Y_t := \underset{(K(p+q)\times 1)}{\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ u_t \\ \vdots \\ u_{t-q+1} \end{bmatrix}}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \\ u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} u_t \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kp \times 1) \\ \left.\vphantom{\begin{matrix} u_t \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kq \times 1) \end{matrix}$$

and

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad [K(p+q) \times K(p+q)],$$

where

$$\mathbf{A}_{11} := \underset{(Kp\times Kp)}{\begin{bmatrix} A_1 & \ldots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \ldots & I_K & 0 \end{bmatrix}},$$

$$\mathbf{A}_{12} := \underset{(Kp\times Kq)}{\begin{bmatrix} M_1 & \ldots & M_{q-1} & M_q \\ 0 & \ldots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 0 \end{bmatrix}},$$

$$\underset{(Kq\times Kp)}{\mathbf{A}_{21}} := 0, \quad \mathbf{A}_{22} := \underset{(Kq\times Kq)}{\begin{bmatrix} 0 & \dots & 0 & 0 \\ I_K & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & \dots & I_K & 0 \end{bmatrix}}.$$

With this notation, we get the VAR(1) representation of Y_t,

$$Y_t = \mathbf{A}Y_{t-1} + U_t. \qquad (11.3.8)$$

If the VAR order is zero ($p = 0$), we choose $p = 1$ and set $A_1 = 0$ in this representation.

The $K(p+q)$-dimensional VAR(1) process in (11.3.8) is stable if and only if y_t is stable. This result follows because

$$\begin{aligned} \det(I_{K(p+q)} - \mathbf{A}z) &= \det(I_{Kp} - \mathbf{A}_{11}z)\det(I_{Kq} - \mathbf{A}_{22}z) \\ &= \det(I_K - A_1 z - \cdots - A_p z^p). \end{aligned} \qquad (11.3.9)$$

Here the rules for the determinant of a partitioned matrix from Appendix A.10 have been used and we have also used that $I_{Kq} - \mathbf{A}_{22}z$ is a lower triangular matrix with ones on the main diagonal which has determinant 1. Furthermore, $\det(I_{Kp} - \mathbf{A}_{11}z) = \det(I_K - A_1 z - \cdots - A_p z^p)$ follows as in Section 2.1.1.

From Chapter 2, we know that if y_t and, hence, Y_t is stable, the latter process has an MA representation

$$Y_t = \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}.$$

Premultiplying by the $(K \times K(p+q))$ matrix $J := [I_K : 0 : \cdots : 0]$ gives

$$y_t = \sum_{i=0}^{\infty} J\mathbf{A}^i U_{t-i} = \sum_{i=0}^{\infty} J\mathbf{A}^i HJU_{t-i} = \sum_{i=0}^{\infty} J\mathbf{A}^i Hu_{t-i} = \sum_{i=0}^{\infty} \Phi_i u_{t-i},$$

where

$$H = \begin{bmatrix} I_K \\ 0 \\ \vdots \\ 0 \\ I_K \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} I_K \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kp \times K) \\ \left.\vphantom{\begin{matrix} I_K \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kq \times K) \end{matrix} .$$

Thus,

$$\Phi_i = J\mathbf{A}^i H. \qquad (11.3.10)$$

As an example, consider the zero mean VARMA(1, 1) process from (11.3.7),

$$y_t = A_1 y_{t-1} + u_t + M_1 u_{t-1}.$$

For this process

$$Y_t = \begin{bmatrix} y_t \\ u_t \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} A_1 & M_1 \\ 0 & 0 \end{bmatrix}, \quad U_t = \begin{bmatrix} u_t \\ u_t \end{bmatrix},$$

$$J = [I_K : 0] \quad (K \times 2K),$$

and

$$H = \begin{bmatrix} I_K \\ I_K \end{bmatrix} (2K \times K).$$

Hence,

$$\begin{aligned} \Phi_0 &= JH = I_K, \\ \Phi_1 &= J\mathbf{A}H = [A_1 : M_1]H = A_1 + M_1, \\ \Phi_2 &= J\mathbf{A}^2 H = J\begin{bmatrix} A_1^2 & A_1 M_1 \\ 0 & 0 \end{bmatrix} H = A_1^2 + A_1 M_1, \\ &\vdots \\ \Phi_i &= J\mathbf{A}^i H = J\begin{bmatrix} A_1^i & A_1^{i-1} M_1 \\ 0 & 0 \end{bmatrix} H = A_1^i + A_1^{i-1} M_1, \quad i = 1, 2, \ldots . \end{aligned} \tag{11.3.11}$$

This, of course, is precisely the same formula obtained from the recursions in (11.3.5).

The foregoing method of computing the MA matrices is just another way of computing the coefficient matrices of the power series

$$I_K + \sum_{i=1}^{\infty} \Phi_i L^i = (I_K - A_1 L - \cdots - A_p L^p)^{-1}(I_K + M_1 L + \cdots + M_q L^q).$$

Therefore, it can just as well be used to compute the Π_i coefficient matrices of the pure VAR representation of a VARMA process. Recall that

$$I_K - \sum_{i=1}^{\infty} \Pi_i L^i = (I_K + M_1 L + \cdots + M_q L^q)^{-1}(I_K - A_1 L - \cdots - A_p L^p).$$

