DATA ANALYTICS

Unit 3: ACF and PACF and Forecasting using AR, MA and ARMA

Jyothi R.

Department of Computer Science and Engineering

Contents

- ARIMA
  - Stationarity
  - AR process
  - MA process
  - Main steps in ARIMA
  - Forecasting using ARIMA model
  - Goodness of fit

Drawbacks of the use of traditional models

- There is no systematic approach for the identification and selection of an appropriate model, and therefore, the identification process is mainly trial-and-error
- There is difficulty in verifying the validity of the model
  - Most traditional methods were developed from intuitive and practical considerations rather than from a statistical foundation

ARIMA Models

- Autoregressive Integrated Moving-average
- A “stochastic” modeling approach that can be used to calculate the probability of a future value lying between two specified limits

Autoregressive Models

- In a multiple regression model, we forecast the variable of interest using a linear combination of predictors.
- In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable.
- The term autoregression indicates that it is a regression of the variable against itself.

Thus, an autoregressive model of order pp can be written as

$$\boxed{y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,}$$

where ε_t is white noise.

This is like a multiple regression but with lagged values of y_t as predictors.

Autoregressive Models

- Autoregressive models are remarkably flexible at handling a wide range of different time series patterns.
- The two series in Figure 1. show series from an AR(1) model and an AR(2) model.
- Changing the parameters $\phi_1,\ldots,\phi_p$ $\phi_1,\ldots,\phi_p$ results in different time series patterns.
- The variance of the error term ε_t will only change the scale of the series, not the patterns.

Autoregressive Models

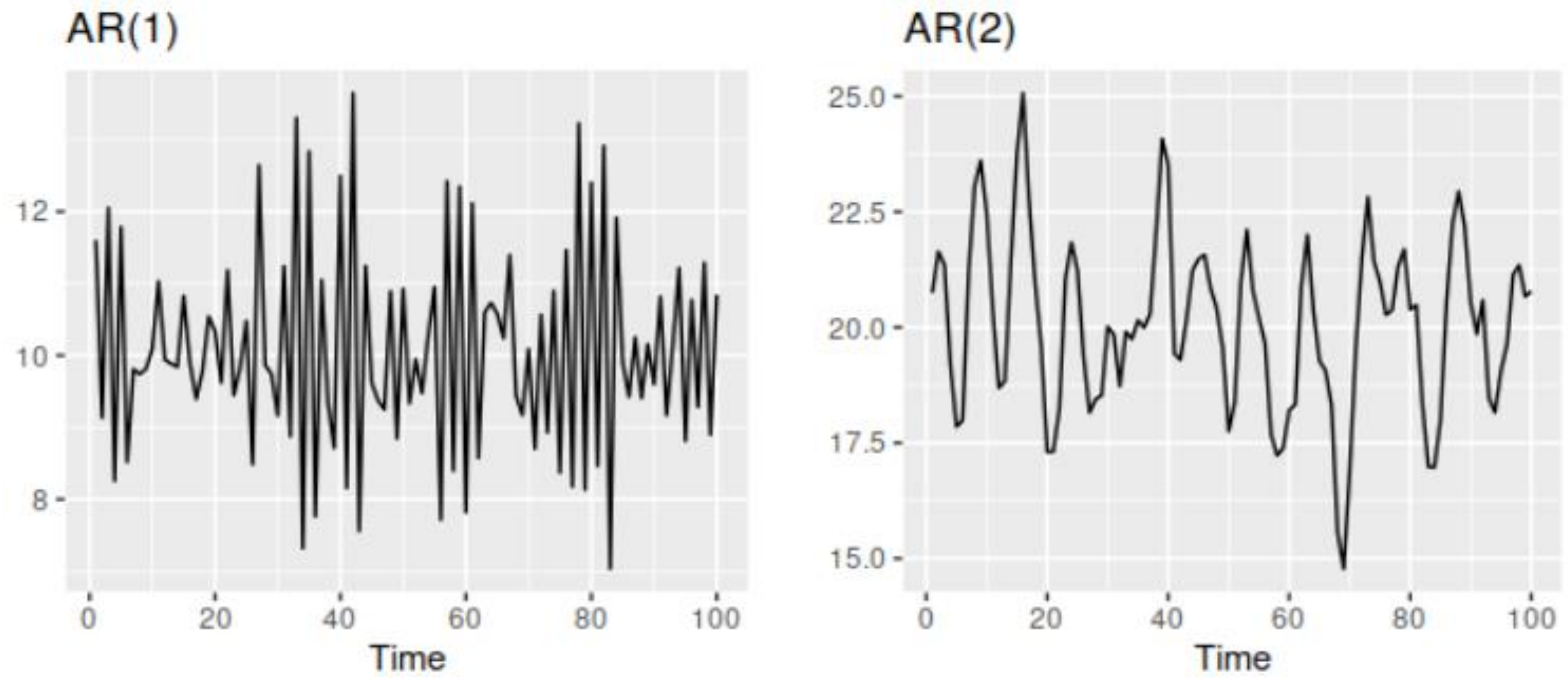


Figure 1: Two examples of data from autoregressive models with different parameters.

