

Notes on ARMA Processes

Brian Borchers and Rick Aster

November 11, 2008

Some notation

In the following, we will make use of forward and backward shifts in time. The B operator is defined by

$$Bz_n = z_{n-1} \quad (1)$$

while the F operator is

$$Fz_n = z_{n+1}. \quad (2)$$

Note that B and F are not numbers but rather operators that act on a time series z_n . We can extend this notation to include powers of B and F

$$B^k z_n = z_{n-k} \quad (3)$$

and

$$F^k z_n = z_{n+k}. \quad (4)$$

Also, we can build polynomials from the B and F operators. For example,

$$(1 - B + 0.5B^2)z_n = z_n - z_{n-1} + 0.5z_{n-2}. \quad (5)$$

Gaussian White Noise

We will frequently make use of a **Gaussian white noise**. The white noise process has A_n normally distributed with mean 0, variance σ_A^2 , and autocovariance $\gamma_k = 0$, $k = 1, 2, \dots$ and autocorrelation $\rho_k = 0$, $k = 0, 1, \dots$. White noise can easily be generated in MATLAB using the **randn** command.

Using our formula for the spectrum of a stationary process from its autocovariance, it's easy to show that the white noise process should have $I(f) = 2\sigma_A^2$, $0 \leq f \leq 1/2$. In the limit as n goes to infinity, the spectrum is constant for all frequencies. However, for any actual realization of the white noise process, the sample spectrum will contain considerable noise.

The ARMA process

An **autoregressive moving average (ARMA)** process is obtained by applying a recursive filter to Gaussian white noise. In terms of the elements of the z_n and a_n sequences,

$$z_n = \phi_1 z_{n-1} + \phi_2 z_{n-2} + \dots + \phi_p z_{n-p} + a_n - \theta_1 a_{n-1} - \dots - \theta_q a_{n-q}. \quad (6)$$

The terms $\phi_1 z_{n-1}$ through $\phi_p z_{n-p}$ are the autoregressive portion of the filter. The terms a_n through $\theta_q a_{n-q}$ are a moving average of the white noise input process. Notice that this has the form of the recursive IIR filter that we previously considered, except that the first coefficients have been normalized to 1.

In terms of our operator notation,

$$\phi(B)z_n = \theta(B)a_n \quad (7)$$

where

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (8)$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q. \quad (9)$$

Note the unusual notational convention of minus signs in front of each coefficient.

Let

$$\psi(B) = \frac{\theta(B)}{\phi(B)}. \quad (10)$$

$$\psi(B) = \frac{1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p}. \quad (11)$$

Then we can write

$$z_n = \psi(B)a_n. \quad (12)$$

Earlier, when we found the z-transform transfer function for a filter, we wrote the transfer functions in powers of z^{-1} . The z-transform transfer function for our filter would be

$$\Phi(z) = \frac{1 - \theta_1 z^{-1} - \theta_2 z^{-2} - \dots - \theta_q z^{-q}}{1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \dots - \phi_p z^{-p}} \quad (13)$$

It's apparent that in our new notation, $\Psi(B) = \theta(B)/\phi(B)$ is equivalent to $\Phi(z)$ in the z-transform notation, with $B = 1/z$.

The power spectrum can be obtained by substituting $B = e^{-2\pi i f}$ in the transfer function

$$I(f) = 2\sigma_a^2 \left| \frac{\theta(e^{-2\pi i f})}{\phi(e^{-2\pi i f})} \right|^2. \quad (14)$$

Note that the minus sign in the exponent of $B = e^{-2\pi i f}$ is again because of the difference in notation between the Z transform and the B notation- $B = 1/z$.

We can expand $\psi(B)$ as

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (15)$$

where the coefficients ψ_k can be obtained by Taylor series expansion. This allows us to write z_n in terms of the inputs at time n and previous times.

$$z_n = a_n + \psi_1 a_{n-1} + \psi_2 a_{n-2} + \dots \quad (16)$$

Note that the constant coefficient is always 1 and that this time (only) we've used positive signs in front of the coefficients. From time to time it will be helpful to use the notation $\psi_0 = 1$, so that we don't have to treat the first term in this power series as a special case.

