Introduction to ARMA Models

Overview

1. Modeling paradigm
2. Review stationary linear processes
3. ARMA processes
4. Stationarity of ARMA processes
5. Identifiability of ARMA processes
6. Invertibility of ARMA processes
7. ARIMA processes

Modeling paradigm

Modeling objective A common measure used to assess many statistical models is their ability to reduce the input data to random noise. For example, we often say that a regression model "fits well" if its residuals ideally resemble *iid* random noise. We often settle for uncorrelated processes with data.

Filters and noise Model the observed time series as the output of an unknown process (or model) M "driven by" an input sequence composed of *independent* random errors $\{\epsilon_t\} \stackrel{\text{iid}}{\sim} Dist(0, \sigma^2)$ (not necessarily normal),

$$\epsilon_t \rightarrow \boxed{\text{Process M}} \rightarrow X_t$$

From observing the output, say $X_1, \ldots, X_n$, the modeling task is to characterize the process (and often predict its course). This "signal processing" may be more appealing in the context of, say, underwater acoustics rather than macroeconomic processes.

Prediction rationale If a model reduces the data to iid noise, then the model captures all of the relevant structure, at least in the sense that we obtain the decomposition

$$X_t = \mathbb{E}\left(X_t | X_{t-1}, ...\right) + \epsilon_t = \hat{X}_t + \epsilon_t \ .$$

Causal, one-sided Our notions of time and causation imply that the current value of the process cannot depend upon the future (nonanticipating), allowing us to express the process M as

$$X_t = M(\epsilon_t, \epsilon_{t-1}, \ldots) \ .$$

Volterra expansion is a general (too general?) expansion (like the infinite Taylor series expansion of a function) that expresses the process in terms of prior values of the driving input noise.

Differentiating M with respect to each of its arguments, we obtain the one-sided expansion (Wiener 1958),

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + \sum_{j,k} \psi_{jk} \epsilon_{t-j} \epsilon_{t-k} + \sum_{j,k,m} \psi_{jkm} \epsilon_{t-j} \epsilon_{t-k} \epsilon_{t-m} + \cdots ,$$

where, for example, $\psi_{jk} = \frac{\partial^2 M}{\partial \epsilon_{t-j} \partial \epsilon_{t-k}}$ evaluated at zero. The first summand on the right gives the linear expansion.

Linearity The resulting process is *linear* if X_t is a linear combination (weighed sum) of the inputs,

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{1}$$

Other processes are *nonlinear*. The process is also said to be *causal* (Defn 3.7) if there exists an white noise sequence ϵ_t and an absolutely summable sequence (or sometimes an ℓ_2 sequence) $\{\psi_j\}$ such that (1) holds. The key notion of causality is that the current observation is a function of current and past white noise terms (analogous to a random variable that is adapted to a filtration).

Invertibility The linear representation (1) suggests a big problem for identifying and then estimating the process: it resembles a regression in which all of the explanatory variables are functions of the unobserved errors. The *invertibility condition* implies that we can also express the errors as weighted sum of current and prior observations,

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}.$$

Thinking toward prediction, we will want to have an equivalence of the form (for any r.v. Y)

$$\mathbb{E}\left[Y|X_t, X_{t-1}, \ldots\right] = \mathbb{E}\left[Y|\epsilon_t, \epsilon_{t-1}, \ldots\right]$$

This equivalence implies that information in the current and past errors is equivalent to information in the current and past data (*i.e.*, the two sigma fields generated by the sequences are equivalent). Notice that the conditioning here relies on the *infinite* collection of prior values, not a finite collection back to some fixed point in time, such as $t = 1$.

Implication and questions The initial goal of time series modeling using the class of ARMA models to be defined next amounts to finding a *parsimonious, linear* model which can reduce $\{X_t\}$ to *iid* noise. Questions remain about how to do this:

1. Do such infinite sums of random variables exist, and how are they to be manipulated?
2. What types of stationary processes can this approach capture (*i.e.*, which covariance functions)?
3. Can one express these models using few parameters?

