INTRODUCTION TO TIME SERIES ANALYSIS

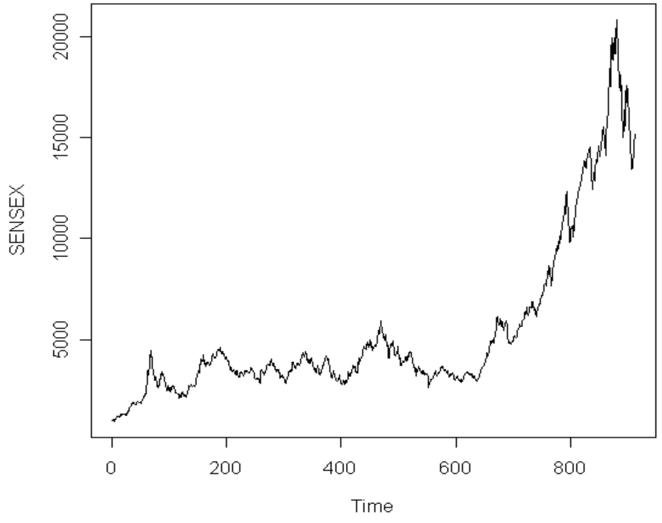
DATA SCIENCE INSTITUTE

What is Time Series?

- Time Series is a sequence of values measured over time.

- Time series can be

- Annual- (GDP, Company Turnover)
- Quarterly-(GDP, Company Turnover)
- Monthly-(Inflation Rates)
- Daily-(Stock prices, Gold Prices)



DATA SCIENCE INSTITUTE

Components of Time Series?
Trend, Seasonality and Cyclic Pattern

- **Trend** refers to the long-term increase or decrease in the time series.
- **Seasonality** in a time series refers to predictable and recurring trends and patterns over a period of time, normally a year. An example of a seasonal time series is retail data, which sees spikes in sales during holiday seasons like Christmas or Diwali.
- A **cyclic pattern** exists when data exhibits rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least 2 years.

Time Series Data Analysis

- **Analyze trend and seasonality** present in the data.
- **Decompose** time series into its components.
- Forecast future values of the time series.

Stationary Time Series

- Time series process is called **stationary** if the statistical properties of the process **remain unchanged over time**.

i.e if Y_t is a time series t=1,2,3,... then

$E(Y_t)= \mu_t = \mu$ (constant) $\forall$ t=1,2,...

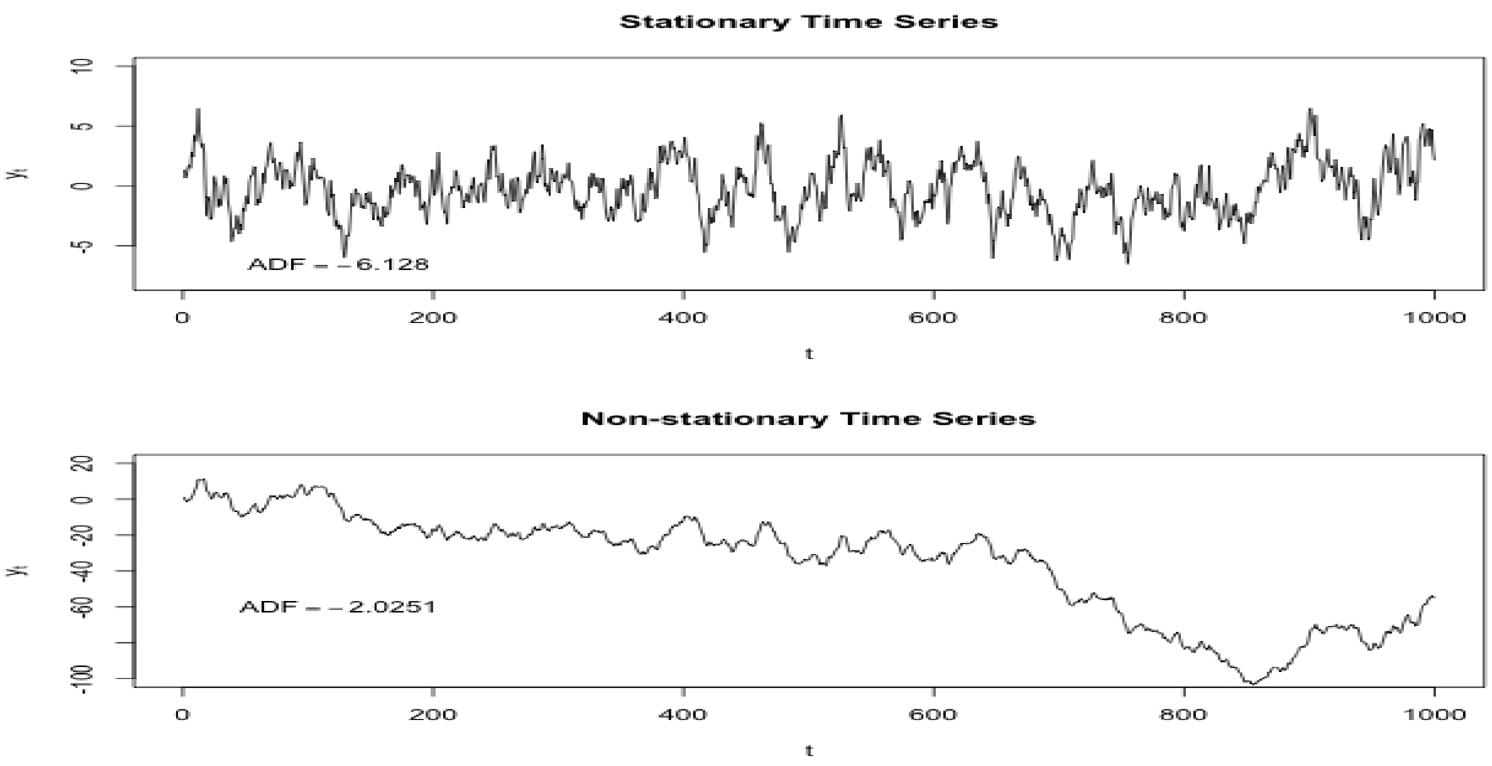
$Var(Y_t)= \sigma_t^2 = \sigma^2$ (constant) $\forall$ t=1,2...

