Chapter 6

Multivariate time series

6.1 Introduction

Usually, there are *several* time series of interest running *concurrently*, which influence each other. The collection is a *multivariate time series*. Each univariate time series is referred to as a *component*.

6.2 Weak Stationarity and Cross-Correlation

We denote an m-variate time series $\{\underline{X}_t : t \in \mathbb{Z}\}$ by:

$$\underline{X}_t := \begin{pmatrix} X_{t1} \\ \vdots \\ X_{tm} \end{pmatrix}, \quad t \in \mathbb{Z},$$

where each component $\{(X_{tj})_{t\in\mathbb{Z}}\}$ is a univariate time series. The mean and covariance of $\{\underline{X}_t\}$ are specified by:

$$\underline{\mu}_t := \mathbb{E}\left[\underline{X}_t\right] = \begin{pmatrix} \mu_{t1} \\ \vdots \\ \mu_{tm} \end{pmatrix} = \begin{pmatrix} \mathbb{E}\left[X_{t1}\right] \\ \vdots \\ \mathbb{E}\left[X_{tm}\right] \end{pmatrix}, \quad t \in \mathbb{Z},$$

and the covariance matrices $\Gamma(t+h,t)$, where

$$\Gamma_{jk}(t+h,t) = \mathbb{E}\left[(X_{t+h,j} - \mu_{t+h,j})(X_{t,k} - \mu_{t,k})\right] = \mathrm{Cov}(X_{t+h,j}, X_{t,k}).$$

Definition 6.1. *The m-variate time series* $\{\underline{X}_t,\ t \in \mathbb{Z}\}$ *is said to be* (weakly) stationary *if*

1. $\underline{\mu}_t = \underline{\mu}$ *for all* $t \in \mathbb{Z}$,
2. $\Gamma(r,s) = \Gamma(r+t, s+t)$ *for all* $r, s, t \in \mathbb{Z}$.

The second condition implies that $\Gamma(r,s)$ is a function of $r-s$ and therefore, as with univariate time series, the ACVF is defined as:

$$\Gamma(h) := \Gamma(h, 0).$$

Definition 6.2 (Correlation Matrix Function)**.** *For a stationary multivariate time series, the* correlation matrix function R *is defined as:*

$$R_{ij}(h) = \frac{\Gamma_{ij}(h)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}}.$$

Lemma 6.3. *The multivariate ACVF has the following properties:*

1. $\Gamma(h) = \Gamma^t(-h)$,
2. $|\Gamma_{ij}(h)| \leq \sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}$.
3. $\Gamma_{ii}(\cdot)$ *is a univariate ACVF,*
4. $\sum_{i,j=1}^n \underline{a}_i^t \Gamma(i-j)\underline{a}_j \geq 0$ *for all* n *and* $\underline{a}_1, \ldots, \underline{a}_n \in \mathbb{R}^n$.

The correlation matrix function satisfies:

1. $R(h) = R^t(-h)$,
2. $|R_{ij}(h)| \leq 1$
3. $R_{ii}(\cdot)$ *is a univariate ACF (autocorrelation function),*
4. $\sum_{i,j=1}^n \underline{a}_i^t R(i-j)\underline{a}_j \geq 0$ *for all* n *and* $\underline{a}_1, \ldots, \underline{a}_n \in \mathbb{R}^n$.
5. $R_{ii}(0) = 1$.

The first of these implies that the multivariate ACVF satisfies: $\Gamma_{ij}(h) = \Gamma_{ji}(-h)$.

