

# Quantitative Analytics tools for financial decisions

*Time Series Analytics Part 2*

**Module 2 Session 8 & 9**

# Some Important news!

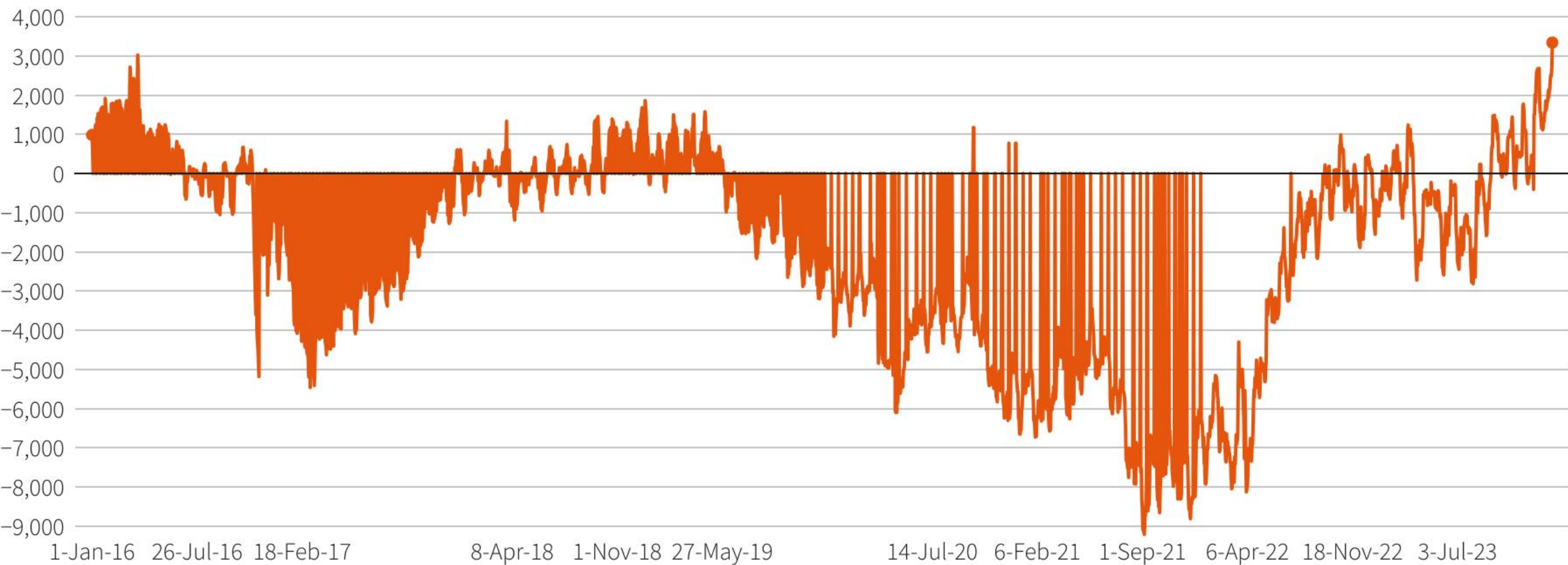
- Liquidity deficit in banking system at highest-ever, tops Rs 3 lakh cr after eight years.

What could be the reason????

# India's banking system liquidity deficit jumps to widest on record

Figures in billion Indian rupees

— Banking system liquidity



Note: Positive figure indicates deficit

Source: RBI data

# Some Important news!

- Banks are facing pressure on the liquidity front as there were **outflows** last month on account of GST (goods and service tax) payments and advances tax outflows, and the central government has accumulated cash balances (estimated at about ₹2-lakh crore) with the RBI.
- They are **also up against stiff competition from non-banking finance companies** for garnering resources. Non-convertible debentures floated by NBFCs are offering relatively higher returns (up to 10 per cent) against Bank term deposit rates **of 6.50-7.25 for over a one year tenor.**
- With many IPO's people are investing in IPO's

The goal is to turn data into information, and  
information into insight.

