Introduction to Time Series

Lecture 4: Linear time series models and the algebra of ARMA models

Matthew F. Dixon

Outline

1. Stationary causal linear processes
   - The autocovariance of a linear process
   - Causal and non-causal AR(1) models
2. ARMA models written using the backshift operator
   - The general ARMA model
   - Causal, invertible ARMA models
   - Reducible and irreducible ARMA models
3. AR(2) models and oscillating behavior
   - The deterministic skeleton of an ARMA model
   - A second order linear differential equation for the AR(2) model

Definition: Stationary causal linear process

- A **stationary causal linear process** is a time series models that can be written as
  [M7] $Y_n = \mu + g_0\epsilon_n + g_1\epsilon_{n-1} + g_2\epsilon_{n-2} + g_3\epsilon_{n-3} + g_4\epsilon_{n-4} + \dots$
  where $\{\epsilon_n, n = \dots, -2, -1, 0, 1, 2, \dots\}$ is a white noise process, defined for all integer timepoints, with variance $\mathrm{Var}(\epsilon_n) = \sigma^2$.
- We do not need to define any initial values. The doubly infinite noise process $\{\epsilon_n, n = \dots, -2, -1, 0, 1, 2, \dots\}$ is enough to define Y_n for every n as long as the infinite sum in [M7] converges.

Question 4.1. When does "stationary" here mean weak stationarity, and when does it mean strict stationary?

- **causal** in [M7] refers to $\{\epsilon_n\}$ being a causal driver of $\{Y_n\}$. The value of Y_n depends only on noise process values already determined by time n.
- This matching a requirement that causes must precede effects (`wikipedia.org/wiki/Bradford_Hill_criteria`).
- **linear** refers to linearity of Y_n as a function of $\{\epsilon_n\}$.

The autocovariance function for a linear process

$$\begin{aligned}
\gamma_h &= \mathrm{Cov}\big(Y_n, Y_{n+h}\big) &(1)\\
&= \mathrm{Cov}\left(\sum_{j=0}^{\infty} g_j \epsilon_{n-j}, \sum_{k=0}^{\infty} g_k \epsilon_{n+h-k}\right) &(2)\\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} g_j g_k \mathrm{Cov}\big(\epsilon_{n-j}, \epsilon_{n+h-k}\big) &(3)\\
&= \sum_{j=0}^{\infty} g_j g_{j+h}\sigma^2, \text{for } h \geq 0. &(4)
\end{aligned}$$

- For the autocovariance function to be finite, we need

$$\sum_{j=0}^{\infty} g_j^2 < \infty. \tag{5}$$

- We assumed we can move $\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}$ through Cov, discussed below.

- The interchange of expectation and infinite sums cannot be taken for granted. $\mathrm{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{Cov}(X_i, Y_j)$ is true for finite m and n, but not necessarily for infinite sums.
- In this course, we do not focus on interchange issues, but we try to notice when we make assumptions.
- The interchange of $\sum_0^\infty$ and Cov can be justified by requiring a stronger condition,

$$\sum_{j=0}^{\infty} |g_j| < \infty. \tag{6}$$

- The MA(q) model that we defined in equation M3 is an example of a stationary, causal linear process.
- The general stationary, causal linear process model, M7, can also be called the MA(∞) model.

A stationary causal linear solution to the AR(1) model, and a non-causal solution

The stochastic difference equation defining the AR(1) model,

[M8] $\quad Y_n = \phi Y_{n-1} + \epsilon_n.$

This has a causal solution,

[M8.1] $Y_n = \sum_{j=0}^{\infty} \phi^j \epsilon_{n-j}.$

It also has a non-causal solution,

[M8.2] $Y_n = \sum_{j=0}^{\infty} \phi^{-j} \epsilon_{n+j}.$

Question 4.2. Work through the algebra to check that M8.1 and M8.2 both solve equation M8.

Convergence of the infinite sums in M8.1 and M8.2

Question 4.3. For what values of ϕ is the causal solution M8.1 a convergent infinite sum, meaning that it converges to a random variable with finite variance? For what values is the non-causal solution M8.2 a convergent infinite sum?

Using the MA(∞) representation to compute the autocovariance of an ARMA model

Question 4.4. The linear process representation can be a convenient way to calculate autocovariance functions. Use the linear process representation in M8.1, together with our expression for the autocovariance of the general linear process M7 in equation 4, to get an expression for the autocovariance function of the AR(1) model.

