# *Covariances of ARMA Processes*

## Overview

1. Review ARMA models: causality and invertibility
2. AR covariance functions
3. MA and ARMA covariance functions
4. Partial autocorrelation function
5. Discussion

## Review of ARMA processes

**ARMA process** A *stationary* solution $\{X_t\}$ (or if its mean is not zero, $\{X_t - \mu\}$) of the linear difference equation

$$\begin{aligned} X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} &= w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \\ \phi(B) X_t &= \theta(B) w_t \end{aligned} \tag{1}$$

where $w_t$ denotes white noise, $w_t \sim WN(0, \sigma^2)$. **Definition 3.5** adds the

- identifiability condition that the polynomials $\phi(z)$ and $\theta(z)$ have no zeros in common and the
- normalization condition that $\phi(0) = \theta(0) = 1$.

**Causal process** A stationary process $\{X_t\}$ is said to be *causal* if there exists a summable sequence (some require square summable ($\ell_2$), others want more and require absolute summability ($\ell_1$)) sequence $\{\psi_j\}$ such that $\{X_t\}$ has the one-sided moving average representation

$$X_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B) w_t \; . \tag{2}$$

**Proposition 3.1** states that a stationary ARMA process $\{X_t\}$ is causal if and only if (iff) the zeros of the autoregressive polynomial

$\phi(z)$ lie *outside* the unit circle (*i.e.*, $\phi(z) \neq 0$ for $|z| \leq 1$). Since $\phi(0) = 1$, $\phi(z) > 0$ for $|z| \leq 1$. (The *unit circle* in the complex plane consists of those $x \in \mathbb{C}$ for which $|z| = 1$; the closed *unit disc* includes the interior of the unit circle as well as the circle; the open unit disc consists of $|z| < 1$.)

*If the zeros of $\phi(z)$ lie outside the unit circle*, then we can invert each of the factors $(1 - B/z_j)$ that make up $\phi(B) = \prod_{j=1}^{p}(1 - B/z_j)$ one at a time (as when back-substituting in the derivation of the AR(1) representation). Owing to the geometric decay in $1/z_j$, the coefficients in the resulting expression are summable. (Complex zeros are more interesting.)

*Suppose, on the other hand, that the process is causal.* Then by taking expectations substituting $X_t = \psi(B)w_t$ for $X_t$ in the definition of the ARMA process, it follows that for all $k = 0, 1, \ldots$ that

$$\mathbb{E}\,\phi(B)\psi(B)w_t\,w_{t-k} = \mathbb{E}\,\theta(B)w_t\,w_{t-k} \quad \Rightarrow \quad \phi(z)\,\psi(z) = \theta(z)$$

By definition, $\phi(z)$ and $\theta(z)$ share no zeros and $\psi(z) < \infty$ for $|z| \leq 1$ (since the $\psi_j$ are summable). The zeros of $\phi(z)$ must be outside the unit circle else we get a contradiction because the existence of such a zero implies that $\phi(\tilde{z}) = 0$ and $\theta(\tilde{z}) = 0$. (For $|z| > 1$, $\psi(z)$ can grow arbitrarily large, balancing the zero of $\phi(z)$.)

**Covariance generating function** The covariances of the ARMA process $\{X_t\}$ are

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+|h|}\psi_j \; . \tag{3}$$

Equivalently, the covariance $\gamma(h)$ is the coefficient of $z^{|h|}$ in the polynomial

$$G(z) = \sigma^2\psi(z)\psi(z^{-1}) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}, \tag{4}$$

which is called the covariance generating function of the process. The constant in the polynomial $z^{-h}G(z)$ is $\gamma(h)$. Note the ambivalence about the direction of time.