Review: Stationary Linear Processes

Notation of S&S uses $\{w_t\}$ as the canonical mean zero, finite variance white-noise process (which is not necessarily normally distributed),

$$w_t \sim WN(0, \sigma^2)$$

Convergence. For the linear process defined as $X_t = \sum_j \psi_j w_{t-j}$ to exist, we need assumptions on the weights ψ_j. An infinite sum is a limit,

$$\sum_{j=0}^{\infty} \psi_j w_{t-j} = \lim_n \sum_{j=0}^{n} \psi_j w_{t-j},$$

and limits require a notion of convergence (how else do you decide if you are close)? Modes of convergence for r.v.s include:

- Almost sure, almost everywhere, with probability one, w.p. 1, $X_n \overset{a.s.}{\to} X$,
$$P\{\omega : \lim X_n = X\} = 1.$$
- In probability, $X_n \overset{P}{\to} X$,
$$\lim_n P\{\omega : |X_n - X| > \epsilon\} = 0$$
- In mean square or ℓ_2, the variance goes to zero:
$$E\,(X_n - X)^2 \to 0.$$

ℓ_2 convergence of linear process requires that

$$\sum_j \psi_j^2 < \infty \text{ or } \{\psi_j\} \in \ell_2.$$

Consequently

$$\sum_j \psi_j \psi_{j+k} \le \sum \psi_j^2 < \infty.$$

In general, if $\{\psi_j\} \in \ell_2$ and $\{X_t\}$ is stationary, then the linear filter

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$$

defines a stationary process with covariance function

$$\text{Cov}(Y_{t+h}, Y_t) = \gamma_Y(h) = \sum_{j,k} \psi_j \psi_k \gamma_X(h - j + k).$$

Informally, $\text{Var}(Y_t) = \sum_{j,k} \psi_j \psi_k \gamma(k-j) \le \gamma(0) \sum_j \psi_j^2$.

Covariances When the "input" is white noise, then the covariances are infinite sums of the coefficients of the white noise,

$$\gamma_Y(h) = \sigma_X^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}. \tag{2}$$

Absolutely summable S&S often assume that $\sum |\psi_j| < \infty$ (absolutely summable). This is a stronger assumption that simplifies proofs of a.s. convergence. For example, $\frac{1}{j}$ is not absolutely summable, but is square summable. We will not be too concerned with a.s. convergence and will focus on mean-square convergence. (The issue is moot for ARMA processes.)

Question Does the collection of linear processes as given define a vector space that allows operations like addition? The answer is yes, using the concept of a *Hilbert space* of random variables.

ARMA Processes

Conflicting goals Obtain models that possess a *wide range* of covariance functions (2) and that characterize ψ_j as functions of a *few* parameters that are reasonably easy to estimate. We have seen several of these parsimonious models previously, *e.g.*,

- Finite moving averages: $\psi_j = 0, j > q > 0$.
- First-order autoregression: $\psi_j = \phi^j, |\phi| < 1$.

ARMA processes also arise when sampling a continuous time solution to a stochastic differential equation. (The sampled solution to a pth degree SDE is an ARMA(p,$p-1$) process.)

Definition 3.5 The process $\{X_t\}$ is an *ARMA(p,q)* process if

1. It is stationary.
2. It (or the deviations $X_t - E\,X_t$) satisfies the *linear* difference equation written in "regression form" (as in S&S, with negative signs attached to the ϕs) as

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \quad (3)$$

where $w_t \sim WN(0, \sigma^2)$.

Backshift operator Abbreviate the equation (3) using the so-called backshift operator defined as $B^k X_t = X_{t-k}$. Using B, write (3) as

$$\phi(B) X_t = \theta(B) w_t$$

where the polynomials are (note the differences in signs)

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

Closure The backshift operator shifts the stochastic process in time. Because the process is stationary (having a distribution that is invariant of the time indices), this transformation maps a stationary process into another stationary process. Similarly, scalar multiplication and finite summation preserve stationarity; that is, the vector space of stationary processes is closed in the sense that if X_t is stationary, then so too is $\Theta(B)\,X_t$ so long as $\sum_j \theta_j^2 < \infty$.

Aside: Shifts elsewhere in math The notion of shifting a stationary process $\{\ldots, X_t, X_{t+1}, \ldots$ to $\{\ldots, X_{t-1}, X_t, \ldots$ has parallels. For example, suppose that $p(x)$ is a polynomial. Define the operator S on the space of polynomials defined as $S\,p(x) = p(x-1)$. If the space of polynomials is finite dimensional (up to degree m), then we can write

$$S = I + D + D_2/2 + D^3/3! + \cdots D^m/m!$$

where I is the identity ($I\,p = p$) and D is the differntiation operator. (The proof is a direct application of Taylor's theorem.)

Value of backshift notation

1. Compact way to write difference equations and avoid backsubstitution, Backsubstitution becomes the conversion of $1/(1-x)$ into a geometric series; If we manipulate B algebraically in the conversion of the AR(1) to moving average form, we obtain the same geometric representation without explicitly doing the tedious backsubstitution:

$$\begin{aligned} \phi(B)X_t = w_t \Rightarrow X_t &= \frac{w_t}{1-\phi B} \\ &= (1 + \phi B + \phi^2 B^2 + \ldots)w_t \\ &= w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots\,. \end{aligned} \tag{4}$$

   Which clearly requires (as in a geometric series) that $|\phi| < 1$.