Hence, if we define

$$\mathbf{M} := \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}, \tag{11.3.12}$$

where

$$\mathbf{M}_{11} := \underset{(Kq\times Kq)}{\begin{bmatrix} -M_1 & \dots & -M_{q-1} & -M_q \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I_K & 0 \end{bmatrix}},$$

$$\mathbf{M}_{12} := \underset{(Kq\times Kp)}{\begin{bmatrix} -A_1 & \dots & -A_{p-1} & -A_p \\ 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix}},$$

$$\underset{(Kp\times Kq)}{\mathbf{M}_{21} := 0}, \quad \mathbf{M}_{22} := \underset{(Kp\times Kp)}{\begin{bmatrix} 0 & \dots & 0 & 0 \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I_K & 0 \end{bmatrix}},$$

we get

$$-\Pi_i = J\mathbf{M}^i H \tag{11.3.13}$$

with

$$H := \begin{bmatrix} I_K \\ 0 \\ \vdots \\ 0 \\ I_K \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} I_K \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kq \times K) \\ \left.\vphantom{\begin{matrix} I_K \\ 0 \\ \vdots \\ 0 \end{matrix}}\right\} (Kp \times K) \end{matrix} .$$

11.4 The Autocovariances and Autocorrelations of a VARMA(p, q) Process

For the K-dimensional, zero mean, stable VARMA(p, q) process

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \dots + M_q u_{t-q}, \tag{11.4.1}$$

the autocovariances can be obtained formally from its pure MA representation as in Section 2.1.2. For instance, if y_t has the canonical MA representation

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i},$$

the autocovariance matrices are

$$\Gamma_y(h) := E(y_t y'_{t-h}) = \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi'_i.$$

For the actual computation of the autocovariance matrices, the following approach is more convenient. Postmultiplying (11.4.1) by y'_{t-h} and taking expectations gives

$$\begin{aligned} E(y_t y'_{t-h}) &= A_1 E(y_{t-1} y'_{t-h}) + \cdots + A_p E(y_{t-p} y'_{t-h}) + E(u_t y'_{t-h}) + \cdots \\ &\quad + M_q E(u_{t-q} y'_{t-h}). \end{aligned}$$

From the pure MA representation of the process, it can be seen that $E(u_t y'_s) = 0$ for $s < t$. Hence, we get for $h > q$,

$$\Gamma_y(h) = A_1 \Gamma_y(h-1) + \cdots + A_p \Gamma_y(h-p). \tag{11.4.2}$$

If $p > q$ and $\Gamma_y(0), \dots, \Gamma_y(p-1)$ are available, this relation can be used to compute the autocovariances recursively for $h = p, p+1, \dots$.

The initial matrices can be obtained from the VAR(1) representation (11.3.8), just as in Chapter 2, Section 2.1.4. In that section, we obtained the relation

$$\Gamma_Y(0) = \mathbf{A} \Gamma_Y(0) \mathbf{A}' + \Sigma_U \tag{11.4.3}$$

for the covariance matrix of the VAR(1) process Y_t. Here $\Sigma_U = E(U_t U'_t)$ is the covariance matrix of the white noise process in (11.3.8). Applying the vec operator to (11.4.3) and rearranging terms gives

$$\operatorname{vec} \Gamma_Y(0) = (I_{K^2(p+q)^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \operatorname{vec}(\Sigma_U), \tag{11.4.4}$$

where the existence of the inverse follows again from the stability of the process, as in Section 2.1.4, by appealing to the determinantal relation (11.3.9).

Having computed $\Gamma_Y(0)$ as in (11.4.4), we may collect $\Gamma_y(0), \dots, \Gamma_y(p-1)$ from

$$\Gamma_Y(0) = \begin{bmatrix} \boldsymbol{\Gamma}_{11}(0) & \boldsymbol{\Gamma}_{12}(0) \\ \boldsymbol{\Gamma}_{12}(0)' & \boldsymbol{\Gamma}_{22}(0) \end{bmatrix},$$

where

$$\boldsymbol{\Gamma}_{11}(0) = \begin{bmatrix} \Gamma_y(0) & \Gamma_y(1) & \dots & \Gamma_y(p-1) \\ \Gamma_y(-1) & \Gamma_y(0) & \dots & \Gamma_y(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_y(-p+1) & \Gamma_y(-p+2) & \dots & \Gamma_y(0) \end{bmatrix},$$

$$\mathbf{\Gamma}_{12}(0) = \begin{bmatrix} E(y_t u_t') & E(y_t u_{t-1}') & \cdots & E(y_t u_{t-q+1}') \\ 0 & E(y_{t-1} u_{t-1}') & \cdots & E(y_{t-1} u_{t-q+1}') \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & E(y_{t-p+1} u_{t-q+1}') \end{bmatrix},$$

and

$$\mathbf{\Gamma}_{22}(0) = \begin{bmatrix} \Sigma_u & 0 & \dots & 0 \\ 0 & \Sigma_u & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_u \end{bmatrix}.$$

As mentioned previously, the recursions (11.4.2) are valid for $h > q$ only. Thus, this way of computing the autocovariances requires that $p > q$. If the VAR order is less than q, then it may be increased artificially by adding lags of y_t with zero coefficient matrices until the VAR order p exceeds the MA order q. Then the aforementioned procedure can be applied. A computationally more efficient method of computing the autocovariances of a VARMA process is described by Mittnik (1990).

The autocorrelations of a VARMA(p,q) process are obtained from its autocovariances as in Chapter 2, Section 2.1.4. That is,

$$R_y(h) = D^{-1}\Gamma_y(h)D^{-1}, \tag{11.4.5}$$

where D is a diagonal matrix with the square roots of the diagonal elements of $\Gamma_y(0)$ on the main diagonal.

