Time Series Forecasting - 4

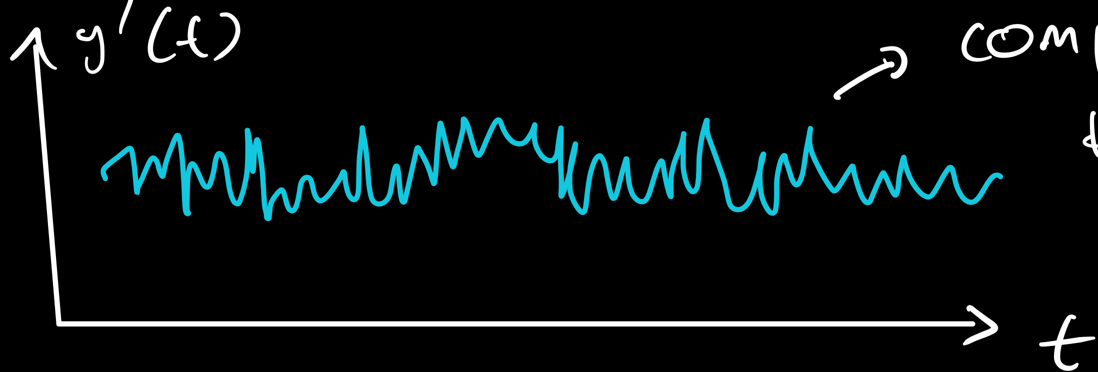
- → Stationarity
- → Autocorrelation / PACF
- → ARIMA Models

For the next family of models we will need to study a new concept.

ARIMA models:

They are used in signals such as financial signals.

Stock return series



completely w/o trend or seasonality

Financial analysts like to predict change in stock instead of stock.

→ Diff stocks series looks diff, but return series looks similar

Another application:

$$y(t) = b(t) + s(t) + \underbrace{e(t)}_{\substack{\text{no trend} \\ \text{no seasonality}}}$$

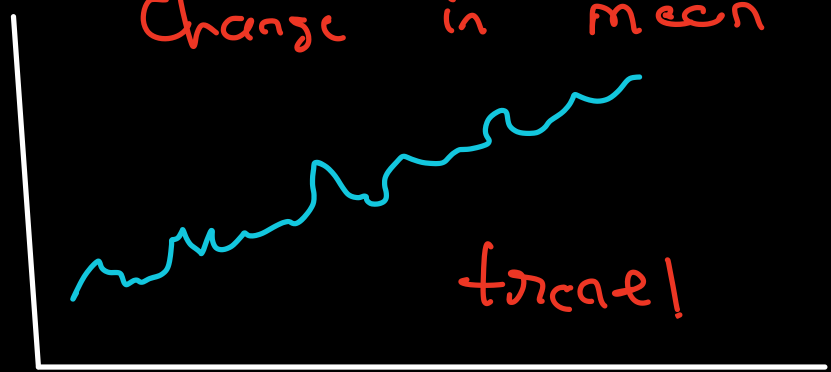
Does it have any predictable information?

Stationarity

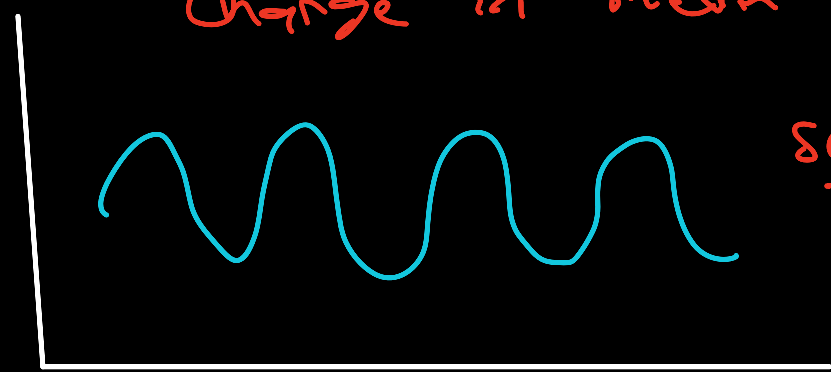
→ A signal is said to be stationary if its parameters such as mean, variance, amplitude, frequency do not change with time.