2. Expression of constraints that assure stationarity and identifiability.
3. Express effects of operations on a process:
   - Adding uncorrelated observation random noise to an AR process produces an ARMA process.
   - A weighted mixture of lags of an AR(p) model is ARMA.

Consider the claim that an average of several lags of an autoregression forms an ARMA process. Backshift polynomials make it trival to show this claim holds:

$$\phi(B)X_t = w_t \quad \Rightarrow \quad \theta(B)X_t = \frac{\theta(B)}{\phi(B)} w_t \ ,$$

which has the rational form of an ARMA process.

Converting to MA form, in general In order to determine $\psi(z)$, notice that $\theta(z)/\phi(z) = \psi(z)$ implies that

$$\phi(z)\psi(z) = \theta(z) \ .$$

Given the normalization $\phi_0 = \theta_0 = \psi_0 = 1$, one solves for the ψ_j by equating coefficients in the two polynomials (recursively).

Stationarity of ARMA Processes

Moving averages If $p = 0$, the process is a moving averge of order q, abbreviated an *MA(q)* process,

$$X_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \ . \tag{5}$$

This is a special case of the general linear process, having a finite number of nonzero coefficients (*i.e.*, $\psi_j = \theta_j, j = 1, \ldots, q$). Thus the MA($q$) process must be *stationary* with covariances of the form (2):

$$\gamma_X(h) = \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, \quad |h| \le q,$$

and zero otherwise.

All moving averages are stationary Under the typical constraint of S&S that the coefficients of a moving average are absolutely summable, all moving averages are stationary — even moving averages of other moving averages.

Proof. Suppose $\sum_j \theta_j^2 < \infty$ and that X_t is stationary with covariance function $\gamma_X(h)$. The covariance function of $Y_t = \sum_j \theta_j X_{t-j}$ is

$$\mathrm{Cov}(Y_{t+h}, Y_t) \quad = \quad \mathrm{Cov}\left(\sum_{j=0}^{\infty} \theta_j Y_{t+h-j}, \sum_{k=0}^{\infty} \theta_k Y_{t-k} \right)$$

$$
\begin{aligned}
&= \sum_{j,k=0}^{\infty} \theta_j \theta_k \,\mathrm{Cov}(Y_{t+h-j}, Y_{t-k}) \\
&= \sum_{j,k=0}^{\infty} \theta_j \theta_k \gamma_X(h-j+k) \\
&\le \left(\sum_j \theta_j^2\right)\left(\sum_j \gamma_X(j)\right)
\end{aligned}
$$

The covariances are summable and invariant of t.

Constraints remain An MA(q) process of finite order models a process that becomes uncorrelated beyond time separation q. There may be other limitations on the structure of the covariances. For example, consider the MA(1) model, $X_t = w_t + \theta_1 w_{t-1}$. The covariances of this process are are

$$\gamma(0) = \sigma^2(1+\theta_1^2),\ \gamma(1) = \sigma^2\theta_1,\ \gamma(h) = 0, h > 1.$$

Hence,

$$\rho(1) = \frac{\theta_1}{1+\theta_1^2} < \tfrac{1}{2}$$

which we can see from a graph or by noting that the maximum occurs where the derivative

$$\partial\rho(1)/\partial\theta_1 = \frac{1-\theta^2}{(1+\theta^2)^2} = 0\ .$$

Don't try to model the covariance function $\{\gamma(h)\} = (1, 0.8, 0, \ldots)$ with an MA(1) process! Other ARMA models place similar types of constraints on the possible covariances.

Autoregressions If $q = 0$, the process is an autoregression, or AR(p),

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t \tag{6}$$

The stationarity of a solution to (6) is less obvious because of the presence of "feedback" (beyond the AR(1) case considered previously). To investigate initially we make the AR process resemble a linear process (a weighted sum of past white noise) since we know that such a process is stationary.

Factor the polynomial $\phi(z)$ as using its zeros ($\phi(z_j) = 0$) as

$$\phi(z) = (1 - z/z_1) \cdots (1 - z/z_p) = \prod_j (1 - z/z_j) \ .$$

Some of the zeros z_j are likely to be complex. (Complex zeros come in conjugate pairs (say $z_j = \overline{z}_k$) since the coefficients ϕ_j are real).

As long as *all of the zeros are greater than one* in modulus ($|z_j| > 1$), we can repeat the method used in (4) to convert $\{X_t\}$ into a moving average, one term at a time. Since at each step we form a linear filtering of a stationary process with square-summable weights (indeed, absolutely summable weights), the steps are valid.