To illustrate the computation of the covariance matrices, we consider the VARMA$(1,1)$ process (11.3.7). Because $p = q$, we add a second lag of y_t so that

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t + M_1 u_{t-1}$$

with $A_2 := 0$. Thus, in this case,

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ u_t \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} A_1 & 0 & M_1 \\ I_K & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$U_t = \begin{bmatrix} u_t \\ 0 \\ u_t \end{bmatrix}, \quad \Sigma_U = \begin{bmatrix} \Sigma_u & 0 & \Sigma_u \\ 0 & 0 & 0 \\ \Sigma_u & 0 & \Sigma_u \end{bmatrix}.$$

With this notation, we get from (11.4.4),

$$\operatorname{vec}\begin{bmatrix} \Gamma_y(0) & \Gamma_y(1) & \Sigma_u \\ \Gamma_y(-1) & \Gamma_y(0) & 0 \\ \Sigma_u & 0 & \Sigma_u \end{bmatrix} = (I_{9K^2} - \mathbf{A}\otimes\mathbf{A})^{-1}\operatorname{vec}(\Sigma_U).$$

Now, because we have the starting-up matrices $\Gamma_y(0)$ and $\Gamma_y(1)$, the recursions (11.4.2) may be applied, giving

$$\Gamma_y(h) = A_1 \Gamma_y(h-1) \quad \text{for } h = 2, 3, \ldots.$$

In stating the assumptions for the VARMA(p, q) process at the beginning of this section, invertibility has not been mentioned. This is no accident because this condition is actually not required for computing the autocovariances of a VARMA(p, q) process. The same formulas may be used for invertible and noninvertible processes. On the other hand, the stability condition is essential here, because it ensures invertibility of the matrix $I - \mathbf{A} \otimes \mathbf{A}$.

11.5 Forecasting VARMA Processes

Suppose the K-dimensional zero mean VARMA(p, q) process

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q} \tag{11.5.1}$$

is stable and invertible. As we have seen in Section 11.3.1, it has a pure VAR representation,

$$y_t = \sum_{i=1}^{\infty} \Pi_i y_{t-i} + u_t, \tag{11.5.2}$$

and a pure MA representation,

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}. \tag{11.5.3}$$

Formulas for optimal forecasts can be given in terms of each of these representations.

Assuming that u_t is *independent* white noise and applying the conditional expectation operator E_t, given information up to time t, to (11.5.1) gives an optimal h-step forecast

$$y_t(h) = \begin{cases} A_1 y_t(h-1) + \cdots + A_p y_t(h-p) \\ \qquad + M_h u_t + \cdots + M_q u_{t+h-q} & \text{for } h \leq q, \\ \\ A_1 y_t(h-1) + \cdots + A_p y_t(h-p) & \text{for } h > q, \end{cases} \tag{11.5.4}$$

where, as usual, $y_t(j) := y_{t+j}$ for $j \leq 0$. Analogously, we get from (11.5.2),

$$y_t(h) = \sum_{i=1}^{\infty} \Pi_i y_t(h-i), \tag{11.5.5}$$

and, in Chapter 2, Section 2.2.2, we have seen that the optimal forecast in terms of the infinite order MA representation is

$$y_t(h) = \sum_{i=h}^{\infty} \Phi_i u_{t+h-i} = \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-i} \tag{11.5.6}$$

(see (2.2.10)). Although in Chapter 2 this result was derived in the slightly more special setting of finite order VAR processes, it is not difficult to see that it carries over to the present situation. All three formulas (11.5.4)–(11.5.6) result, of course, in equivalent predictors or forecasts. They are different representations of the *linear* minimum MSE predictors if u_t is uncorrelated but not necessarily independent white noise.

A forecasting formula can also be obtained from the VAR(1) representation (11.3.8) of the VARMA(p, q) process. From Section 2.2.2, the optimal h-step forecast of a VAR(1) process at origin t is known to be

$$Y_t(h) = \mathbf{A}^h Y_t = \mathbf{A} Y_t(h-1). \tag{11.5.7}$$

Premultiplying with the $(K \times K(p+q))$ matrix $J := [I_K : 0 : \cdots : 0]$ results precisely in the recursive relation (11.5.4) (see Problem 11.4).

The forecasts at origin t are based on the information set

$$\Omega_t = \{y_s | s \le t\}.$$

This information set has the drawback of being unavailable in practice. Usually a finite sample of y_t data is given only and, hence, the u_t cannot be determined exactly. Thus, even if the parameters of the process are known, the prediction formulas (11.5.4)–(11.5.6) cannot be used. However, the invertibility of the process implies that the Π_i coefficient matrices go to zero exponentially with increasing i and we have the approximation

$$\sum_{i=1}^{\infty} \Pi_i y_t(h-i) \approx \sum_{i=1}^{n} \Pi_i y_t(h-i)$$

for large n. Consequently, in practice, if the information set is

$$\{y_1, \dots, y_T\} \tag{11.5.8}$$

and T is large, then the forecast

$$\breve{y}_T(h) = \sum_{i=1}^{T+h-1} \Pi_i \breve{y}_T(h-i), \tag{11.5.9}$$

where $\breve{y}_T(j) := y_{T+j}$ for $j \le 0$, will be almost identical to the optimal forecast. For a low order process, as it is commonly used in practice, for which the roots of

$$\det(I_K + M_1 z + \cdots + M_q z^q)$$

are not close to the unit circle, $T > 50$ will usually result in forecasts that cannot be distinguished from the optimal forecasts. It is worth noting, however, that the optimal forecasts based on the finite information set (11.5.8) can be determined. The resulting forecast formulas are, for instance, given by Brockwell & Davis (1987, Chapter 11, §11.4). A similar problem is not encountered in forecasting finite order VAR processes because there the optimal forecast depends on a finite string of past variables only.

In the presently considered theoretical setting, the forecast MSE matrices are most easily obtained from the representation (11.5.6). The forecast error is

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i}$$

and, hence, the forecast MSE matrix turns out to be

$$\begin{aligned} \Sigma_y(h) &:= E[(y_{t+h} - y_t(h))(y_{t+h} - y_t(h))'] \\ &= \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i', \end{aligned} \tag{11.5.10}$$

as in the finite order VAR case. Note, however, that, in the present case, the M_i coefficient matrices enter in computing the Φ_i matrices. Because the forecasts are unbiased, that is, the forecast errors have mean zero, the MSE matrix is the forecast error covariance matrix. Consequently, if the process is Gaussian, i.e., for all t and $h, y_t, \dots, y_{t+h}$ have a multivariate normal distribution and also the u_t's are normally distributed, then the forecast errors are normally distributed,

$$y_{t+h} - y_t(h) \sim \mathcal{N}(0, \Sigma_y(h)). \tag{11.5.11}$$

This result may be used in the usual fashion in setting up forecast intervals.