—Carly Fiorina

## Getting Deeper Understanding...



*Deeper  
understanding  
of Finance using  
Time Series  
Analytics!*

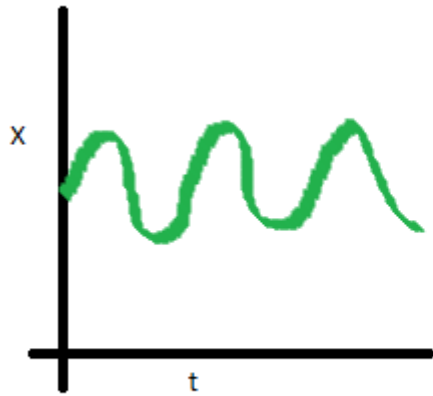
# Today's Agenda

---

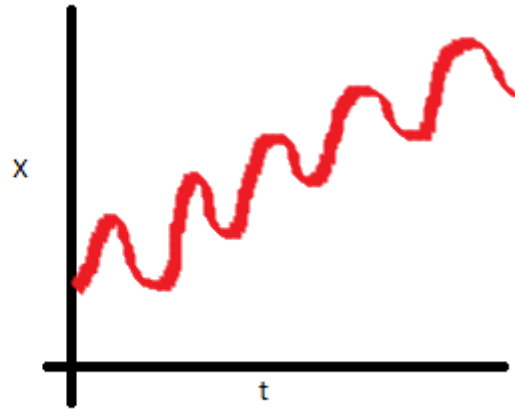
Time Series Analytics

# Stationary Series

Mean should be constant



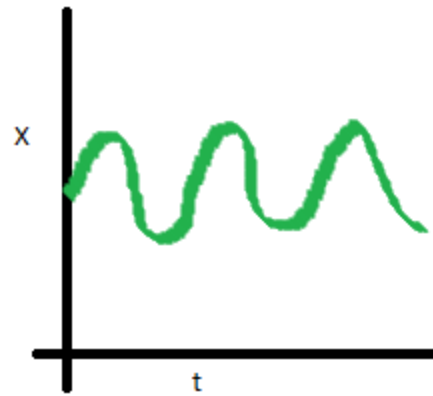
Stationary series



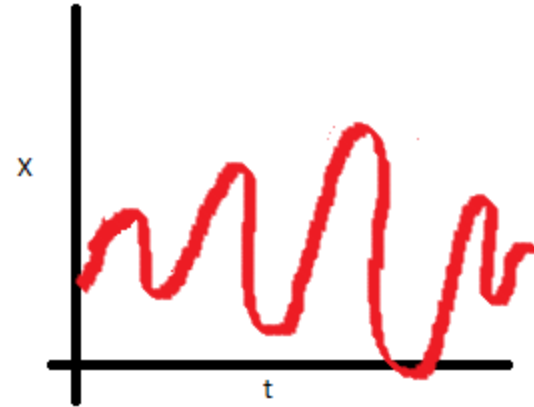
Non-Stationary series

# Stationary Series

Variance should be constant



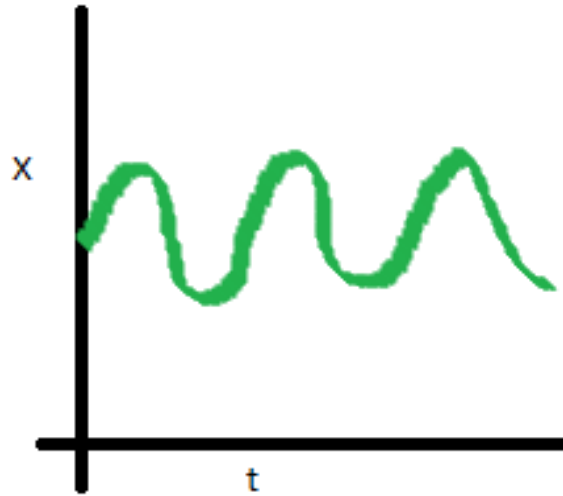
Stationary series



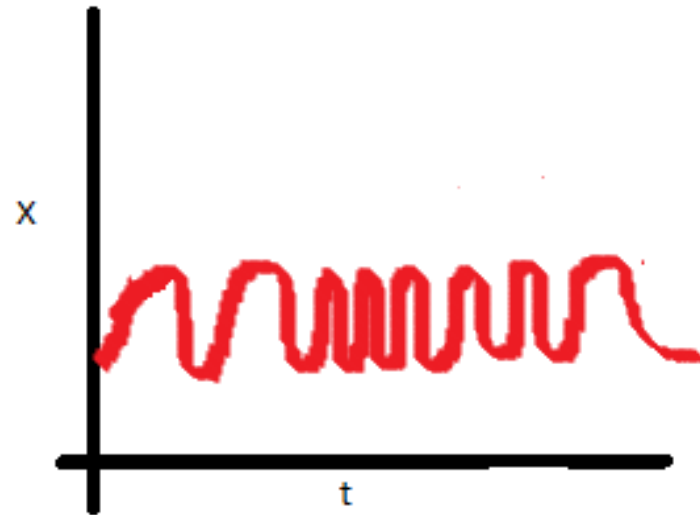
Non-Stationary series

# Stationary Series

Covariance should be constant



Stationary series



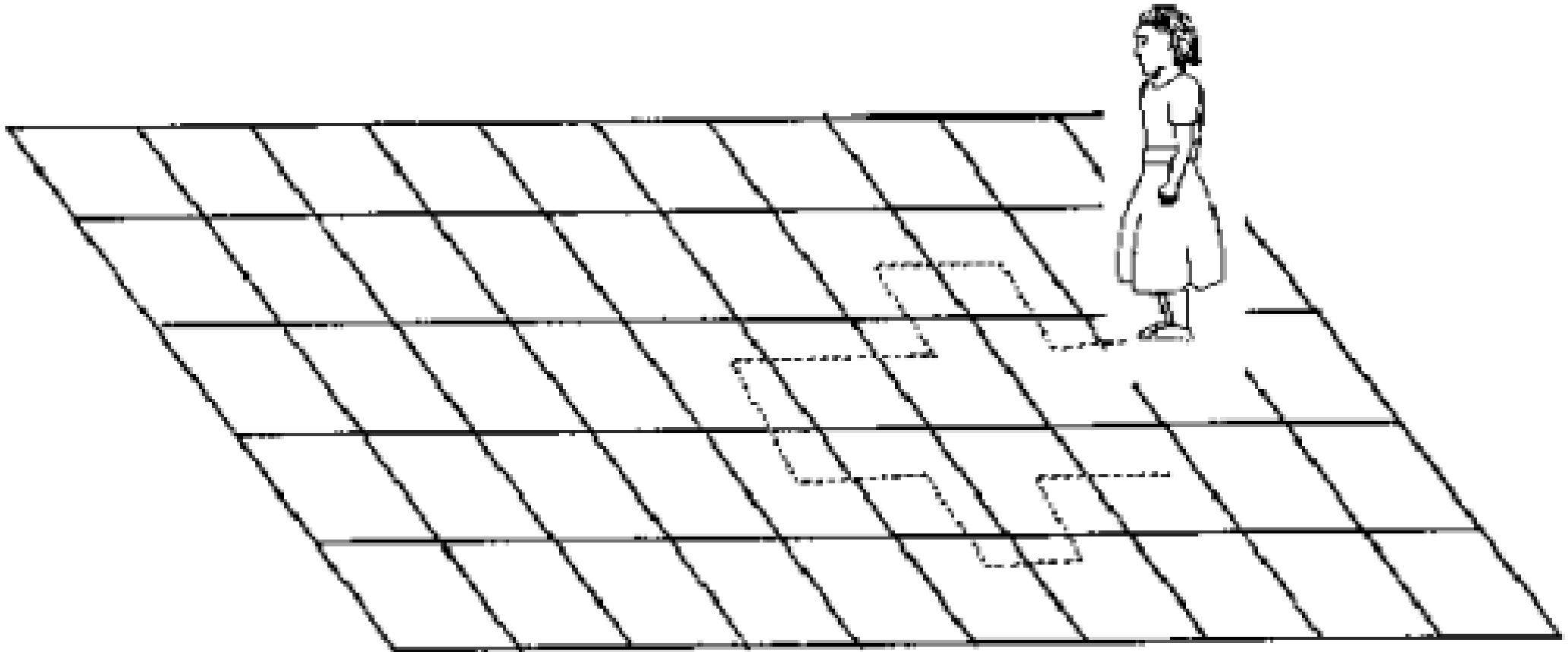
Non-Stationary series

**Why do I care about 'stationarity' of a time series?**

# Reason is.....

- We want stochastic models to predict this time series

