STAT 153 & 248 - Time Series
Lecture Twenty Two

Spring 2025, UC Berkeley

Aditya Guntuboyina

April 15, 2025

In the past few lectures, we studied the MA(q) and AR(p) models which are reviewed below.

1 MA(q)

The MA(q) model is given by:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

where, as always, $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This process is always stationary. Its ACF $\rho(h)$ equals 0 when $|h| > q$. The ACF is therefore considered a signature for the MA(q) model. Given an observed time series dataset $y_1, \dots, y_n$, one can decide whether to fit an MA model to the data by looking at the sample acf of the data. If the sample acf seems to become negligible after a certain lag q, then the MA(q) would be a good model for the dataset.

2 AR(p)

The AR(p) model is defined by the equation:

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \phi_0 + \epsilon_t \tag{1}$$

This is an implicit definition in the sense that y_t is defined as any set of random variables that satisfy (1). There are, in fact, multiple solutions to (1).

Whether AR(p) is stationary or not depends on the exact values of the parameters $\phi_1, \dots, \phi_p$, and also on which solution to (1) is being considered.

2.1 $p = 1$

Consider $p = 1$ when the equation is given by:

$$y_t - \phi_1 y_{t-1} = \phi_0 + \epsilon_t. \tag{2}$$

Here it is quite easy to answer questions of stationarity by separately considering the following three regimes:

1. $|\phi_1| < 1$: Here (2) admits a unique stationary solution that is given by the formula:

$$y_t = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}. \tag{3}$$

   Note that the infinite series above is well-defined because $|\phi_1| < 1$ which means that the powers ϕ_1^j decay rapidly. The process (3) has the property that ϵ_t is independent of $y_{t-1}, y_{t-2}, \dots$. The process (3) is referred to as **causal stationary** AR(1).

2. $|\phi_1| > 1$: Here (2) admits a unique stationary solution that is given by the formula:

$$y_t = \frac{\phi_0}{1-\phi_1} - \sum_{j=0}^{\infty} \frac{\epsilon_{t+j}}{\phi_1^j}. \tag{4}$$

   This infinite series is also well-defined because the powers ϕ_1^{-j} decay rapidly as $|\phi_1| > 1$. For (4), it is no longer true that ϵ_t is independent of $y_{t-1}, y_{t-2}, \dots$. Instead it is true that ϵ_t is independent of $y_{t+1}, y_{t+2}, \dots$. The process (4) is referred to as the **non-causal stationary** AR(1).

3. $|\phi_1| = 1$: Here (2) does not have a stationary solution. $|\phi_1| = 1$ refers to either $\phi_1 = 1$ or $\phi_1 = -1$. Of these two, the case $\phi_1 = 1$ is more commonly used. When ϕ_1, the equation (2) becomes

$$y_t - y_{t-1} = \phi_0 + \epsilon_t. \tag{5}$$

   This means that the **differenced series** $y_t - y_{t-1}$ is Gaussian white noise (plus a constant ϕ_0). When $\phi_0 = 0$, (5) is called the Random Walk model.

The AR(1) stationary solutions (3) (causal) and (4) (non-causal) can be derived directly from the defining equation (2) using Backshift calculus as follows. B denotes the Backshift operator satisfying $B^k y_t = y_{t-k}$, k here can be any integer positive or negative or zero (when $k = 0$, we denote B^0 by simply 1). The AR(1) equation, in backshift notation, becomes $\phi(B)y_t = \phi_0 + \epsilon_t$ where $\phi(z) = 1 - \phi_1 z$ and $\phi(B) = 1 - \phi_1 B$ (here $I = B^0$ is the identity operator). This means that

$$y_t = \frac{1}{\phi(B)} (\phi_0 + \epsilon_t).$$

We now make sense of $1/\phi(B)$. For this, we use the following:

$$\frac{1}{1-r} = 1 + r + r^2 + r^3 + \dots$$

which gives

$$\frac{1}{\phi(B)} = \frac{1}{1-\phi_1 B} = 1 + \phi_1 B + \phi_1^2 B^2 + \dots = \sum_{j=0}^{\infty} \phi_1^j B^j$$

so that

$$y_t = \frac{1}{\phi(B)} (\phi_0 + \epsilon_t) = \sum_{j=0}^{\infty} \phi_1^j B^j (\phi_0 + \epsilon_t) = \sum_{j=0}^{\infty} \phi_1^j B^j (\phi_0) + \sum_{j=0}^{\infty} \phi_1^j B^j (\epsilon_t) = \sum_{j=0}^{\infty} \phi_1^j + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}.$$

The above infinite sums only make sense when $|\phi_1| < 1$, and we get

$$y_t = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$$

which gives us the causal stationary solution (3).