Ex: of non-stationary

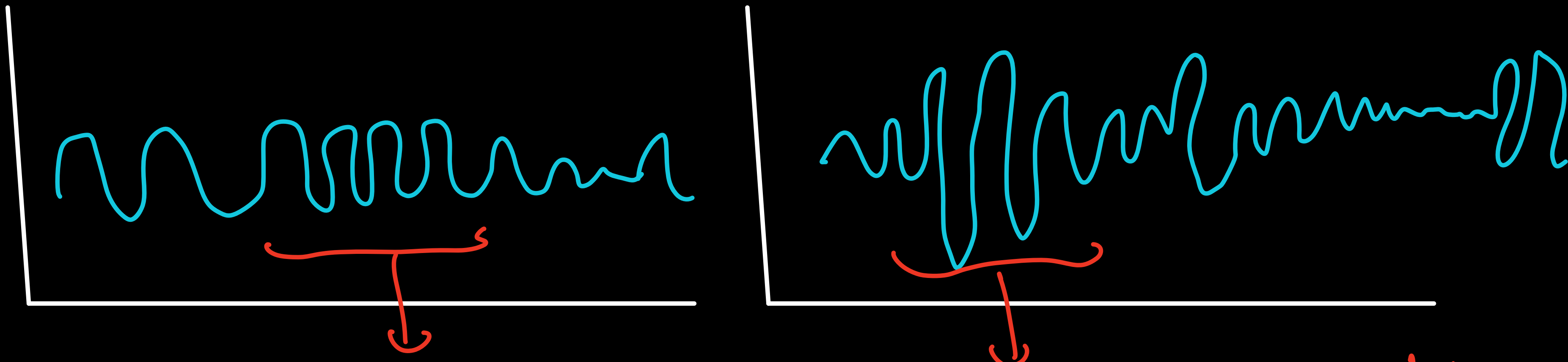
Change in mean



Change in mean



Seasonal



change in frequency

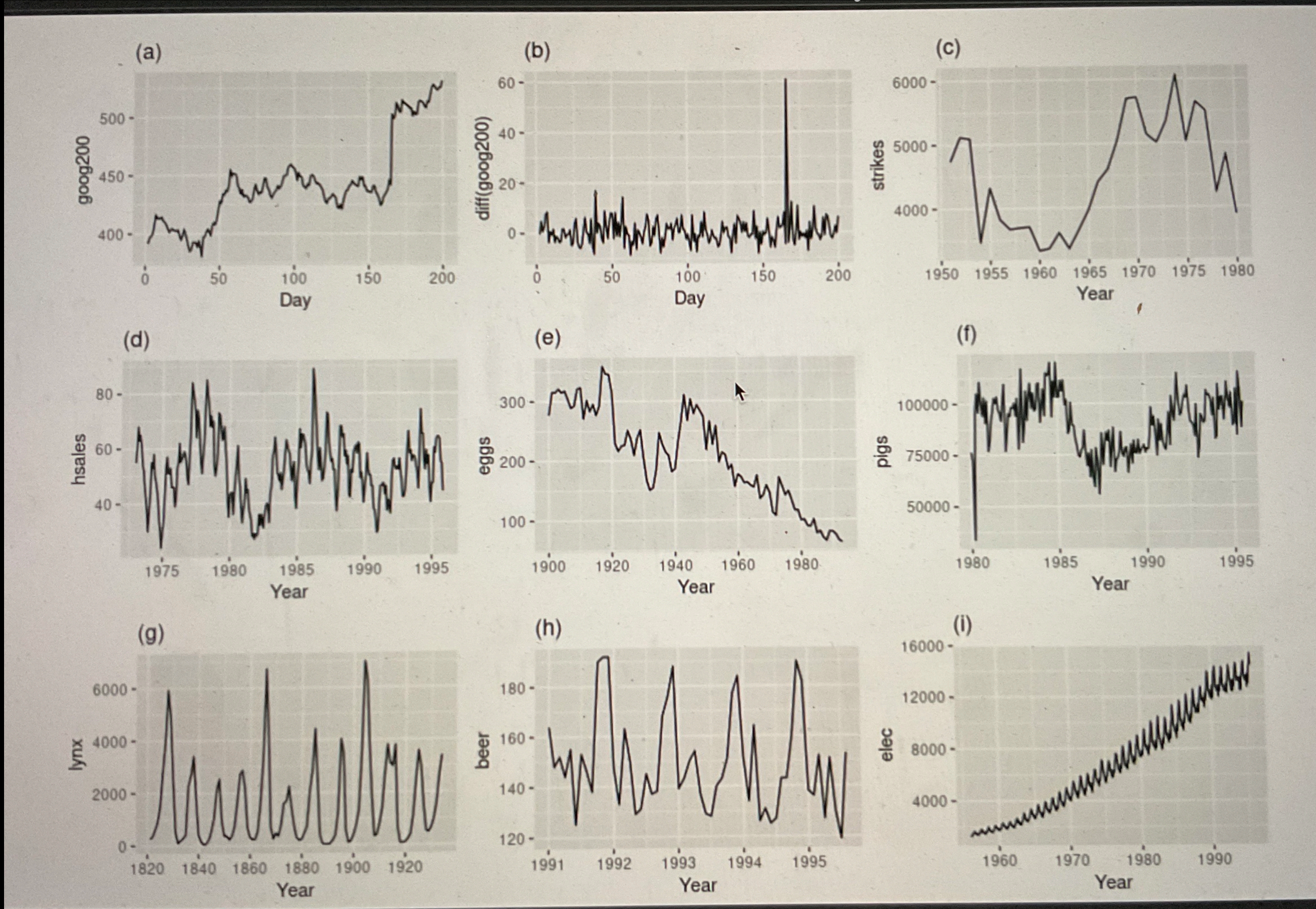
change in amplitude, variance.