# Random Walk

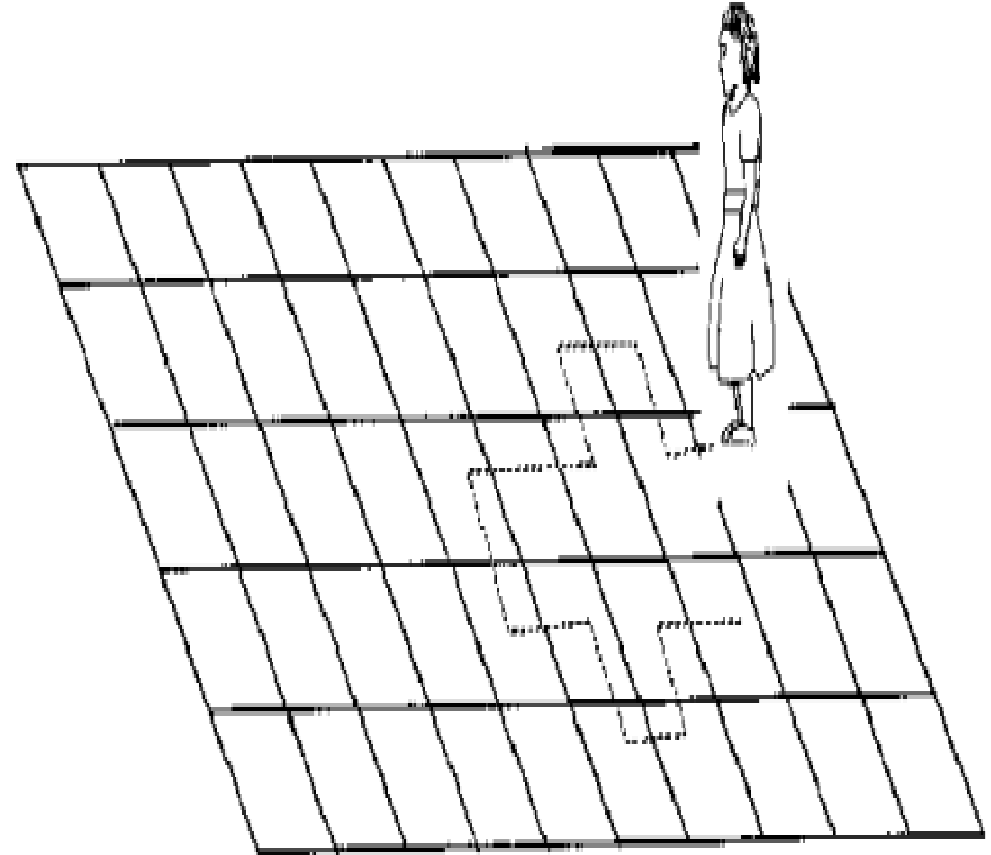


# Random Walk

- Now imagine, you are sitting in another room and are not able to see the girl. You want to predict the position of the girl with time.

$$X(t) = X(t-1) + Er(t)$$

- where  $Er(t)$  is the error at time point  $t$ . This is the randomness the girl brings at every point in time.



# Infosys Ltd.



# Let's spice up things a bit!

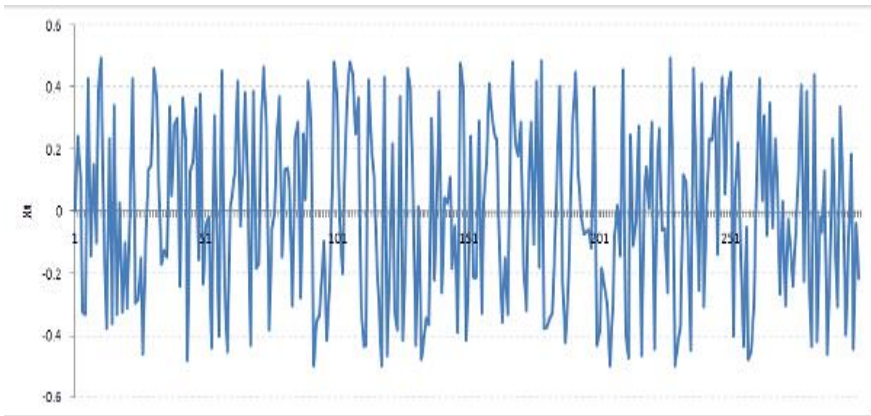
- We already know that a random walk is a non-stationary process.
- Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.

- **Introduced coefficient: Rho**

$$X(t) = \text{Rho} * X(t-1) + \text{Er}(t)$$

# Let's spice up things a bit,

- Rho = 0



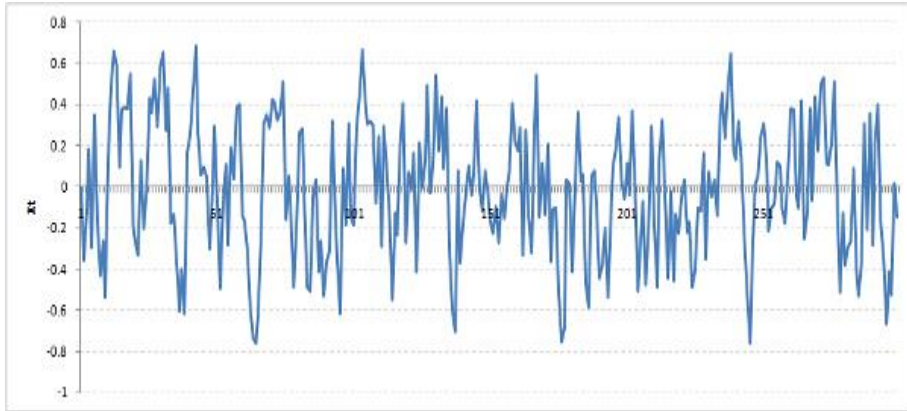
- Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.