When $|\phi_1| > 1$, this definition of $1/\phi(B)$ does not lead to anything meaningful. Here, there is a different formula that can be used for $1/\phi(B)$. This comes from:

$$\frac{1}{1-r} = -\frac{1}{r}\frac{1}{1-(1/r)} = -\frac{1}{r}\left(1 + \frac{1}{r} + \frac{1}{r^2} + \dots\right) = -\sum_{j=1}^{\infty} r^{-j}.$$

This gives

$$\frac{1}{\phi(B)} = -\sum_{j=1}^{\infty} \frac{B^{-j}}{\phi_1^j}$$

and so

$$\frac{1}{\phi(B)}(\phi_0 + \epsilon_t) = -\phi_0 \sum_{j=1}^{\infty} \frac{1}{\phi_1^j} - \sum_{j=1}^{\infty} \frac{B^{-j}\epsilon_t}{\phi_1^j} = \frac{\phi_0}{1-\phi_1} - \sum_{j=1}^{\infty} \frac{\epsilon_{t+j}}{\phi_1^j}$$

which is the non-causal stationary AR(1) in (4).

To recap, we have the following two formulae:

$$\frac{1}{1-\phi_1 B} = \sum_{j=0}^{\infty} \phi_1^j B^j \quad \text{and} \quad \frac{1}{1-\phi_1 B} = -\sum_{j=1}^{\infty} \frac{B^{-j}}{\phi_1^j}. \tag{6}$$

When $|\phi_1| < 1$, we use the first formula because it results in the powers ϕ_1^j which decay rapidly with j. When $|\phi_1| > 1$, we use the second formula because it results in powers ϕ_1^{-j} which again decay rapidly with j.

When $|\phi_1| = 1$, then neither of the two formulae in (6) lead to rapidly decaying coefficients (in other words, neither (3) nor (4) make sense). This is reasonable because, as was mentioned last class, (2) does not have a stationary solution when $|\phi_1| = 1$.

2.2 $p \geq 1$

The backshift method actually works for every $p \geq 1$ and gives us correct answers for causal, non-causal stationary regimes of AR(p). In backshift notation, the AR(p) difference equation becomes

$$\phi(B) y_t = \phi_0 + \epsilon_t \qquad \text{where } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

We can therefore 'solve' it by writing

$$y_t = \frac{1}{\phi(B)}\epsilon_t = \frac{1}{1 - \phi_1 B - \dots - \phi_p B^p}\epsilon_t$$

The next step is to make sense of $1/(1 - \phi_1 B - \dots - \phi_p B^p)$. It is natural here to factorize the polynomial $1 - \phi_1 B - \dots - \phi_p B^p$ into monomials, and then use (6). So we write

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = (1 - a_1 z)\dots(1 - a_p z). \tag{7}$$

so that

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

The numbers $a_1, \dots, a_p$ appearing in (7) are simply the reciprocals of the roots of $\phi(z)$ i.e., the roots of $\phi(z)$ are given by $1/a_1, \dots, 1/a_p$. Note here that some of the a_j's can be complex because the polynomial $1 - \phi_1 z - \dots - \phi_p z^p$ can have complex roots (even though all its coefficients are real).

We then get

$$y_t = \frac{1}{(1-a_1B)\dots(1-a_pB)}(\phi_0+\epsilon_t) = \prod_{k=1}^{p}\frac{1}{1-a_kB}(\phi_0+\epsilon_t).$$

For $1/(1-a_jB)$, we use (6). Specifically, if $|a_j|<1$, we use the first formula in (6) and when $|a_j|>1$, we use the second formula in (6) (when a_j is complex, $|a_j|$ denotes its modulus). Thus we get

$$y_t = \prod_{k:|a_k|<1}\left(\sum_{j=0}^{\infty}a_k^jB^j\right)\prod_{k:|a_k|>1}\left(\sum_{j=1}^{\infty}\frac{B^{-j}}{a_k^j}\right)(\phi_0+\epsilon_t). \tag{8}$$

Suppose now that every $|a_k|$ is strictly smaller than 1. Then

$$\begin{aligned}
y_t &= \prod_k\left(\sum_{j=0}^{\infty}a_k^jB^j\right)(\phi_0+\epsilon_t)\\
&= \left(\sum_{j_1=0}^{\infty}a_1^{j_1}B^{j_1}\right)\dots\left(\sum_{j_p=0}^{\infty}a_p^{j_p}B^{j_p}\right)(\phi_0+\epsilon_t)\\
&= \left(\sum_{j_1=0}^{\infty}\dots\sum_{j_p=0}^{\infty}a_1^{j_1}\dots a_p^{j_p}B^{j_1+\dots+j_p}\right)(\phi_0+\epsilon_t)\\
&= \phi_0\sum_{j_1=0}^{\infty}\dots\sum_{j_p=0}^{\infty}a_1^{j_1}\dots a_p^{j_p} + \sum_{j_1=0}^{\infty}\dots\sum_{j_p=0}^{\infty}a_1^{j_1}\dots a_p^{j_p}\epsilon_{t-j_1-\dots-j_p}.
\end{aligned}$$

Because each a_k above was assumed to have modulus strictly smaller than 1, the series above involves rapidly decaying powers of a_k so it makes sense. The formula above writes y_t in terms of $\epsilon_t, \epsilon_{t-1}, \dots$. It can be checked that this is a causal, stationary solution. By collecting terms where $j_1+\dots+j_p=j$ for each $j=0,1,\dots$, we can write this solution as

$$y_t = \mu + \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$$

for some $\mu, \psi_1, \psi_2, \dots$.