How to check for stationarity using code?

→ Dickey-Fuller Test.

if p-value < significance level

→ stationary.


(a)
goog200
400
450
500
0
50
100
150
200
Day
(b)
diff(goog200)
0
20
40
60
0
50
100
150
200
Day
(c)
strikes
4000
5000
6000
1950
1955
1960
1965
1970
1975
1980
Year
(d)
hsales
40
60
80
1975
1980
1985
1990
1995
Year
(e)
eggs
100
200
300
1900
1920
1940
1960
1980
Year
(f)
pigs
50000
75000
100000
1980
1985
1990
1995
Year
(g)
lynx
0
2000
4000
6000
1820
1840
1860
1880
1900
1920
Year
(h)
beer
120
140
160
180
1991
1992
1993
1994
1995
Year
(i)
elec
4000
8000
12000
16000
1960
1970
1980
1990
Year

Which is not stationary?

- `a, c, e, f, :` Not stationary
  - either have a trend, or
  - mean changing with time.
- `d, h:` Not stationary
  - Seasonality
- `i:` Not stationary
  - Has a trend
  - Variance is also not stable
  - Season
- `b:` Stationary
  - There is 1 outlier
  - Can't say anything about mean; seems to be just noise.
- `g:` Stationary
  - Predicting this is dicey, so we assume it to be stationary, and try building model, and seeing if it performs.
  - Looks like a cyclic time series. But these are not at regular intervals.
  - So, even though there is some seasonality, it can't be predicted.

How to make a signal stationary?

1) Decomposition:

$$y(t) = \underbrace{b(t) + s(t) + e(t)}_{}$$

$$\underbrace{e(t)}_{\text{DF test}}$$

$$e(t) = y(t) - [b(t) + s(t)]$$

2) Differencing: [De-trending]

$$y(t) = b(t) + s(t) + e(t)$$

$$y(t) = m \times t + s(t) + e(t)$$

slope (assuming linear trend)

$$y'(t) = \underbrace{m}_{\text{no trend any more.}} + s'(t) + \underbrace{e_2(t)}_{\text{diff noise = another noise}}$$

3) m-difference [De-seasonalisation]