$Cov(Y_t,Y_{t-s})$ depends only on **s**(lag),

and is independent of **t** (time).

DATA SCIENCE INSTITUTE

Stationary Time Series

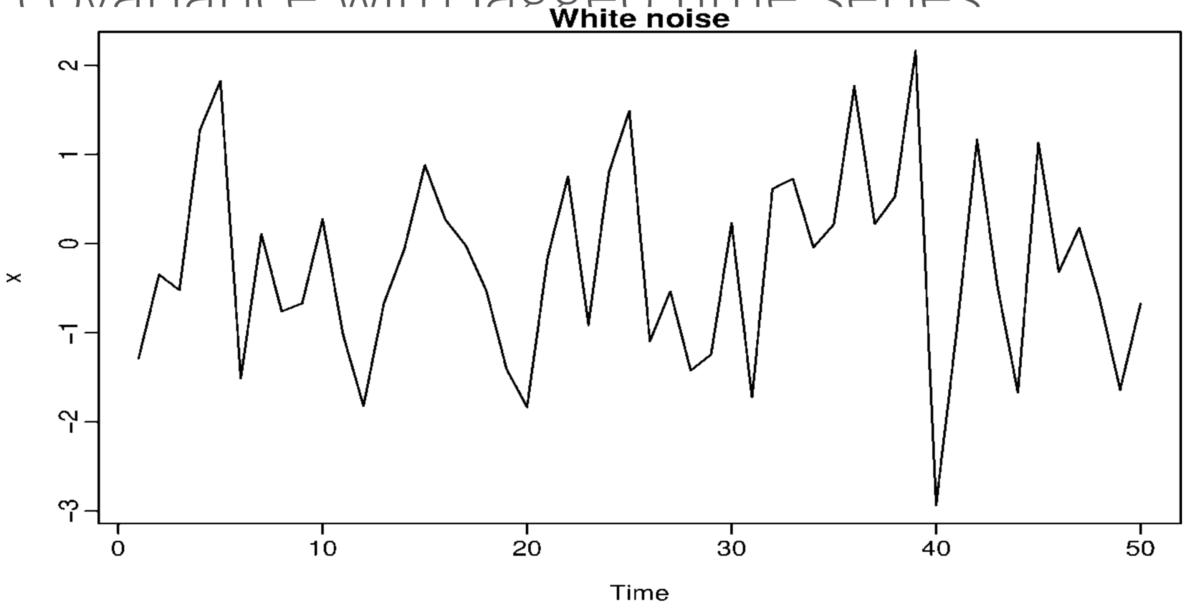


- Examples: Stock returns can be stationary although stock prices are non-stationary
- Macroeconomic data such as GDP, Inflation Rates are also non stationary.

