STAT 153 - Introduction to Time Series Lecture Seventeen

Fall 2022, UC Berkeley

Aditya Guntuboyina

October 26, 2022

1 Recap from last class: Regimes of $AR(1)$

In the last class, we considered the AR(1) model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + Z_t. \tag{1}$$

The above difference equation needs to be supplemented with an initialization for the full specification of the model. One option is to consider the equation for $t = 1, 2, \dots$ with an initial condition $Y_0 = y_0$ at time $t = 0$. One can then explicitly solve Y_t for all $t \geq 1$ leading to the equation

$$Y_t = \phi_0 \left(1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{t-1}\right) + \left(Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \dots + \phi_1^{t-1} Z_1\right) + \phi_1^t y_0. \tag{2}$$

We then considered the following three regimes for the parameter ϕ_1:

Case One: $|\phi_1| < 1$: Here the values ϕ_1^t decay rapidly as t increases. Thus the dependence of (2) on the initial value y_0 decreases quickly with t. Further, when t is not too small, the first term is approximately

$$\phi_0 \left(1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{t-1}\right) \approx \phi_0 \left(1 + \phi_1 + \phi_1^2 + \dots\right) = \frac{\phi_0}{1 - \phi_1},$$

and the second term is approximately

$$Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \dots + \phi_1^{t-1} Z_1 \approx Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \dots + \phi_1^{t-1} Z_1 + \phi_1^t Z_0 + \phi_1^{t+1} Z_{-1} + \dots$$

where we assume that $Z_0, Z_{-1}, Z_{-2}, \dots$ are also i.i.d $N(0, \sigma^2)$ which are also independent from $Z_1, Z_2, \dots$. Therefore in this case, the behaviour of Y_t would be very close to the behaviour of

$$Y_t^* = \frac{\phi_0}{1 - \phi_1} + \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} \tag{3}$$

for all t that are not too close to zero. We have seen that the process Y_t^* is stationary with constant mean $\phi_0/(1 - \phi_1)$ and its covariance and correlation functions are given by

$$\mathrm{cov}(Y_t^*, Y_{t+h}^*) = \frac{\sigma^2 \phi_1^{|h|}}{1 - \phi_1^2} \quad \text{and} \quad \mathrm{corr}(Y_t^*, Y_{t+h}^*) = \phi_1^{|h|}$$

Case Two: $|\phi_1| > 1$: Here the right hand side of (2) will explode when t becomes large. This is thus an explosive process which will be appropriate for some datasets (see, for example, the housing price dataset that we used in the code of Lecture Fourteen).

Instead of initializing the recursion (1) at $t = 0$ and looking at Y_t for large t, suppose we initialize the recursion into the future at $t = N$ (for a large N) and then look at Y_t for t significantly smaller than N. Then it turns out that Y_t can be approximated by a stationary process. To see this, note that (1) is equivalent to

$$Y_{t-1} = -\frac{\phi_0}{\phi_1} + \frac{Y_t}{\phi_1} - \frac{Z_t}{\phi_1}.$$

Recursive using this for $t = N, N-1, \dots$, we obtain

$$Y_{N-k} = -\frac{\phi_0}{\phi_1}\left(1 + \frac{1}{\phi_1} + \frac{1}{\phi_1^2} + \dots + \frac{1}{\phi_1^{k-1}}\right) - \frac{Z_{N-k+1}}{\phi_1} - \frac{Z_{N-k+2}}{\phi_1^2} - \dots - \frac{Z_N}{\phi_1^k} + \frac{Y_N}{\phi_1^k}$$

for $k \geq 1$. As $1/\phi_1$ is smaller than 1 in absolute value, one can see that when k gets large, the right hand side above will be close to

$$\begin{aligned} &-\frac{\phi_0}{\phi_1}\left(1 + \frac{1}{\phi_1} + \frac{1}{\phi_1^2} + \dots + \frac{1}{\phi_1^{k-1}} + \dots\right) - \frac{Z_{N-k+1}}{\phi_1} - \frac{Z_{N-k+2}}{\phi_1^2} - \dots - \frac{Z_N}{\phi_1^k} - \frac{Z_{N+1}}{\phi_1^{k+1}} + \dots \\ &= \frac{\phi_0}{1-\phi_1} - \frac{Z_{N-k+1}}{\phi_1} - \frac{Z_{N-k+2}}{\phi_1^2} - \dots - \frac{Z_N}{\phi_1^k} - \frac{Z_{N+1}}{\phi_1^{k+1}} + \dots \end{aligned}$$

