Chapter 8: ARIMA models

- ARIMA models are based on the autocorrelation in the data.
- AR I MA

2 1 3

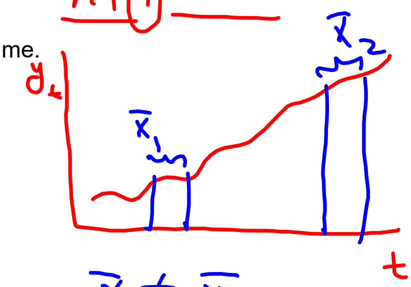
1. I: Stationarity and unit root
2. AR: Autoregressive models ; p
3. MA: Moving average models : q
4. Non-seasonal ARIMA: ARIMA(p, d, q)
5. How to select p & q using ACF and PACF?
6. Seasonal ARIMA models: ARIMA(p, d, q) (P, D, Q)m

ARIMA

What is stationarity of a time series?

(mean, var)

- A stationary time series is the one whose properties do not depend on the time.
- If a time series has a trend or seasonality, it is non-stationary.
- A white noise series is stationary.
- To apply ARIMA models, our time series should be stationary to begin with.
- How to test if a time series is stationary?
- What to do if a time series is non-stationary?

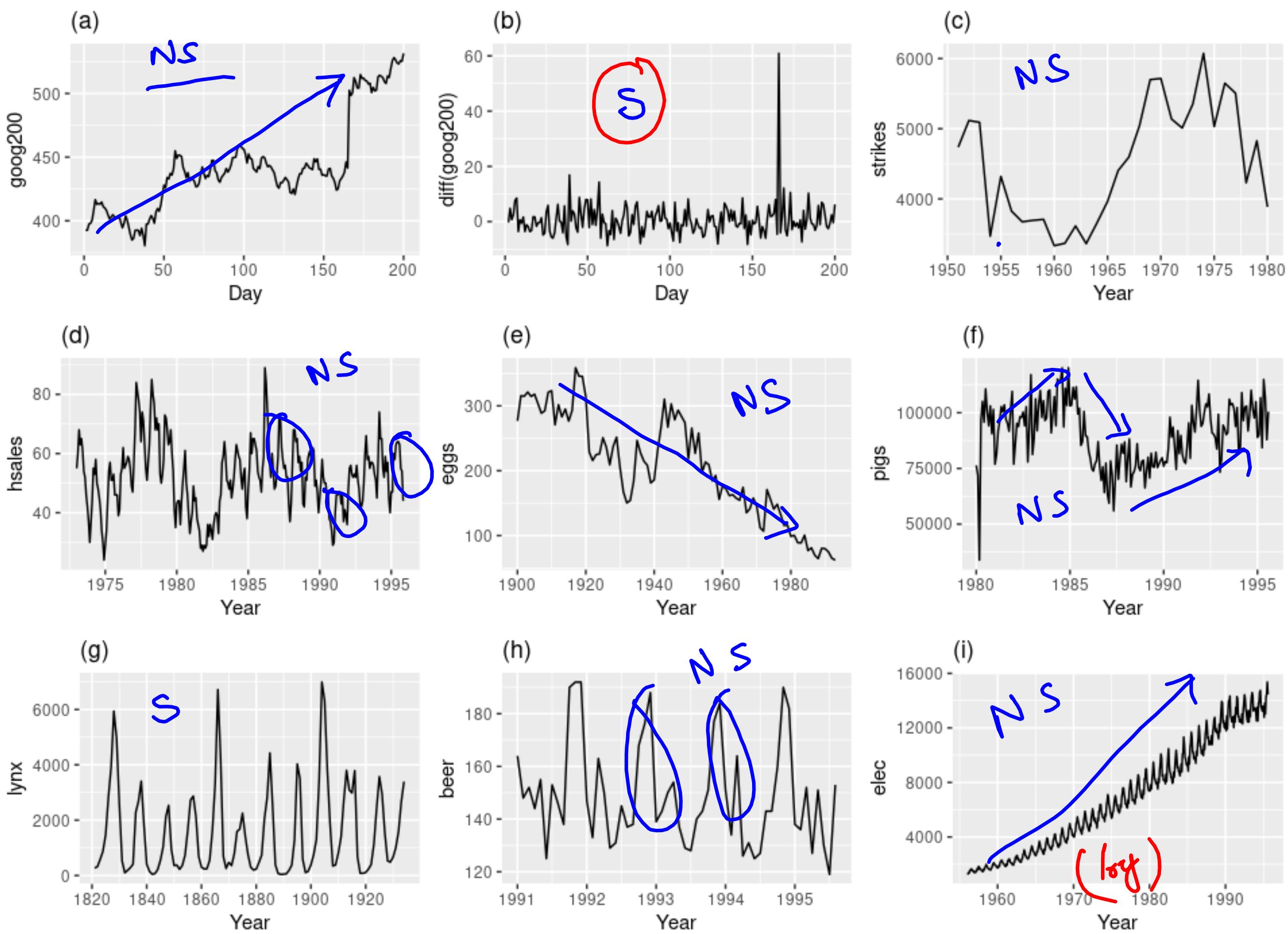


$\bar{x}_1 \neq \bar{x}_2$

y_t — stationarity → ARIMA

y_t → non-stationary → force stationary → ARIMA

Stationarity examples



Differencing

Difference to eliminate trend or seasonality,

$y'_t = y_t - y_{t-1} \rightarrow$ first difference. $I(1)$

$y''_t = y'_t - y'_{t-1} \rightarrow$ Second order difference $I(2)$

$\rightarrow$ stationary.

$y_t = I(0)$ level stationary.

In practice, we never to beyond the second-order difference.

Seasonal difference: It is the difference between an observation at time t and the previous observation from the same season.

$y'_t = y_t - y_{t-m}$