**Connection to spectrum** Recall that the spectral density function also "generates" the covariances in the sense that

$$\gamma(h) = \int_{-1/2}^{1/2} e^{-i2\pi\omega h} f(\omega) d\omega \ .$$

Compare this expression to the following that uses the covariance generating function. Replace $z$ in (4) by $e^{-i2\pi\omega}$ and integrate over $-\frac{1}{2} \le \omega \le \frac{1}{2}$ to get

$$\gamma(h) = \int_{-1/2}^{1/2} e^{-2\pi\omega h} \underbrace{\left( \sigma^2 \frac{|\theta(e^{-i2\pi\omega})|^2}{|\phi(e^{-i2\pi\omega})|^2} \right)}_{f(\omega)} d\omega \ .$$

The spectral density of an ARMA process is the shown ratio of polynomials. Since polynomials are dense in the space of continuous functions on bounded intervals, ARMA processes can approximate any stationary process with continuous spectrum.

**Invertible** The *ARMA* process $\{X_t\}$ is invertible if there exists an $\ell_2$ sequence $\{\pi_j\}$ such that we obtain

$$w_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \ . \tag{5}$$

**Proposition 3.2** states that the process $\{X_t\}$is invertible if and only if the zeros of the moving average polynomial lie *outside* the unit circle.

**Common assumption: invertible and causal** In general when dealing with covariances of ARMA processes, we assume that the process is causal and invertible so that we can move between the two one-sided representations (5) and (2).

**Example 3.6** shows what happens with common zeros in $\phi(z)$ and $\theta(z)$. The process is

$$X_t = 0.4X_{t-1} + 0.45X_{t-2} + w_t + w_{t-1} + 0.25w_{t-2}$$

for which

$$\phi(z) = (1 + 0.5z)(1 - 0.9z), \quad \theta(z) = (1 + 0.5z)^2 \ .$$

Hence, the two share a common factor. The initial ARMA(2,2) reduces to a causal, invertible ARMA(1,1) model.

**Calculations in R** R includes the function `polyroot` for finding the zeros of polynomials. Other symbolic software (*e.g.*, Mathematica) do this much better, giving you a formula for the roots in general (when possible). S&S supply some additional utilities, particularly `ARMAtoMA` which finds the moving average coefficients for an arbitrary ARMA process and `ARMAacf` and `ARMApacf`.

## AR covariance functions

**Estimation** Given the assumption of stationarity, in most cases we can easily obtain consistent estimates of the process covariances, such as

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{T-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{n} \; . \tag{6}$$

What should such a covariance function resemble? Does the "shape" of the covariance function or estimated correlation function

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

distinguish the underlying process? (There are problems with this approach, suggested by our discussion of the dependence in covariance estimates.)

**Yule-Walker equations** These equations expose the relationship between the covariances and coefficients of an autoregression: the covariances satisfy the difference equation that defines the autoregression. (This comment helps to explain the huge dependence in estimates of the covariances.) The one-sided moving-average representation of the process (2) implies that $w_t$ is uncorrelated with prior observations $X_s, s < t$. Hence, for lags $k = 1, 2, \ldots$ (assuming $\mathbb{E}\, X_t = 0$)

$$\begin{aligned} \mathbb{E}\left[X_{t-k}(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})\right] &= E[X_{t-k} w_t] \\ &\Rightarrow \\ \gamma(k) - \phi_1\gamma(k-1) \cdots - \phi_p\gamma(k-p) &= \phi(B)\gamma(k) = 0 \; . \end{aligned}$$

Rearranging the terms, we obtain the *Yule-Walker equations* (where $\delta_k = 0$ for $k \neq 0$),

$$\delta_k \sigma^2 = \gamma(k) - \sum_{j=1}^{p} \gamma(k-j)\phi_j \ , \quad k = 0, 1, 2, \ldots . \tag{7}$$

Written out, these give an equation that provides $\sigma^2$ (for $k = 0$)

$$\sigma^2 = \gamma(0) - \gamma(1)\phi_1 - \gamma(2)\phi_2 - \cdots - \gamma(p)\phi_p$$

and $p$ additional equations that give the coefficients:

$$\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \gamma(3) \\ \vdots \\ \gamma(p) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \ldots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \ldots & \gamma(p-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \ldots & \gamma(p-3) \\ \vdots & \vdots & & \ddots & \\ \gamma(p-1) & \gamma(p-2) & \gamma(p-3) & \ldots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{pmatrix}$$