In other words, Y_t will be approximately equal to

$$Y_t' = \frac{\phi_0}{1-\phi_1} - \sum_{j=1}^{\infty} \frac{Z_{t+j}}{\phi_1^j} \tag{4}$$

for t sufficiently smaller than N. It can be easily checked that Y_t' is a stationary process. Thus when $|\phi_1| > 1$, there exists a stationary solulution to (1). However this solution is non-causal as Y_t depends on future values $Z_{t+1}, Z_{t+2}, \dots$. Further, this solution will be obtained only when the recursion is initalized at some future value. If the recursion is initialized in the past, then the process will be explosive and not stationary.

Case Three: $|\phi_1| = 1$. There are two cases here: $\phi_1 = 1$ and $\phi_1 = -1$. When $\phi_1 = 1$, (2) becomes

$$Y_t = t\phi_0 + (Z_t + Z_{t-1} + \dots + Z_1) + y_0.$$

This is a nonstationary process (its variance increases as $t\sigma^2$). It will not be possible to get a stationary solution to (1) with alternative initializations. The same thing will be true for $\phi_1 = -1$ as well.

2 Backshift Notation

A convenient piece of notation used while working with AR and MA models is the Backshift notation. Let B denote the *backshift operator* defined by

$$BY_t = Y_{t-1}, B^2 Y_t = Y_{t-2}, B^3 Y_t = Y_{t-3}, \dots$$

and similarly

$$BZ_t = Z_{t-1}, B^2 Z_t = Z_{t-2}, B^3 Z_t = Z_{t-3}, \dots.$$

Also let I denote the identity operator: $IY_t = Y_t$. More generally, we can define polynomial functions of the Backshift operator by, for example,

$$(I + B + 3B^2)Y_t = IY_t + BY_t + 3B^2Y_t = Y_t + Y_{t-1} + 3Y_{t-2}.$$

In general, for every polynomial $f(z)$, we can define $f(B)$. One can even extend this notation to negative powers of B which correspond to forward shifts. For example, $B^{-1}Y_t = Y_{t+1}, B^{-5}Y_t = Y_{t+5}$ and $(B^3 + 9B^{-2})Y_t = Y_{t-3} + 9Y_{t+2}$ etc.

In this notation, the defining equation $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t$ for the $AR(p)$ model can be written as $\phi(B)Y_t = \phi_0 + Z_t$ for the polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$.

The defining equation $Y_t = Z_t + \theta Z_{t-1}$ for the MA(1) model can be written as $Y_t = \theta(B)Z_t$ for the polynomial $\theta(z) = 1 + \theta_1 z$.

The defining equation $Y_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$ for the MA(q) model becomes $Y_t = \theta(B)Z_t$ for the polynomial $\theta(z) = 1 + \theta_1 z + \dots \theta_q z^q$.

3 $AR(1)$ solutions using Backshift Calculus

The two stationary solutions (3) and (4) to the AR(1) difference equation (1) for the two cases $|\phi_1| < 1$ and $|\phi_1| > 1$ can also be derived using formal operations that are sometimes known as Backshift Calculus. This is described in this section. First note that (1) can be written as

$$\phi(B)Y_t = \phi_0 + Z_t \qquad \text{where } \phi(z) = 1 - \phi_1 z.$$

Thus we can formally write

$$Y_t = \frac{1}{\phi(B)}(\phi_0 + Z_t).$$

Using

$$\frac{1}{\phi(z)} = \frac{1}{1 - \phi_1 z} = 1 + \phi_1 z + \phi_1^2 z^2 + \phi_1^3 z^3 + \dots,$$

we obtain

$$\begin{aligned} Y_t &= \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right)(\phi_0 + Z_t) \\ &= \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right)\phi_0 + \left(I + \phi_1 B + \phi_1^2 B^2 + \dots\right) Z_t \\ &= \left(1 + \phi_1 + \phi_1^2 + \dots\right)\phi_0 + \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} = \frac{\phi_0}{1 - \phi_1} + \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} = Y_t^*. \end{aligned}$$