Autoregressive Models

- Two examples of data from autoregressive models with different parameters.
- Left: AR(1) with $Y_t = 18 - 0.8Y_{t-1} + Y_t = 18 - 0.8Y_{t-1} + \varepsilon_t$.
- Right: AR(2) with $Y_t = 8 + 1.3Y_{t-1} - 0.7Y_{t-2} + \varepsilon_t$ $Y_t = 8 + 1.3Y_{t-1} - 0.7Y_{t-1} + \varepsilon_t$.
- In both cases, ε_t is normally distributed white noise with mean zero and variance one.

Autoregressive Models

- For an AR(1) model:
- when ϕ_1 =0ϕ_1=0, ytyt is equivalent to white noise;
- when ϕ_1=1ϕ_1=1 and c=0c=0, ytyt is equivalent to a random walk;
- when ϕ_1 =1 ϕ_1 =1 and c≠0c≠0, ytyt is equivalent to a random walk with drift;
- when ϕ_1<0 ϕ_1<0, ytyt tends to oscillate around the mean.
- We normally restrict autoregressive models to stationary data.
- For an AR(1) model: −1< ϕ_1<1−1< ϕ_1 <1.
- For an AR(2) model: −1< ϕ_2<1−1< ϕ_2<1, ϕ_1+ ϕ_1<1 ϕ_1+ ϕ_2<1, ϕ_2− ϕ_1<1 ϕ_2− ϕ_1<1.
- When p≥3p≥3,
- R takes care of these restrictions when estimating a model.

Moving average models

- Rather than using past values of the forecast variable in a regression,
- A moving average model uses past forecast errors in a regression-like model.
- $$y_{t} = c + \varepsilon_t + \theta_{1}\varepsilon_{t-1} + \theta_{2}\varepsilon_{t-2} + \dots + \theta_{q}\varepsilon_{t-q},$$
- where ε_t is white noise.
- **MA(q) model**, a moving average model of order q.
- Of course, we do not *observe* the values of ε_t,
- so it is not really a regression in the usual sense.
- A moving average model is used for forecasting future values, while moving average smoothing is used for estimating the