Proof For the ACVF: The first property follows immediately from the definition. The second follows directly from the Cauchy Schwartz inequality. The third follows from the observation that Γ_{ii} is the ACVF of the stationary series $\{X_{ti} : t \in \mathbb{Z}\}$. The fourth follows from the fact that

$$\sum_{i,j=1}^n \underline{a}_i^t \Gamma(i-j)\underline{a}_j = \sum_{pqij} a_{pi} a_{qj} \Gamma_{pq}(i-j) = \mathbb{E}\left[\left(\sum_p \sum_{j=1}^n a_{pj}\left(X_{pj} - \mu_j\right)\right)^2\right].$$

For the correlation matrix function, properties 1,2,3 and 5 are obvious. For property 4,

$$\sum_{i,j=1}^n \underline{a}_i^t R(i-j)\underline{a}_j = \sum_{i,j=1}^n \sum_{pq} a_{ip} a_{jq} R_{pq}(i-j) = \mathbb{E}\left[\left(\sum_p \sum_{j=1}^n \frac{a_{pj}}{\sqrt{\Gamma_{pp}(0)}}(X_{pj} - \mu_j)\right)^2\right]$$

from which the property follows. □

Example 6.1.

Consider the following bivariate stationary process $\{\underline{X}_t\}$ defined by:

$$\begin{cases} X_{t1} = Z_t \\ X_{t2} = Z_t + 0.75 Z_{t-10} \quad \{Z_t\} \sim \text{WN}(0,1) \end{cases}$$

VMA stands for 'vector moving average'.

Then $\underline{\mu} = \mathbb{E}[\underline{X}_t] = \underline{0}$,

$$\Gamma(-10) = \begin{pmatrix} 0 & 0.75 \\ 0 & 0.75 \end{pmatrix}, \qquad \Gamma(0) = \begin{pmatrix} 1 & 1 \\ 1 & 1.5625 \end{pmatrix} \qquad \Gamma(10) = \begin{pmatrix} 0 & 0 \\ 0.75 & 0.75 \end{pmatrix}$$

$$\Gamma(h) = 0 \qquad h \neq 0, \pm 10.$$

Using $\sqrt{1.5625} = 1.25$, the correlation matrix function is therefore:

$$R(-10) = \begin{pmatrix} 0 & 0.6 \\ 0 & 0.48 \end{pmatrix} \qquad R(0) = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \qquad R(10) = \begin{pmatrix} 0 & 0 \\ 0.6 & 0.48 \end{pmatrix}$$

$$R(h) = 0 \qquad h \neq 0, \pm 10.$$

□

Definition 6.4 (Multivariate white noise)**.** *An m-variate process*

$$\{\underline{\epsilon}_t : t \in \mathbb{Z}\}$$

is said to be a white noise with covariance matrix Σ, written

$$\{\underline{\epsilon}_t\} \sim WN(\underline{0}, \Sigma),$$

if $\mathbb{E}[\underline{\epsilon}_t] = \underline{0}$ *and* $\Gamma(h) = \begin{cases} \Sigma & \text{if } h = 0, \\ 0 & \text{if } h \neq 0. \end{cases}$

The notation $\{\underline{\epsilon}_t\} \sim \text{IID}(\underline{0}, \Sigma)$ denotes that the random vectors $\{\underline{\epsilon}_t : t \in \mathbb{Z}\}$ are independent identically distributed with expectation vector $\underline{0}$ and covariance matrix Σ.

Definition 6.5 (Multivariate Linear Process)**.** *A stationary multivariate time series $\{\underline{X}_t : t \in \mathbb{Z}\}$ is a* linear process *if and only if*

$$\underline{X}_t = \sum_{j=-\infty}^{\infty} \Psi_j \underline{\epsilon}_{t-j} \qquad \{\epsilon_t\} \sim WN(\underline{0}, \Sigma)$$

where $(\Psi_j)_{j\in\mathbb{Z}}$ is a sequence of matrices satisfying $\sum_{j;p,q} |\Psi_{j;p,q}| < +\infty$.

Lemma 6.6. *The covariance matrix of the multivariate linear process of Definition 6.5 is given by:*

$$\Gamma(h) = \sum_{j=-\infty}^{\infty} \Psi_{j+h} \Sigma \Psi_j'$$

Proof Exercise. □

6.3 Estimating the Mean Vector and Covariance Matrix Function

We give the multivariate results, which are similar to the univariate setting; the proofs use the same techinques. The following proposition deals with estimating the mean vector.

