STAT 153 & 248 - Time Series
Lecture Twenty One

Fall 2025, UC Berkeley

Aditya Guntuboyina

November 13, 2025

Last week, we mainly discussed AR models, and we briefly saw MA models. In the next few lectures, we will complete our discussion of MA and AR models (and also ARMA and ARIMA models). Let us start today with MA models.

1 MA(q) models

Given a positive integer $q \geq 1$, the Moving Average model with order q (denoted by MA(q)) is defined by the equation:

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{1}$$

where $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The MA($q$) model has $q+2$ unknown parameters which are estimated from observed data: $\mu, \theta_1, \dots, \theta_q, \sigma$.

The MA(q) 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 ϵ_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 time 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 $\theta_1, \dots, \theta_q$. Note that successive observations y_t share some common causes leading to dependence between the successive values of y_t.

The MA(q) model is always causal stationary (no matter what specific values its parameters $\mu, \theta_1, \dots, \theta_q, \sigma$ take; in this respect, the MA(q) is different from AR(p) which can be causal-stationary or not depending on specific values of the parameters).

Here is how the causal stationarity of MA(q) is proved. The causality follows straight from the definition because y_t is written in terms of the present and past ϵ-values $\epsilon_t, \epsilon_{t-1}, \dots$. For stationarity, we need to compute $\mathbb{E} y_t$ and $\text{cov}(y_t, y_{t+h})$:

1. The mean of y_t is clearly equal to μ (so it is constant in t).

2. The covariance between y_t and y_{t+h} is given by:

$$
\begin{aligned}
&\text{cov}(y_t, y_{t+h}) \\
&= \text{cov}\left(\mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}, \mu + \epsilon_{t+h} + \theta_1\epsilon_{t+h-1} + \theta_2\epsilon_{t+h-2} + \cdots + \theta_q\epsilon_{t+h-q}\right) \\
&= \text{cov}\left(\mu + \sum_{j=0}^{q}\theta_j\epsilon_{t-j}, \mu + \sum_{k=0}^{q}\theta_k\epsilon_{t+h-k}\right) \\
&= \sum_{j=0}^{q}\sum_{k=0}^{q}\theta_j\theta_k\text{cov}\left(\epsilon_{t-j}, \epsilon_{t+h-k}\right).
\end{aligned}
$$

Note that, in the sum $\sum_{j=0}^{q}\theta_j\epsilon_{t-j}$, we take $\theta_0 = 1$. Because $\{\epsilon_t\}$ is Gaussian white noise, the covariance $\text{cov}\left(\epsilon_{t-j}, \epsilon_{t+h-k}\right)$ equals zero unless $t-j = t+h-k$ i.e., $k = j+h$. So we need the three conditions $0 \le j \le q$, $0 \le k \le q$ as well as $k = j+h$. If $h > q$, it is clear that this is not possible for any j, k. So we have $\text{cov}(y_t, y_{t+h})$ equals zero when $h > q$. When $0 \le h \le q$, we have $0 \le j \le q$ and $0 \le j+h \le q$ which implies $0 \le j \le q-h$. We then get

$$\text{cov}(y_t, y_{t+h}) = \sigma^2\sum_{j=0}^{q-h}\theta_j\theta_{j+h}.$$

We thus have:

$$\text{cov}(y_t, y_{t+h}) = \begin{cases} \sigma^2\sum_{j=0}^{q-h}\theta_j\theta_{j+h} & : 0 \le h \le q \\ 0 & : h > q \end{cases}$$

The above covariance does not depend on t which shows that the MA(q) model is stationary.

Based on the above calculation, the ACVF of MA(q) is:

$$\gamma(h) = \begin{cases} \sigma^2\sum_{j=0}^{q-h}\theta_j\theta_{j+h} & : 0 \le h \le q \\ 0 & : h > q \end{cases}$$

The ACF $\rho(h) = \gamma(h)/\gamma(0)$ equals:

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h}\theta_j\theta_{j+h}}{\sum_{j=0}^{q}\theta_j^2} & : 0 \le h \le q \\ 0 & : h > q \end{cases}$$

The simplest of these MA(q) models is MA(1) (i.e., $q = 1$):

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}. \tag{2}$$

The ACF of MA(1) is:

$$\rho(h) = \begin{cases} 1 & : h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & : h = 1 \\ 0 & : h > 1 \end{cases}$$

In order to assess whether the MA(q) model is suitable for a particular dataset, a simple way is to compute the sample correlation between y_t and y_{t+h} for each value of h (in other words, the correlation between $(a_t, b_t), t = 1, \ldots, n-h$ where $a_t = y_t$ and $b_t = y_{t+h}$). The sample ACF is defined as a slight modification of this correlation.