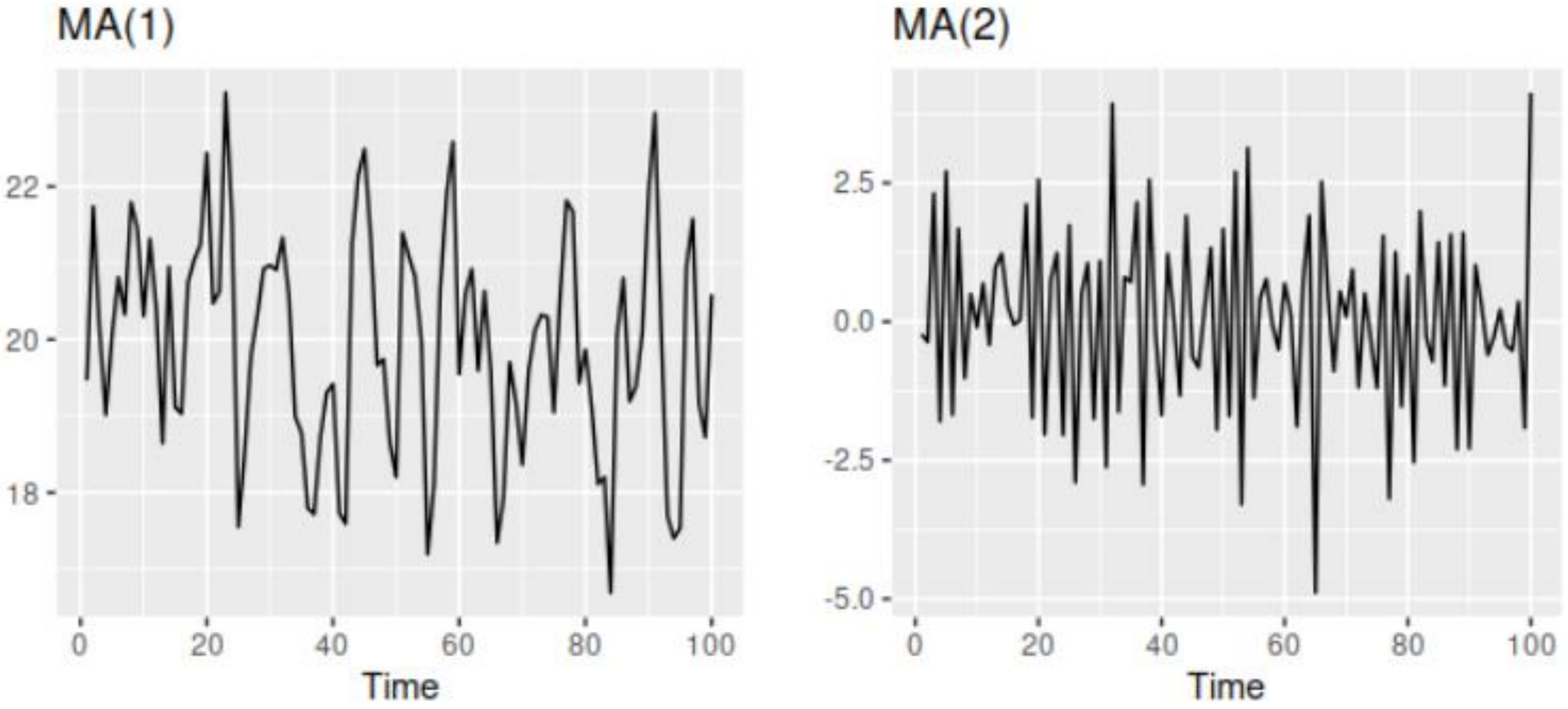
Moving average models

- A moving average model is used for forecasting future values,
- Moving average smoothing is used for estimating the trend-cycle of past values.
- Figure 2. shows some data from an MA(1) model and an MA(2) model.
- Changing the parameters $\theta_1,\dots,\theta_q$ results in different time series patterns.
- As with autoregressive models, the variance of the error term ε_t will only change the scale of the series, not the patterns.

Moving average models

- Figure 2: Two examples of data from moving average models with different parameters.
- Left: MA(1) with $y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$.
- Right: MA(2) with $y_t = \varepsilon_t - \varepsilon_{t-1}+0.8\varepsilon_{t-2}$.
- In both cases, ε_t is normally distributed white noise with mean zero and variance one.

Moving average models



- Figure 2. Two examples of data from moving average models with different parameters