Proposition 6.7. *Let $\{\underline{X}_t\}$ be a stationary multivariate time series defined by:*

$$\underline{X}_t = \underline{\mu} + \sum_{k=-\infty}^{\infty} \Psi_k \underline{\epsilon}_{t-k} \qquad \{\underline{Z}_t\} \sim IID(\underline{0}, \Sigma).$$

where the sequence $(\Psi_k)_{k\in\mathbb{Z}}$ satisfies $\sum_{k;p,q} |\Psi_{k;pq}| < +\infty$. Then

$$\overline{\underline{X}}_n := \frac{1}{n}\sum_{t=1}^{n} \underline{X}_t \sim AN\left(\underline{\mu}, \frac{1}{n}\left(\sum_{k=-\infty}^{\infty} \Psi_k\right) \Sigma \left(\sum_{k=-\infty}^{\infty} \Psi_k^t\right)\right)$$

Proof Omitted - the basic method is similar to the univariate setting. □

Definition 6.8 (Impulse Response Function)**.** *For a stationary linear time series, the operator Ψ is often referred to as the* impulse response function*. The matrix Ψ_k indicates the effect that the noise at lag k has on the process.*

Estimating the Covariance and Correlation Function As with the univariate case, the estimator of the covariance matrix function is:

$$\widehat{\Gamma}(h) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-h}\left(\underline{X}_{t+h} - \overline{\underline{X}}_n\right)\left(\underline{X}_t - \overline{\underline{X}}_n\right)^t & 0 \le h \le n-1 \\ \frac{1}{n}\sum_{t=-h+1}^{n}\left(\underline{X}_{t+h} - \overline{\underline{X}}_n\right)\left(\underline{X}_t - \overline{\underline{X}}_n\right)^t & -n+1 \le h < 0 \end{cases}$$

The estimator of the correlation matrix function is:

$$\widehat{R}_{ij}(h) = \frac{\widehat{\Gamma}_{ij}(h)}{\sqrt{\widehat{\Gamma}_{ii}(0)\widehat{\Gamma}_{jj}(0)}}.$$

These estimators converge in probability; for each i, j, h, $\widehat{\Gamma}_{ij}(h) \rightarrow_{(p)} \Gamma_{ij}(h)$ and $\widehat{R}_{ij}(h) \rightarrow_{(p)} R_{ij}(h)$.

In general, there are not so many asymptotic distributional results for $\widehat{\Gamma}$. The following result indicates the magnitude one should expect from a sample cross correlation if two processes are independent.

Theorem 6.9. *Let*

$$\begin{cases} X_{t1} = \sum_{j=-\infty}^{\infty} \alpha_j \epsilon_{t-j,1} & \{\epsilon_{t1}\} \sim IID(0, \sigma_1^2) \\ X_{j2} = \sum_{j=-\infty}^{\infty} \beta_j \epsilon_{t-j,2} & \{\epsilon_{t2}\} \sim IID(0, \sigma_2^2) \end{cases}$$

where the two sequences $\{\epsilon_{t1}\}$ and $\{\epsilon_{t2}\}$ are independent, $\sum_j |\alpha_j| < +\infty$ and $\sum_j |\beta_j| < +\infty$. Then, for $h \ge 0$,

$$\widehat{R}_{12}(h) \sim AN\left(0, \frac{1}{n}\sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j)\right).$$

If $h, k \geq 0$ and $h \neq k$, then

$$\begin{pmatrix} \widehat{R}_{12}(h) \\ \widehat{R}_{12}(k) \end{pmatrix} \sim AN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} \sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j) & \sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j+k-h) \\ \sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j+k-h) & \sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j) \end{pmatrix}\right).$$