The backshift operator and the difference operator

- The **backshift** operator B, also known as the **lag** operator, is given by
$$BY_n = Y_{n-1}. \tag{7}$$
- The **difference** operator $\Delta = 1 - B$ is
$$\Delta Y_n = (1-B)Y_n = Y_n - Y_{n-1}. \tag{8}$$
- Powers of the backshift operator correspond to different time shifts, e.g.,
$$B^2 Y_n = B(BY_n) = B(Y_{n-1}) = Y_{n-2}. \tag{9}$$
- We can also take a second difference,
$$\begin{aligned}\Delta^2 Y_n &= (1-B)(1-B)Y_n \\ &= (1-2B+B^2)Y_n = Y_n - 2Y_{n-1} + Y_{n-2}.\end{aligned} \tag{10}$$
- The backshift operator is linear, i.e.,
$$B(\alpha X_n + \beta Y_n) = \alpha B X_n + \beta B Y_n = \alpha X_{n-1} + \beta Y_{n-1} \tag{11}$$

- Backshift operators and their powers can be added, multiplied by each other, and multiplied by a scalar.
- Mathematically, backshift operators follow the same rules as the algebra of polynomial functions.
- For example, a distributive rule for $\alpha + \beta B$ is

$$(\alpha+\beta B)Y_n = (\alpha B^0+\beta B^1)Y_n = \alpha Y_n+\beta BY_n = \alpha Y_n+\beta Y_{n-1}. \quad (12)$$

- Mathematical properties we know about polynomials can be used to work with backshift operators.
- The AR, MA and linear process model equations can all be written in terms of polynomials in the backshift operator.
- Write $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p$, an order p polynomial, The equation M1 for the AR(p) model can be rearranged to give

$$Y_n - \phi_1 Y_{n-1} - \phi_2 Y_{n-2} - \cdots - \phi_p Y_{n-p} = \epsilon_n, \quad (13)$$

which can be written using the backshift operator as

[M1′] $\qquad \phi(B)Y_n = \epsilon_n.$

- Write $\psi(x)$ for a polynomial of order q,

$$\psi(x) = 1 + \psi_1 x + \psi_2 x^2 + \cdots + \psi_q x^q. \tag{14}$$

- The MA(q) equation M3 is equivalent to
  [M3′] $\qquad Y_n = \psi(B)\epsilon_n.$
- If $g(x)$ is a function defined by the Taylor series

$$g(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 + g_4 x^4 + \ldots, \tag{15}$$

  we can write the stationary causal linear process equation [M7] as
  [M7′] $\qquad Y_n = \mu + g(B)\epsilon_n.$
- Whatever you know or learn about working with Taylor series expansions helps you understand AR, MA and ARMA models.

The general ARMA model

- Putting together M1 and M3 suggests an **autoregressive moving average** ARMA(p,q) model given by

[M9] $Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + \cdots + \phi_p Y_{n-p} + \epsilon_n + \psi_1 \epsilon_{n-1} + \cdots + \psi_q \epsilon_{n-q}$,

  where $\{\epsilon_n\}$ is a white noise process. Using the backshift operator, we can write this more succinctly as

  [M9′] $$\phi(B) Y_n = \psi(B) \epsilon_n.$$

- Experience with data analysis suggests that models with both AR and MA components often fit data better than a pure AR or MA process.
- The general stationary ARMA(p,q) also has a mean μ so we get

  [M9″] $$\phi(B)(Y_n - \mu) = \psi(B) \epsilon_n.$$

Obtaining the MA(∞) representation and autocovariance of the ARMA(1,1) model, $Y_n = \phi Y_{n-1} + \epsilon_n + \psi \epsilon_{n-1}$.

Step 1. Put the model in the form $Y_n = g(B)\epsilon_n$.
Formally, we can write

$$(1-\phi B)Y_n = (1+\psi B)\epsilon_n, \tag{16}$$

which algebraically is equivalent to

$$Y_n = \left(\frac{1+\psi B}{1-\phi B}\right)\epsilon_n. \tag{17}$$

We can write this as

$$Y_n = g(B)\epsilon_n, \tag{18}$$

where

$$g(x) = \left(\frac{1+\psi x}{1-\phi x}\right). \tag{19}$$

Step 2. Work out the Taylor series expansion,

$$g(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 + \dots \tag{20}$$

You can do this either by hand or using your favorite math software.

Step 3. Obtain the MA(∞) representation, by putting (20) into (18).

Step 4. Obtain the autocovariance function, by using the general formula for an MA(∞) process.

Carrying out this calculation is an exercise.