If a time series has strong seasonality, it could be necessary to take both the seasonal difference and the first difference to obtain a stationary series. In this case, take the seasonal difference first; you may not need to take the first difference.

Unit root tests to detect stationarity

A non-stationary series possesses a unit root. The presence of unit root require differencing the series to make it stationary.

1- ACF and the Ljung-Box test. → white noise → strong.

2- ADF (Augmented Dickey-Fuller) test

3- KPSS test

Augmented Dickey-Fuller (ADF) test

$y_t - y_{t-1}$

$$y'_t = \alpha + \beta t + \phi y_{t-1} + \gamma_1 y'_{t-1} + \dots + \gamma_k y'_{t-k} + \varepsilon_t$$

if $\hat{\phi} = 0 \rightarrow$ non-stationary $\rightarrow$ need differencing

$\hat{\phi} < 0 \rightarrow$ stationary,

$H_0: \phi = 0 \rightarrow$ non-stationary,

if p-value is large $\rightarrow$ fail to reject H_0

" " small $\rightarrow$ reject $H_0 \rightarrow$ stationary.

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

The null hypothesis is that the data are stationary, and we look for evidence that the null hypothesis is false.
A small p-values (e.g., less than 0.05) suggest that differencing is required.

H_0: Stationary

Algorithm to test stationarity of a time series

Series has no unit root (stationary)

No

yes

Take the first difference

$y'_t = y_t - y_{t-1}$

Stationary of I(0)

y'_t series has no unit root (stationary)

yes

Stationary of I(1)

NO

Take the second difference

$y''_t = y'_t - y'_{t-1}$

y''_t series has no unit root (stationary)