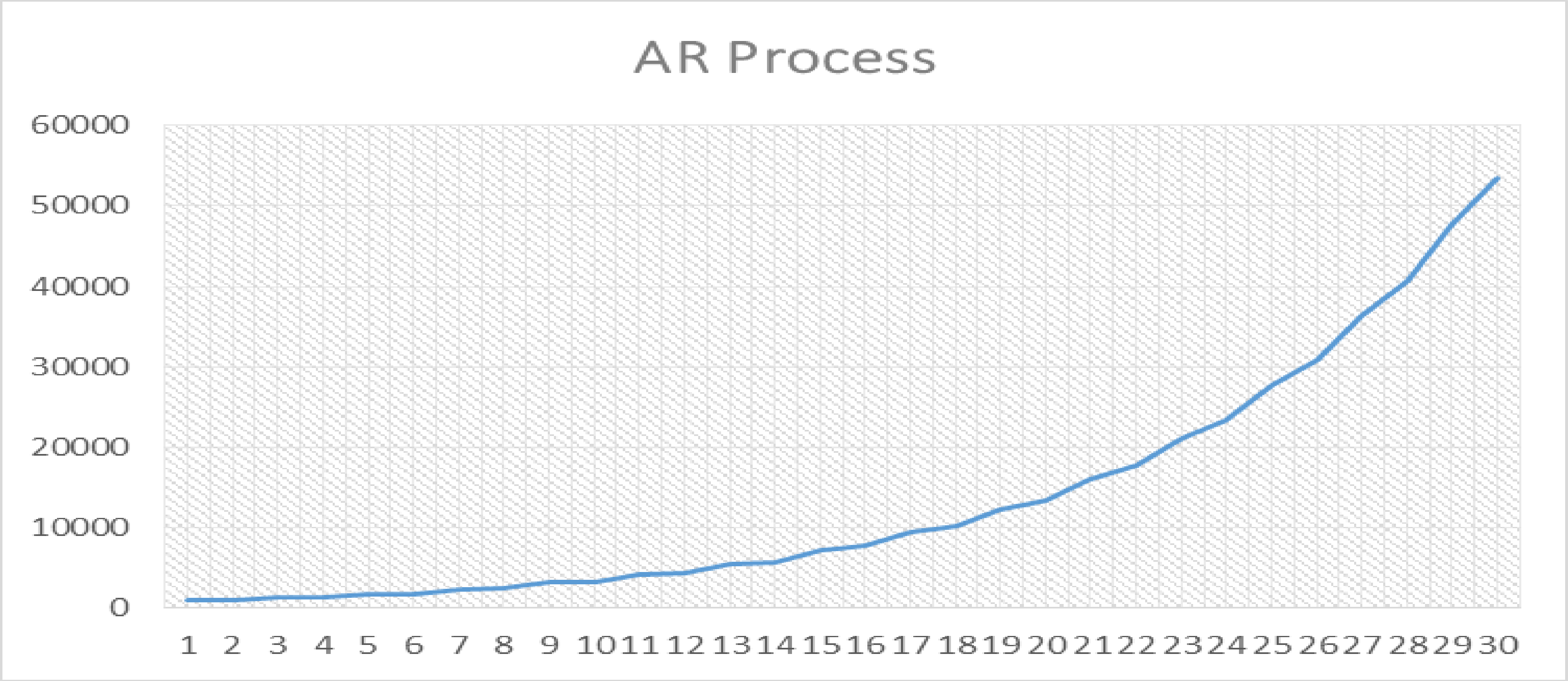
AR & MAModels

- It is possible to write any stationary AR(p) model as an MA(∞) model.
- The reverse result holds if we impose some constraints on the MA parameters.
- Then the MA model is called **invertible**.
- That is, we can write any invertible MA(q) process as an AR(∞) process.
- Invertible models are not simply introduced to enable us to convert from MA models to AR models.
- They also have some desirable mathematical properties.

AR & MAModels

- Autoregressive AR process:
  - Series current values depend on its own previous values
  - AR(p) - Current values depend on its own p-previous values
  - P is the order of AR process
- Moving average MA process:
  - The current deviation from mean depends on previous deviations
  - MA(q) - The current deviation from mean depends on q- previous deviations
  - q is the order of MA process
- Autoregressive Moving average ARMA process