AR(2) example These processes are interesting because they allow for complex-valued zeros in the polynomial $\phi(z)$. The presence of 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

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = (1 - z/z_1)(1 - z/z_2). \tag{7}$$

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 (7), $\phi_1^2 + 4\phi_2 < 0$ 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*.

ARMA(p,q) case In general, the process has the representation (again, $|z_j| > 1$)

$$X_t = \frac{\theta(B)}{\phi(B)} w_t = \frac{\prod_j (1 - B/s_j)}{\prod_j (1 - B/z_j)} w_t = \psi(B) w_t \qquad (8)$$

where $s_j, j = 1, \ldots, q$ are the zeros of $\theta(B)$ and $\psi(B) = \theta(B)/\phi(B)$. This is a sum of q stationary processes, and thus stationary. Stationarity does not require that $|s_j| > 1$; that's required for invertibility (defined below).

Identifiability of ARMA processes

Identifiable A model with likelihood $L(\theta)$ is identified if different parameters produce different likelihoods, $\theta_1 \neq \theta_2 \rightarrow L(\theta_1) \neq L(\theta_2)$. For Gaussian time series, this condition amounts to having a 1-to-1 correspondence between the parameters and the covariances of the process.

Analogy to regression The most common example of a poorly identified model is a regression model with collinear explanatory variables. If $X_1 + X_2 = 1$, say, then

$$Y = \beta_0 + \beta_1 X_1 + 0\ X_2 + \epsilon \quad \Leftrightarrow \quad Y = (\beta_0 + \beta_1) + 0\ X_1 - \beta_1 X_2 + \epsilon$$

Both models obtain the same fit, but with very different coefficients. Least squares can find many fits that all obtain the same R^2 (the coefficients lie in a subspace).

Non-causal process. These "odd" processes hint at how models are not identifiable. Suppose that $|\tilde{\phi}| > 1$. Is there a stationary solution to $X_t - \tilde{\phi}X_{t-1} = Z_t$ for some white-noise process Z_t? The surprising answer is yes, but it's weird because it runs *backwards* in time. The *hint* that this might happen lies in the symmetry of the covariances, $\text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t+h})$.

To arrive at this representation, forward-substitute rather than back-substitute. This flips the coefficient from $\tilde{\phi}$ to $1/\tilde{\phi} < 1$. Start with the process at time $t+1$

$$X_{t+1} = \tilde{\phi}X_t + w_{t+1} \quad \Rightarrow \quad X_t = (1/\tilde{\phi})X_{t+1} - (1/\tilde{\phi})w_{t+1} \,.$$

Continuing recursively,

$$\begin{aligned} X_t &= -w_{t+1}/\tilde{\phi} + (1/\tilde{\phi})X_{t+1} \\ &= -w_{t+1}/\tilde{\phi} + (1/\tilde{\phi})\left(-w_{t+2}/\tilde{\phi} + (1/\tilde{\phi})X_{t+2}\right) \\ &= -w_{t+1}/\tilde{\phi} - w_{t+2}/\tilde{\phi}^2 + X_{t+2}/\tilde{\phi}^2 \\ &\cdots \\ &= -\sum_{j=1}^{k} w_{t+j}/\tilde{\phi}^j + X_{t+k}/\tilde{\phi}^k \end{aligned}$$

which in the limit becomes

$$X_t = -\frac{w_{t+1}}{\tilde{\phi}} - \frac{w_{t+2}}{\tilde{\phi}^2} - \cdots = -\sum_{j=1}^{\infty} \tilde{\phi}^{-j} w_{t+j} = -\sum_{j=0}^{\infty} \tilde{\phi}^{-j} \tilde{w}_{t+1+j},$$

where $\tilde{w}_t = w_t/\tilde{\phi}$. This is the *unique* stationary solution to the difference equation $X_t - \tilde{\phi}X_{t-1} = w_t$. The process is said to be non-causal since X_t depends on "future" errors w_s, $s > t$, rather than those in the past. If $|\tilde{\phi}| = 1$, no stationary solution exists.

Non-uniqueness of covariance The covariance formula (2) implies that

$$\text{Cov}(-\sum_{j=0}^{\infty} \tilde{\phi}^{-j} \tilde{w}_{t+1+h+j}, -\sum_{j=0}^{\infty} \tilde{\phi}^{-j} \tilde{w}_{t+1+j}) = \frac{\sigma^2}{\tilde{\phi}^2} \frac{(1/\tilde{\phi})^{|h|}}{1 - (1/\tilde{\phi})^2}$$

Thus, the non-causal process also has *same correlation* function as the more familiar process with coefficient $|1/\tilde{\phi}| < 1$ (the non-causal version has smaller error variance).