2 Sample ACF

For a fixed value of h, the correlation coefficient between $(a_t, b_t), t = 1, \dots, n-h$ where $a_t = y_t$ and $b_t = y_{t+h}$ is given by:

$$\frac{\sum_{t=1}^{n-h}(a_t - \bar{a})(b_t - \bar{b})}{\sqrt{\sum_{t=1}^{n-h}(a_t - \bar{a})^2}\sqrt{\sum_{t=1}^{n-h}(b_t - \bar{b})^2}} = \frac{\sum_{t=1}^{n-h}(y_t - \bar{a})(y_{t+h} - \bar{b})}{\sqrt{\sum_{t=1}^{n-h}(y_t - \bar{a})^2}\sqrt{\sum_{t=1}^{n-h}(y_{t+h} - \bar{b})^2}}$$

where

$$\bar{a} = \frac{1}{n-h}\sum_{t=1}^{n-h} a_t = \frac{1}{n-h}\sum_{t=1}^{n-h} y_t \quad \text{and} \quad \bar{b} = \frac{1}{n-h}\sum_{t=1}^{n-h} b_t = \frac{1}{n-h}\sum_{t=1}^{n-h} y_{t+h}$$

This correlation can be simplified slightly by making the following approximations:

$$\bar{a} \approx \bar{y} \quad \bar{b} \approx \bar{y} \quad \sum_{t=1}^{n-h}(y_t - \bar{a})^2 \approx \sum_{t=1}^{n}(y_t - \bar{y})^2 \quad \text{and} \sum_{t=1}^{n-h}(y_{t+h} - \bar{b})^2 \approx \sum_{t=1}^{n}(y_t - \bar{y})^2$$

which are reasonable when h is very small compared to n. Making these approximations lead to the following definition of the sample ACF:

$$r_h := \frac{\sum_{t=1}^{n-h}(y_t - \bar{y})(y_{t+h} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad \text{for } h = 0, 1, 2, \dots$$

Note that r_0 is always equal to 1.

The sample ACF $r_h, h \geq 0$ can be computed for any time series dataset, although it is only useful for data for which stationary models are appropriate.

The sample ACF is particularly useful for determining the order q for fitting an MA(q) model. For an MA(q) model, we have seen in the previous section that the theoretical ACF $\rho(h)$ becomes exactly zero when $h > q$. This suggests that if the sample ACF r_h for a particular dataset becomes small (not exactly zero because of randomness) when h exceeds a particular q, then MA(q) is probably a good model for that dataset. This diagnostic is very commonly used when working with MA models.

3 On Parameter Estimation in MA(q)

Even though MA(q) models are, in some ways, simpler than AR(p) models, parameter estimation is much more complicated compared to AR(p). We will not study this topic (and simply rely on the `ARIMA` function for fitting these models to data). But here, I will just illustrate the difficulties involved in the simplest case $q = 1$ i.e., the MA(1) model.

For estimation, the main step is to write down the likelihood. 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 Σ equals the $n \times n$ matrix whose $(i,j)^{th}$ entry is given by

$$\Sigma(i,j) = \begin{cases} \sigma^2\left(1+\theta^2\right) & \text{when } i = j \\ \sigma^2\theta & \text{when } |i-j| = 1 \\ 0 & \text{for all other } (i,j) \end{cases}$$

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 $y_1, \dots, y_n$. This is a function of the unknown parameters μ, θ, σ which can be estimated by maximizing the logarithm of the likelihood. The presence of Σ^{-1} makes this computationally expensive. Some (exact or approximate) formula should be used for Σ^{-1} so that one does not need to invert an $n \times n$ matrix every time the log-likelihood is to be computed.

4 AR(p) models

The AR(p) model is:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t. \tag{3}$$

Unlike the MA(q) model, the AR(p) model can be stationary or not depending on the specific values of the parameters $\phi_1, \dots, \phi_p$. If $\phi_1, \dots, \phi_p$ are such that every root of the AR polynomial:

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

has modulus strictly larger than 1, then (3) has a causal stationary solution. This solution can be written in the form:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{4}$$

for some μ and coefficients $\psi_1, \psi_2, \dots$ satisfying $\sum_j |\psi_j| < \infty$. There are inbuilt functions in `statsmodels` which output values of $\mu, \psi_0, \psi_1, \dots$ given $\phi_0, \phi_1, \dots, \phi_p$.

The representation (4) resembles the MA equation (1) except now $q = \infty$. However the coefficients ψ_j are all explicit functions of only the $p+1$ parameters $\phi_0, \dots, \phi_p$.

5 The Box-Jenkins Modeling Philosophy

G. Box and G. Jenkins (two researchers who developed ARIMA models) recommended avoiding working with AR models which are **NOT** causal-stationary. Specifically, their recommendation is:

1. Only work with MA models (which are always causal stationary) and causal stationary AR models.
2. If the data are such that stationary models are not a good fit, then preprocess the data using differencing (possibly after applying a preliminary transformation such as log). After appropriate differencing, fit MA models or causal stationary AR models.

More generally Box and Jenkins recommended working with causal stationary ARMA models (after preprocessing the data). We shall look at ARMA models in the next lecture but they are given by the formula:

$$y_t - \phi_0 - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$

The left hand side resembles the AR(p) model and the right hand side resembles the MA(q) model. This is the reason why this model is called ARMA(p, q). Clearly ARMA(p, 0) = AR(p) and ARMA(0, q) = MA(q).

We say that y_t satisfies the ARIMA(p, d, q) model if x_t satisfies the ARMA(p, q) model where x_t is the time series obtained by differencing y_t d times (first difference is given by $y_t - y_{t-1}$, second difference is $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$ and so on).

Note therefore that ARIMA(p, 0, q) = ARMA(p, q), ARIMA(p, 0, 0) = AR(p) and ARIMA(0, 0, q) = MA(q).