- **Introduced coefficient: Rho**

$$X(t) = \text{Rho} * X(t-1) + \text{Er}(t)$$

# Let's spice up things a bit,

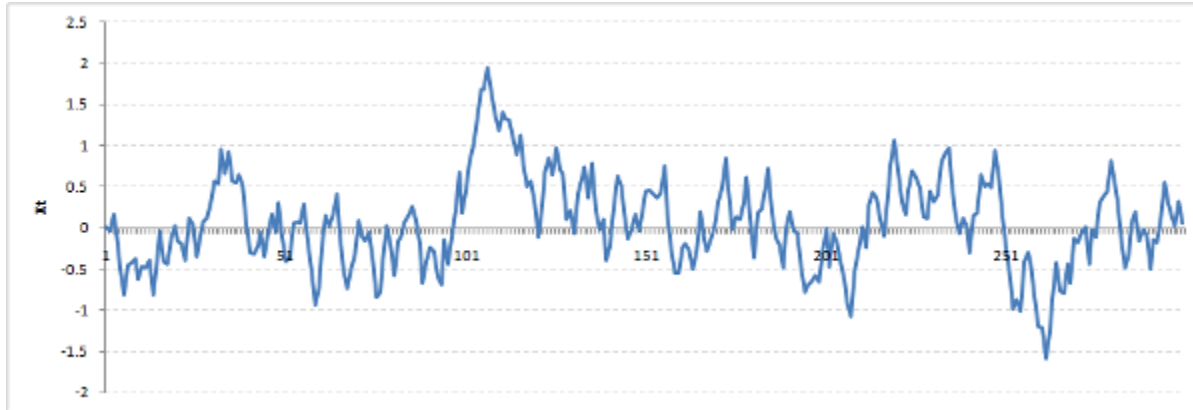
- Rho =0.5



- Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.
- **Introduced coefficient: Rho**  
$$X(t) = \text{Rho} * X(t-1) + E_r(t)$$

# Let's spice up things a bit,

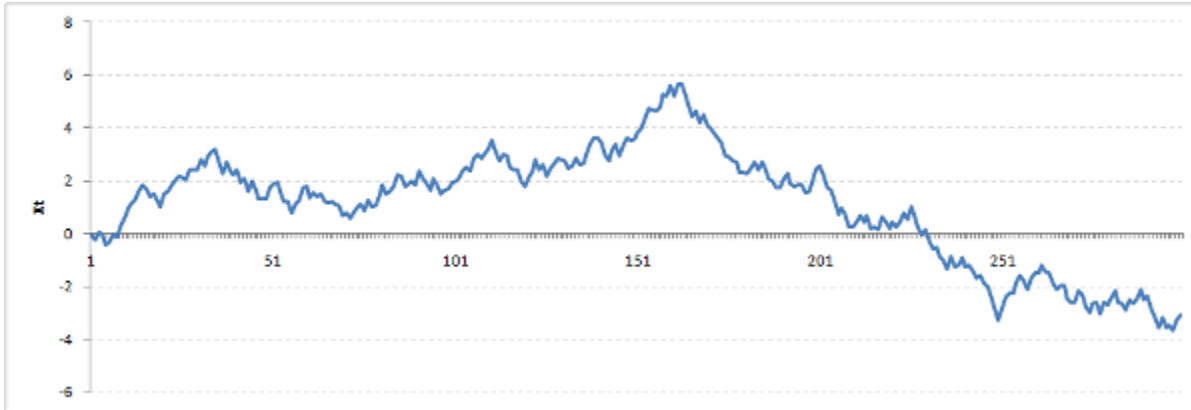
- Rho =0.9



- Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.
- **Introduced coefficient: Rho**  
$$X(t) = \text{Rho} * X(t-1) + E_r(t)$$

# Let's spice up things a bit,

- Rho = 1



- Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.

- **Introduced coefficient: Rho**

$$X(t) = \text{Rho} * X(t-1) + \text{Er}(t)$$

Tool to understand the behaviour of a time series ...



**Correlogram ...**

# Correlogram

Date: 04/13/12 Time: 10:12

Sample: 1996 2010

Included observations: 15

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.729	0.729	9.6842	0.002
		2	0.429	-0.218	13.302	0.001
		3	0.168	-0.122	13.900	0.003
		4	-0.025	-0.075	13.914	0.008
		5	-0.069	0.118	14.035	0.015
		6	-0.104	-0.117	14.340	0.026
		7	-0.134	-0.072	14.913	0.037
		8	-0.170	-0.081	15.961	0.043
		9	-0.209	-0.055	17.822	0.037
		10	-0.248	-0.116	20.959	0.021
		11	-0.283	-0.107	26.048	0.006
		12	-0.264	-0.006	31.965	0.001

# ACF and PACF

- Autocorrelation function (ACF)

$$\rho(p \text{ or } q) = \frac{\text{cov}(y_t, y_{t-p \text{ or } q})}{\text{var}(y_t)}$$

- Partial ACF (PACF)

$$\phi(p) = \frac{\rho_p - \sum_{j=1}^{p-1} (\phi_{p-2,j} - \phi_{pp} \phi_{p-2,p-j}) \rho_{p-j}}{1 - \sum_{j=1}^{p-1} (\phi_{p-2,j} - \phi_{pp} \phi_{p-2,p-j}) \rho_j}, \quad p \geq 3$$

## Autocorrelation and Partial Autocorrelation...

Autocorrelation is simple correlation between  $X_t$  and, say,  $X_{t+h}$ , it is a correlation between a series but with a lag.

While Partial Autocorrelation is a correlation between observations  $X_t$  and  $X_{t+h}$  after removing the linear relationship of all observations that fall between  $X_t$  and  $X_{t+h}$ .