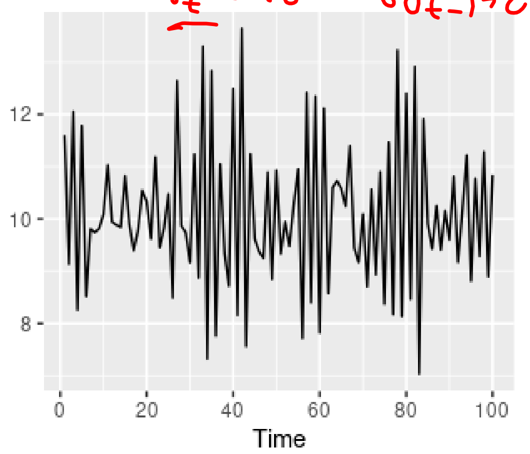
Stationary of I(2)

Apply ARIMA

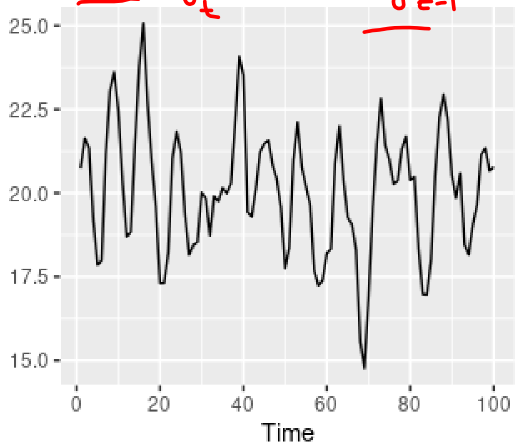
Autoregressive (AR) models

- MLP : $y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \varepsilon_t$
- AR models: $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$: AR(P) → order of AR model
- Why use a univariant model?
  - > Other explanatory variables not available.
  - > Other explanatory variables not directly observable.
  - > Examples: inflation rate, unemployment rate, exchange rate, firm's sales, gold prices, interest rate, etc.

AR(1): $y_t = 18 - 0.8 y_{t-1} + \varepsilon$



AR(2): $y_t = 8 + 1.3 y_{t-1} - 0.7 y_{t-2} + \varepsilon_t$



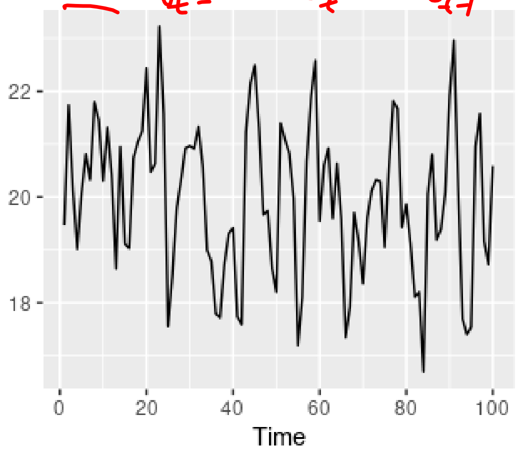
Moving average (MA) models

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

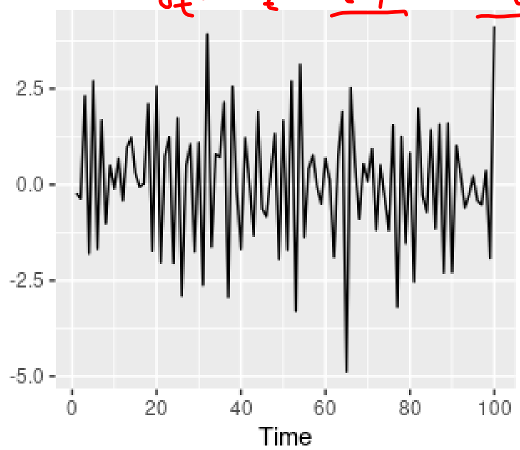
MA(q) → moving average of order q.

$\boxed{q = ?}$

MA(1) : $y_t = 20 + \varepsilon_t + 0.8\,\varepsilon_{t-1}$



MA(2): $y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\,\varepsilon_{t-2}$



Non-seasonal AutoRegressive Integrated Moving Average (ARIMA) model

ARIMA(p, d, q)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

ARIMA(p, d, q)

Table 8.1: Special cases of ARIMA models.