If a process with nonzero mean vector μ is considered, the mean vector may simply be added to the prediction formula for the mean-adjusted process. For example, if y_t has zero mean and $x_t = y_t + \mu$, then the optimal h-step forecast of x_t is

$$x_t(h) = y_t(h) + \mu.$$

The forecast MSE matrix is not affected, that is, $\Sigma_x(h) = \Sigma_y(h)$.

11.6 Transforming and Aggregating VARMA Processes

In practice, the original variables of interest are often transformed before their generation process is modelled. For example, data are often seasonally adjusted prior to an analysis. Also, sometimes they are temporally aggregated. For instance, quarterly data may have been obtained by adding up the corresponding monthly values or by taking their averages. Moreover, contemporaneous aggregation over a number of households, regions or sectors of the economy is quite common. For example, the GNP (gross national product) value for some period is the sum of private consumption, investment expenditures, net exports, and government spending for that period. It is often of interest to see what these transformations do to the generation processes of the variables in order to assess the consequences of transformations for forecasting and structural analysis. In the following, we assume that the original data are generated by a VARMA process and we study the consequences of linear transformations. These results are of importance because many temporal as well as contemporaneous aggregation procedures can be represented as linear transformations.

11.6.1 Linear Transformations of VARMA Processes

We shall begin with the result that a linear transformation of a process possessing an MA(q) representation gives a process that also has a finite order MA representation with order not greater than q.

Proposition 11.1 (*Linear Transformation of an MA(q) Process*)
Let u_t be a K-dimensional white noise process with nonsingular covariance matrix Σ_u and let

$$y_t = \mu + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}$$

be a K-dimensional invertible MA(q) process. Furthermore, let F be an $(M \times K)$ matrix of rank M. Then the M-dimensional process $z_t = F y_t$ has an invertible MA($\breve{q}$) representation,

$$z_t = F\mu + v_t + N_1 v_{t-1} + \cdots + N_{\breve{q}} v_{t-\breve{q}},$$

where v_t is M-dimensional white noise with nonsingular covariance matrix Σ_v, the N_i are $(M \times M)$ coefficient matrices and $\breve{q} \le q$. ■

We will not give a proof of this result here but refer the reader to Lütkepohl (1984) or Lütkepohl (1987, Chapter 4). The proposition is certainly not surprising because considering the autocovariance matrices of z_t, it is seen that

$$\begin{aligned} \Gamma_z(h) &= E[(Fy_t - F\mu)(Fy_{t-h} - F\mu)'] = F\Gamma_y(h)F' \\ &= \begin{cases} \sum_{i=0}^{q-h} FM_{i+h}\,\Sigma_u M_i' F', & h = 0, 1, \ldots, q, \\ 0, & h = q+1, q+2, \ldots, \end{cases} \end{aligned}$$

by (11.2.7). Thus, the autocovariances of z_t for lags greater than q are all zero. This result is a necessary requirement for the proposition to be true. It also helps to understand that the MA order of z_t may be lower than that of y_t because $\Gamma_z(h) = F\Gamma_y(h)F'$ may be zero even if $\Gamma_y(h)$ is nonzero.

The proposition has some interesting implications. As we will see in the following (Corollary 11.1.1), it implies that a linearly transformed VARMA(p, q) process has again a finite order VARMA representation. Thus, the VARMA class is closed with respect to linear transformations. The same is not true for the class of finite order VAR processes because, as we will see shortly, a linearly transformed VAR(p) process may not admit a finite order VAR representation. This, of course, is an argument in favor of considering the VARMA class rather than restricting the analysis to finite order VAR processes.

Corollary 11.1.1
Let y_t be a K-dimensional, stable, invertible VARMA(p, q) process and let F be an $(M \times K)$ matrix of rank M. Then the process $z_t = Fy_t$ has a VARMA$(\breve{p}, \breve{q})$ representation with

$$\breve{p} \leq Kp$$

and

$$\breve{q} \leq (K-1)p + q.$$

■

Proof: We write the process y_t in lag operator notation as

$$A(L)y_t = M(L)u_t, \tag{11.6.1}$$

where the mean is set to zero without loss of generality as y_t may represent deviations from the mean. Premultiplying by the adjoint $A(L)^{adj}$ of $A(L)$ gives

$$|A(L)|y_t = A(L)^{adj}M(L)u_t, \tag{11.6.2}$$

where $A(L)^{adj}A(L) = |A(L)|$ has been used. It is easy to check that $|A(z)^{adj}| \neq 0$ for $|z| \leq 1$. Thus, (11.6.2) is a stable and invertible VARMA representation of y_t. Premultiplying (11.6.2) with F results in

$$|A(L)|z_t = FA(L)^{adj}M(L)u_t. \tag{11.6.3}$$

The operator $A(L)^{adj}M(L)$ is easily seen to have degree at most $p(K-1)+q$ and, thus, the right-hand side of (11.6.3) is just a linearly transformed finite order MA process which, by Proposition 11.1, has an MA$(\breve{q})$ representation with

$$\breve{q} \leq p(K-1) + q.$$

The degree of the AR operator $|A(L)|$ is at most Kp because the determinant is just a sum of products involving one operator from each row and each column of $A(L)$. This proves the corollary. ■

The corollary gives upper bounds for the VARMA orders of a linearly transformed VARMA process. For instance, if y_t is a VAR(p)=VARMA($p,0$) process, a linear transformation $z_t = Fy_t$ has a VARMA($\breve{p},\breve{q}$) representation with $\breve{p} \leq Kp$ and $\breve{q} \leq (K-1)p$. For some linear transformations, $\breve{q}$ will be zero. We will see in the following, however, that generally there are transformations for which the upper bounds for the orders are attained and a representation with lower orders does not exist. This result implies that a linear transformation of a finite order VAR(p) process may not admit a finite order VAR representation. Specifically, the subprocesses or marginal processes of a K-dimensional process y_t are obtained by using transformation matrices such as $F = [I_M : 0]$. Hence, a subprocess of a VAR(p) process may not have a finite order VAR but just a mixed VARMA representation.