Example 1 Consider the filter in which

$$\phi(B) = 1 - 0.5B \quad (17)$$

and

$$\theta(B) = 1. \quad (18)$$

In terms of the filter equations,

$$z_n - 0.5z_{n-1} = a_n \quad (19)$$

or

$$z_n = a_n + 0.5z_{n-1}. \quad (20)$$

We can recursively apply the equation to write this as

$$z_n = a_n + 0.5(a_{n-1} + 0.5z_{n-2}). \quad (21)$$

$$z_n = a_n + 0.5(a_{n-1} + 0.5(a_{n-2} + 0.5z_{n-3})). \quad (22)$$

We end up with

$$z_n = a_n + 0.5a_{n-1} + 0.25a_{n-2} + 0.125a_{n-3} + \dots \quad (23)$$

Using the operator notation, it is much simpler to get the same result by doing a Taylor series expansion.

$$\Psi(B) = \frac{1}{1 - 0.5B} = 1 + 0.5B + 0.25B^2 + 0.125B^3 + \dots \quad (24)$$

An alternative is to let

$$\pi(B) = \frac{1}{\psi(B)} = \frac{\phi(B)}{\theta(B)}. \quad (25)$$

In this case,

$$a_n = \pi(B)z_n. \quad (26)$$

Clearly,

$$\pi(B) = \frac{1}{\psi(B)}. \quad (27)$$

We can expand $\pi(B)$ in a Taylor's series as

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots \quad (28)$$

Example 2 Consider the ARMA process

$$Z_n - 0.5Z_{n-1} = A_n - 0.3A_{n-1} + 0.2A_{n-2}. \quad (29)$$

Here $\phi(B) = 1 - 0.5B$ and $\theta(B) = 1 - 0.3B + 0.2B^2$. Using Maple to compute the Taylor's series, we obtain

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = 1 + 0.2B + 0.3B^2 + 0.15B^3 + \dots \quad (30)$$

Thus $\psi_1 = 0.2$, $\psi_2 = 0.3$, and $\psi_3 = 0.15$. Similarly,

$$\pi(B) = \frac{\phi(B)}{\theta(B)} = 1 - 0.2B - 0.26B^2 - 0.038B^3 - \dots \quad (31)$$

Thus $\pi_1 = 0.2$, $\pi_2 = 0.26$, and $\pi_3 = 0.038$.

Stationarity and Invertibility

Unfortunately, it is easy to write down an ARMA process which is not covariance stationary. For example, let

$$\psi(B) = \frac{1}{1-B} = 1 + B + B^2 + \dots \quad (32)$$

Then

$$Z_n = \sum_{k=-\infty}^n A_k \quad (33)$$

and

$$\text{Var}(Z_n) = \sum_{k=-\infty}^n \text{Var}(A_k) = \infty. \quad (34)$$

It can be shown that if

$$\sum_{j=1}^{\infty} |\psi_j| < \infty \quad (35)$$

then the ARMA process is stationary. This happens if the series $\psi(B)$ converges for every B with $|B| \leq 1$. Since $\psi(B)$ is a rational function, it can also be shown that the series converges for every B with $|B| \leq 1$ if the complex zeros of $\phi(B)$ lie outside the unit circle.

Recall that when we worked with the z transform of a digital filter, the stability condition was that the poles of the transfer function must lie within the unit circle. Why is the stability condition for an ARMA process that the zeros of $\phi(B)$ must lie outside the unit circle? The problem here is one of notation. In the digital filtering case, the transfer function was a rational function of $1/z$. Here in the ARMA case, the transfer function is a rational function of B . Thus B and $1/z$ are effectively playing the same role. When a pole lies inside the unit circle in the z plane, the corresponding pole lies outside of the unit circle in the B plane, where $B = 1/z$.

If we have a stationary ARMA process, then since $Z_n = \psi(B)A_n$, and the expected values of A_n are all 0, the expected value of Z_n is also 0.

A related issue is that of **invertibility**. Recall that we can write z_n in terms of a_n and previous values of z_{n-k} . That is,

$$\pi(B)z_n = a_n \quad (36)$$

or

$$z_n = a_n + \pi_1 z_{n-1} + \pi_2 z_{n-2} + \dots \quad (37)$$

This **inverted form** of the process provides a very useful way of generating a random sequence according to our ARMA process. However, this infinite sum must be truncated in practice. If the π_j coefficients do not decay to zero, then it isn't possible to approximate this infinite sum by truncating it.