# Look...Think...Infer

Date: 04/27/16 Time: 19:28  
 Sample: 1/01/2014 3/19/2014  
 Included observations: 54

Autocorrelation	Partial Correlation	AC	PAC	Q-Sta...	Prob
██████	██████	1 0.921	0.921	48.435	0.000
██████	██████	2 0.854	0.035	90.871	0.000
██████	██████	3 0.770	-0.14...	126.06	0.000
██████	██████	4 0.697	0.011	155.46	0.000
██████	██████	5 0.636	0.055	180.42	0.000
██████	██████	6 0.520	-0.41...	197.43	0.000
██████	██████	7 0.420	-0.01...	208.75	0.000
██████	██████	8 0.310	-0.03...	215.08	0.000
██████	██████	9 0.224	0.006	218.45	0.000
██████	██████	1... 0.142	-0.07...	219.84	0.000
██████	██████	1... 0.078	0.195	220.26	0.000
██████	██████	1... 0.009	-0.13...	220.27	0.000
██████	██████	1... -0.05...	0.004	220.46	0.000
██████	██████	1... -0.10...	-0.03...	221.24	0.000
██████	██████	1... -0.14...	0.051	222.75	0.000
██████	██████	1... -0.15...	-0.04...	224.69	0.000
██████	██████	1... -0.20...	-0.19...	228.03	0.000
██████	██████	1... -0.24...	-0.10...	232.94	0.000
██████	██████	1... -0.26...	0.182	239.03	0.000
██████	██████	2... -0.28...	-0.09...	246.06	0.000
██████	██████	2... -0.29...	-0.11...	253.79	0.000
██████	██████	2... -0.29...	0.180	262.23	0.000
██████	██████	2... -0.31...	-0.07...	271.88	0.000
██████	██████	2... -0.28...	0.195	280.09	0.000

**Dependent Variable: CLOSE\_PRICE**  
 Method: Least Squares  
 Date: 04/27/16 Time: 19:28  
 Sample (adjusted): 1/02/2014 3/19/2014  
 Included observations: 53 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.36246	20.49878	1.432400	0.1581
CLOSE_PRICE_1	0.922970	0.054046	17.07743	0.0000
R-squared	0.851155	Mean dependent var		379.0009
Adjusted R-squared	0.848237	S.D. dependent var		18.93627
S.E. of regression	7.376970	Akaike info criterion		6.871608
Sum squared resid	2775.404	Schwarz criterion		6.945959
Log likelihood	-180.0976	Hannan-Quinn criter.		6.900200
F-statistic	291.6384	Durbin-Watson stat		2.059933
Prob(F-statistic)	0.000000			

**Dependent Variable: CLOSE\_PRICE**  
 Method: Least Squares  
 Date: 04/27/16 Time: 19:28  
 Sample (adjusted): 1/03/2014 3/19/2014  
 Included observations: 52 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	28.66232	21.27551	1.347198	0.1841
CLOSE_PRICE_1	0.889037	0.142718	6.229324	0.0000
CLOSE_PRICE_2	0.035967	0.142684	0.252076	0.8020
R-squared	0.851589	Mean dependent var		379.1365
Adjusted R-squared	0.845531	S.D. dependent var		19.09502
S.E. of regression	7.504823	Akaike info criterion		6.924930
Sum squared resid	2759.796	Schwarz criterion		7.037502
Log likelihood	-177.0482	Hannan-Quinn criter.		6.968087
F-statistic	140.5819	Durbin-Watson stat		1.978242
Prob(F-statistic)	0.000000			



# Correlogram...

Date: 04/13/12 Time: 10:12

Sample: 1996 2010

Included observations: 15

*What is this?*

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.729	0.729	9.6842	0.002
		2	0.429	-0.218	13.302	0.001
		3	0.168	-0.122	13.900	0.003
		4	-0.025	-0.075	13.914	0.008
		5	-0.069	0.118	14.035	0.015
		6	-0.104	-0.117	14.340	0.026
		7	-0.134	-0.072	14.913	0.037
		8	-0.170	-0.081	15.961	0.043
		9	-0.209	-0.055	17.822	0.037
		10	-0.248	-0.116	20.959	0.021
		11	-0.283	-0.107	26.048	0.006
		12	-0.264	-0.006	31.965	0.001

# A discussion on Ljung-Box Q statistic

- **Ljung-Box Q statistic:** The Ljung-Box Q-test is a quantitative way to test for autocorrelation at multiple lags jointly.
- **The null hypothesis** for this test is that the first  $m$  autocorrelations are jointly zero,  $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ .
- Alternatively, at least one of  $\rho_i$  is not zero.
- Note that it is a test related to AUTOCORRELATION.
- **Ljung-Box Q statistic** is having asymptotic  $\chi^2$  distribution

# Correlogram...

*Now, interpret the results!*

Date: 04/13/12 Time: 10:12  
 Sample: 1996 2010  
 Included observations: 15

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.729	0.729	9.6842	0.002
		2	0.429	-0.218	13.302	0.001
		3	0.168	-0.122	13.900	0.003
		4	-0.025	-0.075	13.914	0.008
		5	-0.069	0.118	14.035	0.015
		6	-0.104	-0.117	14.340	0.026
		7	-0.134	-0.072	14.913	0.037
		8	-0.170	-0.081	15.961	0.043
		9	-0.209	-0.055	17.822	0.037
		10	-0.248	-0.116	20.959	0.021
		11	-0.283	-0.107	26.048	0.006
		12	-0.264	-0.006	31.965	0.001



Let's work with these ways of forecasting...

We have data about...

Sales Data of ONGC

Price Data of Maruti

Price Data of Maruti and Sensex

# Augmented Dickey Fuller Test

$$X(t) = \text{Rho} * X(t-1) + \text{Er}(t)$$

- Null Hypothesis  $\text{rho} = 1$  = non- stationary
- Alternate  $\text{Rho}$  not equal to 1 hence series is stationary.