For some transformations the result in Corollary 11.1.1 can, in fact, be tightened. Generally, tighter bounds for the VARMA orders are available if $M > 1$, as is seen in the following corollary.

Corollary 11.1.2
Let y_t be a K-dimensional, stable, invertible VARMA(p,q) process and let F be an $(M \times K)$ matrix of rank M. Then the process $z_t = Fy_t$ has a VARMA($\breve{p},\breve{q}$) representation with

$$\breve{p} \leq (K - M + 1)p$$

and

$$\breve{q} \leq (K - M)p + q.$$

■

Proof: We first consider the case where z_t is a subprocess of y_t consisting of the first M components. To treat this case, we denote the first M and last $K - M$ components of the process y_t by y_{1t} and y_{2t}, respectively, and we partition the VAR and MA operators as well as the white noise process u_t accordingly. Thus, we can write the process as

$$A_{11}(L)y_{1t} + A_{12}(L)y_{2t} = M_{11}(L)u_{1t} + M_{12}(L)u_{2t}, \tag{11.6.4}$$

$$A_{21}(L)y_{1t} + A_{22}(L)y_{2t} = M_{21}(L)u_{1t} + M_{22}(L)u_{2t}. \tag{11.6.5}$$

Premultiplying (11.6.5) by the adjoint of $A_{22}(L)$ gives

$$\begin{aligned} |A_{22}(L)|y_{2t} &= -A_{22}(L)^{adj}A_{21}(L)y_{1t} + A_{22}(L)^{adj}M_{21}(L)u_{1t} \\ &\quad +A_{22}(L)^{adj}M_{22}(L)u_{2t}. \end{aligned} \tag{11.6.6}$$

Moreover, premultiplying (11.6.4) by $|A_{22}(L)|$, replacing $|A_{22}(L)|y_{2t}$ by the right-hand side of (11.6.6) and rearranging terms, we get

$$\begin{aligned}
&[|A_{22}(L)|A_{11}(L) - A_{12}(L)A_{22}(L)^{adj}A_{21}(L)]y_{1t} \\
&\qquad = [|A_{22}(L)|M_{11}(L) - A_{12}(L)A_{22}(L)^{adj}M_{21}(L)]u_{1t} \qquad (11.6.7)\\
&\qquad\quad +[|A_{22}(L)|M_{12}(L) - A_{12}(L)A_{22}(L)^{adj}M_{22}(L)]u_{2t}.
\end{aligned}$$

The VAR part of this representation has order

$$\breve{p} \le \max\{(K-M)p + p, (K-M-1)p + p + p\} = (K-M+1)p$$

and, by Proposition 11.1, the right-hand side of (11.6.7) has an MA representation with order

$$\breve{q} \le \max\{(K-M)p + q, p + (K-M-1)p + q\} = (K-M)p + q.$$

Hence, we have established the corollary for transformations $F = [I_M : 0]$.

For a general $(M \times K)$ transformation matrix F with $\text{rk}(F) = M$, we choose a $((K-M) \times K)$ matrix C such that the $(K \times K)$ matrix

$$\mathfrak{F} = \begin{bmatrix} F \\ C \end{bmatrix}$$

is nonsingular and we consider the process $x_t = \mathfrak{F} y_t$. Because nonsingular transformations do not increase the orders of a VARMA process, x_t also has a VARMA(p,q) representation. Now we get the result of the corollary by considering the transformation $z_t = F y_t = [I_M : 0]x_t$. ■

Other bounds for the VARMA orders than those provided in Corollaries 11.1.1 and 11.1.2 for linearly transformed VARMA processes and bounds for special linear transformations are given in various articles in the literature. For further results and references see Lütkepohl (1987, Chapter 4; 1986, Kapitel 2).

To illustrate Corollaries 11.1.1 and 11.1.2, we consider the bivariate VAR(1) process

$$\begin{bmatrix} 1-0.5L & 0.66L \\ 0.5L & 1+0.3L \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \quad \text{with } \Sigma_u = I_2. \qquad (11.6.8)$$

Here $K = 2$, $p = 1$, and $q = 0$. Thus, $z_t = [1,0]y_t = y_{1t}$ as a univariate $(M = 1)$ marginal process has an ARMA representation with orders not greater than $(2,1)$. The precise form of the process can be determined with the help of the representation (11.6.3). Using that representation gives

$$\begin{aligned}
&[(1+0.3L)(1-0.5L) - 0.66 \cdot 0.5L^2]z_t \\
&\quad = [1,0]\begin{bmatrix} 1+0.3L & -0.66L \\ -0.5L & 1-0.5L \end{bmatrix}\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \qquad (11.6.9)\\
&\quad = (1+0.3L)u_{1t} - 0.66Lu_{2t}.
\end{aligned}$$