AR & MAModels

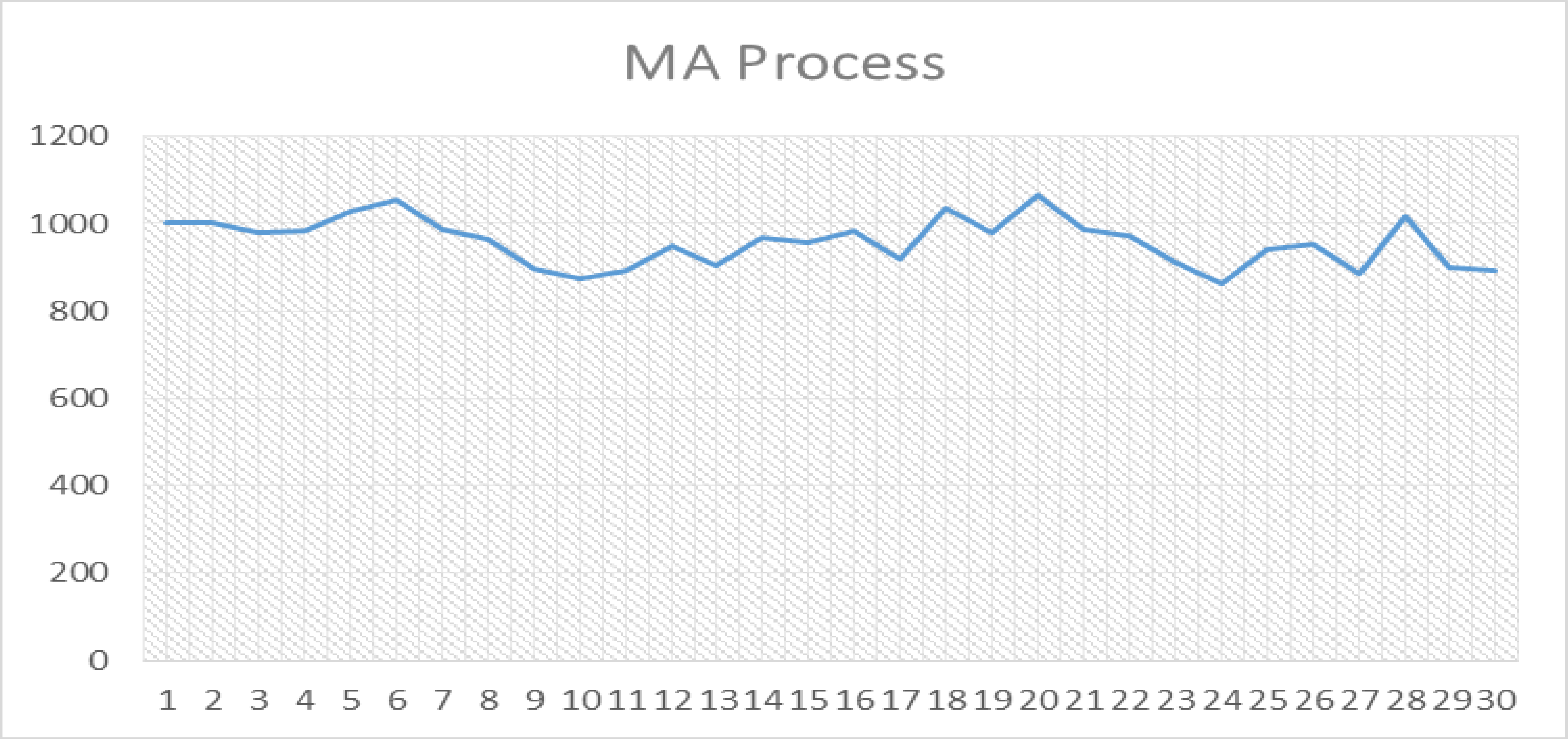


AR(1) $y_t = a1^* \, y_{t-1}$

AR(2) $y_t = a1^* \, y_{t-1} + a2^* \, y_{t-2}$

AR(3) $y_t = a1^* \, y_{t-1} + a2^* \, y_{t-2} + a3^* \, y_{t-2}$

AR & MAModels



MA(1) $\varepsilon_t = b1*\varepsilon_{t-1}$

MA(2) $\varepsilon_t = b1*\varepsilon_{t-1} + b2*\varepsilon_{t-2}$

MA(3) $\varepsilon_t = b1*\varepsilon_{t-1} + b2*\varepsilon_{t-2} + b3*\varepsilon_{t-3}$

ARIMA Models

- Autoregressive (AR) process:
  - Series current values depend on its own previous values
- Moving average (MA) process:
  - The current deviation from mean depends on previous deviations
- Autoregressive Moving average (ARMA) process
- Autoregressive Integrated Moving average (ARIMA)process.
- ARIMA is also known as Box-Jenkins approach. It is popular because of its generality;
- It can handle any series, with or without seasonal elements, and it has well-documented computer programs

ARIMA Model

$y_t \longrightarrow$ AR filter $\longrightarrow$ Integration filter $\longrightarrow$ MA filter $\longrightarrow$ ε_t

(long term) (stochastic trend) (short term) (white noise error)

ARIMA (2,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \varepsilon_{t-1}$

ARIMA (3,0,1) $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + b_1 \varepsilon_{t-1}$ ARIMA (1,1,0) $\Delta y_t = a_1 \Delta y_{t-1} + \varepsilon_t$, where $\Delta y_t = y_t - y_{t-1}$

ARIMA (2,1,0) $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + \varepsilon_t$ where $\Delta yt = yt - yt-1$

To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q

ARIMA equations

- ARIMA(1,0,0)
  - $y_t = a_1 y_{t-1} + \varepsilon_t$
- ARIMA(2,0,0)
  - $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$
- ARIMA (2,1,1)
  - $\Delta yt = a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + b_1 \varepsilon_{t-1}$ where

    $\Delta yt = yt - yt-1$

Overall Time series Analysis & Forecasting Process

- Prepare the data for model building- Make it stationary
- Identify the model type
- Estimate the parameters
- Forecast the future values

ARIMA (p,d,q) modeling

To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q

- **Ensuring Stationarity**
  - Determine the appropriate values of d
- **Identification:**
  - Determine the appropriate values of p & q using the ACF, PACF, and unit root tests
  - p is the AR order, d is the integration order, q is the MA order

ARIMA (p,d,q) modeling

- **Estimation :**
  - Estimate an ARIMA model using values of p, d, & q you think are appropriate.
- **Diagnostic checking:**
  - Check residuals of estimated ARIMA model(s) to see if they are white noise; pick best model with well behaved residuals.
- **Forecasting:**
  - Produce out of sample forecasts or set aside last few data points for in-sample forecasting.

Identifying the numbers of AR or MA terms in an ARIMA model

- **ACF and PACF plots:** After a time series has been stationarized by differencing,
- The next step in fitting an ARIMA model is to determine whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series.
- With software like Statgraphics, we could just try some different combinations of terms.