Proof Firstly,

$$\widehat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}\underline{X}_{t+h}\underline{X}_t^t - \frac{1}{n}\overline{\underline{X}}_n\sum_{t=1}^{n-h}\underline{X}_t^t - \frac{1}{n}\sum_{t=1}^{n-h}\underline{X}_{t+h}\overline{\underline{X}}_n^t + \left(1-\frac{h}{n}\right)\overline{\underline{X}}_n\overline{\underline{X}}_n^t.$$

Using this, together with Proposition 6.7, that

$$n^{1/2}\left|\widehat{\Gamma}_{12}(h) - \Gamma_{12}^*(h)\right| \stackrel{n\to+\infty}{\longrightarrow}_p 0$$

where

$$\Gamma_{12}^*(h) = \frac{1}{n}\sum_{t=1}^{n}X_{t+h,1}X_{t,2} = \frac{1}{n}\sum_{t=1}^{n}\sum_i\sum_j\alpha_{i+h}\beta_j Z_{t-i,1}Z_{t-j,2}.$$

Since $\mathbb{E}\left[\Gamma_{12}^*(h)\right] = 0$, it follows that

$$n\mathrm{Var}\left(\Gamma_{12}^*(h)\right) = \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\sum_{i,j,k,l}\alpha_{i+h}\beta_j\alpha_{k+h}\beta_l\mathbb{E}\left[Z_{s-i,1}Z_{s-j,2}Z_{t-k,1}Z_{t-l,2}\right].$$

From the independence assumptions, it follows that:

$$\mathbb{E}\left[Z_{s-i,1}Z_{s-j,2}Z_{t-k,1}Z_{t-l,2}\right] = \begin{cases} \sigma_1^2\sigma_2^2 & s-i=t-k, \quad s-j=t-l \\ 0 & \text{otherwise} \end{cases}$$

It follows that

$$n\mathrm{Var}(\Gamma_{12}^*(h)) = \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\left(\sum_i\alpha_{i+h}\alpha_{t-s+i+h}\right)\sigma_1^2\left(\sum_j\beta_j\beta_{t-s+j}\right)\sigma_2^2 = \sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\Gamma_{11}(k)\Gamma_{22}(k).$$

It now follows by elementary analysis that

$$\lim_{n\to+\infty} n\mathrm{Var}\left(\Gamma_{12}^*(h)\right) = \sum_{j=-\infty}^{\infty}\Gamma_{11}(j)\Gamma_{22}(j).$$

To show that Γ_{12}^* is asymptotically normal, it is necessary to show that it can be approximated by sums of independent identically distributed variables, which have finite variance. For fixed m, consider the $(2m+h)$-dependent, strictly stationary time series

$$\left\{\sum_{|i|\leq m}\sum_{|j|\leq m}\alpha_i\beta_j Z_{t+h-i,1}Z_{t-j,2} : t \in \mathbb{Z}\right\}.$$

Then

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{|i|\leq m}\sum_{|j|\leq m}\alpha_i\beta_j Z_{t+h-i,1}Z_{t-j,2} \sim \mathrm{AN}\left(0, a_m\right)$$

$$a_m = \sum_{|k|\leq m}\left(\sum_{|i|\leq m}\alpha_i\alpha_{i+|k|}\right)\sigma_1^2\left(\sum_{|j|\leq m}\beta_j\beta_{j+|m|}\right)\sigma_2^2 \stackrel{m\to+\infty}{\longrightarrow}\sum_j\Gamma_{11}(j)\Gamma_{22}(j).$$

From this, the remaining convergence arguments are straightforward and hence

$$\Gamma_{12}^*(h) \sim \mathrm{AN}\left(0, \frac{1}{n}\sum_{k=-\infty}^{\infty}\Gamma_{11}(k)\Gamma_{22}(k)\right).$$