When $|\phi_1| > 1$, the above process does not make sense. So we expand $1/\phi(z)$ in the following alternative way:

$$\begin{aligned} \frac{1}{\phi(z)} &= \frac{1}{1 - \phi_1 z} \\ &= \frac{-1}{\phi_1 z}\left(1 - \frac{1}{\phi_1 z}\right)^{-1} \\ &= \frac{-1}{\phi_1 z}\left(1 + \frac{1}{\phi_1 z} + \frac{1}{\phi_1^2 z^2} + \dots\right) = -\frac{z^{-1}}{\phi_1} - \frac{z^{-2}}{\phi_1^2} - \frac{z^{-3}}{\phi_1^3} - \dots. \end{aligned}$$

We thus get

$$
\begin{aligned}
Y_t &= \frac{1}{\phi(B)}\,(\phi_0 + Z_t) \\
&= \left(-\frac{B^{-1}}{\phi_1} - \frac{B^{-2}}{\phi_1^2} - \frac{B^{-3}}{\phi_1^3} - \dots\right)(\phi_0 + Z_t) = \frac{\phi_0}{1-\phi_1} - \sum_{j=1}^{\infty} \frac{Z_{t+j}}{\phi_1^j} = Y_t'.
\end{aligned}
$$

This formal method is called Backshift Calculus and it works for higher order AR models as well.

4 Causal stationary AR models

An $AR(p)$ model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$$

is said to be causal stationary if the characteristic equation

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

has all roots with modulus strictly larger than 1. For such an $AR(p)$ model, we can represent Y_t in terms of a linear combination of present and past values of Z_t (in a manner that is analogous to (3)). To do this, first write the characteristic polynomial $\phi(z)$ in terms of its roots:

$$\phi(z) = (1 - a_1 z)(1 - a_2 z)\dots(1 - a_p z).$$

This means that the roots of ϕ are $1/a_1, 1/a_2, \dots, 1/a_p$. In the causal stationary regime, each $|a_i| < 1$ so that the modulus of every root is strictly bigger than 1. Using the Backshift notation, the $AR(p)$ model becomes

$$\phi(B)Y_t = \phi_0 + Z_t$$

so that

$$
\begin{aligned}
Y_t &= \frac{1}{\phi(B)}\,(\phi_0 + Z_t) \\
&= \frac{1}{(I - a_1 B)(I - a_2 B)\dots(I - a_p B)}\,(\phi_0 + Z_t) \\
&= (I - a_1 B)^{-1}(I - a_2 B)^{-1}\dots(I - a_p B)^{-1}\,(\phi_0 + Z_t) \\
&= (I + a_1 B + a_1^2 B^2 + \dots)(I + a_2 B + a_2^2 B^2 + \dots)\dots(I + a_p B + a_p^2 B^2 + \dots)\,(\phi_0 + Z_t)
\end{aligned}
$$

We now multiply things out and simplify to obtain an explicit expression for Y_t in terms of $Z_t, Z_{t-1}, Z_{t-2}, \dots$. This is illustrated in the following example.

Example 4.1 (Simple $AR(2)$)**.** *Consider the AR(2) difference equation:*

$$Y_t - Y_{t-1} + 0.5 Y_{t-2} = Z_t \tag{5}$$

where Z_t is i.i.d $N(0, \sigma^2)$. Show that this equation has a solution $\{Y_t\}$ that is causal and stationary. Also, explicitly represent the solution in terms of $\{Z_t\}$.