Define vectors $\gamma = (\gamma(1), \ldots, \gamma(p))'$, $\phi = (\phi_1, \ldots, \phi_p)'$ and the matrix $\Gamma_p = [\gamma(j-k)]_{j,k=1,\ldots,p}$. The matrix form of these last $p$ equations of the Yule-Walker equations is

$$\gamma = \Gamma_p \phi \ .$$

(The Yule-Walker equations are analogous to the normal equations from least-squares regression, with $Y = X_t$ and the explanatory variables $X_{t-1}, \ldots, X_{t-p}$). The sample version of these equations is used in some methods for estimation, most often to obtain initial estimates required for an iterative estimator.

**Difference equations** Backshift notation is handy in appreciating these characteristics of the covariances. When we multiply both sides of $\phi(B)X_t = w_t$ by $X_{t-h}, h > 0$ and take expections we obtain

$$\phi(B)\gamma(h) = 0 \Rightarrow \ \gamma(h) \text{ solves the difference eqn defined by } \phi(B)$$

A solution of this homogeneous difference equation is a sequence $\{c_h\}$ for which

$$c_h - \phi_1 c_{h-1} - \phi_2 c_{h-2} - \cdots - \phi_p c_{h-p} \phi(B)c = 0 \ .$$

As a solution, try (as in finding the eigenvectors of the circulant matrix) $c_h = z_j^{-h}$, where $\phi(z_j) = 0$. Then observe that

$$z_j^{-h}(1 - \phi_1 z_j - \phi_2 z_j^2 - \cdots - \phi_p z_j^p) = 0$$

In general, any linear combination of the zeros of $\phi(z)$ is a solution. To find the particular solution, one must take into account boundary conditions implied by the first few values of $\gamma(h)$ given by the Yule-Walker equations. The solution is of the form (ignoring duplicated zeros)

$$\gamma(k) = \sum_j a_j z_j^{-k}, \quad |z_j| < 1 \ ,$$

where the $a_j$ are constants (determined by boundary conditions).

**Boundary conditions** The Yule-Walker equations can be solved for $\gamma(0)$, $\gamma(1), \ldots, \gamma(p-1)$ given $\phi_1, \ldots, \phi_p$. This use of these equations (*i.e.*, to solve for the covariances from the coefficients rather than expressing the coefficients in terms of the covariances) is the "inverse" of how they are used in estimation.

## MA and ARMA covariance functions

**Moving average case** For an *MA(q)* process, we have ($\theta_0 = 1$)

$$\gamma(h) = \sigma^2 \sum_j \theta_{j+|h|}\theta_j$$

where $\theta_j = 0$ for $j < 0$ and $j > q$. In contrast to the geometric decay of an autoregression, the covariances of a moving average "cut off" abruptly. Such covariance functions are necessary and sufficient to identify a moving average process.

**Calculation** of the covariances via the infinite MA representation and equation (3) proceeds by solving system of equations, defined by the relation

$$\psi(z) = \frac{\theta(z)}{\phi(z)} \quad \Rightarrow \quad \psi(z)\phi(z) = \theta(z) \ .$$

The idea is to match the coefficients of like powers of $z$ in

$$(1 + \psi_1 z + \psi_2 z^2 + \ldots)(1 + \phi_1 z + \phi_2 z^2 + \ldots) = (1 + \theta_1 z + \ldots)$$

For example, the equating the coefficients of $z$ implies that

$$\psi_1 + \phi_1 = \theta_1$$

But this only leads to the collection of $\psi$'s, not the covariances. The covariances require summing the resulting expressions.