Since $\widehat{\Gamma}_{11}(0) \stackrel{n\to+\infty}{\longrightarrow}_{(p)} \Gamma_{11}(0)$ and $\widehat{\Gamma}_{22}(0) \stackrel{n\to+\infty}{\longrightarrow}_{(p)} \Gamma_{22}(0)$, it follows that

$$\widehat{R}_{12}(h) = \frac{\widehat{\Gamma}_{12}(h)}{\sqrt{\widehat{\Gamma}_{11}(0)\widehat{\Gamma}_{22}(0)}} \sim \mathrm{AN}\left(0, \frac{1}{n}\sum_{j=-\infty}^{\infty}\rho_{11}(j)\rho_{22}(j)\right).$$

Similarly,

$$n\mathrm{Cov}\left(\Gamma_{12}^*(h), \Gamma_{12}^*(k)\right) \stackrel{n\to+\infty}{\longrightarrow}\sum_{j=-\infty}^{\infty}\Gamma_{11}(j)\Gamma_{22}(j+k-h).$$

and a similar argument, using $(2m+h)$-dependent bivariates may be used to conclude the result. □

6.4 The Multivariate Portmanteau Tests

The univariate Ljung-Box statistic $Q(m)$ has been generalised to the multivariate setting by Hosking (1980) and Hosking (1981), also Li and McLeod (1981). Here it is stated without proof, but the paper by Hosking from 1980 is worth reading.

For a stationary multivariate series, where $\text{Var}(X_t) = \Sigma$, consider

$$H_0 : R(1) = \ldots = R(m) = 0$$

versus H_1, the alternative (at least one of $R(1), \ldots, R(m)$ is not zero). Here $R(h)$ is the correlation *matrix* for time lag h. For a k-variate time series, the test statistic is:

$$Q_k(m) = n^2 \sum_{l=1}^{m} \frac{1}{n-l} \text{tr}\left(\widehat{\Gamma}_l^t \widehat{\Gamma}_0^{-1} \widehat{\Gamma}_l \widehat{\Gamma}_0^{-1}\right)$$

where n is the sample size. Under the null hypothesis, asymptotically $Q_k(m) \sim \chi^2_{k^2 m}$. This test statistic may be written as:

$$Q_k(m) = n^2 \sum_{l=1}^{m} \underline{b}_l^t \left(\widehat{R}_0^{-1} \otimes \widehat{R}_0^{-1}\right) \underline{b}_l$$

where $\underline{b}_l = \text{vec}(\widehat{R}_l^t)$. The test statistic proposed by Li and McLeod (1981) is:

$$Q_k(m) = n^2 \sum_{l=1}^{m} \underline{b}_l^t \left(\widehat{R}_0^{-1} \otimes \widehat{R}_0^{-1}\right) \underline{b}_l + \frac{k^2 m(m+1)}{2n},$$

which is asymptotically equivalent.

6.5 The VARMA (vector ARMA) Process

Definition 6.10 (The VARMA(p,q) process)**.** *The process $\{\underline{X}_t,\ t \in \mathbb{Z}\}$ is a mean* 0 *m-VARMA(p,q)* process *if $\underline{X}$ is an m-vector and it is a stationary solution of the difference equations*

$$\underline{X}_t - \Phi_1 \underline{X}_{t-1} - \ldots - \Phi_p \underline{X}_{t-p} = \underline{\epsilon}_t + \Theta_1 \underline{\epsilon}_{t-1} + \ldots + \Theta_q \underline{\epsilon}_{t-q}, \tag{6.1}$$

where $\{\underline{\epsilon}_t\} \sim WN(\underline{0}, \Sigma)$ and this representation cannot be reduced further.

A process $\{\underline{X}_t\}$ is an m-VARMA(p,q) process with mean $\underline{\mu}$ if $\{\underline{X}_t - \underline{\mu}\}$ is an m-VARMA(p,q) process.

Equations (6.1) can be written as

$$\Phi(B)\underline{X}_t = \Theta(B)\underline{\epsilon}_t, \quad t \in \mathbb{Z},$$

where

$$\Phi(z) = I - z\Phi_1 - \ldots - z^p \Phi_p,$$

$$\Theta(z) = I + z\Theta_1 + \ldots + z^q\Theta_q,$$

are matrix-valued polynomials.