The right-hand side, say w_{1t}, is the sum of an MA(1) process and a white noise process. Thus, by Proposition 11.1, it is known to have an MA(1) representation, say $w_{1t} = v_{1t} + \gamma v_{1,t-1}$. To determine γ and $\sigma_1^2 = \text{Var}(v_{1t})$, we use

$$\begin{aligned} E(w_{1t}^2) &= E(v_{1t} + \gamma v_{1,t-1})^2 = (1+\gamma^2)\sigma_1^2 \\ &= E[(1+0.3L)u_{1t} - 0.66Lu_{2t}]^2 = 1.53 \end{aligned}$$

and

$$\begin{aligned} E(w_t w_{t-1}) &= E[(v_{1t} + \gamma v_{1,t-1})(v_{1,t-1} + \gamma v_{1,t-2})] = \gamma\sigma_1^2 \\ &= E[((1+0.3L)u_{1t} - 0.66u_{2,t-1}) \\ &\qquad \times((1+0.3L)u_{1,t-1} - 0.66u_{2,t-2})] \\ &= 0.3. \end{aligned}$$

Solving this nonlinear system of two equations for γ and σ_1^2 gives

$$\gamma = 0.204 \quad \text{and} \quad \sigma_1^2 = 1.47.$$

Note that we have picked the invertible solution with $|\gamma| < 1$. Thus, from (11.6.9), we get a marginal process

$$(1 - 0.2L - 0.48L^2)y_{1t} = (1 + 0.204L)v_{1t} \quad \text{with} \quad \sigma_1^2 = 1.47.$$

In other words, y_{1t} has indeed an ARMA(2, 1) representation and it is easy to check that cancellation of the AR and MA operators is not possible. Hence, the ARMA orders are minimal in this case.

As another example, consider again the bivariate VAR(1) process (11.6.8) and suppose we are interested in the process $z_t := y_{1t}+y_{2t}$. Thus, $F = [1, 1]$ is again a (1×2) vector. Multiplying (11.6.8) by the adjoint of the VAR operator gives

$$(1 - 0.2L - 0.48L^2)\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1+0.3L & -0.66L \\ -0.5L & 1-0.5L \end{bmatrix}\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

Hence, multiplying by F gives

$$(1 - 0.2L - 0.48L^2)(y_{1t} + y_{2t}) = (1 - 0.2L)u_{1t} + (1 - 1.16L)u_{2t}.$$

Using similar arguments as for (11.6.9), it can be shown that the right-hand side of this expression is a process with MA(1) representation $v_t - 0.504v_{t-1}$, where $\sigma_v^2 := \text{Var}(v_t) = 2.70$. Consequently, the process of interest has the ARMA(2, 1) representation

$$(1 - 0.2L - 0.48L^2)z_t = (1 - 0.504L)v_t \quad \text{with} \quad \sigma_v^2 = 2.70. \qquad (11.6.10)$$

The following result is of interest if forecasting is the objective of the analysis.

Proposition 11.2 (*Forecast Efficiency of Linearly Transformed VARMA Processes*)
Let y_t be a stable, invertible, K-dimensional VARMA(p,q) process, let F be an $(M \times K)$ matrix of rank M, and let $z_t = Fy_t$. Furthermore, denote the MSE matrices of the optimal h-step predictors of y_t and z_t by $\Sigma_y(h)$ and $\Sigma_z(h)$, respectively. Then

$$\Sigma_z(h) - F\Sigma_y(h)F'$$

is positive semidefinite. ■

This result means that $Fy_t(h)$ is generally a better predictor of z_{t+h} with smaller (at least not greater) MSEs than $z_t(h)$. In other words, forecasting the original process y_t and transforming the forecasts is generally better than forecasting the transformed process directly. A proof and references for related results were given by Lütkepohl (1987, Chapter 4). To see the point more clearly, consider again the example process (11.6.8) and suppose we are interested in the sum of its components $z_t = y_{1t} + y_{2t}$. Forecasting the bivariate process one step ahead results in a forecast MSE matrix $\Sigma_y(1) = \Sigma_u = I_2$. Thus, the corresponding 1-step ahead forecast of z_t has MSE

$$[1,1]\Sigma_y(1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2.$$

In contrast, if a univariate forecast is obtained on the basis of the ARMA$(2,1)$ representation (11.6.10), the 1-step ahead forecast MSE becomes $\sigma_v^2 = 2.70$. Clearly, the latter forecast is inferior in terms of MSE.

Of course, these results hold for VARMA processes for which all the parameters are known. They do not necessarily carry over to estimated processes, a case which was also investigated and reviewed by Lütkepohl (1987).

11.6.2 Aggregation of VARMA Processes

There is little to be added to the foregoing results for the case of contemporaneous aggregation. Suppose $y_t = (y_{1t}, \dots, y_{Kt})'$ consists of K variables. If all or some of them are contemporaneously aggregated by taking their sum or average, this just means that y_t is transformed linearly and the foregoing results apply directly. In particular, the aggregated process has a finite order VARMA representation if the original process does. Moreover, if forecasts for the aggregated variables are desired it is generally preferable to forecast the disaggregated process and aggregate the forecasts rather than forecast the aggregated process directly.

The foregoing results are also helpful in studying the consequences of temporal aggregation. Suppose we wish to aggregate the variables y_t generated by

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t + M_1 u_{t-1}$$

over, say, $m = 3$ subsequent periods. To be able to use the previous framework, we construct a process

$$\begin{bmatrix} I_K & 0 & 0 \\ -A_1 & I_K & 0 \\ -A_2 & -A_1 & I_K \end{bmatrix} \begin{bmatrix} y_{m(\tau-1)+1} \\ y_{m(\tau-1)+2} \\ y_{m\tau} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & A_2 & A_1 \\ 0 & 0 & A_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{m(\tau-2)+1} \\ y_{m(\tau-2)+2} \\ y_{m(\tau-1)} \end{bmatrix} + \begin{bmatrix} I_K & 0 & 0 \\ M_1 & I_K & 0 \\ 0 & M_1 & I_K \end{bmatrix} \begin{bmatrix} u_{m(\tau-1)+1} \\ u_{m(\tau-1)+2} \\ u_{m\tau} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & M_1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_{m(\tau-2)+1} \\ u_{m(\tau-2)+2} \\ u_{m(\tau-1)} \end{bmatrix}.$$

