Moving Average

Arif Nurwahid

Seminar on 24 Jan 2023

Moving Average Models (MA models)

Loosely speaking, a **moving average** term in a time series model is a **past error** (multiplied by a coefficient).

Definition 1

Let $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. The following equation is called the **Moving Average** model of order *q* and denote by **MA(*q*)**:

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

where $\mu, \theta_1, \ldots, \theta_q \in \mathbb{R}$ and $\theta_q \neq 0$.

The **1st order moving average model**, denoted by MA(1) is: $X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1}$

The **2nd order moving average model**, denoted by MA(2) is:
$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$

Theoretical Properties of a Time Series with an MA(1) Model

- Mean is $E[X_t] = \mu$
- Variance is $Var[X_t] = \sigma_\epsilon^2(1 + \theta_1^2)$
- Autocorrelation function (ACF) is

  $\rho_1 = \dfrac{\theta_1}{1 + \theta_1^2}$, and $\rho_k = 0$ for $k \geq 2$

> Note!

That the only nonzero value in the theoretical ACF is for lag 1. All other autocorrelations are 0. Thus a sample ACF with a significant autocorrelation only at lag 1 is an indicator of a possible MA(1) model.

Proof of Properties of MA(1)

- Mean is $E[X_t] = E[\mu + \epsilon_t + \theta_1 \epsilon_{t-1}] = \mu + 0 + (\theta_1)(0) = \mu$
- Variance is $Var[X_t] = Var[\mu + \epsilon_t + \theta_1 \epsilon_{t-1}] = 0 + Var[\epsilon_t] + Var[\theta_1 \epsilon_{t-1}] = \sigma_\epsilon^2 (1 + \theta_1^2)$
- ACF: Consider the covariance of X_t and X_{t-k}, that is $E[(X_t - \mu)(X_{t-k} - \mu)]$, which equals

$$E[(\epsilon_t + \theta_1 \epsilon_{t-1})(\epsilon_{t-k} + \theta_1 \epsilon_{t-k-1})] = E[\epsilon_t \epsilon_{t-k} + \theta_1 \epsilon_{t-1} \epsilon_{t-k} + \theta_1 \epsilon_t \epsilon_{t-k-1} + \theta_1^2 \epsilon_{t-1} \epsilon_{t-}$$

When $k = 1$, the previous equation equals to $\theta_1 \sigma_\theta^2$. For any $k \geq 2$, the previous expression = 0. The reason is that, by definition of independence of the ϵ_t, $E[\epsilon_j \epsilon_k] = 0$ for any $j \neq k$. Further, because the ϵ_t have mean 0, $E[\epsilon_j \epsilon_j] = E[\epsilon_j^2] = \sigma_\epsilon^2$.

For a time series, $\rho_k = \dfrac{\text{Covariance for lag } k}{\text{Variance}}$.

Apply this to get the desired ACF above.

PACF of MA(1)

Lag 1

$\phi_{11} = \rho_1 = \theta_1 1 + \theta_1^2$

Lag 2

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

Lag = k

$$\phi_{kk} = \frac{(-1)^{k+1} \theta_1^k}{\sum_{i=0}^{k} \theta_1^{2i}}$$

Example

Suppose that an MA(1) model is $X_t = \epsilon_t + .7\epsilon_{t-1}$, where $\epsilon_t \sim WN(0, 1)$.

Thus the coefficient $\theta_1 = 0.7$

The theoritical ACF is given by:

$\rho_1 = \dfrac{0.7}{1 + 0.7^2} = 0.4698$, and $\rho_k = 0$ for all lags $k \geq 2$.

In practice, a sample will not usually provude a clear pattern of ACF.

Here we will simulate 500 sample values using above model.

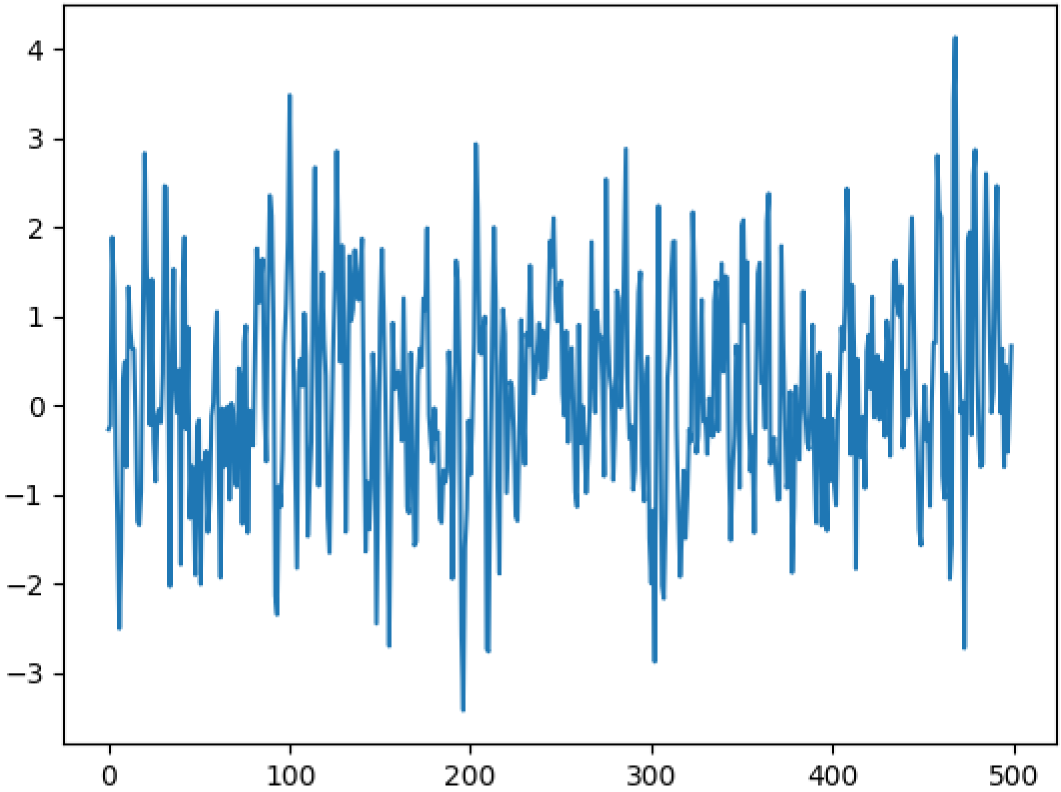
In [2]:

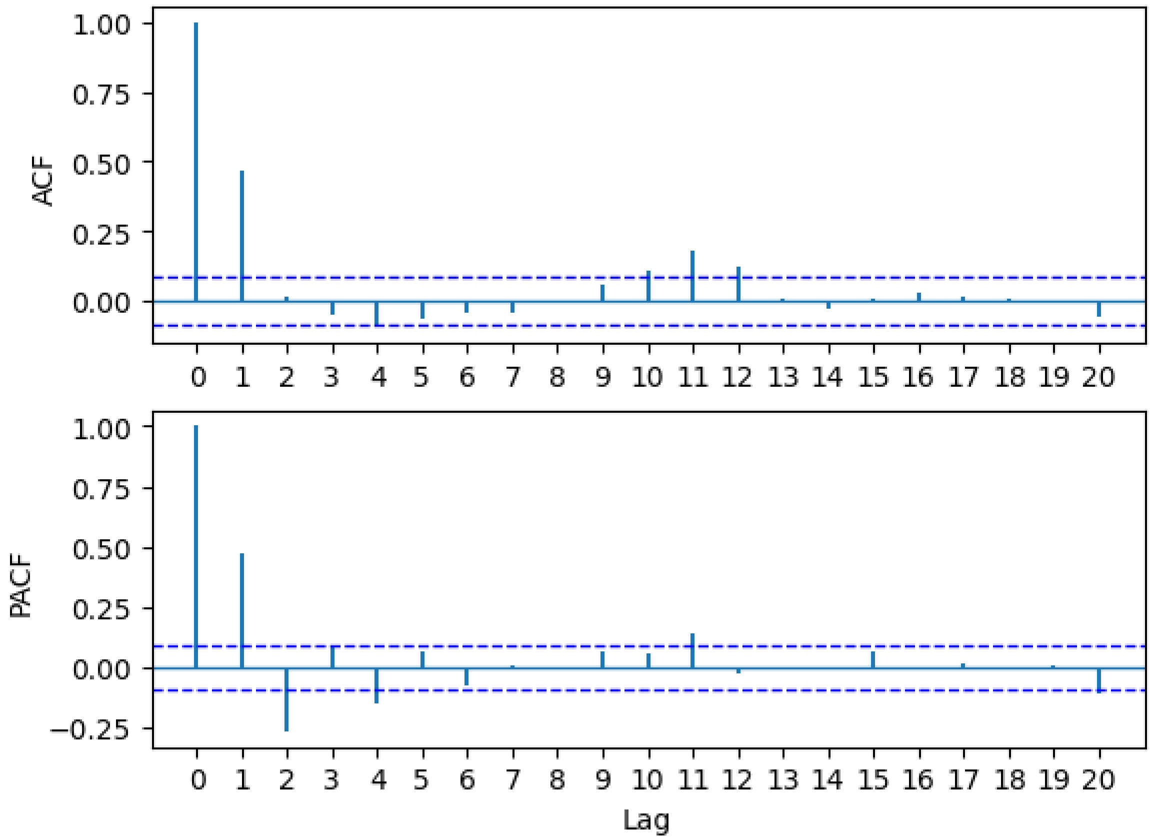
```
import numpy as np
import pandas as pd
from statsmodels.tsa.arima_process import arma_generate_sample
import matplotlib.pyplot as plt
from PythonTsa.plot_acf_pacf import acf_pacf_fig
from pandas.plotting import lag_plot

ma = np.array([1, 0.7])
np.random.seed(123457)
x = arma_generate_sample(ar = [1], ma = ma, nsample = 500)# ar = [1] means
type(x) # x is not a series

x = pd.Series(x)
type(x)# now x is a series
x.plot()

acf_pacf_fig(x, both = True, lag = 20)
```





We cannot tell much from the plot of time series above.

As the sample ACF for the simulated data, we see a **"spike"** at lag 1 followed by generally non-significant values for lags past 1.

Note that the sample ACF does not match the theoretical pattern of the underlying MA(1), which is that all autocorrelations for lags past 1 will be 0.

A different sample would have a slightly different sample ACF shown above, but would likely have the same broad features.