ARMA Models

Overview

1. Modeling paradigm
2. Stationary linear processes
3. ARMA and ARIMA processes
4. Identifiability
5. Backshift operator, polynomials
6. Causal and invertible models

Paradigm

Modeling paradigm 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 resemble *iid* random noise.

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 $\{\epsilon_t\} \overset{\text{iid}}{\sim} Dist(0, \sigma^2)$ (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).

Prediction rationale If a model can reduce the data to iid noise, then the model captures all of the relevant structure, at least in the sense that it captures $\mathbb{E}\, X_t | X_{t-1}, X_{t-2}, \ldots$, so that $X_t = \mathbb{E}\,(X_t | X_{t-1}, ...) + \epsilon_t$.

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 a Taylor series) that expresses the process in terms of prior values. 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,

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

Other processes are *nonlinear*.

Invertibility Form of model suggests problems for estimation; it resembles a regression in which all of the explanatory variables are simply functions of the unobserved errors. The invertibility condition implies that we can also write 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

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

so that information in the current and past errors is equivalent to information in the current and past data (*i.e.*, the sigma fields 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.*, what n.n.d 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, $X_n \stackrel{a.s.}{\to} X$,
$$P\{\omega : \lim X_n = X\} = 1.$$
- In probability, $X_n \stackrel{P}{\to} X$,
$$\lim_n P\{\omega : |X_n - X| > \epsilon\} = 0$$
- 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

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

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

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

$$\gamma_Y(h) = \sigma_X^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}. \tag{1}$$

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?

ARMA Processes

Conflicting goals Obtain models that possess a *wide range* of covariance functions (1) and that characterize ψ_j as functions of a *few* parameters. We have seen several of these parsimonious models previously, *e.g.*,

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

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

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

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \quad (2)$$

   where $w_t \sim WN(0, \sigma_2)$.

Backshift operator Abbreviate the equation (2) using the backshift operator defined as $B^k X_t = X_{t-k}$. Using B, write (2) 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 is a dependence-preserving mapping that 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 sums of stationary processes also remain stationary.

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

$$X_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \ . \tag{3}$$

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 (1):

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

and zero otherwise.

Scope of MA(q) processes. An MA(q) process of finite order models a process that becomes uncorrelated beyond time separation q. Many of these have 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,

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

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),

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + w_t \tag{4}$$

The stationarity of a solution to (4) is less obvious because of the presence of "feedback" (beyond the AR(1) case considered previously). To investigate initially (more coming) we attempt to *construct* one. Idea of the construction is to make the AR process resemble a linear process since we know that a linear process is stationary.

Rationales for ARMA models ARMA models occur in several situations:

1. Wider range of covariance functions with fewer parameters;
2. Sampling a pth order continuous linear SDE produces a discrete-time ARMA$(p, p-1)$ model;
3. Adding uncorrelated observation random noise to an AR process produces an ARMA process;
4. A weighted mixture of lags of an AR(p) model is ARMA.

ARIMA Processes

Nonstationary processes are quite common in practice, and would at first appear outside the scope of ARMA models. 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.

Identifiability of ARMA processes

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

The most common example of a poorly identified model (in practice) 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.

Non-causal process. 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? The answer is yes, but it's weird because it runs *backwards* in time.

The idea is to 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,

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

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 (1) 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 much 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 likilihood is a function of the covariances. Since these two have the same correlations, the model is not identified.

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 $\psi_0 = 1$.

Backshift polynomials

Generalize We need to extend the identifiability ideas from the AR(1) problem to more complex processes. The key to this extension are the polynomials $\phi(z)$ and $\theta(z)$.

AR(1) case If we manipulate B algebraically in the conversion of the AR(1) to moving average form ($|\phi| < 1$), we obtain the same geometric representation:

$$\begin{aligned} \phi(B)X_t = w_t \Rightarrow X_t &= \frac{w_t}{1-\phi B} \\ &= (1 + \phi B + \phi^2 B^2 + \ldots) w_t \\ &= w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \ldots . \end{aligned} \tag{5}$$

ARMA(1,1) case In order to determine $\psi(z)$, notice that $\theta(z)/\phi(z) = \psi(z)$ implies that

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

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

AR(p) case Factor the polynomial $\phi(z)$ 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 (these come in conjugate pairs 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 (5) 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.

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

$$X_t = \frac{\theta(B)}{\phi(B)} w_t = \frac{\prod_j (1 - B/s_j)}{\prod_j (1 - B/z_j)} w_t = \psi(B) w_t \tag{6}$$

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.

Value of notation Consider the claim that an average of several lags of an autoregression forms an ARMA process. Backshift polynomials make it trival to show in the general case:

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

which has the rational form of an ARMA process.

Covariance generating function The representation of the ARMA(p,q) process given by (6) combined with our fundamental result (1) implies that the covariances of $\{X_t\}$ are

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+|h|}\psi_j \tag{7}$$

where ψ_j is the coefficient of z^j in $\psi(z)$. The sum in (7) 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

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

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

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.

Common factors Suppose that the polynomials share a common root (zero) r,

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

Then notice that 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 $\phi(B)X_t = \theta(B)w_t$.

Causal and invertible ARMA processes

Causal (Defn 3.7) A mean zero *ARMA* process $\{X_t\}$ is causal if there exists an absolutely summable sequence (or sometimes just an ℓ_2 sequence) $\{\psi_j\}$ such that

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

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

Conditions for causal ARMA Assume that the polynomials $\phi(z)$ and $\theta(z)$ have no common zeros. The process $\{X_t\}$ is stationary and causal if and only if the zeros of the autoregression polynomial $\phi(z)$ lie *outside* the unit circle (ie, $\phi(z) \neq 0$ for $|z| \leq 1$).

Invertible (Defn 3.8) An *ARMA* process $\{X_t\}$ is invertible if there exists an absolutely summable sequence (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.