Example 6.2 (VAR(1) (vector AR(1)) Process)**.**

The VAR(1) process satisfies

$$\underline{X}_t = \Phi \underline{X}_{t-1} + \underline{\epsilon}_t, \qquad \{\underline{\epsilon}_t\} \sim \mathrm{WN}(\underline{0}, \Sigma).$$

This process may be expressed as

$$\underline{X}_t = \sum_{j=0}^{\infty} \Phi^j \underline{\epsilon}_{t-j}$$

provided all the eigenvalues of Φ have absolute value less than 1. This is equivalent to:

$$\det(I - z\Phi) \neq 0 \qquad \forall \{z \in \mathbb{C} | |z| \leq 1\}.$$

□

In the multivariate setting, causality and invertibility are defined as follows:

Definition 6.11 (Multivariate Causality and Invertibility)**.** *A stationary multivariate time series* $\underline{X}_t$ *is said to be* causal *if it can be expressed as:*

$$\underline{X}_t = \sum_{j=0}^{\infty} \Psi_j \underline{\epsilon}_{t-j}$$

where $\sum_{j,k} |\Psi_{j,k}| < +\infty$ *and is said to be* invertible *if* $\underline{\epsilon}_t$ *can be expressed as:*

$$\underline{\epsilon}_t = \sum_{j=0}^{\infty} \Pi_j \underline{X}_{t-j}$$

where $\sum_{j,k} |\Pi_{jk}| < +\infty$.

For the VARMA process, causality and invertibility are characterised as follows:

Theorem 6.12 (Causality for VARMA(p,q))**.** *Let* $\underline{X}_t$ *satisfy Equation* (6.1)*. If*

$$det\Phi(z) \neq 0 \qquad \forall \{z \in \mathbb{C} : |z| \leq 1\} \tag{6.2}$$

then $\underline{X}_t$ *has exactly one stationary solution*

$$\underline{X}_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j} \tag{6.3}$$

where the matrices Ψ_j *are determined uniquely by:*

$$\Psi(z) := \sum_{j=0}^{\infty} z^j \Psi_j = \Phi^{-1}(z)\Theta(z). \tag{6.4}$$

Proof The condition given by (6.2) implies that

$$\exists \epsilon > 0 \quad : \quad \Phi^{-1}(z) \quad \text{well defined} \quad \forall |z| < 1+\epsilon.$$

Each element of $\Phi^{-1}(z)$ is a rational function of z with no singularities in $\{|z| < 1+\epsilon\}$. It follows that Φ^{-1} may be expanded as a power series:

$$\Phi^{-1}(z) = \sum_{j=0}^{\infty} z^j A_j = A(z).$$

It follows that for each m, n,

$$\lim_{j\to+\infty} |A_{j;m,n}| \left(1+\frac{\epsilon}{2}\right)^j = 0$$

and hence that there is a $K \in (0, +\infty)$ such that

$$\max_{j;m,n} \left(1+\frac{\epsilon}{2}\right)^j |A_{j;m,n}| \le K.$$

This implies that $\sum_{j;m,n} |A_{j;m,n}| < +\infty$. Furthermore,

$$A(z)\Phi(z) = I \qquad \forall |z| \le 1.$$

It now follows directly that

$$\underline{X}_t = A(B)\Theta(B)\underline{\epsilon}_t$$

which is the required representation, with $\Psi(B) = A(B)\Theta(B)$.

Conversely, if $\underline{X}_t = \sum_{j=0}^{\infty} \Psi_j \underline{\epsilon}_{t-j}$ with Ψ defined by Equation (6.4), then

$$\Phi(B)\underline{X}_t = \Phi(B)\Psi(B)\underline{Z}_t = \Theta(B)\underline{\epsilon}_t$$

and hence $\{\Psi(B)\underline{\epsilon}_t\}$ is a stationary solution of Equation (6.1).