DATA SCIENCE INSTITUTE

White Noise Process

- **White noise** is the simplest example of stationary time series.
- White Noise time series has zero mean, constant variance and zero covariance with lagged time series



DATA SCIENCE INSTITUTE

Random Walk
Non-Stationary Time Series

Random walk is the simplest case of non-stationary time series. It is of the form

$$Y_t = Y_{t-1} + U_t \quad t=1,2,3..... \quad (1)$$

We assume that U_t is a random series with constant mean μ and constant variance σ^2. Also it is serially uncorrelated .

From 1 it is clear that value of Y at time t is equal to its value at time (t-1) plus a random shock.

Let $Y_0=0$ at time $t=0$, hence

$$Y_1 = U_1$$
$$Y_2 = Y_1 + U_2 = U_1 + U_2$$
$$Y_3 = Y_2 + U_3 = U_1 + U_2 + U_3$$
$$\text{so,} \quad Y_t = \Sigma\, U_t \qquad t=1,2,3.....$$

DATA SCIENCE INSTITUTE

Why Random Walk is Non-Stationary?

- $Y_t = \Sigma U_t$
- $E(Y_t) = E(\Sigma U_t) = t.\mu$
- i.e Mean is not constant

- $Y_t = \Sigma U_t$
- $var(Y_t) = E(\Sigma U_t) = t. \sigma^2$
- Variance is not constant

Conclusion:

Random Walk is a non stationary time series.

Importance of Stationary Time Series

- **Calibration** (Estimation of model parameters using historical data.) is an important concept in the **forecasting** of time series values.
- In the **calibration** of time series models we need a stationary time series.
- With a non stationary time series we get into **spurious** regression, which badly affects forecasting.

How to Make a Non Stationary Time Series Stationary?

- There are 2 methods to make a non stationary time series stationary.
  - Differencing
  - De-trending

- Differencing: A non stationary time series can be made stationary by differencing. Consider the following non- stationary process

$$Y_t=Y_{t-1}+U_t \quad t=1,2,3.....$$

We assume that U_t is a random series with constant mean μ and constant variance σ^2 also it is serially uncorrelated i.e (U_t is stationary).

- Hence $Y_t-Y_{t-1}=\Delta Y_t = U_t$, which is a stationary time series.

- Differencing can be well applied in case of stochastic time series.

DATA SCIENCE INSTITUTE

How to Make a Non Stationary Time Series Stationary? (Contd.)

- Detrending:

Consider the following regression:

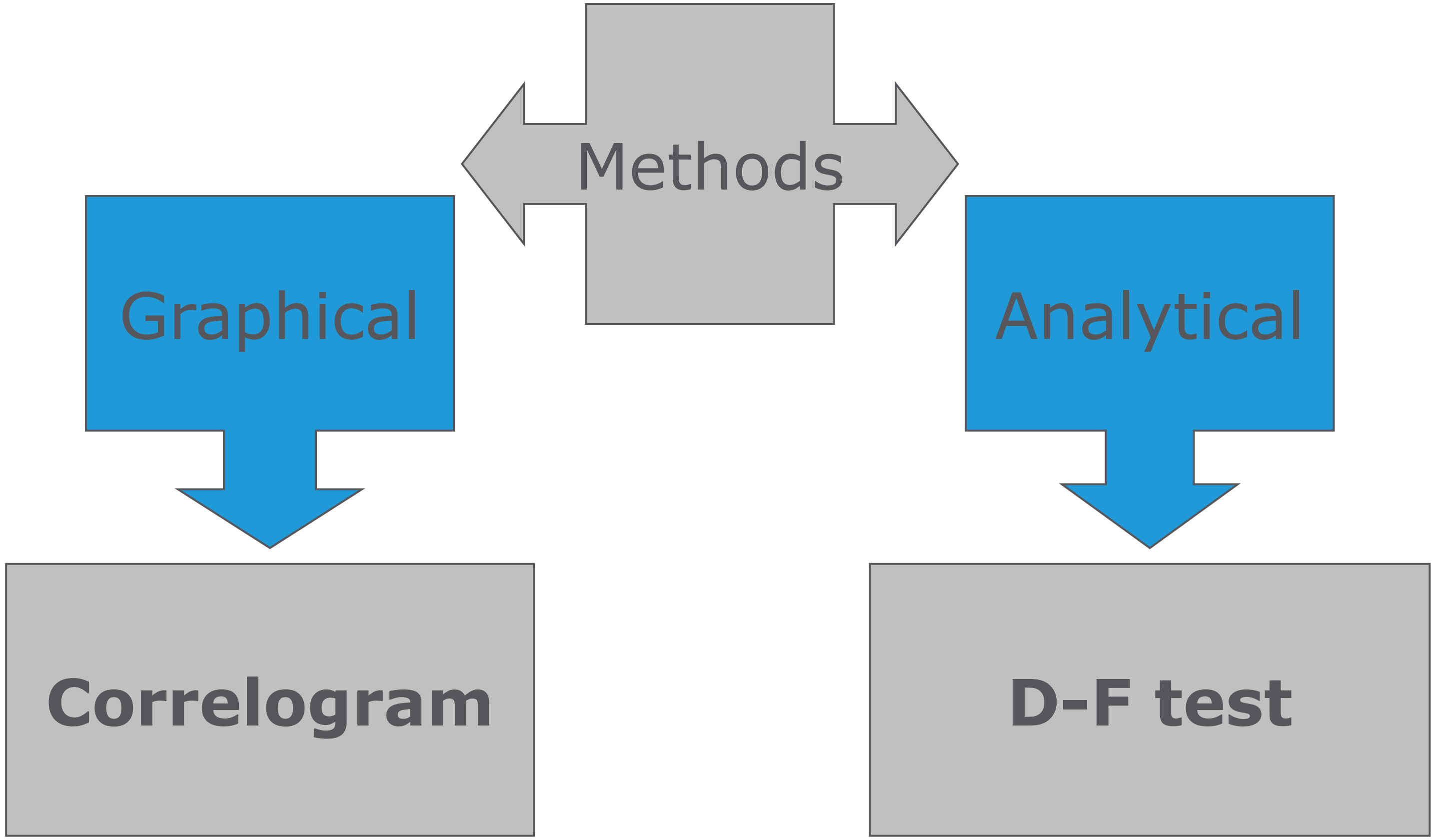
$$Y_t = \beta_1 + \beta_2 t + U_t$$

where U_t is stationary with, say mean zero and variance σ^2.

- When trend element $(\beta_1 + \beta_2 t)$ is subtracted from the above equation, the result is a stationary process.
- De-trending is useful when trend is deterministic.
- Forecasts made from a Trend-Stationary Process will be more reliable in the long-run.

DATA SCIENCE INSTITUTE

Identifying Stationary Time Series



DATA SCIENCE INSTITUTE

Concept of Autocorrelation

- A mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals.

- Autocorrelation is a correlation between a time series (Y_t) and another time series representing lagged values of the same time series.(Y_{t-k})

- Autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, the correlation is between two values of the same variable at times Y_t and Y_{t-k}

DATA SCIENCE INSTITUTE

Autocorrelation of Lag 1

Year	Sales(m)
Jan	110
Feb	113
Mar	121
Apr	123
May	126
Jun	120
Jul	119
Aug	127
Sep	129
Oct	131
Nov	130
Dec	132

Year	Yt	Yt-1
Jan	110	
Feb	113	110
Mar	121	113
Apr	123	121
May	126	123
Jun	120	126
Jul	119	120
Aug	127	119
Sep	129	127
Oct	131	129
Nov	130	131
Dec	132	130

$$\sum_{t=2}^{n} \frac{(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

$$r_1 = 0.588$$

Autocorrelation of Lag 2

Year	Sales(m)
Jan	110
Feb	113
Mar	121
Apr	123
May	126
Jun	120
Jul	119
Aug	127
Sep	129
Oct	131
Nov	130
Dec	132

Year	Yt	Yt-2
Jan	110	
Feb	113	
Mar	121	110
Apr	123	113
May	126	121
Jun	120	123
Jul	119	126
Aug	127	120
Sep	129	119
Oct	131	127
Nov	130	129
Dec	132	131

$$r_2 = 0.204$$

Correlogram

- Plot of the sample **Autocorrelation** function against lag is called **correlogram**.

- ACF (autocorrelation function) is a general expression for lag k autocorrelation.

- Correlogram is mainly used in deciding stationarity of a series.