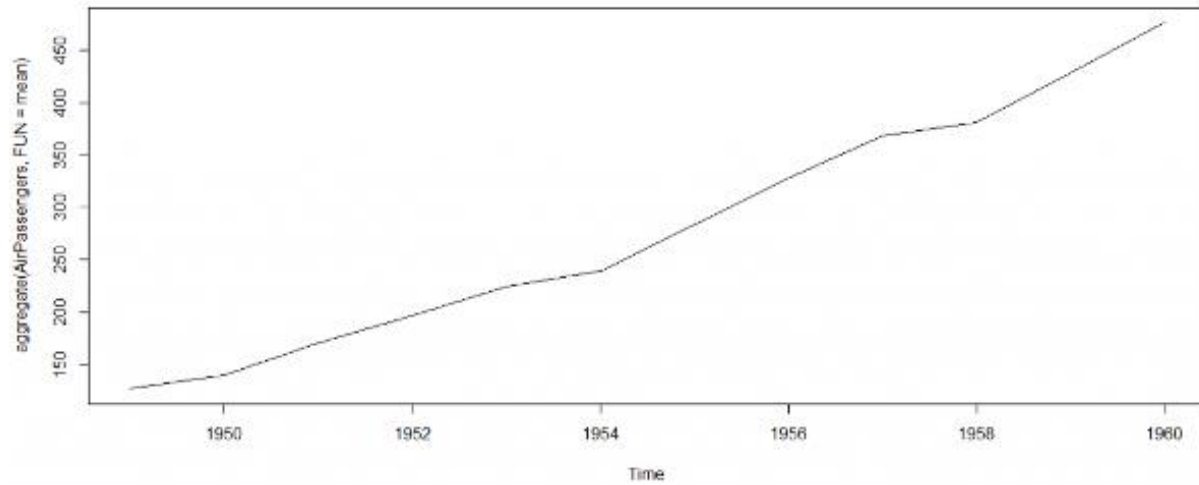
# R Code for running test

)

```
>data(AirPassengers)
```

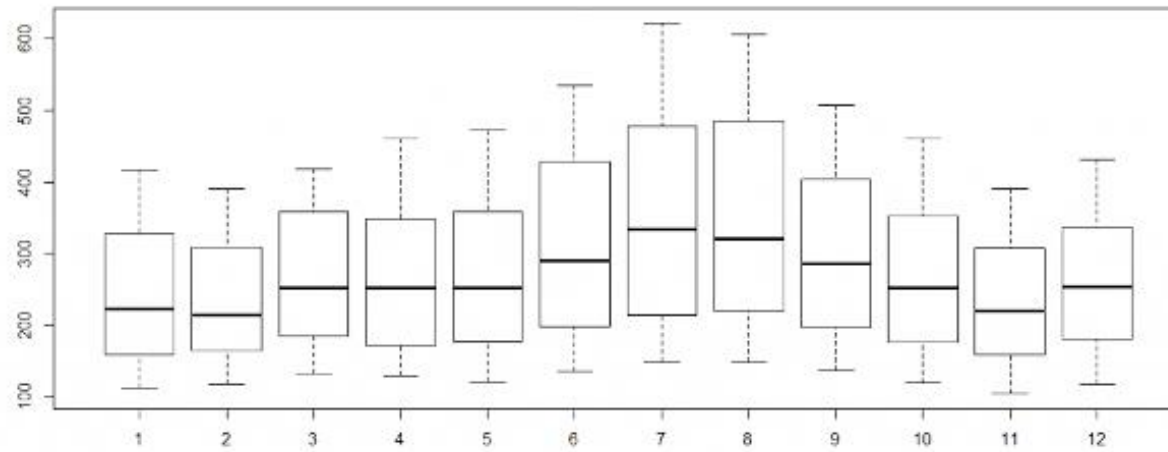
- `class(AirPassengers)`
- `summary(AirPassengers)`
- `plot(AirPassengers)`
  
- `abline(reg=lm(AirPassengers~time(AirPassengers)))` This will fit a line
  
- `cycle(AirPassengers)`
- `plot(aggregate(AirPassengers,FUN=mean)`
- 
  
- `boxplot(AirPassengers~cycle(AirPassengers))`
  
-

# Plot



- `plot(aggregate(AirPassengers, FUN = mean))`

# Box plot

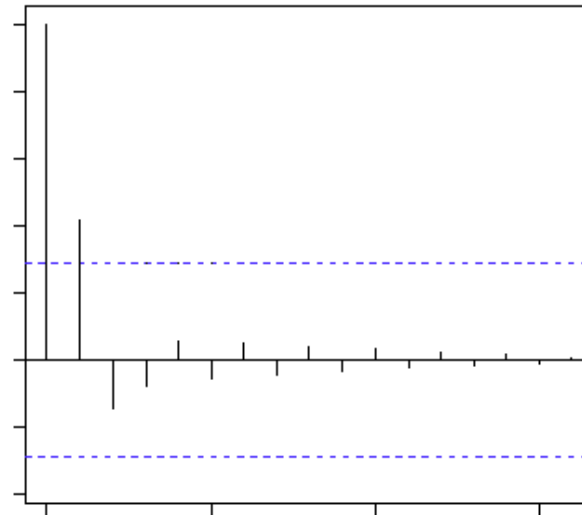


- `boxplot(AirPassengers~cycle(AirPassengers))`

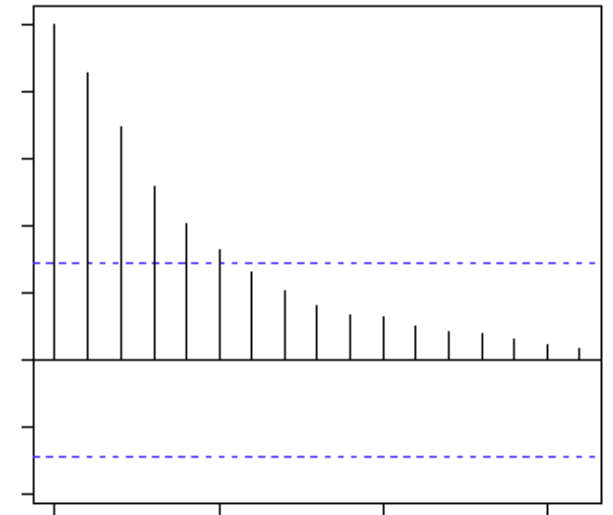
-

# ACF and PACF plot

- ACF and PACF plot
- To identify AR terms ACF should be declining and PACF should show a cut-off point. In this case PACF cut-off is at 2 lags hence 2 AR terms should be used.