We say that the process is **invertible** if

$$\sum_{j=1}^{\infty} |\pi_j| < \infty \quad (38)$$

Since $\pi(B)$ is a rational function, the series is invertible if the complex zeros of $\theta(B)$ lie outside of the unit circle.

Example 3 Recall the ARMA process of example 5. In this case, since $\phi(B) = 1 - 0.5B$, the only zero of $\phi(B)$ is at $B = 2$, which is outside of the unit circle, so the process is stationary. The zeros of $\theta(B)$ are at $B = 0.75 \pm 2.1i$, so the process is also invertible.

Finding the autocovariance and autocorrelation of an ARMA process

The $\psi(\cdot)$ form of the ARMA model can be used to find $Var(Z_n)$. Since

$$Z_n = \sum_{k=0}^{\infty} \psi_k A_{n-k} \quad (39)$$

and the A_i are independent with mean 0 and variance σ_A^2 , we can compute

$$Var(Z_n) = \sum_{k=0}^{\infty} \psi_k^2 Var(A_{n-k}) = \sigma_A^2 \sum_{k=0}^{\infty} \psi_k^2. \quad (40)$$

If we know the value of this infinite sum, then we're all set. If we don't know the infinite sum, but the ψ_k coefficients decay quickly to 0, then we can truncate the infinite series and get a good approximation to $Var(Z_n)$.

Similarly, we can use the $\psi()$ form of the ARMA process to find covariances between Z_n and A_{n-k} for $k = 0, \dots$

$$Cov(Z_n, A_{n-k}) = Cov\left(\sum_{k=0}^{\infty} \psi_k A_{n-k}, A_{n-k}\right). \quad (41)$$

Since the A_i are independent of each other,

$$Cov(Z_n, A_{n-k}) = Cov(\psi_k A_{n-k}, A_{n-k}) = \psi_k \sigma_A^2. \quad (42)$$

In order to find the autocovariance of an ARMA process, we start with the model in the recursive filter form.

$$Z_n = \phi_1 Z_{n-1} + \phi_2 Z_{n-2} + \dots + \phi_p Z_{n-p} + A_n - \theta_1 A_{n-1} - \dots - \theta_q A_{n-q}. \quad (43)$$

Next, we multiply both sides by Z_{n-k} and take expected values. Since $E[Z_n] = E[z_{n-k}] = E[A] = 0$, $Cov(Z_n, Z_{n-k}) = E[Z_n Z_{n-k}]$. Thus

$$\begin{aligned} Cov(Z_n, Z_{n-k}) &= \phi_1 Cov(Z_{n-1}, Z_{n-k}) + \dots + \phi_p Cov(Z_{n-p}, Z_{n-k}) \\ &\quad + Cov(Z_{n-k}, A_n) - \theta_1 Cov(Z_{n-k}, A_{n-1}) - \dots - \theta_q Cov(Z_{n-k}, A_{n-q}) \end{aligned}$$

So,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \gamma_{ZA}(k) - \theta_1 \gamma_{ZA}(k-1) - \dots - \theta_q \gamma_{ZA}(k-q) \quad (44)$$

where

$$\gamma_{ZA}(k-j) = Cov(Z_{n-k}, A_{n-j}) \quad (45)$$

Since Z_{n-k} is independent of the white noise at times after $n-k$, these covariances are 0. Also, since $Z_{n-k} = \sum_{j=0}^{\infty} \psi_j A_{n-k-j}$, the remaining covariances are given by $Cov(Z_{n-k}, A_{n-k-j}) = \psi_j \sigma_A^2$. Thus

$$\gamma_{ZA}(j) = \begin{cases} 0 & j > 0 \\ \psi_{-j} \sigma_A^2 & j \leq 0 \end{cases} \quad (46)$$

So, we can express the autocovariance at lag k as

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \sigma_A^2 (-\theta_k \psi_0 - \theta_{k+1} \psi_1 - \dots - \theta_q \psi_{q-k}) \quad (47)$$

When $k \geq q+1$, this simplifies to

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}. \quad (48)$$

Another important case is $k = 0$. The variance γ_0 is given by

$$\gamma_0 = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_A^2 (1 - \theta_1 \psi_1 - \dots - \theta_q \psi_q) \quad (49)$$

These recurrence relations can be solved to obtain the autocovariance and autocorrelation.