It follows that if $\det(\Phi(z)) \neq 0$ for $|z| < 1$, then the unique stationary solution of equaiton of equation (6.1) is the causal solution (6.3). □

The criteria and result for invertible VARMA are similar:

Theorem 6.13 (Invertibility for VARMA(p,q))**.** *Let* $\underline{X}_t$ *satisfy Equation* (6.1) *and suppose that:*

$$\det(\Theta(z)) \neq 0 \qquad \forall z \in \mathbb{C} \qquad \text{such that} \qquad |z| < 1.$$

Then

$$\underline{\epsilon}_t = \sum_{j=0}^{\infty} \Pi_j \underline{X}_{t-j} \tag{6.5}$$

where the matrices Π_j *are determined uniquely by:*

$$\Pi(z) = \sum_{j=0}^{\infty} z^j \Pi_j = \Theta^{-1}(z)\Phi(z) \qquad |z| \le 1. \tag{6.6}$$

Proof Exercise. The proof follows in the same way as the proof for causality. □

Example 6.3 (VARMA(1,1)).

Let

$$\Phi = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{pmatrix}$$

and

$$\Theta = \Phi^t = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{pmatrix}.$$

Let $\underline{X}_t$ solve:

$$\underline{X}_t - \Phi \underline{X}_{t-1} = \underline{\epsilon}_t + \Theta \underline{\epsilon}_{t-1}.$$

Then, using $\Phi(z) = (I - z\Phi)$ and $\Theta(z) = (I + z\Theta)$,

$$\Phi(z) = \begin{pmatrix} 1-0.5z & -0.5z \\ 0 & 1-0.5z \end{pmatrix} \Rightarrow \Phi(z)^{-1} = \frac{1}{(1-0.5z)^2} \begin{pmatrix} 1-0.5z & 0 \\ 0.5z & 1-0.5z \end{pmatrix}$$

so that

$$\begin{aligned} \Psi(z) &= \Phi(z)^{-1}\Theta(z) = \frac{1}{(1-0.5z)^2} \begin{pmatrix} 1-0.5z & 0 \\ 0.5z & 1-0.5z \end{pmatrix} \begin{pmatrix} 1+0.5z & 0 \\ 0.5z & 1+0.5z \end{pmatrix} \\ &= \frac{1}{(1-0.5z)^2} \begin{pmatrix} 1 & 0.5z(1+0.5z) \\ 0.5z(1-0.5z) & 1-0.25z^2 \end{pmatrix}. \end{aligned}$$

To compute Ψ_j:

$$\Psi(z) = \left(\sum_{j=0}^{\infty} \frac{(j+1)z^j}{2^j} \right) \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + z \begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix} + z^2 \begin{pmatrix} 0 & 0.25 \\ -0.25 & -0.25 \end{pmatrix} \right\}$$

giving

$$\Psi_j = \frac{1}{2^j} \begin{pmatrix} j+1 & 2j-1 \\ 1 & 2 \end{pmatrix} \qquad j = 1, 2, \ldots$$

□

The Covariance Matrix Function of a Causal VARMA Process From the representation of Equation (6.4), the covariance matrix $\Gamma(h) := \mathbb{E}\left[\underline{X}_{t+h}\underline{X}_t^t\right]$ of the causal process has representation:

$$\Gamma(h) = \sum_{k=0}^{\infty} \Psi_{h+k} \Sigma \Psi_k^t \qquad h \in \mathbb{Z}$$

If $\{\underline{X}_t\}$ is a causal VARMA(p,q) process, then the covariance has exponential decay; there is an $\alpha \in (0, 1)$ and a constant K such that

$$|\Gamma_{ij}(h)| \leq K\epsilon^{|h|} \qquad \forall(i,j,h). \tag{6.7}$$

Equations for the acvf (in terms of Φ) can be computed from:

$$(\Phi(B)\underline{X}_t)\underline{X}'_{t-h} = (\Theta(B)\underline{\epsilon}_t)\underline{X}'_{t-h}$$

which give (plugging in $\underline{X}'_{t-h} = \sum_{j=0}^{\infty} \underline{\epsilon}'_{t-h-j}\Psi'_j$:

$$\Gamma(h) - \sum_{r=1}^{p} \Phi_r\Gamma(h-r) = \sum_{h\leq r\leq q} \Theta_r\Sigma\Psi'_{r-h} \tag{6.8}$$