Defining

$$\mathfrak{y}_\tau := \begin{bmatrix} y_{m(\tau-1)+1} \\ y_{m(\tau-1)+2} \\ y_{m\tau} \end{bmatrix} \quad \text{and} \quad \mathfrak{u}_\tau := \begin{bmatrix} u_{m(\tau-1)+1} \\ u_{m(\tau-1)+2} \\ u_{m\tau} \end{bmatrix},$$

we get

$$\mathfrak{A}_0 \mathfrak{y}_\tau = \mathfrak{A}_1 \mathfrak{y}_{\tau-1} + \mathfrak{M}_0 \mathfrak{u}_\tau + \mathfrak{M}_1 \mathfrak{u}_{\tau-1}, \tag{11.6.11}$$

where $\mathfrak{A}_0$, $\mathfrak{A}_1$, $\mathfrak{M}_0$, and $\mathfrak{M}_1$ have the obvious definitions. This form is a VARMA$(1,1)$ representation of the $3K$-dimensional process $\mathfrak{y}_\tau$. Our standard form of a VARMA$(1,1)$ process can be obtained from this form by premultiplying with $\mathfrak{A}_0^{-1}$ and defining $\mathfrak{v}_\tau = \mathfrak{A}_0^{-1}\mathfrak{M}_0\mathfrak{u}_\tau$ which gives

$$\mathfrak{y}_\tau = \mathfrak{A}_0^{-1}\mathfrak{A}_1\mathfrak{y}_{\tau-1} + \mathfrak{v}_\tau + \mathfrak{A}_0^{-1}\mathfrak{M}_1\mathfrak{M}_0^{-1}\mathfrak{A}_0\mathfrak{v}_{\tau-1}.$$

Now temporal aggregation over $m = 3$ periods can be represented as a linear transformation of the process $\mathfrak{y}_\tau$. Clearly, it is not difficult to see that this method generalizes for higher order processes and temporal aggregation over more than three periods. Moreover, different types of temporal aggregation can be handled. For instance, the aggregate may be the sum of subsequent values or it may be their average. Furthermore, temporal and contemporaneous aggregation can be dealt with simultaneously. In all of these cases, the aggregate has a VARMA representation if the original variables are generated by a finite order VARMA process and its structure can be studied using the foregoing framework. Moreover, by Proposition 11.2, if forecasts of the aggregate are of interest, it is in general preferable to forecast the original disaggregated process and aggregate the forecasts rather than forecast the aggregate directly. A detailed discussion of these issues and also of forecasting with estimated processes can be found in Lütkepohl (1987).

11.7 Interpretation of VARMA Models

The same tools and concepts that we have used for interpreting VAR models may also be applied in the VARMA case. We will consider Granger-causality and impulse response analysis in turn.

11.7.1 Granger-Causality

To study Granger-causality in the context of VARMA processes, we partition y_t in two groups of variables, z_t and x_t, and we partition the VAR and MA operators as well as the white noise process u_t accordingly. Hence, we get

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} M_{11}(L) & M_{12}(L) \\ M_{21}(L) & M_{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \qquad (11.7.1)$$

where again a zero mean is assumed for simplicity and without loss of generality. The results derived in the following are not affected by a nonzero mean term. The process (11.7.1) is assumed to be stable and invertible and its pure, canonical MA representation is

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

From Proposition 2.2, we know that x_t is not Granger-causal for z_t if and only if $\Phi_{12}(L) \equiv 0$. Although the proposition is stated for VAR processes, it is easy to see that it remains correct for the presently considered VARMA case. We also know that

$$\begin{aligned} &\begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \\ &\quad = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix}^{-1} \begin{bmatrix} M_{11}(L) & M_{12}(L) \\ M_{21}(L) & M_{22}(L) \end{bmatrix} \\ &\quad = \begin{bmatrix} D(L) & -D(L)A_{12}(L)A_{22}(L)^{-1} \\ -A_{22}(L)^{-1}A_{21}(L)D(L) & A_{22}(L)^{-1} + A_{22}(L)^{-1}A_{21}(L)D(L)A_{12}(L)A_{22}(L)^{-1} \end{bmatrix} \\ &\qquad \times \begin{bmatrix} M_{11}(L) & M_{12}(L) \\ M_{21}(L) & M_{22}(L) \end{bmatrix}, \end{aligned}$$

where

$$D(L) := [A_{11}(L) - A_{12}(L)A_{22}(L)^{-1}A_{21}(L)]^{-1}$$

and the rules for the partitioned inverse have been used (see Appendix A.10). Consequently, x_t is not Granger-causal for z_t if and only if

$$0 \equiv D(L)M_{12}(L) - D(L)A_{12}(L)A_{22}(L)^{-1}M_{22}(L)$$

or, equivalently,

$$M_{12}(L) - A_{12}(L)A_{22}(L)^{-1}M_{22}(L) \equiv 0.$$

Moreover, it follows as in Proposition 2.3 that there is no instantaneous causality between x_t and z_t if and only if $E(u_{1t}u'_{2t}) = 0$. We state these results as a proposition.