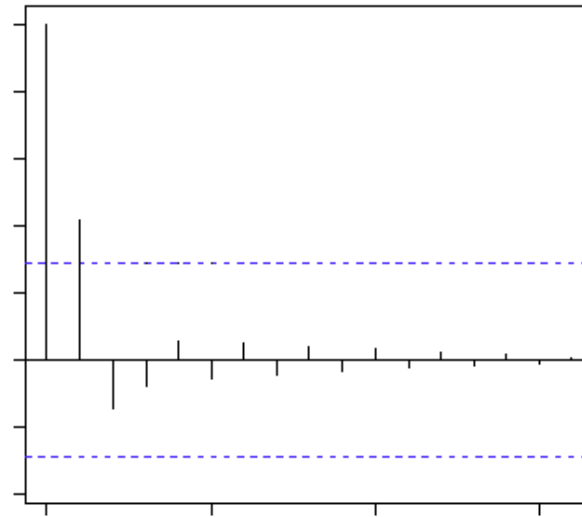
PACF



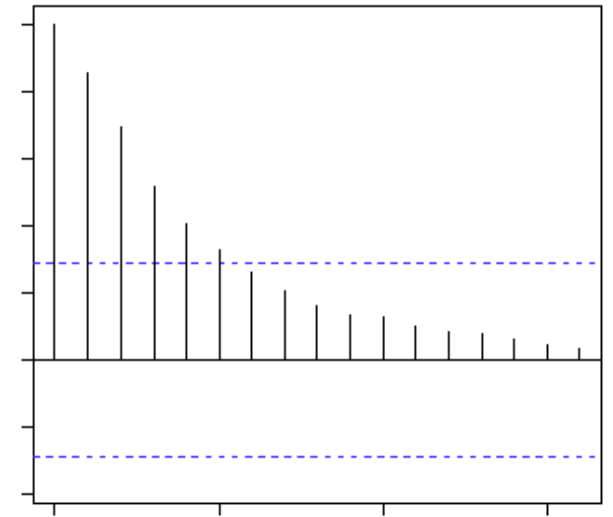
ACF

# MA terms

- MA terms , PACF should be declining and ACF should show cut-off



ACF



PACF

# Steps in ARIMA process MA terms

1. Visualize the time series

```
graph TD; A[1. Visualize the time series] --> B[2. Stationarize the series]; B --> C[3. Plot ACF/PACF charts and find optimal parameters]; C --> D[4. Build the ARIMA model]; D --> E[5. Make Predictions];
```

2. Stationarize the series

3. Plot ACF/PACF charts and find optimal parameters

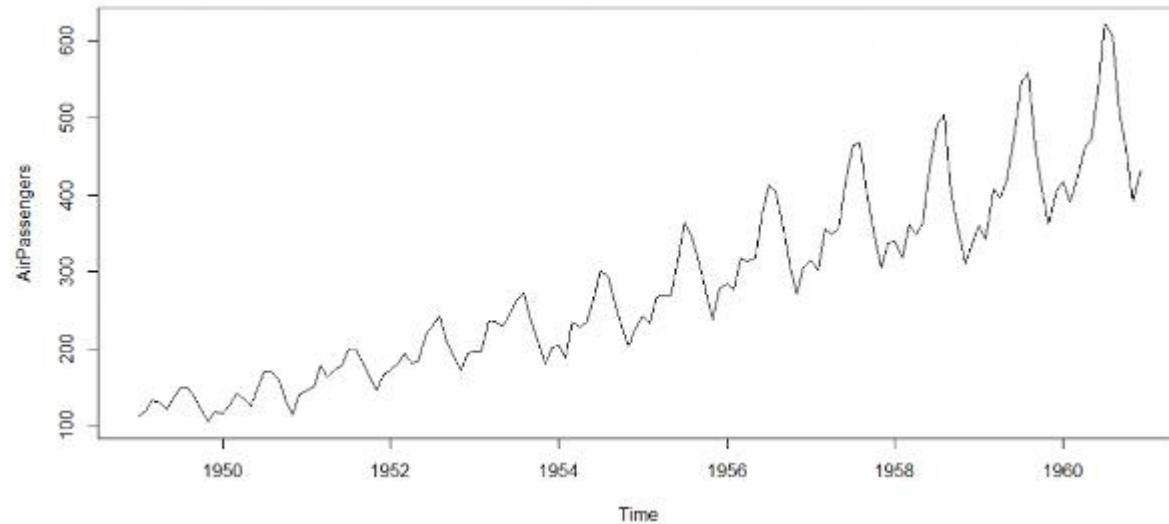
4. Build the ARIMA model

5. Make Predictions

## 2. Converting series to Stationary

- $x(t) - x(t-1) = \text{ARMA}(p, q)$  = just différence the séries

# Let's Begin forecasting



- There is a trend component which grows the passenger year by year.
- There looks to be a seasonal component which has a cycle less than 12 months.
- The variance in the data keeps on increasing with time.

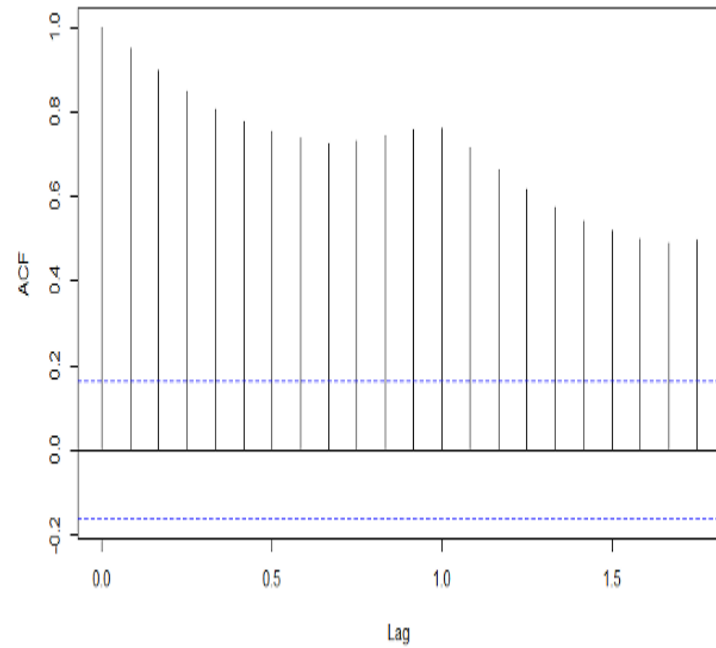
# We need to address these problems

- We remove unequal variance by taking log.
- We need to remove trend by taking difference.
- Last we need to check for stationarity.
  