For $h \geq q+1$, we have:

$$\Gamma(h) - \sum_{r=1}^{p} \Phi_r\Gamma(h-r) = 0$$

from which the bounds of (6.7) follow.

6.5.1 Yule-Walker Estimation

Note that Yule-Walker estimation can be carried out for VAR(p) processes. The equations are:

$$\begin{cases} \Gamma(h) - \sum_{r=1}^{p} \Phi_r\Gamma(h-r) = 0 & h \geq 1 \\ \Gamma(0) - \sum_{r=1}^{p} \Phi_r\Gamma(-r) = \Sigma \end{cases}$$

As for the univariate setting, the estimates $\widehat{\Gamma}(1), \ldots, \widehat{\Gamma}(p)$ and equations for $h = 1, \ldots, p$ are used to estimate $\Phi_1, \ldots, \Phi_p$, while the equation for $h = 0$ is then used to estimate Σ.

6.6 Reduced and Structural Forms for the VAR(p) Model

The form

$$\Phi(B)\underline{X}_t = \underline{\alpha} + \underline{\epsilon}_t \qquad \underline{\epsilon}_t \sim \text{WN}(\underline{0}, \Sigma)$$

for a VAR(p) process is referred to in the economics literature as the *reduced* form of the VAR(p) model. (Note that $\mu = \mathbb{E}[X_t]$ satisfies $\phi(B)\underline{\mu} = \underline{\alpha}$). It is so called because it does not show explicitly the concurrent dependence between the component series; the matrix Σ may have $\Sigma_{ij} \neq 0$ for some $i \neq j$.

The *structural equation* is the form of the equation provides a transformation such that the covariance matrix of the innovations in the structural form is a diagonal matrix. It is obtained by constructing a *lower* triangular matrix L where $L_{jj} = 1$ for each j such that

$$L^{-1}\Sigma(L^{-1})^t = D$$

where D is diagonal. Let $\underline{\widetilde{\epsilon}}_t = L^{-1}\underline{\epsilon}_t$ so that

$$\mathrm{Cov}(\underline{\widetilde{\epsilon}}_t) = D$$

and hence $\underline{\widetilde{\epsilon}}_t \sim \mathrm{WN}(\underline{0}, D)$. Then

$$L^{-1}\Phi(B)\underline{X}_t = L^{-1}\underline{\alpha} + \underline{\widetilde{\epsilon}}_t.$$

This is known as the *structural form* of the equation.

6.7 Granger Causality

For a k-variate time series, process $X_i(t)$ is said to *Granger cause* (or G-cause) $X_j(t)$ if the lagged observations of $X_i(t-s) : s \geq 1$ help to predict $X_j(t)$ in the sense that the null hypothesis H_0 : $\Phi_{ji}(s) = 0$ for all $s \geq 1$ is rejected, which is equivalent to a significant reduction in the estimated mean squared prediction error for series j.

Model selection criteria, such as the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC), can be used to determine the appropriate model order p.

The entire k-variate series should be considered; pairwise considerations can lead to an overly complicated model. For example, suppose that the causal dependencies are $X_1 \to X_2$ and $X_2 \to X_3$; with X_2 included in the model, X_1 gives no further information about X_3. If attention were restricted to pairwise considerations, we would also have $X_1 \to X_3$.

This formulation of G-causality requires two important assumptions about the data:

1. The series is wide-sense stationary (i.e., the mean and covariance structure of k-variate time series does not change over time);

2. it can be adequately described by a VAR(p) model, for some p.