If $|a_k|>1$ for even one k, then it is easy to see that (8) would give

$$y_t = \mu + \sum_{j=-\infty}^{\infty}\psi_j\epsilon_{t-j}$$

for some $\psi_j, j=\dots,-3,-2,-1,0,1,2,3,\dots$. In this case, the formula for y_t involves future values of ϵ_t so this is a non-causal stationary solution.

Finally if $|a_k|=1$ even for one k, then we cannot make sense of $1/(1-a_kB)$ (both the formulae in (6) fail to work). This hints that a stationary solution might not exist in this regime. This indeed turns out to be true.

Here is a summary of the discussion above: To determine the nature of the solutions of the AR(p) equation (1), first compute the roots $z_1,\dots,z_p$ of $\phi(z)=1-\phi_1z-\dots-\phi_pz^p$ and set $a_j=1/z_j$ for $j=1,\dots,p$.

1. If $|a_k|\neq 1$ for every k (i.e., no root of $\phi(z)$ has modulus equal to 1), then there exists a unique stationary solution to (1).

2. If $|a_k| < 1$ for every k, then the unique stationary solution is causal i.e., it can be written as $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ (the right hand side of this formula does not involve ϵ_s for any future time $s > t$).

3. If $|a_k| < 1$ for some k and $|a_k| > 1$ for some other k, then the unique stationary solution will be of the form $y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$. This is called a non-causal solution as the formula for y_t involves future values ϵ_s for $s > t$.

2.3 The Box-Jenkins Modeling Philosophy

G. Box and G. Jenkins (two researchers who developed ARIMA models) recommended the following:

1. Only work with causal stationary AR models (actually also ARMA models that we shall study soon after).

2. If the data are such that stationary models are not a good fit, then preprocess the data using differencing. After appropriate differencing, fit causal stationary ARMA models.

We shall see some real examples of this approach.

3 ARMA(p, q) models

The combination of ideas behind the AR and MA models are generalized to obtain ARMA models. The ARMA(p, q) model is defined by the equation:

$$(y_t - \mu) - \phi_1 (y_{t-1} - \mu) - \cdots - \phi_p (y_{t-p} - \mu) = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

where, as usual, $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The unknown parameters in this model are $\mu, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and the noise standard deviation σ. In Backshift notation, we can write

$$\phi(B)(y_t - \mu) = \theta(B)\epsilon_t$$

where $\phi(B)$ and $\theta(B)$ are the AR and MA polynomials:

$$\phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

applied to the Backshift operator B.

It can be shown that if the AR polynomial $\phi(z)$ has all roots with modulus strictly larger than 1, then the ARMA(p, q) difference equation has a stationary and causal solution. This solution can be written in the form:

$$y_t = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \ldots.$$

These coefficients $\{\psi_j\}$ can be explicitly determined in terms of $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ by solving the equation: $\psi(z) = \theta(z)/\phi(z)$ (here $\psi(z) := \psi_0 + \psi_1 z + \psi_2 z^2 + \ldots$) which can be done by writing

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = \phi(z) \times \psi(z) = (1 - \phi_1 z - \cdots - \phi_p z^p)\left(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots\right)$$

and then equating the coefficients of z^j on both sides for $j = 0, 1, \ldots$ to get

$$1 = \psi_0, \quad \theta_1 = \psi_1 - \psi_0 \phi_1, \quad \theta_2 = \psi_2 - \psi_1 \phi_1 - \psi_0 \phi_2, \quad \theta_3 = \psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 - \phi_3 \psi_0, \quad \ldots$$

ARMA(p, q) models generalize both AR(p) and MA(q) models. When $p = q = 0$ (i.e., when $\phi(z) = 1$ and $\theta(z) = 1$), we obtain the white noise model. When $p = 0$ (i.e., when $\phi(z) = 1$), we get the MA(q) model. When $q = 0$ (i.e., when $\theta(z) = 1$), we get the AR(p) model.

The ACF and PACF functions of causal stationary ARMA(p, q) models when both p and q are nonzero are more complicated compared to those of AR(p) and MA(q) models. In particular, neither the ACF nor the PACF cuts off after a certain lag for ARMA(p, q) models with both $p, q \geq 1$.

4 Additional Optional Reading

1. Sections 3.1, 3.2, 3.3, 3.4 of Shumway-Stoffer 4th edition.