As an example, consider the second order autoregressive process

$$Z_n = \phi_1 Z_{n-1} + \phi_2 Z_{n-2} + A_n \quad (50)$$

It can be show that this process is stationary if $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, and $-1 < \phi_2 < 1$. Because $\theta(B) = 1$ this process is always invertible.

To compute the autocovariance, we multiply the above formula by Z_{n-k} and take expected values.

$$Cov(Z_n, Z_{n-k}) = \phi_1 Cov(Z_{n-1}, Z_{n-k}) + \phi_2 Cov(Z_{n-2}, Z_{n-k}) + Cov(A_n, Z_{n-k}) \quad (51)$$

When $k = 0$, we get

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + Cov(A_n, Z_n) \quad (52)$$

But $Z_n = \phi_1 Z_{n-1} + \phi_2 Z_{n-2} + A_n$, and A_n is independent of Z_{n-1} and Z_{n-2} , so

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + Cov(A_n, A_n) \quad (53)$$

or

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_A^2 \quad (54)$$

When $k > 0$, $Cov(A_n, Z_{n-k}) = 0$, and we get

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad (55)$$

In terms of the autocorrelation function, we have

$$\rho_0 = 1 \quad (56)$$

and

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1. \quad (57)$$

Solving this equation for ρ_1 , we get

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad (58)$$

For $k > 2$, we get

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k > 2 \quad (59)$$

Example 4 Consider an AR(2) process with $\phi_1 = 0.5$ and $\phi_2 = 0.3$. We generated a random sequence according to this process. Figure 1a shows the first 1000 points of this random process. Figure 1b shows the theoretical autocorrelation. Figure 1 shows the autocorrelation as estimated from the 20,000 point sequence.

Next, consider the ARMA(1,1) process

$$Z_n - \phi_1 Z_{n-1} = A_n - \theta_1 A_{n-1} \quad (60)$$

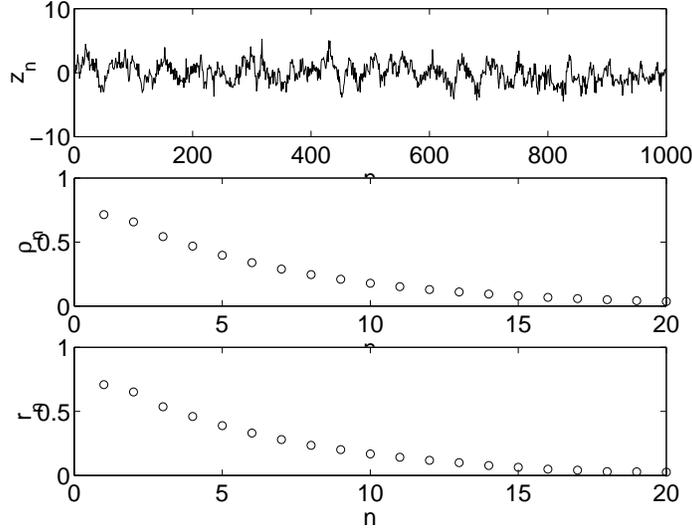


Figure 1: An AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.3$.

Here $\phi(B) = 1 - \phi_1 B$ and $\theta(B) = 1 - \theta_1 B$. We need to make sure that the roots of $\phi(B)$ and $\theta(B)$ are outside of the unit circle. This process is stationary if $-1 < \phi_1 < 1$ and invertible when $-1 < \theta_1 < 1$. We can also compute $\psi(B) = 1 + (\phi_1 - \theta_1)B + \dots$. The recurrence relations for the autocovariance give

$$\gamma_0 = \phi_1 \gamma_1 + \sigma_A^2 (1 - \theta_1 \psi_1) \quad (61)$$

$$\gamma_1 = \phi_1 \gamma_0 - \theta_1 \sigma_A^2 \quad (62)$$

$$\gamma_k = \phi_1 \gamma_{k-1} \quad k \geq 2 \quad (63)$$

These equations can be solved for the autocovariance. We can then convert the solution to an autocorrelation function. The result is

$$\rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1} \quad (64)$$

$$\rho_k = \phi_1 \rho_{k-1} \quad k \geq 2 \quad (65)$$

These computations can all be performed for arbitrary ARMA(p,q) processes. However, in practice, the most important processes have p and q quite small, and general solutions for these particular ARMA(p,q) processes have been developed. Box, Jenkins, and Reinsel contains specific solutions for the autocorrelations of a variety of ARMA(p,q) processes with small values of p and q . Table 1 summarizes these formulas. Table 2 gives the stationarity and invertibility conditions for these ARMA models.