White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(p,0,0) → AR(p)
Moving average	ARIMA(0,0,q) → MA(q)

auto.arima() in R will pick p, q, and d.

$$con = c + 0.588\, con_{t-1} - 0.352\, \varepsilon_{t-1} + 0.684\, \varepsilon_{t-2} + 0.173\, \varepsilon_{t-3} + \varepsilon_t$$

Building your own ARIMA models

ARIMA (p, d, q): $y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$

- Using auto.arima() could be dangerous.
- Choosing appropriate values of p and q.
- Use ACF and PACF plots.

ACF vs. PACF

ACF

ACF plot shows the autocorrelations. $r_1 = corr(y_t, y_{t-1})$

$r_2 = corr(y_t, y_{t-2})$

$\vdots$

$r_k \quad corr(y_t, y_{t-k})$

Partial autocorrelations. These measure these relationships after removing the effects of lags

PACF

$y_t \rightarrow y_{t-2}$

$y_t \rightarrow y_{t-1} \rightarrow y_{t-2}$

$corr(y_t, y_{t-2})$

$corr(y_t, y_{t-3})$: after removing

$corr(y_t, y_{t-1})$

$corr(y_t, y_{t-2})$

ARIMA(p, d, 0):

1. The ACF is exponentially decaying or sinusoidal;
2. There is a significant spike at lag p in the PACF, but none beyond lag p.

ARIMA(0, d, q):

1. The PACF is exponentially decaying or sinusoidal;
2. There is a significant spike at lag q in the ACF, but none beyond lag q.

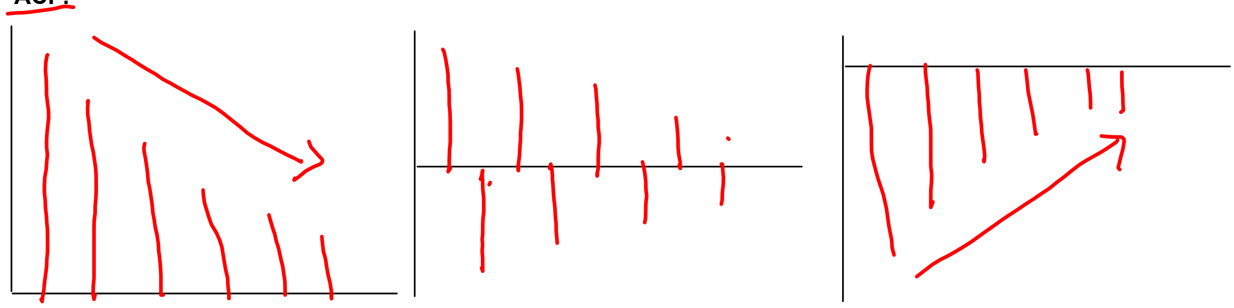
Note: If both p and q are both positive, then the plots do not help in finding suitable values of p and q.

How to pick the order of AR(p) using ACF vs. PACF?

ARIMA(p, d, 0):

1. The ACF is exponentially decaying or sinusoidal;
2. There is a significant spike at lag p in the PACF, but none beyond lag p.

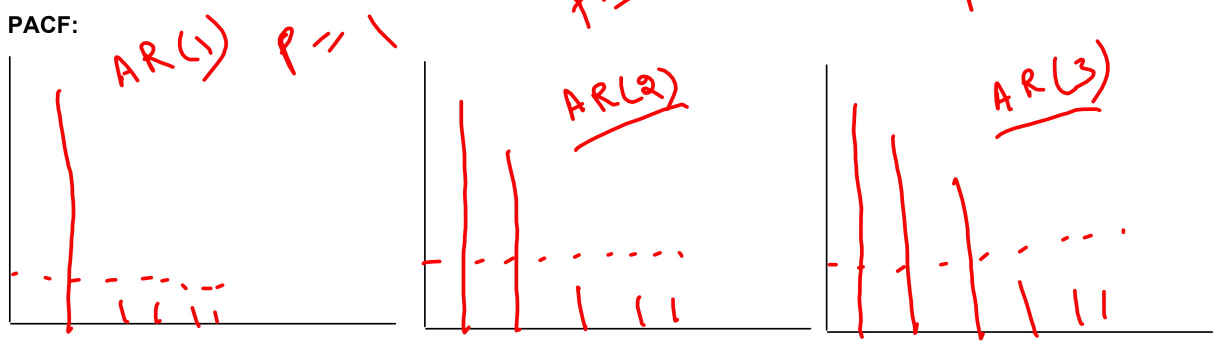
ACF:



p = 2

p = 3

PACF:

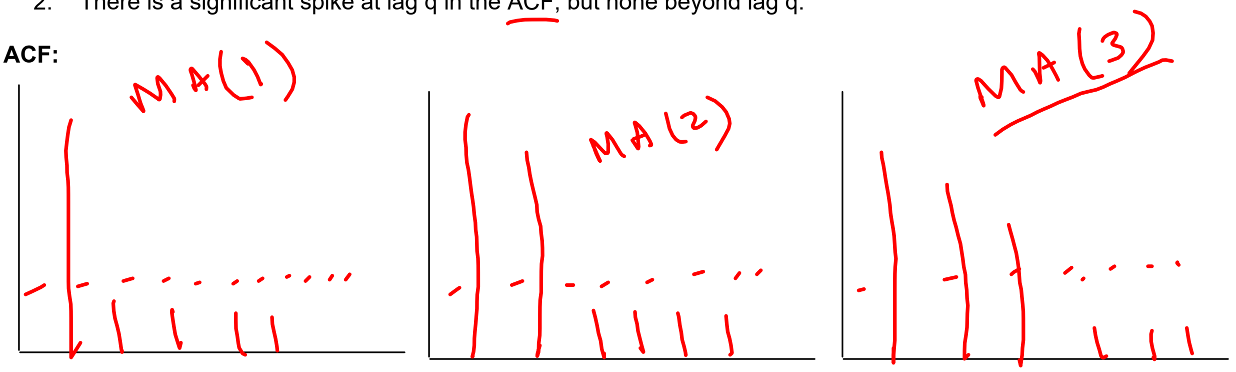


How to pick the order of MA(q) using ACF vs. PACF?

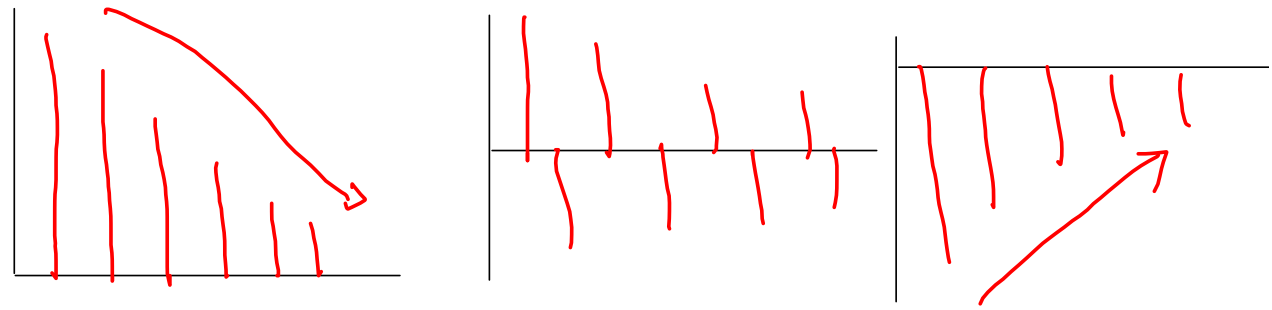
ARIMA(0, d, q):

1. The PACF is exponentially decaying or sinusoidal;
2. There is a significant spike at lag q in the ACF, but none beyond lag q.