- r_K = Sample autocorre

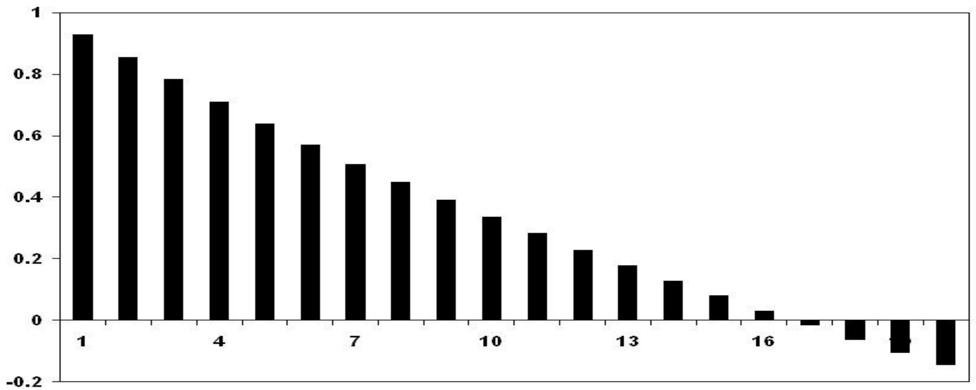
$$\frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-K} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

Correlogram for Checking Stationarity

- For a non-stationary time series showing trend, **slow decay pattern will be observed.**

- If a time series is characterized by seasonal fluctuations, then the correlogram would also exhibit oscillations at the same frequency

GRAPHICAL TECHNIQUES FOR DETECTING NONSTATIONARITY



Sample correlogram of *LGDPI*

This figure presents the sample correlogram.

DATA SCIENCE INSTITUTE

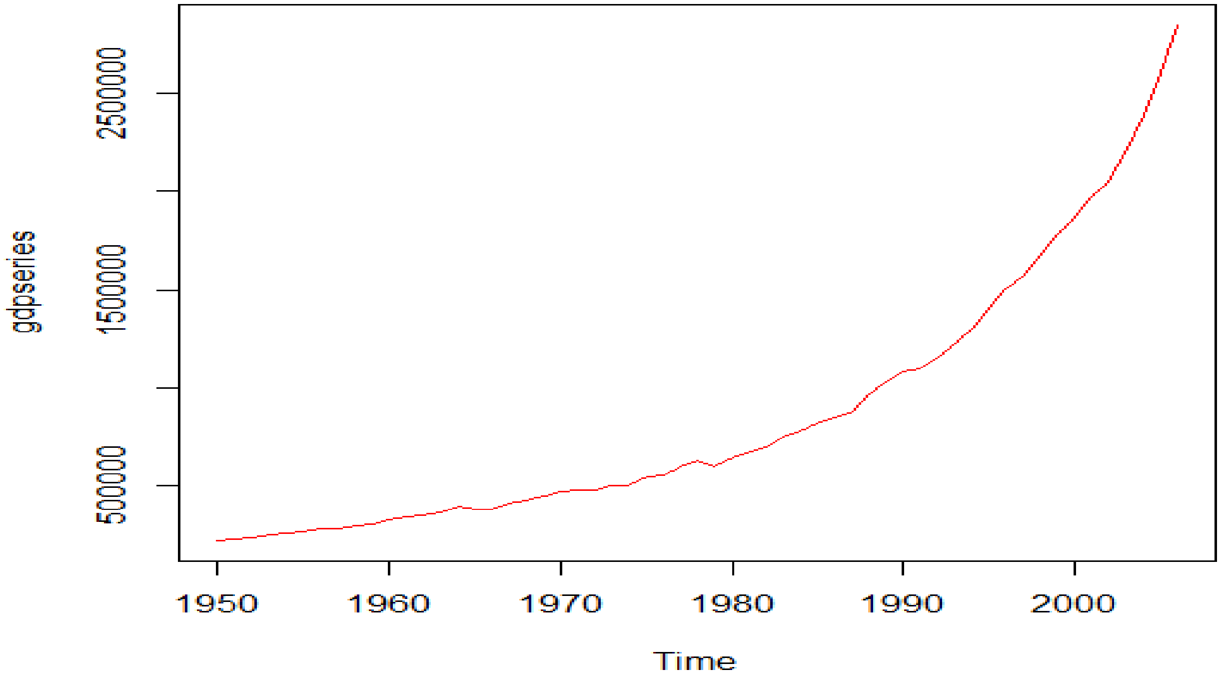
GDP Time Series Data Snapshot

Year	GDP
1950-51	224786
1951-52	230034
1952-53	236562
1953-54	250960
1954-55	261615
1955-56	268316
1956-57	283589
1957-58	280160
1958-59	301422
1959-60	308018
1960-61	329825
1961-62	340060
1962-63	347253
1963-64	364834
1964-65	392503
1965-66	378157
1966-67	382006
1967-68	413094
1968-69	423874