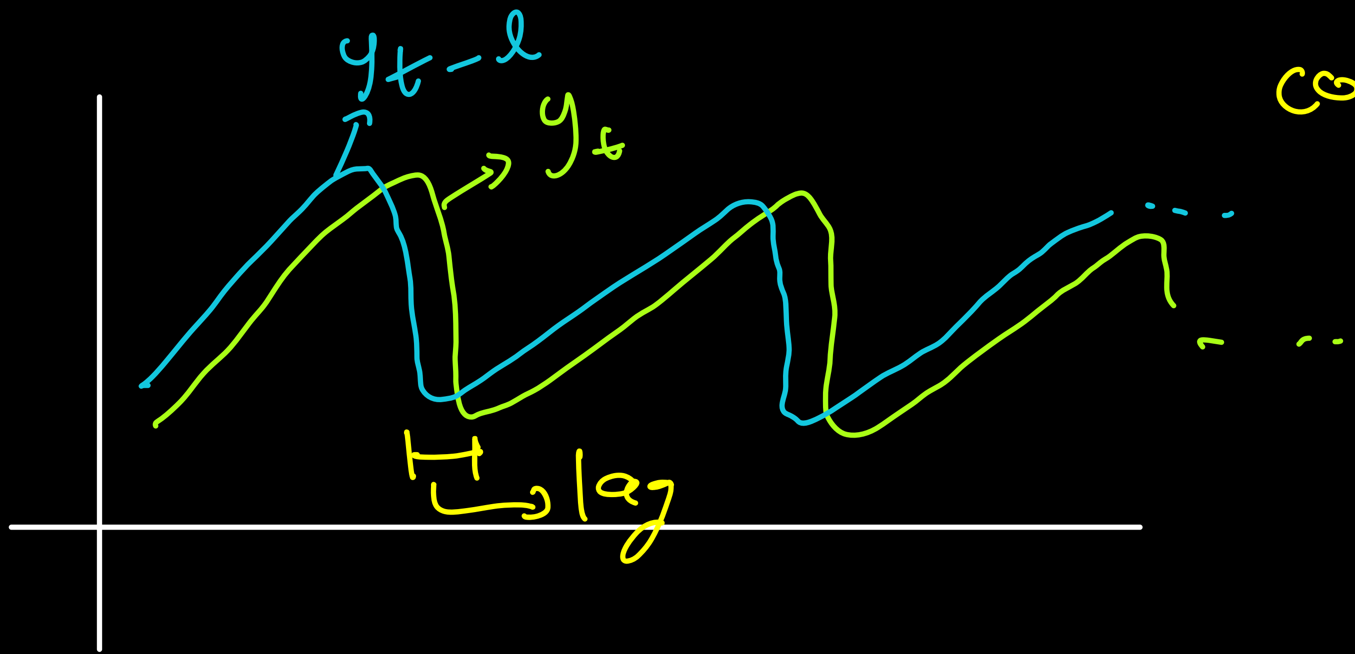
$$y''(t) = m + \underbrace{s(t) - s(t-m)}_{\text{Jan 23 - Jan 22}} + e_3(t)$$

Effect of season will be killed like this.

ACF / PACF

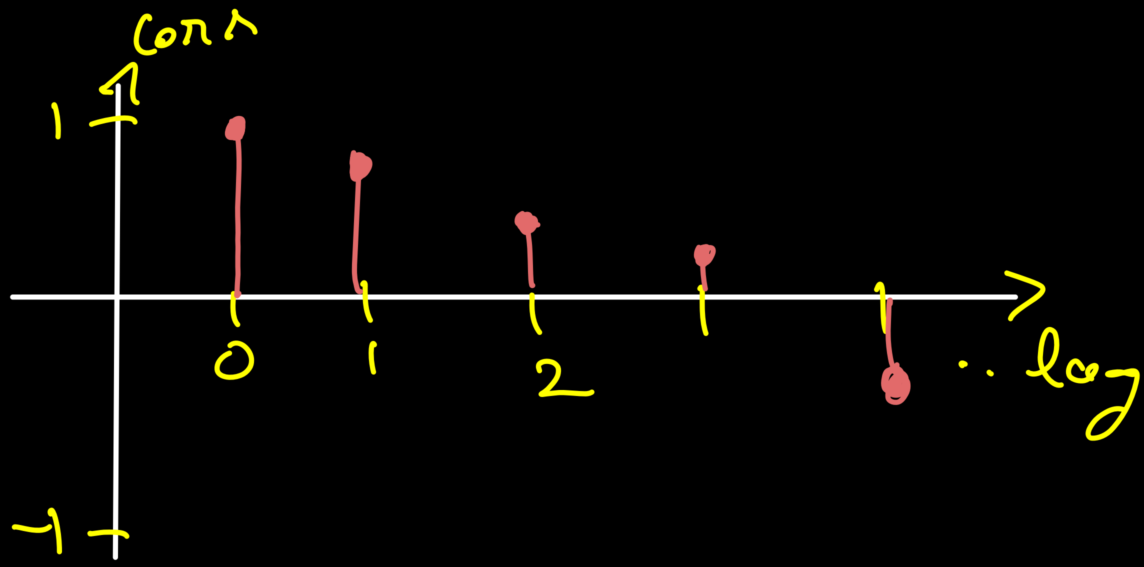
↳ Can I use previous values to forecast future values?