- `adf.test(diff(log(AirPassengers)), alternative="stationary", k=0)`

data: diff(log(AirPassengers)) Dickey-Fuller = -9.6003, Lag order = 0,  
p-value = 0.01 alternative hypothesis: stationary

# Plot ACF and PACF

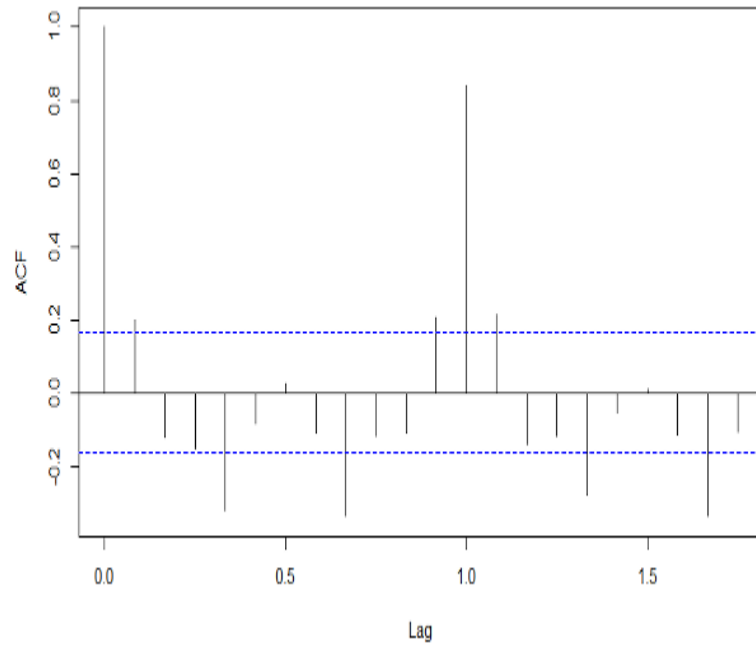
```
acf(log(AirPassengers))
```



Clearly, the decay of ACF chart is very slow This is run on raw series

# Plot ACF and PACF

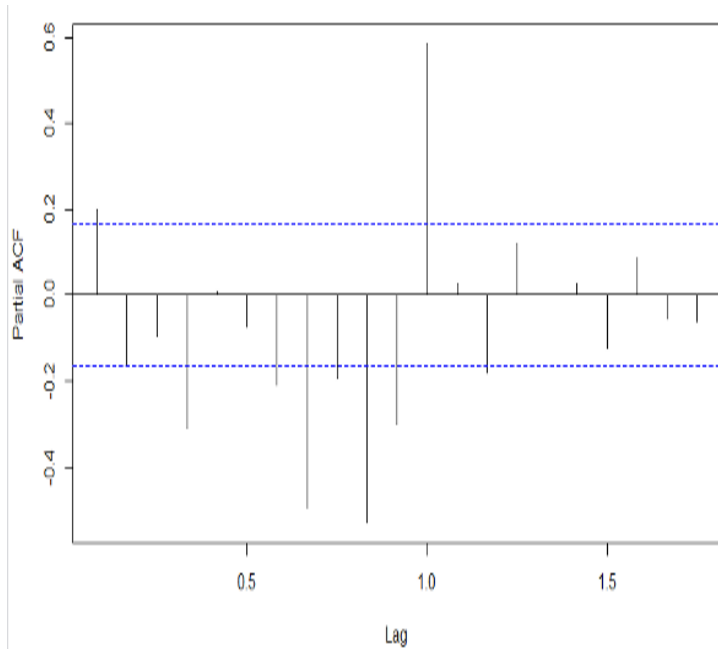
```
acf(log(AirPassengers))
```



Clearly, the decay of ACF chart is very slow. This is run on differenced series

# Plot ACF and PACF

```
pacf(log(AirPassengers))
```



- Clearly, the decay of PACF chart is declining.
- After a few iterations, we found that (0,1,1) as (p,d,q) comes out to be the combination with least AIC and BIC.

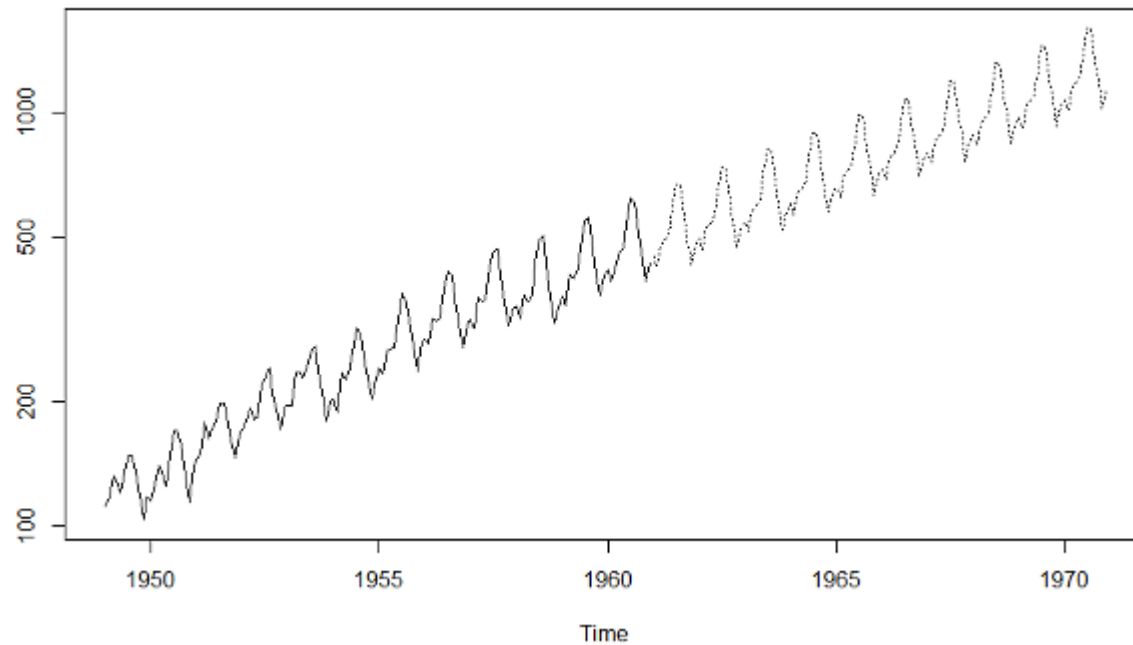
# ARIMA fitting

```
(fit <- arima(log(AirPassengers), c(0, 1, 1)  
,seasonal = list(order = c(0, 1, 1), period = 12
```

# Predicting

```
pred <- predict(fit, n.ahead = 10*12)
```

```
ts.plot(AirPassengers, 2.718^pred$pred, log = "y", lty = c(1,3))
```



What is happening with HDFC bank ?????