Identifying the numbers of AR or MA terms in an ARIMA model

- The **autocorrelation function (ACF)** and **partial autocorrelation (PACF)** plots of the differenced series.
- We can tentatively identify the numbers of AR and/or MA terms that are needed.
- ACF plot: it is merely a bar chart of the coefficients of correlation between a time series and lags of itself.
- The PACF plot is a plot of the *partial* correlation coefficients between the series and lags of itself.

ACF Plot

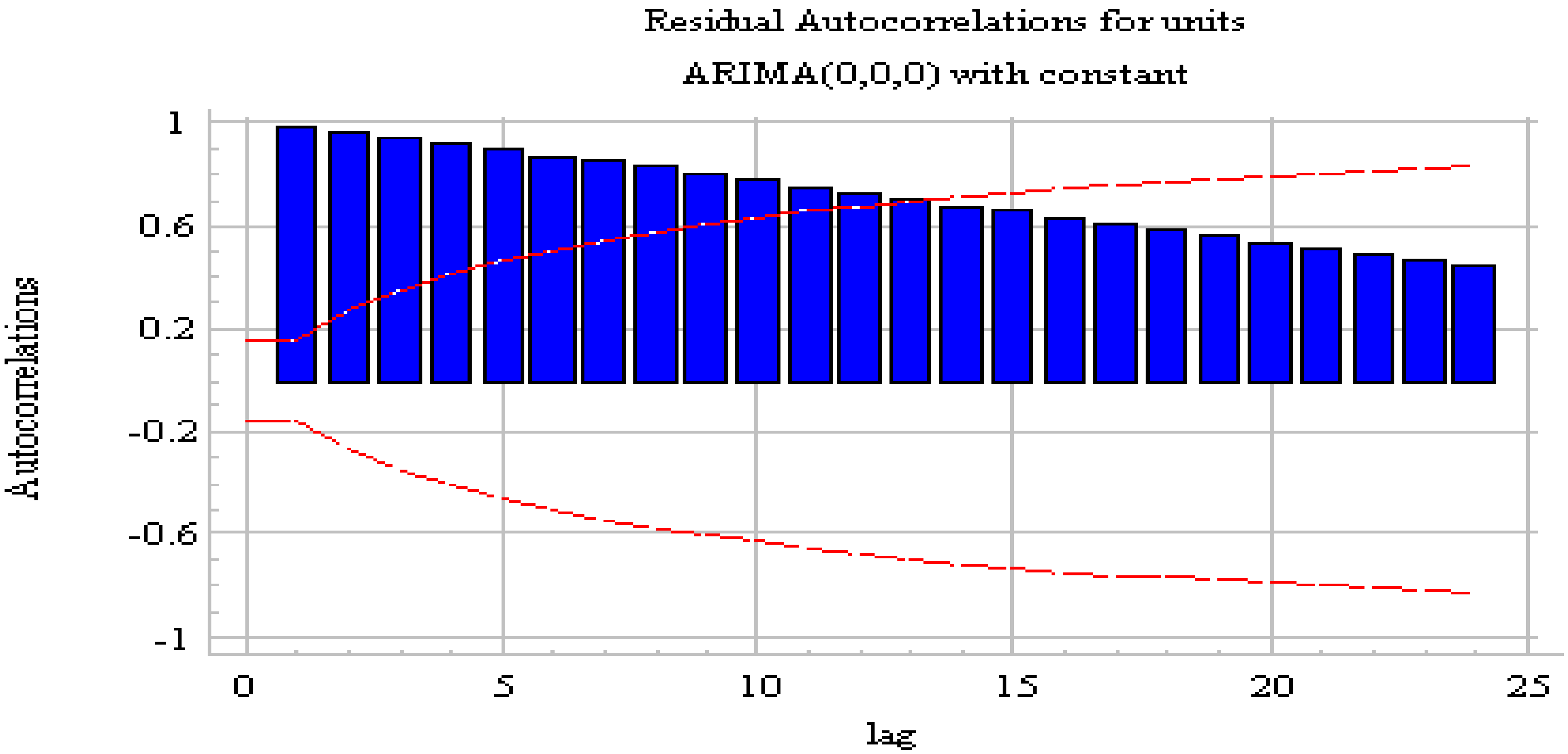


Figure 3:The autocorrelation function (ACF) of the UNITS series, before any differencing is performed:

DATA ANALYTICS

PACF plot

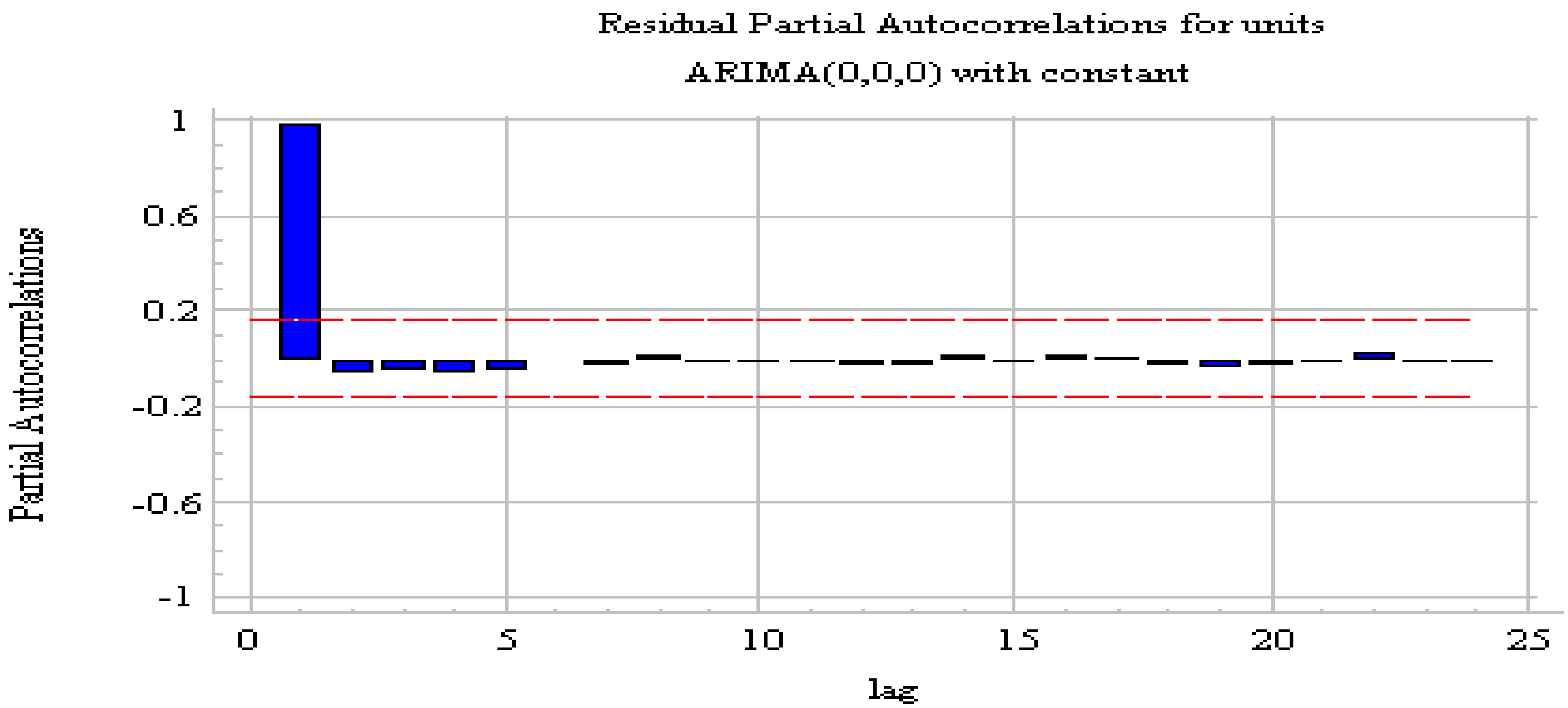


Figure 4: Note that the PACF plot has a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation.

ACF and PACF plots

- The autocorrelations are significant for a large number of lags--but perhaps the autocorrelations at lags 2 and above are merely due to the propagation of the autocorrelation at lag 1. This is confirmed by the PACF plot:
- The partial autocorrelations at all lags can be computed by fitting a succession of autoregressive models with increasing numbers of lags.
- In particular, the partial autocorrelation at lag *k* is equal to the estimated AR(*k*) coefficient in an autoregressive model with *k* terms--i.e., a multiple regression model in which Y is regressed on LAG(Y,1), LAG(Y,2), etc., up to LAG(Y,*k*).
- By mere inspection of the PACF we can determine how many AR terms we need to use to explain the autocorrelation pattern in a time series:
- If the partial autocorrelation is significant at lag *k* and not significant at any higher order lags--i.e., if the PACF "cuts off" at lag *k*--then this suggests that you should try fitting an autoregressive model of order *k*

ACF and PACF plots

- The PACF of the UNITS series provides an extreme example of the cut-off phenomenon:
- It has a very large spike at lag 1 and no other significant spikes, indicating that in the absence of differencing an AR(1) model should be used.
- However, the AR(1) term in this model will turn out to be equivalent to a first difference, because the estimated AR(1) coefficient, which is the height of the PACF spike at lag 1 will be almost exactly equal to 1.