Proposition 11.3 (*Characterization of Noncausality*)
Let

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix}$$

be a stable and invertible VARMA(p,q) process as in (11.7.1) with possibly nonzero mean. Then x_t is not Granger-causal for z_t if and only if

$$M_{12}(L) \equiv A_{12}(L)A_{22}(L)^{-1}M_{22}(L). \qquad (11.7.2)$$

There is no instantaneous causality between z_t and x_t if and only if

$$E(u_{1t}u'_{2t}) = 0.$$

■

Remark 1 Obviously, the restrictions characterizing Granger-noncausality are not quite so easy here as in the VAR(p) case. Consider, for instance, a bivariate VARMA$(1,1)$ process

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} + \begin{bmatrix} m_{11,1} & m_{12,1} \\ m_{21,1} & m_{22,1} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}.$$

For this process, the restrictions (11.7.2) reduce to

$$m_{12,1}L = (-\alpha_{12,1}L)(1 - \alpha_{22,1}L)^{-1}(1 + m_{22,1}L)$$

or

$$(1 - \alpha_{22,1}L)m_{12,1}L = -(1 + m_{22,1}L)\alpha_{12,1}L$$

or

$$m_{12,1} = -\alpha_{12,1} \qquad \text{and} \qquad \alpha_{22,1}m_{12,1} = \alpha_{12,1}m_{22,1}.$$

This, of course, is a set of nonlinear restrictions whereas only linear constraints were required to characterize Granger-noncausality in the corresponding pure VAR(p) case. However, a sufficient condition for (11.7.2) to hold is

$$M_{12}(L) \equiv A_{12}(L) \equiv 0, \qquad (11.7.3)$$

which is again a set of linear constraints. Occasionally, these sufficient conditions may be easier to test than (11.7.2). ■

Remark 2 To turn the arguments put forward prior to Proposition 11.3 into a formal proof requires that we convince ourselves that all the operations performed with the matrices of lag polynomials are feasible and correct. Because we have not proven these results, the arguments should just be taken as an indication of how a proof may proceed. ■

11.7.2 Impulse Response Analysis

The impulse responses and forecast error variance decompositions of a VARMA model are obtained from its pure MA representation, as in the finite order VAR case. Thus, the discussion of Sections 2.3.2 and 2.3.3 carries over to the present case, except that the Φ_i's are computed with different formulas. Also, Propositions 2.4 and 2.5 need modification. We will not give the details here but refer the reader to the exercises (see Problem 11.9).

It may be worth reiterating some caveats of impulse response analysis which may be more apparent now after the discussion of transformations in Section 11.6. In particular, we have seen there that dropping variables (considering subprocesses) or aggregating the components of a VARMA process temporally and/or contemporaneously results in possibly quite different VARMA structures. They will in general have quite different coefficients in their pure MA representations. In other words, the impulse responses may change drastically if important variables are excluded from a system or if the level of aggregation is altered, for instance, if quarterly instead of monthly data are considered. Again, this does not necessarily render impulse response analysis useless. It should caution the reader against over interpreting the evidence from VARMA models, though. Some thought must be given to the choice of variables, the level of aggregation, and other transformations of the variables.

11.8 Exercises

Problem 11.1
Write the MA(1) process $y_t = u_t + M_1 u_{t-1}$ in VAR(1) form, $Y_t = \mathbf{A}Y_{t-1} + U_t$, and determine $\mathbf{A}^i$ for $i = 1, 2$.

Problem 11.2
Suppose $y_t = A_1 y_{t-1} + u_t + M_1 u_{t-1} + M_2 u_{t-2}$ is a stable and invertible VARMA(1, 2) process. Determine the coefficient matrices Π_i, $i = 1, 2, 3, 4$, of its pure VAR representation and the coefficient matrices Φ_i, $i = 1, 2, 3, 4$, of its pure MA representation.

Problem 11.3
Evaluate the autocovariances $\Gamma_y(h)$, $h = 1, 2, 3$, of the bivariate VARMA(2, 1) process

$$y_t = \begin{bmatrix} .3 \\ .5 \end{bmatrix} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} y_{t-2} + u_t + \begin{bmatrix} .6 & .2 \\ 0 & .3 \end{bmatrix} u_{t-1}. \quad (11.8.1)$$

(Hint: The use of a computer will greatly simplify this problem.)

Problem 11.4
Write the VARMA(1, 1) process $y_t = A_1 y_{t-1} + u_t + M_1 u_{t-1}$ in VAR(1) form, $Y_t = \mathbf{A} Y_{t-1} + U_t$. Determine forecasts $Y_t(h) = \mathbf{A}^h Y_t$ for $h = 1, 2, 3$, and compare them to forecasts obtained from the recursive formula (11.5.4).

Problem 11.5
Derive a univariate ARMA representation of the second component, y_{2t}, of the process given in (11.6.8).

Problem 11.6
Provide upper bounds for the ARMA orders of the process $z_t = y_{1t} + y_{2t} + y_{3t}$, where $y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})'$ is a 4-dimensional VARMA(3, 3) process.

Problem 11.7
Write the VARMA(1, 1) process y_t from Problem 11.4 in a form such as (11.6.11) that permits to analyze temporal aggregation over four periods in the framework of Section 11.6.2. Give upper bounds for the orders of a VARMA representation of the process obtained by temporally aggregating y_t over four periods.

Problem 11.8
Write down explicitly the restrictions characterizing Granger-noncausality for a bivariate VARMA(2, 1) process. Is y_{1t} Granger-causal for y_{2t} in the process (11.8.1)?

Problem 11.9
Generalize Propositions 2.4 and 2.5 to the VARMA(p, q) case.
(Hint: Show that for a K-dimensional VARMA(p, q) process,

$$\phi_{jk,i} = 0, \quad \text{for } i = 1, 2, \dots,$$

is equivalent to

$$\phi_{jk,i} = 0, \quad \text{for } i = 1, 2, \dots, p(K-1) + q;$$

and

$$\theta_{jk,i} = 0, \quad \text{for } i = 0, 1, 2, \dots,$$

is equivalent to

$$\theta_{jk,i} = 0, \quad \text{for } i = 0, 1, \dots, p(K-1) + q.)$$

Problem 11.10
Suppose that m is a real number with $|m| > 1$ and u_t is a white noise process. Show that the process

$$v_t = \left(1 + \frac{1}{m}L\right)^{-1} (1 + mL)u_t$$

is also white noise with $\text{Var}(v_t) = m^2\text{Var}(u_t)$.