What is the correlation w.r.t past?



corr(y, y_{t-1})

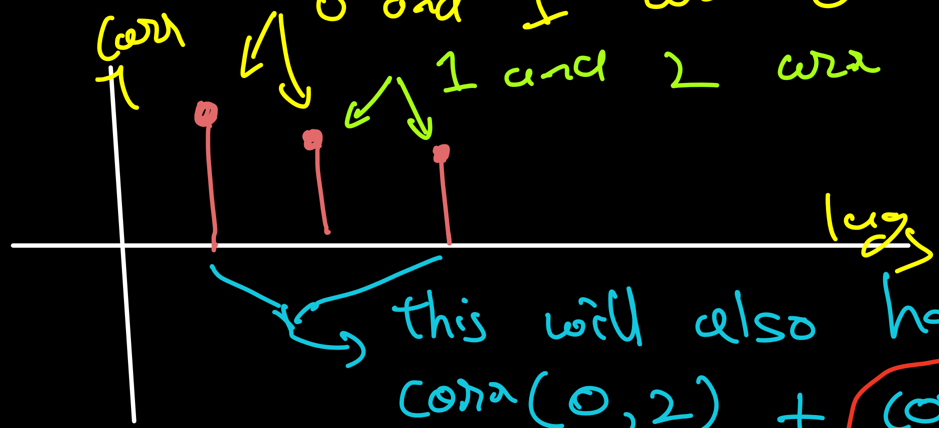
Auto-correlation Function [ACF]



PACF

0 and 1 are correlated

1 and 2 are corr



this will also have

corr(0,2) + corr(0,1) + corr(1,2)

indirect correlation

↓

removed in PACF

Post read for more info

ARIMA FAMILY

Scribble (5)

Auto Regression = Regression with self.

$$\hat{y}_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots a_n y_{t-p}$$

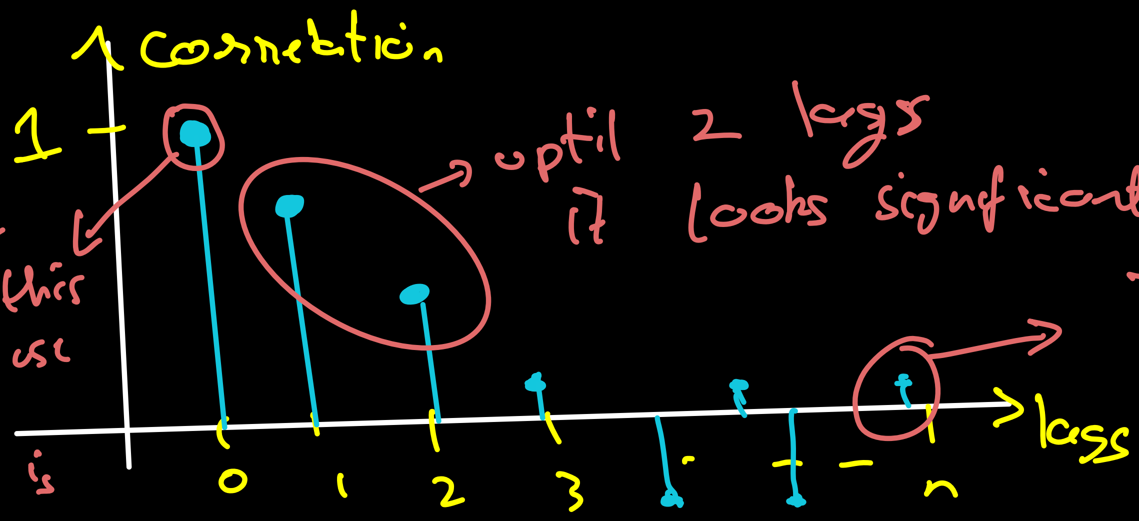
Scribble (4

X				y
y_{t-1}	y_{t-2}	y_{t-3}	...	y_t
y_{t-2}	y_{t-3}	y_{t-4}	..	y_{t-1}
y_{t-3}	y_{t-4}	y_{t-5}	..	y_{t-2}
	$\vdots$			y_{t-3}
				$\vdots$

