Chapter 4. Linear time series models and the algebra of ARMA models

Objectives

1. Putting autoregressive moving average (ARMA) models into the context of linear time series models.
2. Introduce the backshift operator, and see how it can be used to develop an algebraic approach to studying the properties of ARMA models.

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 sequence in [M7] converges.
- **stationary** since the construction is the same for each n.

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

weak stationary: constant mean & covariances shift-invariant.

strict stationary if the white noise process is also strict stationary, e.g. IID white noise.

- **causal** 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 for causation `wikipedia.org/wiki/Bradford_Hill_criteria` that causes must precede effects.
- **linear** refers to linearity of Y_n as a function of $\{\epsilon_n\}$.

The autocovariance function for a linear process

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

assuming this interchange is OK.

need $k=h+j$ otherwise the covariance is zero. we use this to remove the sum over k.

- In order for this autocovariance function to exist, we need

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

- Above, we assumed we can move $\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}$ through Cov.
- The interchange of expectation and infinite sums can't 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,

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

- 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

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

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

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,

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

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

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

- 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

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

- Therefore, 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

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

which is equivalent to

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

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

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

  the MA(q) equation M3 is equivalent to
  [M3′] $\quad Y_n = \psi(B)\epsilon_n.$
- Additionally, if $g(x)$ is a function defined by the Taylor series (`wikipedia.org/wiki/Taylor_series`) expansion

$$g(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 + g_4 x^4 + \dots,$$

  we can write the stationary causal linear process equation [M7] as
  [M7′] $\quad Y_n = \mu + g(B)\epsilon_n.$
- Whatever skills you have acquired, or acquire during this course, about working with Taylor series expansions will help you understand AR and MA models, and ARMA models that combine both autoregressive and moving average features.

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′] $\qquad \phi(B) Y_n = \psi(B) \epsilon_n$.
- 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 stationanary ARMA(p,q) also has a mean μ so we get
  [M9″] $\qquad \phi(B)(Y_n - \mu) = \psi(B) \epsilon_n$.

Question 4.5. Carry out the following steps to obtain the MA(∞) representation and the autocovariance function of the ARMA(1,1) model,

$$Y_n = \phi Y_{n-1} + \epsilon_n + \psi \epsilon_{n-1}.$$

1. Formally, we can write

$$(1 - \phi B) Y_n = (1 + \psi B)\epsilon_n,$$

which algebraically is equivalent to

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

We write this as

$$Y_n = g(B)\epsilon_n,$$

where

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

2. To make sense of $g(B)$ we work out the Taylor series expansion,

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

Do this either by hand or using your favorite math software.

3. From 1. we can get the MA(∞) representation. Then, we can apply the general formula for the autocovariance function of an MA(∞) process.

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 $\{\epsilon_n, \epsilon_{n-1}, \epsilon_{n-2}, \ldots\}$.
- **Invertibility** is about writing ϵ_n in terms of $\{Y_n, Y_{n-1}, Y_{n-2}, \ldots\}$.
- 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. We will state the result without proof.
- The ARMA model is causal if the AR polynomial,

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

  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,

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

  has all its roots (i.e., solutions to $\psi(x) = 0$) outside the unit circle in the complex plane.
- 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

## [1] -0.5+0.5i -0.5-0.5i
```

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

```r
abs(roots)
```

```
## [1] 0.7071068 0.7071068
```

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

Question 4.6. 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

- The ARMA model can be viewed as a ratio of two polynomials,
$$Y_n = \frac{\phi(B)}{\psi(B)}\epsilon_n.$$
- 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
$$(1 - x/\lambda_1) \times (1 - x/\lambda_2) \times \cdots \times (1 - x/\lambda_p),$$
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
  $Y_n = \frac{5}{6}Y_{n-1} - \frac{1}{6}Y_{n-2} + \epsilon_n - \epsilon_{n-1} + \frac{1}{4}\epsilon_{n-2}$
  is **exactly the same model** as
  $Y_n = \frac{1}{3}Y_{n-1} + \epsilon_n - \frac{1}{2}\epsilon_{n-1}$.

1 $\phi(x) = 1 - \frac{1}{3}x$

2 $\psi(x) = 1 - \frac{1}{2}x$

- 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)))
```

```
## $AR_roots
## [1] 2+0i 3+0i
##
## $MA_roots
## [1] 2+0i 2-0i
```

these two roots cancel.