Name	p	q	ρ_1	ρ_2	$\rho_k, k \geq 3$
AR(1)	1	0	$\rho_1 = \phi_1$	$\rho_2 = \phi_1^2$	$\rho_k = \phi_1^k$
AR(2)	2	0	$\rho_1 = \frac{\phi_1}{1-\phi_2}$	$\rho_2 = \phi_1\rho_1 + \phi_2$	$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$
ARMA(1,1)	1	1	$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$	$\rho_2 = \phi_1\rho_1$	$\rho_k = \phi_1\rho_{k-1}$
MA(1)	0	1	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$	$\rho_2 = 0$	$\rho_k = 0$
MA(2)	0	2	$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$	$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\rho_k = 0$

Table 1: Autocorrelations for ARMA(p,q) processes. Based on formulas from Chapter 3 of Box, Jenkins, and Reinsel.

Name	p	q	Stationarity	Invertibility
AR(1)	1	0	$ \phi_1 < 1$	none
AR(2)	2	0	$ \phi_2 < 1, \phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1$	none
ARMA(1,1)	1	1	$ \phi_1 < 1$	$ \theta_1 < 1$
MA(1)	0	1	none	$ \theta_1 < 1$
MA(2)	0	2	none	$ \theta_2 < 1, \theta_1 + \theta_2 < 1, \theta_2 - \theta_1 < 1$

Table 2: Stationarity and Invertibility Conditions ARMA(p,q) processes. Based on formulas from Chapter 3 of Box, Jenkins, and Reinsel.

The Partial Autocorrelation Function

Suppose that our ARMA process is purely autoregressive of order k . That is,

$$Z_n = A_n + \phi_{k1}Z_{n-1} + \phi_{k2}Z_{n-2} + \dots + \phi_{kk}Z_{n-k} \quad (66)$$

In this case, the equations for the autocorrelations $\rho_j, j = 1, 2, \dots, k$ are particularly simple. They take the form

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \dots, k \quad (67)$$

Here we have used the notation ϕ_{kj} for the ϕ_j coefficient in an autoregressive model of order k .

These **Yule-Walker equations** can be written in matrix form as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \dots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_k \end{bmatrix} \quad (68)$$

or

$$P_k \phi_{\mathbf{k}} = \rho_{\mathbf{k}}. \quad (69)$$

In general, there will be nonzero autocorrelations at lags greater than k , and this system of equations doesn't help us determining those autocorrelations.

Example 5 Recall the AR(2) process

$$Z_n = A_n + \phi_1 Z_{n-1} + \phi_2 Z_{n-2} \quad (70)$$

The Yule-Walker equations are

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \quad (71)$$

Solving these equations, we obtain

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad (72)$$

and

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 \quad (73)$$

which matches our earlier calculation.

The Yule-Walker equations can be used in two important ways. If we know the coefficients ϕ_{k1} through ϕ_{kk} , then we can use the equations to compute the autocorrelations ρ_1 through ρ_k . Conversely, if we know (or can estimate) the autocorrelations, we can solve the equations to obtain estimates of the coefficients $\phi_{k1}, \dots, \phi_{kk}$.

The **partial autocorrelation function (PACF)** associated with a sequence z_n consists of the sequence $\hat{\phi}_{11}, \hat{\phi}_{22}, \hat{\phi}_{33}, \dots$ of partial autocorrelations estimated from the sequence z_n . This sequence can be obtained by estimating the autocorrelations, inserting the autocorrelations into the Yule-Walker equations (68), and then solving the Yule-Walker equations for $k = 1, k = 2, \dots$

In practice, a recursive formula due to Durbin is more efficient. The Durbin formula is

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{p,j} - \hat{\phi}_{p+1,p+1} \hat{\phi}_{p,p-j+1} \quad (74)$$

$$\hat{\phi}_{p+1,p+1} = \frac{r_{p+1} - \sum_{j=1}^p \hat{\phi}_{p,j} r_{p+1-j}}{1 - \sum_{j=1}^p \hat{\phi}_{p,j} r_j} \quad (75)$$

where

$$\hat{\phi}_{1,1} = r_1 \quad (76)$$

The PACF is very useful in identifying an autoregressive process. If our original process is autoregressive of order p , then for $k > p$, we should have $\hat{\phi}_{kk} = 0$. This provides a very useful test for whether or not a process is autoregressive. Of course, we need to know when the $\hat{\phi}_{kk}$ are effectively zero. It can be shown that the variance of $\hat{\phi}_{kk}$ is approximately $1/n$ when we have n points from an AR(p) process and $k \geq p + 1$.