**Mixed models** Observe that the covariances satisfy the convolution expression (multiply both sides of (1) by lags $X_{t-j}, j \geq \max(p, q+1)$)

$$\gamma(j) - \sum_{k=1}^{p} \phi_k \gamma(j-k) = 0, \quad j \geq \max(p, q+1),$$

which is again a homogeneous linear difference equation. Thus, for high enough lag, the covariances again decay as a sum of geometric series. The mixed ARMA(1,1) example from the text (**Example 3.11**, p. 105) illustrates these calculations. For initial values, we find

$$\gamma(j) - \sum_{k=1}^{p} \phi_k \gamma(j-k) = \sigma^2 \sum_{j \leq k \leq q} \theta_k \psi_{k-j}$$

a generalization of the Yule-Walker equations. The extra summand arises from the observation that $\mathbb{E}\, X_{t-k} w_t = \sigma^2 \psi_k$.

Essentially, the initial $q$ covariances of an ARMA$(p, q)$ process deviate from the recursion that defines the covariances of the AR$(p)$ components of the process.

## Partial autocorrelation function

**Definition** The partial autocorrelation function $\phi_{hh}$ is the partial correlation between $X_{t+h}$ and $X_t$ conditioning upon the intervening variables,

$$\phi_{hh} = \text{Corr}(X_{t+h}, X_t | X_{t+1}, \ldots, X_{t+h-1}) \; .$$

Consequently, $\phi_{11} = \rho(1)$, the usual autocorrelation. The partial autocorrelations are often called *reflection coefficients*, particularly in signal processing.

**Partial regression** The reasoning behind the use of the partial correlations resembles the motivation for partial regression residual plots which show the impact of a variable in regression. If we have the OLS fit of the two-predictor linear model

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \text{residual}$$

and we form two "partial" regressions by regressing out the effects of $X_1$ from $X_2$ and $Y$,

$$r_2 = X_2 - a_0 - a_1 X_1, \quad r_y = Y - c_0 - c_1 X_1,$$

then the regression coefficient of residual $r_y$ on the other residual $r_2$ is $b_2$, the multiple regression coefficient.

**Defining equation** Since the partial autocorrelation $\phi_{hh}$ is the coefficient of the last lag in the regression of $X_t$ on $X_{t-1}$, $X_{t-2}$, $\ldots, X_{t-h}$, we obtain an equation for $\phi_{hh}$ (assuming that the mean of $\{X_t\}$ is zero) by noting that the normal equations imply that

$$\mathbb{E}\,[X_{t-j}(X_t - \phi_{h1}X_{t-1} - \phi_{h2}X_{t-2} - \cdots - \phi_{hh}X_{t-h})] = 0, \quad j = 1, 2, \ldots, h\,.$$

**Key property** For an AR($p$) process, $\phi_{hh} = 0, h > p$ so that the partial correlation function cuts off after the order $p$ of the autoregression. Also notice that $\phi_{pp} = \phi_p$. Since an invertible moving average can be represented an infinite autoregression, the partial autocorrelations of a moving average process decay geometrically.

Hence, we have the following table of behaviors (**Table 3.1**):

| | AR($p$) | ARMA($p,q$) | MA($q$) |
|---|---|---|---|
| $\gamma(h)$ | geometric decay | geometric after $q$ | cuts off at $q$ |
| $\phi_{hh}$ | cuts off at $p$ | geometric after $p$ | geometric decay |

Once upon a time before the introduction of model selection criteria such as AIC (discussed in S&S), this table was the key to choosing the order of an ARMA($p,q$) process.

**Estimates** Estimates of the partials autocorrelations arise from solving the Yule-Walker equations (7), using a recursive method known as the Levinson recursion. This algorithm is discussed later.

## Discussion

**Weak spots** We have left some problems only partially solved, such as the meaning of infinite sums of random variables. How does one manipulate these expressions? When are such manipulations valid?

**Correlation everywhere** Much of time series analysis is complicated because the observable terms $\{X_t\}$ are correlated. In a sense, time series analysis is a lot like regression with collinearity. Just as regression is simplified by moving to uncorrelated predictors, time series analysis benefits from using a representation in terms of uncorrelated terms.