ACF and PACF plots

- Now, the forecasting equation for an AR(1) model for a series Y with no orders of differencing is:
- $\hat{Y}_t = \mu + \phi_1 Y_{t-1}$
- If the AR(1) coefficient ϕ_1 in this equation is equal to 1, it is equivalent to predicting that the first difference of Y is constant--i.e., it is equivalent to the equation of the random walk model with growth:
- $\hat{Y}_t = \mu + Y_{t-1}$
- The PACF of the UNITS series is telling us that, if we don't difference it, then we should fit an AR(1) model which will turn out to be equivalent to taking a first difference.
- In other words, it is telling us that UNITS really needs an order of differencing to be stationarized.

AR and MA signatures

- If the PACF displays a sharp cutoff while the ACF decays more slowly i.e., has significant spikes at higher lags.
- The stationarized series displays an "AR signature," meaning that the autocorrelation pattern can be explained more easily by adding AR terms than by adding MA terms.
- We find that an AR signature is commonly associated with positive autocorrelation at lag 1--i.e., it tends to arise in series which are slightly under differenced.
- The reason for this is that an **AR term can act like a "partial difference" in the forecasting equation**.
- For example, in an AR(1) model, the AR term acts like a first difference if the autoregressive coefficient is equal to 1, it does nothing if the autoregressive coefficient is zero, and it acts like a partial difference if the coefficient is between 0 and 1.

AR and MA signatures

- Autocorrelation has not completely been eliminated, it will "ask for" a partial difference by displaying an AR signature.
- Hence, the following rule of thumb for determining when to add AR terms:
- **Rule 6: If the PACF of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive--i.e., if the series appears slightly "under differenced"--then consider adding an AR term to the model. The lag at which the PACF cuts off is the indicated number of AR terms.**

AR and MA signatures

- An MA signature is commonly associated with negative autocorrelation
- At lag 1--i.e., it tends to arise in series which are slightly overdifferenced.
- The reason for this is that **an MA term can "partially cancel" an order of differencing in the forecasting equation**.
- To see this, recall that an ARIMA(0,1,1) model without constant is equivalent to a Simple Exponential Smoothing model.
- The forecasting equation for this model is
- $\hat{Y}_t = \mu + Y_{t-1} - \theta_1 e_{t-1}$
- where the MA(1) coefficient θ_1 corresponds to the quantity 1- α in the SES model. If θ_1 is equal to 1, this corresponds to an SES model with α =0, which is just a CONSTANT model because the forecast is never updated.
- This means that when θ_1 is equal to 1, it is actually cancelling out the differencing operation that ordinarily enables the SES forecast to re-anchor itself on the last observation.

AR and MA signatures

- On the other hand, if the moving-average coefficient is equal to 0, this model reduces to a random walk model--i.e., it leaves the differencing operation alone.
- So, if θ_1 is something greater than 0, it is as if we are partially cancelling an order of differencing .
- If the series is already slightly *over*differenced--i.e., if negative autocorrelation has been introduced--then it will "ask for" a difference to be partly cancelled by displaying an MA signature.

AR and MA signatures

- A lot of arm-waving is going on here! A more rigorous explanation of this effect is found in the Mathematical Structure of ARIMA Models.

- Hence the following additional rule of thumb**:**
- **Rule 7: If the ACF of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is negative--i.e., if the series appears slightly "overdifferenced"--then consider adding an MA term to the model. The lag at which the ACF cuts off is the indicated number of MA terms.**

Case Study

References

Text Book:

"Business Analytics, The Science of Data-Driven Making", U. Dinesh Kumar, Wiley 2017

Chapter-13

Lecture 6: 13.10, 13.12, 13.13 in text - AR, MA and ARMA models

Introduction to Time Series and Forecasting, Second Edition
Peter J. Brockwell, Richard A. Davis Springer 2002

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

Lecture 6: 13.10, 13.12, 13.13 in text - AR, MA and ARMA models
(AR https://otexts.com/fpp2/AR.html) + MA (https://otexts.com/fpp2/MA.html) + ARMA (Venkat Reddy's slides on ARIMA

(https://app.box.com/s/nizsfr6pza79nef6gfxkw45ppgp6shcj),
https://people.duke.edu/~rnau/411arim3.htm)

THANK YOU

Dr.Mamatha H R
Professor,Department of Computer Science
mamathahr@pes.edu

+91 80 2672 1983 Extn 834