The characteristic equation corresponding to (5) *is*

$$\phi(z) = 1 - z + 0.5 z^2.$$

Its roots are $1-i$ and $1+i$. The modulus of both roots equals $\sqrt{2}$ which is strictly larger than 1. Thus (5) admits a stationary and causal solution. To represent the solution explicitly in terms of $\{Z_t\}$, we first write

$$\phi(z) = 1 - z + 0.5z^2 = \left(1 - \frac{z}{1-i}\right)\left(1 + \frac{z}{1+i}\right).$$

Then

$$\begin{aligned}
Y_t &= \left(I - \frac{B}{1-i}\right)^{-1}\left(I - \frac{B}{1+i}\right)^{-1} Z_t \\
&= \left(I + \frac{B}{1-i} + \frac{B^2}{(1-i)^2} + \dots\right)\left(I + \frac{B}{1+i} + \frac{B^2}{(1+i)^2} + \dots\right) Z_t \\
&= \left(I + B\left[\frac{1}{1-i} + \frac{1}{1+i}\right] + B^2\left[\frac{1}{(1-i)^2} + \frac{1}{(1+i)^2} + \frac{1}{(1-i)(1+i)}\right] + \dots\right) Z_t \\
&= Z_t + \left[\frac{1}{1-i} + \frac{1}{1+i}\right] Z_{t-1} + \left[\frac{1}{(1-i)^2} + \frac{1}{(1+i)^2} + \frac{1}{(1-i)(1+i)}\right] Z_{t-2} + \dots \\
&= Z_t + Z_{t-1} + 0.5Z_{t-2} + \dots
\end{aligned}$$

5 Moving Average (MA) models

The $MA(q)$ (Moving Average Model of order q) model is given by

$$Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$$

where $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model has $q+2$ parameters: μ, $\theta_1, \dots, \theta_q$ and σ^2. This model has been called the "Summation of Random Causes" by its inventor Slutzky in the original paper titled "The summation of random causes as the source of cyclic processes" published in Econometrica in 1937. Basically the Z_t's can be treated as random causes which are assumed to be independently and identically distributed. The actual observations Y_t's are consequences of these causes. The consequence for tine t depends on the cause for time t as well as the causes for times $t-1, \dots, t-q$. These different causes affect the consequence at time t differently depending on the values of $\psi_1, \dots, \psi_q$. Note that successive observations Y_t share some common causes leading to dependence between the successive values of Y_t. To determine the exact form of the dependence between different Y_t's, we can calculate the covariance function as:

$$\operatorname{Cov}(Y_t, Y_{t+h}) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & \text{for } 0 \le h \le q \\ 0 & \text{for } h > q \end{cases}$$

where $\theta_0 = 1$.

Given data $y_1, \dots, y_n$, how do we estimate the parameters of the $MA(q)$ model? We can write down the likelihood and then use maximum likelihood estimates. The joint density of $Y_1, \dots, Y_n$ is multivariate normal with mean vector $m := (\mu, \dots, \mu)^T$ and covariance matrix Σ where

$$\Sigma(i,j) = \operatorname{Cov}(Y_i, Y_j) = \begin{cases} \sigma^2 \sum_{l=0}^{q-|i-j|} \theta_l \theta_{l+|i-j|} & \text{for } 0 \le |i-j| \le q \\ 0 & \text{for } h > q \end{cases}$$

The likelihood is therefore

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n (\det \Sigma)^{1/2} \exp\left(-\frac{1}{2}(y-m)'\Sigma^{-1}(y-m)\right)$$

where y is the $n \times 1$ vector with components $y_1, \dots, y_n$. We shall study the problem of optimizing this with respect to the parameters $\mu, \theta_1, \dots, \theta_q, \sigma$ in the next class.

6 Recommended Reading for Today

1. Read Chapter 3 of the book by Shumway and Stoffer titled *Time Series Analysis and its applications* (Fourth Edition) for more information on the AR and MA models.