The PACF also turns out to be important in forecasting. It can be shown that the best (least squares) predictor of z_n using the $k - 1$ previous values $z_{n-1}, z_{n-2}, \dots, z_{n-k+1}$ is

$$z_n = \phi_{k-1,1} z_{n-1} + \phi_{k-1,2} z_{n-2} + \dots + \phi_{k-1,k-1} z_{n-k+1} \quad (77)$$

Order	ρ_k	ϕ_{kk}
(1,0)	exp decay	only ϕ_{11} nonzero
(0,1)	only ρ_1 nonzero	exp decay
(2,0)	exp or damped sine wave	only ϕ_{11}, ϕ_{22} nonzero
(0,2)	only $\rho_{1,2}$ nonzero	exp or damped sine wave
(1,1)	exp decay	exp decay

Table 3: Rules for selecting p and q .

ARMA Modeling in Practice

Now that we understand the theoretical behavior of ARMA processes, we will consider how to take an actual observed time series, fit an ARMA model to the data, and forecast future values of the time series.

The stages in our process for ARMA modeling a time series beginning with observed values z_1, z_2, \dots, z_n are:

1. Remove any nonzero mean from the time series.
2. Estimate the autocorrelation and PACF of the time series. Use these to determine the autoregressive order p and the moving average order q .
3. Estimate the coefficients $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$.
4. Estimate a_1, a_2, \dots, a_n .
5. Use the fitted model to forecast z_{n+1}, z_{n+2}, \dots

The first step in our process is removing any nonzero mean from the time series z . This is a very straight forward computation- just compute the mean of the time series, and subtract it from each element of the time series.

The next step in the process is determining the autoregressive order p and the moving average order q . Table 1 (taken from BJR) summarizes the behavior of ARMA(p,q) processes for $p = 0, 1, 2$ and $q = 0, 1, 2$. If the observed autocorrelation and PACF match up with one of the lines in this table, then it's reasonable to fit a model of that type.

Example 6 Recall the time series of yields from a batch chemical process that we previously analyzed. Figure 2 shows the autocorrelations for this data. The autocorrelations seem to follow an exponentially damped sine wave, but they quickly hit a noise level beyond a lag of four or five.

We can also estimate the PACF for this data. We get $\hat{\phi}_{1,1} = -0.3889$, $\hat{\phi}_{2,2} = 0.1797$, $\hat{\phi}_{3,3} = 0.0023$, $\hat{\phi}_{4,4} = -0.0443, \dots$. In this case, since $n = 70$, we'd expect the standard deviation of the estimates to be about $1/\sqrt{70} = 0.12$ once we get out past a lag of p . Thus it appears that only the first two coefficients $\hat{\phi}_{1,1}$ and $\hat{\phi}_{2,2}$ are definitely nonzero.

The autocorrelation and PACF suggests an AR(2) model for this data set.

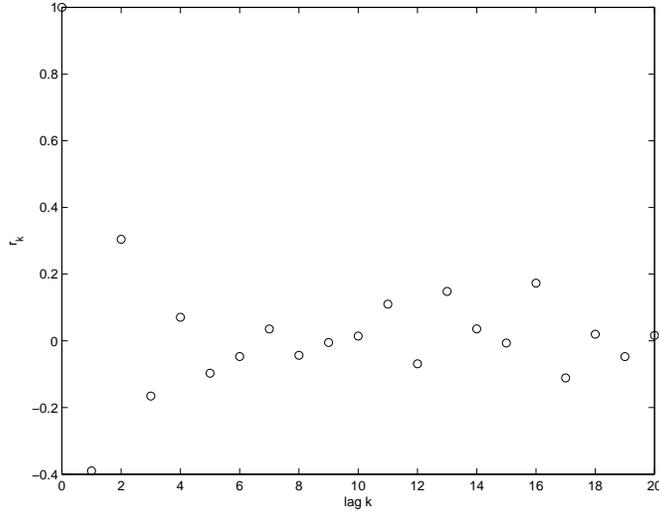


Figure 2: Autocorrelations for the batch data.