we can use the same code on model 2. or, from 2, we can see by inspection that the roots are: AR: 3, MA: 2

Deterministic skeletons: Using differential equations to study ARMA models

- Non-random physical processes evolving through time have been modeled using differential equations ever since the ground-breaking work by Newton (1687).
- We have to attend to the considerable amount of randomness (unpredictability) present in data and systems we want to study.
- However, we want to learn a little bit from the extensive study of 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 work in reverse: we 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, which is fairly straightforward to compute once you know how.
- The solution method is very similar to the method for solving difference equations coming up elsewhere in time series analysis, so 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

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

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

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

2. 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, we have a general solution to [M10] given by

$$x(t) = Ae^{i\omega t} + Be^{-i\omega t}.$$

3. Using the two identities,

$$\sin(\omega t) = \frac{1}{2}\left(e^{i\omega t} - e^{-i\omega t}\right), \qquad \cos(\omega t) = \frac{1}{2}\left(e^{i\omega t} + e^{-i\omega t}\right),$$

we can rewrite the general solution as

$$x(t) = A\sin(\omega t) + B\cos(\omega t),$$

which can also be written as

$$x(t) = A\sin(\omega t + \beta).$$

Physics: simple harmonic motion.

For the solution in the form $x(t) = A\sin(\omega t + \beta)$,

- ω is called the **frequency**
- A is called the **amplitude** of the oscillation
- β is called the **phase**.
- The frequency of the oscillation is determined by [M10], but the amplitude and phase are unspecfied constants.
- Initial conditions can be used to specify A and β.

- 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,

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

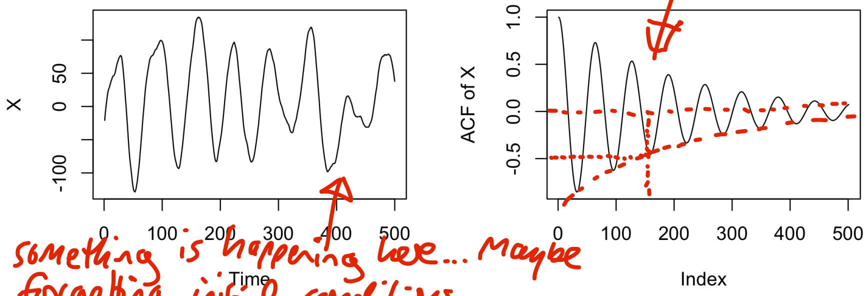
- Adding white noise, and expanding out $\Delta^2 = (1-B)^2$, we get a stochastic model,

  [M11] $\qquad Y_n = \frac{2Y_{n-1}}{1+\omega^2} - \frac{Y_{n-2}}{1+\omega^2} + \epsilon_n.$
- It seems reasonable to hope that model [M11] would be a good candidate to describe systems that have semi-regular but somewhat eratic fluctuations, called **quasi-periodic** behavior. Such behavior is evident in business cycles or wild animal populations.

Let's look at a simulation from [M11] with $\omega = 0.1$ and $\epsilon_n \sim \text{IID } N[0, 1]$. From our exact solution to the deterministic skeleton, we expect that the **period** of the oscillations (i.e., the time for each completed oscillation) should be approximately $2\pi/\omega$.

```
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")
```

the half life of decay of the ACF is about 150 time units.



something is happening here... maybe forgetting initial conditions.

- Quasi-periodic fluctuations are said to be "phase locked" as long as the random peturbations 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 4.7. 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?

Acknowledgments and License

- These notes build on previous versions at `ionides.github.io/531w16` and `ionides.github.io/531w18`.