ACF:



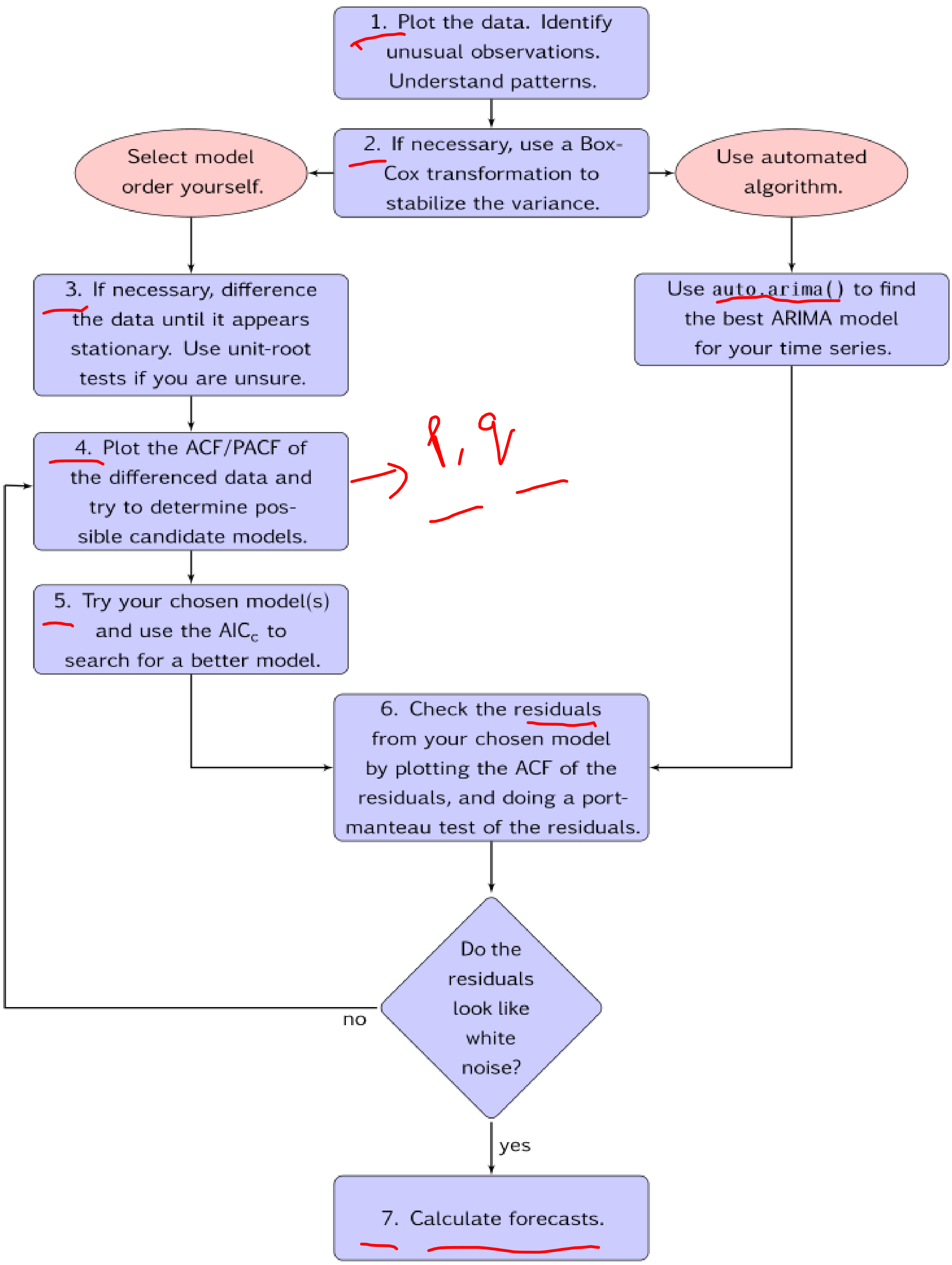
PACF:



Notes:

- Picking p and q is an art.
- ACF and PACF give you a good starting point to pick p and q.
- auto.arima() also gives you a starting point.
- Practically, we experiment with different models in the neighborhood of our starting point to find the best model.