next value is target

Past values are features | y_{t-n} ,

Scribble 19



until 2 lags it looks significant

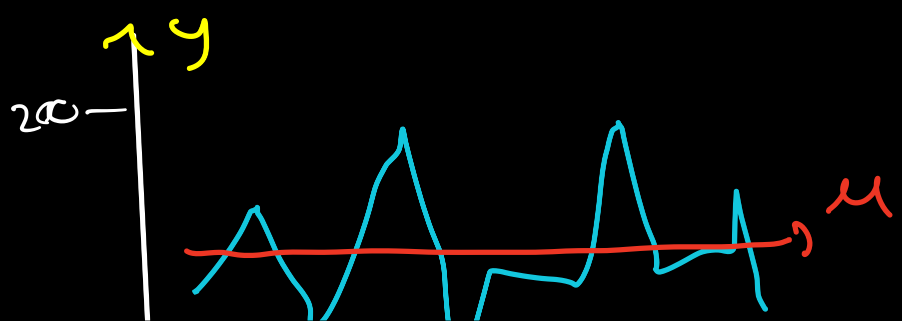
can't use this because this is "present"

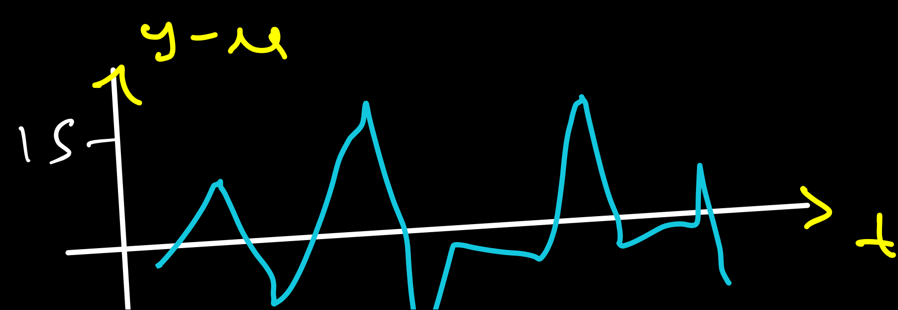
PACF

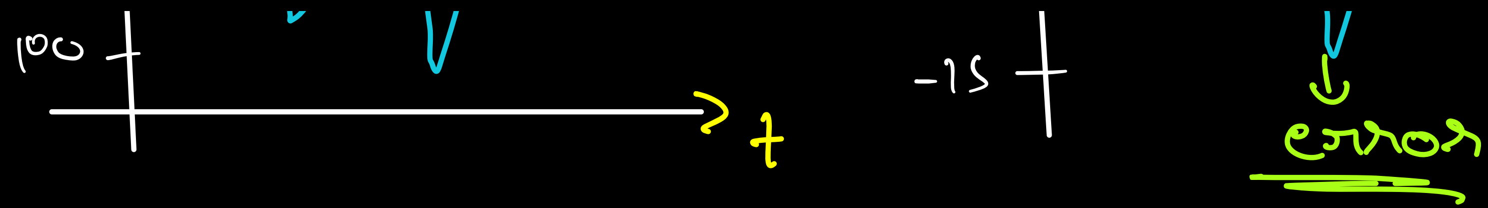
There is no chance of later spike because stationary hence no seasonality

---

20







21

$$e_t = y_t - \mu$$

$$e_{t-1} = y_{t-1} - \mu$$

$$\dots$$

$$\hat{y}_t = \mu + m_1 e_{t-1} + m_2 e_{t-2} \dots m_3 e_{t-q}$$

22

$$\hat{y}_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots a_n y_{t-p}$$
$$+ m_1 e_{t-1} + m_2 e_{t-2} + \dots m_d e_{t-q}$$

23

$$\hat{y}'_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \dots a_n y_{t-p} + m_1 e_{t-1} + m_2 e_{t-2} + \dots m_d e_{t-d}$$

$$\hat{y}_t = \underbrace{y_{t-1} + \hat{y}'_t}_{\text{Many additions over time will result in integration}} \quad : \text{if } d = 1$$

$$\hat{y}_t = y_{t-1} + \underbrace{y'_{t-1} + y''_t}_{\text{anologous to double integration}} \quad : \text{if } d = 2$$

$\vdots$

24

$$\underbrace{p, d, q}_{\text{Regular ARIMA}}, \underbrace{P, D, Q}_{\text{Extra set of params to capture seasonal part}}, \underbrace{S}_{\text{period, Eg: 12 for monthly data}}$$