This is partial data.
The data has GDP values for 1950-51 to 2006-07

Time Series Analysis in R
Plot Time Series

```
gdpdata<-read.csv(file.choose(),header=T)
# Define time series using ts function
gdpseries<-ts(gdpdata$GDP,start=1950,end=2006)
plot(gdpseries,col="red")
```

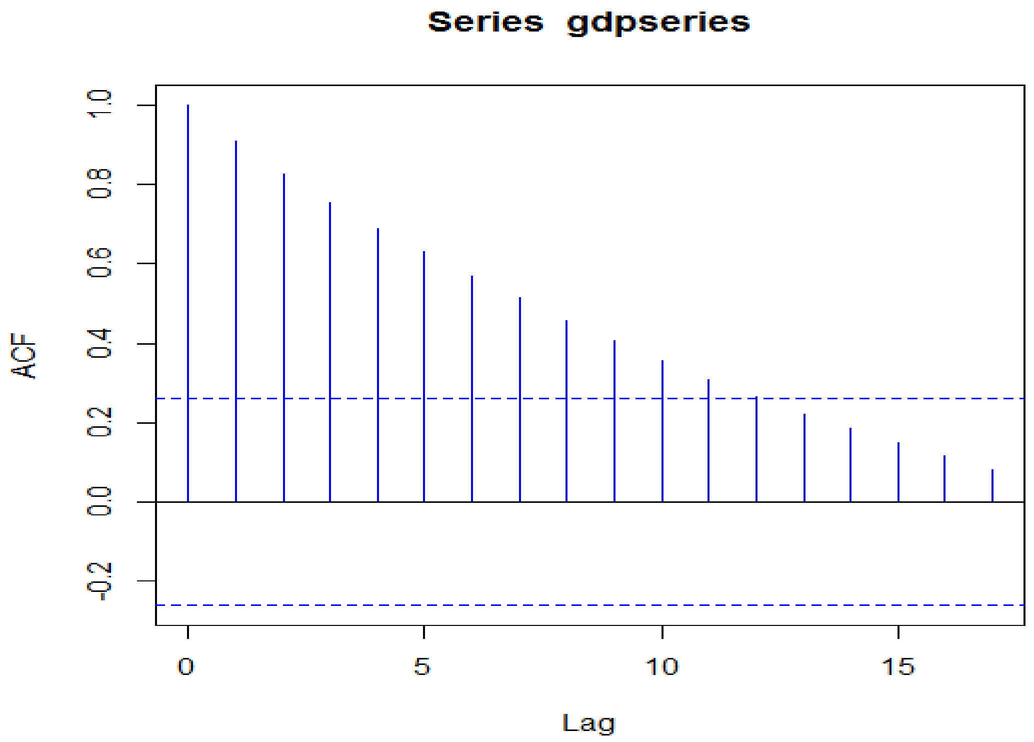


Clearly a non-stationary time series.

