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 ℓ_2, others want more and restrict these to ℓ_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 $|z| \leq 0$. (The *unit circle* in the complex plane consists of those $x \in$ for which $|z| = 1$; the *unit disc* includes the interior of the unit circle.)

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.

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

$$\mathbb{E}\,\phi(B)\psi(B)w_t = \mathbb{E}\,\theta(B)w_t \quad \Rightarrow \quad \phi(z)\;\psi(z) = \theta(z)$$

By definition, $\phi(z)$ and $\theta(z)$ share no zeros and $\psi(z) < \infty$ for $|z| \le 1$ (since the ψ_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 the covariance generating function of the process.

Invertible The *ARMA* process $\{X_t\}$ is invertible if there exists an ℓ_2 sequence $\{\pi_j\}$ such that

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

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.

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

$$\phi(z) = (1 + 0.5z)(1 - 0.9z), \quad \theta(z) = (1 + 0.5z_0)^2 .$$

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 has 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).

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 mvoe between the two one-sided representations (5) and (2).

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?

AR(1) For a first-order autoregression, we know that the covariances $\gamma(h)$, $h = 0, 1, 2, \ldots$ decay geometrically:

$$\gamma(h) = \frac{\sigma^2}{1 - \phi^2}\, \phi^{|h|} .$$

Compare this theoretical property to the properties of realizations of an AR(1). Do the estimated covariances (6) necessarily decay geometrically?

Yule-Walker equations The covariances satisfy the difference equation that defines the autoregression. The one-sided moving-average representation of the process (2) implies that w_t is uncorrelated with past observations $X_s, s < t$. Hence, for lags $k = 1, 2, \ldots$ (assuming $\mathbb{E}\, X_t = 0$)

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

Rearranging the terms, we obtain the *Yule-Walker equations*,

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

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

$$\gamma = \Gamma \phi \;.$$

(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 is used in some methods for estimation.

Geometric decay In the general *AR(p)* case, the covariances decay as a sum of geometric series. To solve a linear difference equation such as (7, with $k > 0$) for the covariances, replace $\gamma(k)$ by z^k, obtaining

$$z^k - \sum_{j=1}^{p} \phi_j z^{k-j} = z^k \phi(1/z) = 0 \;.$$

Note that any zero of the polynomial $\phi(1/z)$ solves the equation, and by linearity, so does any linear combination of such zeros. Thus, the "eventual" solution is of the form (ignoring duplicated zeros)

$$\gamma(k) = \sum_j a_j r_j^k, \quad |r_j| < 1 \;,$$

where the a_j are constants (determined by boundary conditions) and the r_j's are the *reciprocals of the zeros* of the AR polynomial $\phi(z)$.

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

AR(2) example These are interesting because they allow for complex-valued zeros in the polynomial $\phi(z)$. The presence of such complex pairs produces oscillations in the observed process.

For the process to be stationary, we need the zeros of $\phi(z)$ to lie *outside* the unit circle. If the two zeros are z_1 and z_2, then

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = (1 - z/z_1)(1 - z/z_2). \tag{8}$$

Since $\phi_1 = 1/z_1 + 1/z_2$ and $\phi_2 = -1/(z_1 z_2)$, the coefficients lie within the rectangular region

$$-2 < \phi_1 = 1/z_1 + 1/z_2 < +2, \quad -1 < \phi_2 = \frac{-1}{z_1 z_2} < +1 \ .$$

Since $\phi(z) \neq 0$ for $|z| \leq 1$ and $\phi(0) = 1$, $\phi(z)$ is positive for over the unit disc $|z| \leq 1$ and

$$\begin{aligned} \phi(1) = 1 - \phi_1 - \phi_2 > 0 \quad &\Rightarrow \quad \phi_1 + \phi_2 < 1 \\ \phi(-1) = 1 + \phi_1 - \phi_2 > 0 \quad &\Rightarrow \quad \phi_2 - \phi_2 < 1 \end{aligned}$$

From the quadratic formula applied to (8), $\phi_1^2 + 4\phi_2 < 0$ implies that the zeros form a complex conjugate pair,

$$z_1 = re^{i\lambda}, z_2 = re^{-i\lambda} \quad \Rightarrow \quad \phi_1 = 2\cos(\lambda)/r, \phi_2 = -1/r^2.$$

Turning to the covariances of the AR(2) process, these satisfy the difference equation $\gamma(h) - \phi_1\gamma(h-1) - \phi_2\gamma(h-2) = 0$ for $h = 1, 2, \ldots$. We need two initial values to start these recursions. To make this chore easier, work with correlations. Dividing by $\gamma(0)$ gives

$$\rho(h) - \phi_1\rho(h-1) - \phi_2\rho(h-2) = 0 \ ,$$

and we know $\rho(0) = 1$. To find $\rho(1)$, use the equation defined by $\gamma(0)$ and $\gamma(1) = \gamma(-1)$:

$$\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(-1) = \phi_1 + \phi_2 \rho(1),$$

which shows $\rho(1) = \phi_1/(1 - \phi_2)$. When the zeros are a complex pair, $\rho(h) = c\, r^h \cos(h\lambda)$, a damped sinusoid, and realizations exhibit *quasi-periodic behavior*.

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

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

But this only leads to the collection of ψ's, not the covariances.

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 *ARMA(1,1)* example from the text (**Example 3.11**, p. 105) illustrates these calculations. For initial values, we find

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

a generalization of the Yule-Walker equations.

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 ϕ_{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 partial correlation 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 the 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 ϕ_{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 ϕ_{hh} (assuming that the mean of $\{X_t\}$ is zero) by noting that the normal equations imply that

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

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
ϕ_{hh}	cuts off at p	geometric after p	geometric decay

Once upon a time before the introduction of model selection criteria such as AIC, 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 that is more general than the one-sided infinite moving average form.