Not identified Either choice for $\phi \neq 1$ (ϕ and $1/\phi$) generates a solution of the first-order difference equation

$$X_t = \phi X_{t-1} + w_t \ .$$

If $|\phi| < 1$, we can find a solution via back-substitution. If $|\phi| > 1$, we obtain a stationary distribution via forward substitution. For a Gaussian process with mean zero, the likelihood is a function of the covariances. Since these two have the same correlations, the model is not identified. Either way, we cannot allow $|\phi| = 1$.

Moving averages: one more condition Such issues also appear in the analysis of moving averages. Consider the covariances of the two processes

$$X_t = w_t + \theta w_{t-1} \quad \text{and} \quad X_t = w_{t-1} + \theta w_{t-2}$$

The second incorporates a time delay. Since both are finite moving averages, both are stationary. Is the model identified? It is with the added condition that $\psi_0 = 1$.

Covariance generating function This function expresses the covariance of an ARMA process in terms of the polynomials $\phi(z)$ and $\theta(z)$. The moving average representation of the ARMA(p, q) process given by (8) combined with our fundamental result (2) implies that the covariances of $\{X_t\}$ are

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+|h|} \psi_j \tag{9}$$

where ψ_j is the coefficient of z^j in $\psi(z)$. The sum in (9) can be recognized as the coefficient of z^h in the product $\psi(z)\psi(z^{-1})$, implying that the covariance $\gamma(h)$ is the coefficient of z^h in

$$\begin{aligned} \Gamma(z) &= \sigma^2 \psi(z) \psi(z^{-1}) \\ &= \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}, \end{aligned}$$

which is known as the covariance generating function of the process.

Reciprocals of zeros The zeros of $\phi(1/z)$ are the *reciprocals* of those of $\phi(z)$. Hence, as far as the covariances are concerned, it does not matter whether the zeros go inside or outside the unit circle. They cannot lie *on* the unit circle.

Since both $\phi(z)$ (which has zeros outside the unit circle) and $\phi(1/z)$ (which has zeros inside the unit circle) both appear in the definition of $\Gamma(z)$, some authors state the conditions for stationarity in terms of one polynomial *or* the other. In any case, no zero can lie *on* the unit circle.

Further identifiability issue: Common factors Suppose that the polynomials share a common zero r,

$$\theta(z) = \tilde{\theta}(z)(1 - z/r), \quad \phi(z) = \tilde{\phi}(z)(1 - z/r)$$

Then this term cancels in the covariance generating function. Thus, the process $\tilde{\phi}(B)X_t = \tilde{\theta}(B)w_t$ has the same covariances as the process $\phi(B)X_t = \theta(B)w_t$. To avoid this type of non-identifiability, we require that $\phi(z)$ and $\theta(z)$ have distinct zeros.

Invertibility of ARMA processes

Invertible (Defn 3.8) An *ARMA* process $\{X_t\}$ is invertible if there exists an absolutely summable sequence (or perhaps ℓ_2 sequence) $\{\pi_j\}$ such that

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

Conditions for invertible ARMA Assume that the polynomials $\phi(B)$ and $\theta(B)$ have no common zeros. The process $\{X_t\}$ is invertible if and only if the zeros of the moving average polynomial $\theta(B)$ lie *outside* the unit circle.

ARIMA processes

Nonstationary processes are common in many situations, and these would at first appear outside the scope of ARMA models (certainly by the definition of S&S). The use of differencing, via the operator $(1 - B)X_t = X_t - X_{t-1}$ changes this.

Differencing is the discrete-time version of differentiation. For example, differencing a process whose mean function $\mathbb{E}\, X_t = a + b\, t$ is trending in time removes this source of nonstationarity. For example, if $\{X_t\}$ is a stationary process, then differencing

$$Y_t = \alpha + \beta\, t + X_t \quad \Rightarrow \quad (1-B)Y_t = \beta + X_t - X_{t-1}$$

reveals the possibly stationary component of the process. (Note, however, that if X_t were stationary to begin with, the differences of X_t would not be stationary!)

ARIMA(p,d,q) models are simply ARMA(p,q) models with

$$X_t \text{ replaced by } (1-B)^d\, X_t$$

where $(1-B)^d$ is manipulated algebraically.

Long-memory processes are stationary (unlike ARIMA processes) and formed by raising the differencing operator to a fractional power, say $(1-B)^{1/4}$. With time, we will study these later.