Once we've determined p and q , the next step is to estimate the actual parameters $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$. One very simple approach can be used if we have formulas for the autocorrelations in terms of the parameters. For example, for an AR(2) process, we know that

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \quad (78)$$

and

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \quad (79)$$

These equations can be solved for ϕ_1 and ϕ_2 to get

$$\phi_1 = \frac{\rho_1(\rho_2 - 1)}{\rho_1^2 - 1} \quad (80)$$

and

$$\phi_2 = \frac{\rho_1^2 - \rho_2}{\rho_1^2 - 1}. \quad (81)$$

If we use our estimates r_1 and r_2 , we can obtain estimates of ϕ_1 and ϕ_2 .

A similar approach can be used to estimate the parameters of other low order ARMA models. Table 2 summarizes the equations to be solved for the (1,0), (0,1), (2,0), (0,2) and (1,1) cases.

A more sophisticated approach is to use maximum likelihood estimation to obtain the parameters. Unfortunately, there aren't any functions for this in the

ARMA(p,q)	ρ_1	ρ_2
(1,0)	$\rho_1 = \phi_1$	
(0,1)	$\rho_1 = -\theta_1/(1 + \theta_1^2)$	
(2,0)	$\rho_1 = \phi_1/(1 - \phi_2)$	$\rho_2 = (\phi_1^2)/(1 - \phi_2) + \phi_2$
(0,2)	$\rho_1 = -\theta_1(1 - \theta_2)/(1 + \theta_1^2 + \theta_2^2)$	$\rho_2 = -\theta_2/(1 + \theta_1^2 + \theta_2^2)$
(1,1)	$\rho_1 = (1 - \theta_1\phi_1)(\phi_1 - \theta_1)/(1 + \theta_1^2 - 2\phi_1\theta_1)$	$\rho_2 = \rho_1\phi_1$

Table 4: Equations to be solved for ϕ and θ .

MATLAB toolboxes available at NMT. However, this can be done with more sophisticated statistical packages such as Minitab and R.

Example 7 Continuing with the batch process data, using (80) and (81), we estimate that

$$\phi_1 = -0.3198 \quad (82)$$

and

$$\phi_2 = 0.1797. \quad (83)$$

Suppose that we are now at time n , we've found p and q , and fitted the parameters $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$. Now we want to predict the observations at times $n + 1, n + 2, \dots, n + l$. We will use the notation z_{n-k} for the known observations up to time n . As before, we will use Z_{n+k} for unknown (random) future values of the time series. We will use $\hat{z}_n(l)$ for the predicted observation at time $n + l$ based on observations through time n .

In making a forecast $\hat{z}_n(l)$, we want to minimize the expected value of the square of the error in the forecast.

$$\min E[(Z_{n+l} - \hat{z}_n(l))^2] \quad (84)$$

where this expected value is conditioned on all of the observations through time n . It can be shown that picking $\hat{z}_n(l) = E[Z_{n+l}]$ minimizes the expected value of the squared error. Using the $\psi(\cdot)$ function, we know that

$$Z_{n+l} = \sum_{j=0}^{\infty} \psi_j A_{n+l-j} = A_{n+l} + \psi_1 A_{n+l-1} + \dots \quad (85)$$

This infinite sum contains some terms which correspond to times up to time n and other terms which lie in the future and are still random. To get the expected value of Z_{n+l} , we take the expected value of each term on the right hand side. For A_{n+l} and other future inputs from the white noise, this expected value is 0. For A_n and other past white noise inputs, the expected value of A_{n-k} is the actual value a_{n-k} that was observed. Thus our prediction is given by

$$\hat{z}_n(l) = \sum_{j=l}^{\infty} \psi_j a_{n+l-j}. \quad (86)$$

The random error associated with our forecast is

$$e_n(l) = A_{n+l} + \psi_1 A_{n+l-1} + \dots + \psi_{l-1} A_{n+1}. \quad (87)$$

Clearly, the expected value of $e_n(l)$ is 0. Furthermore, we can work out the variance associated with our prediction.