ARIMA modelling procedure

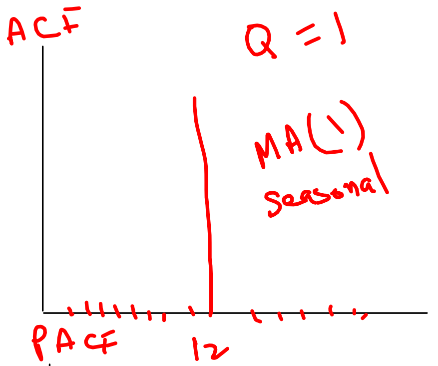


Seasonal ARIMA models (SARIMA)

ARIMA models are also capable of modelling a wide range of seasonal data.

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

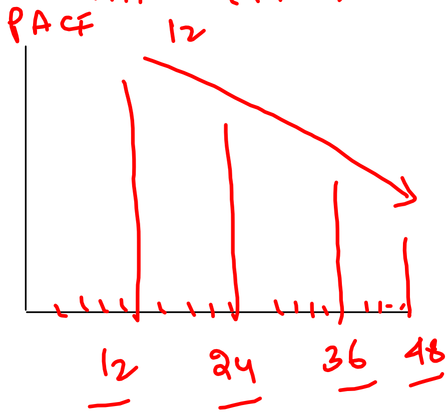
For example, ARIMA(1, 1, 1)(1, 1, 1)4 model is for the quarterly data (m = 4).



ACF/PACF

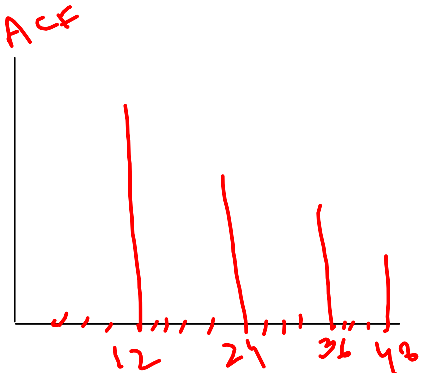
The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF. For example, an ARIMA(0,0,0)(0,0,1)12 model will show: m = 12

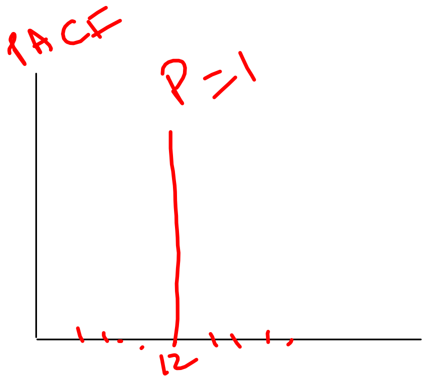
- a spike at lag 12 in the ACF but no other significant spikes;
- exponential decay in the seasonal lags of the PACF (i.e., at lags 12, 24, 36, …).



Similarly, an ARIMA(0,0,0)(1,0,0)12 model will show:

- exponential decay in the seasonal lags of the ACF;
- a single significant spike at lag 12 in the PACF.





Chapter 8: ARIMA models summary

- ARIMA models are based on the autocorrelation in the data.
- AR I MA

1. I: Stationarity and unit root → ADF, KPSS tests → I(0), I(1), I(2)
2. AR: Autoregressive models AR(p)
3. MA: Moving average models MA(q)
4. Non-seasonal ARIMA: ARIMA(p, d, q)
5. How to select p & q using ACF and PACF?
6. Seasonal ARIMA models: ARIMA(p, d, q) (P, D, Q)m

$y_t = y_t$ lags & Error lags.