Causal, invertible ARMA models

- We say that the ARMA model [M9] is **causal** if its MA(∞) representation is a convergent series.
- Recall that **causality** is about writing Y_n in terms of the driving noise process $\{\epsilon_n, \epsilon_{n-1}, \epsilon_{n-2}, \dots\}$.
- **Invertibility** is about writing ϵ_n in terms of $\{Y_n, Y_{n-1}, Y_{n-2}, \dots\}$.
- To assess causality, we consider the convergence of the Taylor series expansion of $\psi(x)/\phi(x)$ in the ARMA representation

$$Y_n = \frac{\psi(B)}{\phi(B)}\epsilon_n.$$

- To assess invertibility, we consider the convergence of the Taylor series expansion of $\phi(x)/\psi(x)$ in the inversion of the ARMA model given by

$$\epsilon_n = \frac{\phi(B)}{\psi(B)}Y_n.$$

- Fortunately, there is a simple way to check causality and invertibility without calculating the Taylor series.
- The ARMA model is causal if the AR polynomial,

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p$$

  has all its roots (i.e., solutions to $\phi(x) = 0$) outside the unit circle in the complex plane.
- The ARMA model is invertible if the MA polynomial,

$$\psi(x) = 1 + \psi_1 x + \psi_2 x^2 + \dots + \psi_q x^q$$

  has all its roots outside the unit circle.
- We can check the roots using the 'polyroot' function in R. For example, consider the MA(2) model, $Y_n = \epsilon_n + 2\epsilon_{n-1} + 2\epsilon_{n-2}$. The roots to $\psi(x) = 1 + 2x + 2x^2$ are

```
roots <- polyroot(c(1,2,2))
roots
```

- Finding the absolute value shows that we have two roots inside the unit circle, so this MA(2) model is not invertible.

  ```
  abs(roots)
  ```

- In this case, you should be able to find the roots algebraically. In general, numerical evaluation of roots is useful.

Question 4.5. It is undesirable to use a non-invertible model for data analysis. Why? Hint: One answer to this question involves diagnosing model misspecification.

Reducible and irreducible ARMA models

- We have seen the ARMA model written as a ratio of two polynomials,

$$Y_n = \frac{\phi(B)}{\psi(B)}\epsilon_n. \quad (21)$$

- If the two polynomials $\phi(x)$ and $\psi(x)$ share a common factor, it can be canceled out without changing the model.
- The **fundamental theorem of algebra** says that every polynomial $\phi(x) = 1 - \phi_1 x - \cdots - \phi_p x^p$ of degree p can be written in the form

$$(1 - x/\lambda_1) \times (1 - x/\lambda_2) \times \cdots \times (1 - x/\lambda_p), \quad (22)$$

  where $\lambda_{1:p}$ are the p roots of the polynomial, which may be real or complex valued.
- The Taylor series expansion of $\phi(B)^{-1}$ is convergent if and only if $(1 - B/\lambda_i)^{-1}$ has a convergent expansion for each $i \in 1:p$. This happens if $|\lambda_i| > 1$ for each i.

- The polynomials $\phi(x)$ and $\psi(x)$ share a common factor if, and only if, they share a common root.
- It is not clear, just from looking at the model equations, that

$$Y_n = \frac{5}{6}Y_{n-1} - \frac{1}{6}Y_{n-2} + \epsilon_n - \epsilon_{n-1} + \frac{1}{4}\epsilon_{n-2} \tag{23}$$

is **exactly the same model** as

$$Y_n = \frac{1}{3}Y_{n-1} + \epsilon_n - \frac{1}{2}\epsilon_{n-1}. \tag{24}$$

- To see this, you have to do the math! We see that the second of these equations is derived from the first by canceling out the common factor $(1 - 0.5B)$ in the ARMA model specification.

```
list(AR_roots=polyroot(c(1,-5/6,1/6)),
     MA_roots=polyroot(c(1,-1,1/4)))
```

The deterministic skeleton: Using differential equation to study ARMA models

- Non-random physical processes evolving through time have been modeled using differential equations ever since the ground-breaking work by Sir Isaac Newton (1687).
- We have to attend to the considerable amount of randomness (unpredictability) present in data and systems we want to study.
- However, it is helpful to study a related deterministic systems.
- The **deterministic skeleton** of a time series model is the non-random process obtained by removing randomness from a stochastic model.
- For a discrete-time model, we can define a continuous-time deterministic skeleton by replacing the discrete-time difference equation with a differential equation.
- Rather than deriving a deterministic skeleton from a stochastic time series model, we can instead add stochasticity to a deterministic model to get a model that can explain non-deterministic phenomena.

Example: Oscillatory behavior modeled using an AR(2) process

- In physics, a basic model for processes that oscillate (springs, pendulums, vibrating machine parts, etc) is simple harmonic motion.
- The differential equation for a simple harmonic motion process $x(t)$ is

[M10] $$\frac{d^2}{dt^2}x(t) = -\omega^2 x(t).$$

- This is a second order linear differential equation with constant coefficients. Such equations have a closed form solution. You may already know that the solution to M10 is **sinusoidal**.
- Finding the solution to a linear differential equation is very similar to the task of solving difference equations which is useful elsewhere in time series analysis. It also gives a chance to review complex numbers. Let's see how it is done.