$$\text{Var}(e_n(l)) = \text{Var}(A_{n+l}) + \psi_1^2 \text{Var}(A_{n+l-1}) + \dots + \psi_{l-1}^2 \text{Var}(A_{n+1}) \quad (88)$$

$$\text{Var}(e_n(l)) = \sigma_A^2 (1 + \psi_1^2 + \dots + \psi_{l-1}^2). \quad (89)$$

There are three important practical issues that we need to resolve before we can actually start computing forecasts. The first problem is that we have a time series z_k , but not the corresponding a_k series. To compute the a sequence, notice that

$$e_n(1) = A_{n+1} \quad (90)$$

Thus

$$z_n - \hat{z}_{n-1}(1) = a_n. \quad (91)$$

We can use this to compute the values a_k for $k \leq n$. Just compute the lag 1 predictions, and subtract them from the actual values. In doing this, we may have to refer to z_k and a_k values from before the start of our observations. Set these to 0. In practice, the 0 initial conditions will have little effect on the forecasts.

The second issue is that we may not know σ_A^2 . In this case, we use the sample variance of the a values that we have computed as an estimate for σ_A^2 .

The third issue is that evaluating the infinite sum

$$\hat{z}(l) = \sum_{j=l}^{\infty} \psi_j a_{n+l-j}. \quad (92)$$

may be impractical. If the ψ_j weights decay rapidly, we can safely truncate the series, but if the ψ_j weights decay slowly this may be impractical. Fortunately, it is also possible to use the two other main forms of the model

$$Z_n = A_n + \pi_1 Z_{n-1} + \pi_2 Z_{n-2} + \dots \quad (93)$$

or

$$\phi(B)Z_n = \theta(B)A_n \quad (94)$$

for forecasting. If the model is purely autoregressive, then the π weights are the way to go. If the model is purely moving average, then it's best to use the ψ weights. For mixed models, the form $\phi(B)Z_n = \theta(B)A_n$ is usually the easiest to work with.

In making a forecast using any of the three forms of the model, we use the same basic idea. We start by computing the previous a_k values. Next, we substitute observed or expected values for all terms in the model to get $\hat{z}_n(l)$. The expected values of all future a_{n+k} values are 0. The expected values of

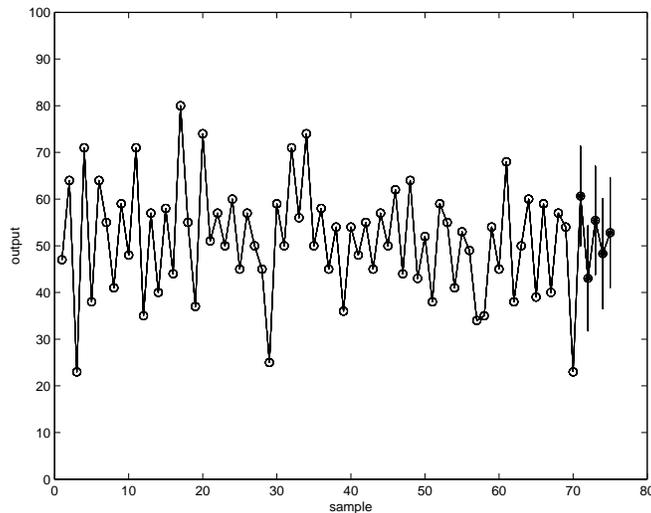


Figure 3: Predicted data points for the batch data.

future z_{n+k} values are given by our predictions $\hat{z}_n(k)$, $k = 1, 2, \dots$. We compute $\hat{z}_n(1), \hat{z}_n(2), \dots, \hat{z}_n(l)$, and then use the variance formula to get confidence intervals for our predictions.

Example 8 Figure 3 shows next five points predicted from the batch data. The general pattern of alternating high and low values is predicted to continue, but the error bars on these predictions are quite broad. The problem is that the original data is quite noisy. The estimated value of σ_A is 10.7 which is about 20% of the typical data points of around 50.

ARMA modeling and a slightly more sophisticated variation called ARIMA modeling are very widely used in time series forecasting. The technique is also known as Box-Jenkins forecasting after its inventors. When a model can be found that fits the data well and σ_A is relatively small, it can provide very good predictions. A huge advantage of this approach is that it produces error bars—many other simple forecasting schemes do not provide any indication of the uncertainty of the predictions. However, in some cases ARIMA modeling can fail, either because the underlying dynamics of the time series are too complicated to be captured by a simple ARIMA model, or because the noise level σ_A is simply too large.