1. Look for solutions of the form $x(t) = e^{\lambda t}$. Substituting this into the differential equation [M10] we get

$$\lambda^2 e^{\lambda t} = -\omega^2 e^{\lambda t}. \tag{25}$$

2. Canceling the term $e^{\lambda t}$, we see that this has two solutions, with

$$\lambda = \pm\omega i, \qquad \text{where } i = \sqrt{-1}. \tag{26}$$

3. The linearity of the differential equation means that if $y_1(t)$ and $y_2(t)$ are two solutions, then $ay_1(t) + by_2(t)$ is also a solution for any a and b. So, the **general solution** to M10 is

$$x(t) = ae^{i\omega t} + be^{-i\omega t}. \tag{27}$$

Here, a and b could be complex numbers.

4. We may suspect that $x(t) = ae^{i\omega t} + be^{-i\omega t}$ is sinusoidal by recalling the identities

$$\sin(\omega t) = \frac{1}{2i}\big(e^{i\omega t} - e^{-i\omega t}\big), \qquad \cos(\omega t) = \frac{1}{2}\big(e^{i\omega t} + e^{-i\omega t}\big). \tag{28}$$

5. Typically for physical systems we want $x(t)$ to be real, so we know that the complex part of (27) is zero. This requires $a = (A/2)e^{i\phi}$ and $b = -(A/2)e^{-i\phi}$ for real A and ϕ. The algebra to show this is not critical for this course, but is a good exercise if you enjoy it. The factor of $1/2$ is arbitrary. This gives

$$x(t) = \frac{A}{2}\left(e^{i(\omega t+\phi)} - e^{-i(\omega t+\phi)}\right). \tag{29}$$

6. Putting together (29) and (28) we get

$$x(t) = A\sin(\omega t + \phi), \tag{30}$$

which explains why the factor of $1/2$ in (29) is convenient.

Frequency, amplitude and phase for $x(t) = A\sin(\omega t + \phi)$

- ω is called the **frequency**, and ϕ is called the **phase**.
- Angle is usually measured in **radians**, so the units of ω are radians per unit time, and units of ϕ are radians.
- The **period** is $2\pi/\omega$, the time for one cycle.
- A is called the **amplitude**.
- The frequency of the oscillation is determined by ω in M10, but the amplitude and phase are unspecified constants which may be determined by initial conditions.
- It may be convenient to rescale to **cycles per unit time**,

$$x(t) = A\sin\left(2\pi(\omega' t + \phi')\right) \tag{31}$$

where $\omega' = \omega/2\pi$, $\phi' = \phi/2\pi$.

- A discrete time version of M10 is a deterministic linear difference equation, replacing $\frac{d^2}{dt^2}$ by the second difference operator, $\Delta^2 = (1-B)^2$. This corresponds to a deterministic model equation,

$$\Delta^2 y_n = -\omega^2 y_n.$$

- Adding white noise, and expanding out $\Delta^2 = (1-B)^2$, we get a stochastic model,

[M11] $$Y_n = \frac{2Y_{n-1}}{1+\omega^2} - \frac{Y_{n-2}}{1+\omega^2} + \epsilon_n.$$

- Model M11 may be appropriate to describe systems that have semi-regular but somewhat erratic fluctuations, called **quasi-periodic** behavior. Such behavior is evident in business cycles or wild animal populations.

We look at a simulation from M11 with $\omega = 0.1$ and $\epsilon_n \sim \mathrm{iid}\, N[0,1]$. From our exact solution to the deterministic skeleton, we expect that a typical period of the oscillations should be $2\pi/\omega \approx 60$.

```
<< quasi_periodic_code,echo=T,eval=F>>=
omega <- 0.1
ar_coefs <- c(2/(1+omega^2), - 1/(1+omega^2))
X <- arima.sim(list(ar=ar_coefs),n=500,sd=1)
par(mfrow=c(1,2))
plot(X)
plot(ARMAacf(ar=ar_coefs,lag.max=500),type="l",ylab="ACF of X")
@
set.seed(8395200)
<<quasi_periodic_code>>
```

- Quasi-periodic fluctuations are **phase locked** when the random perturbations are not able to knock the oscillations away from being close to their initial phase.
- Eventually, the randomness should mean that the process is equally likely to have any phase, regardless of the initial phase.

Question 4.6. What is the timescale on which the simulated model shows phase locked behavior? Equivalently, on what timescale does the phase of the fluctuations lose memory of its initial phase?

Further reading

- Section 1.5 (Eq. 1.5.10) of Brockwell and Davis (course textbook) introduces the backshift (aka backward shift) operator.
- Section 2.2 (see after Example 2.2.1) and, especially, Section 2.3 of Brockwell and Davis (course textbook) first introduces causality and invertibility in ARMA models.
- (optional) Section 3.6 of Time Series: Theory and Methods (by the same authors) gives a difference equation approach to calculating ARMA autocovariance functions which gives an opportunity to practice algebra similar to our study of the AR(2) model.