DATA SCIENCE INSTITUTE

Time Series Analysis in R
Correlogram

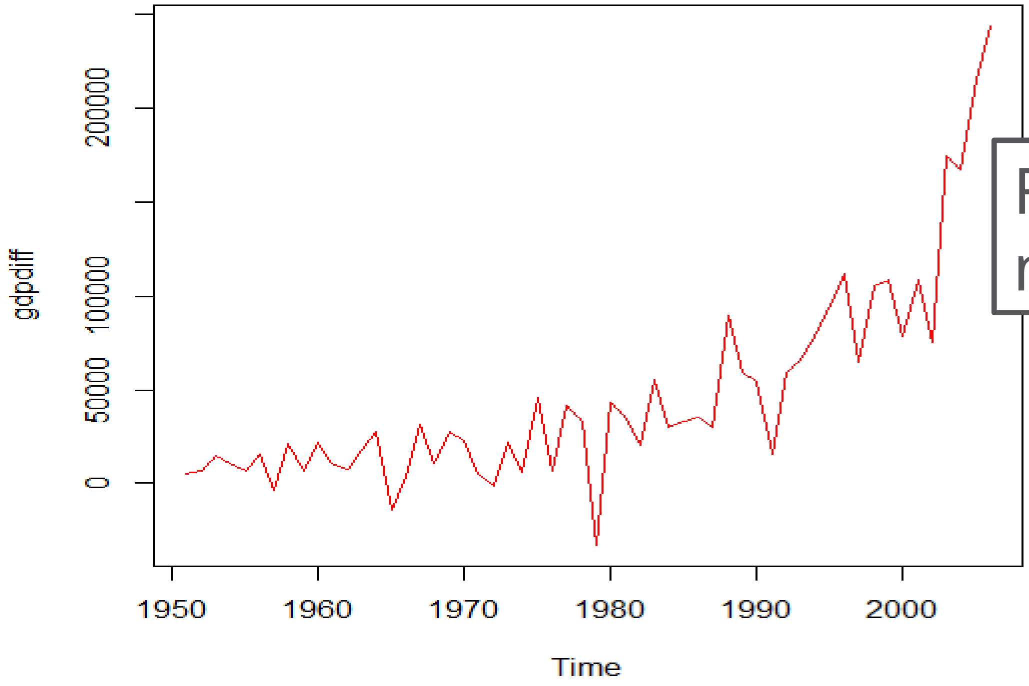
```
acf(gdpseries,col="blue")
```



Very slow decay.
Sign of non-stationarity

Time Series Analysis in R
Plot Difference Time Series

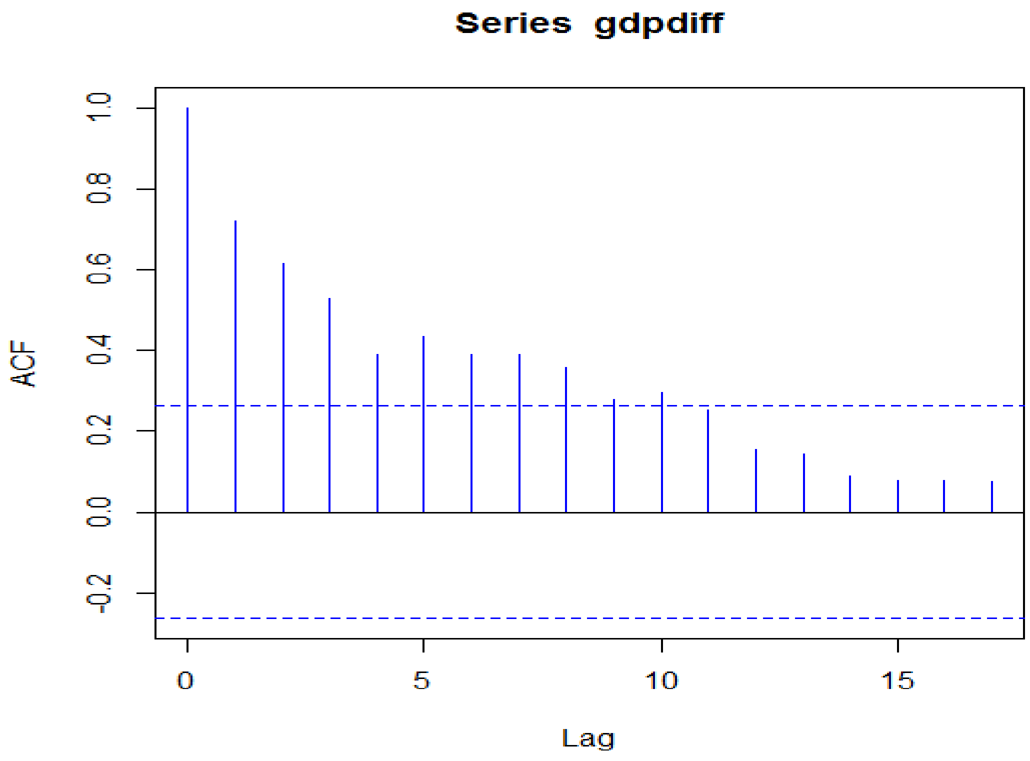
```
gdpdiff<-diff(gdpseries,differences=1)
plot(gdpdiff)
```



First difference also looks non-stationary time series.

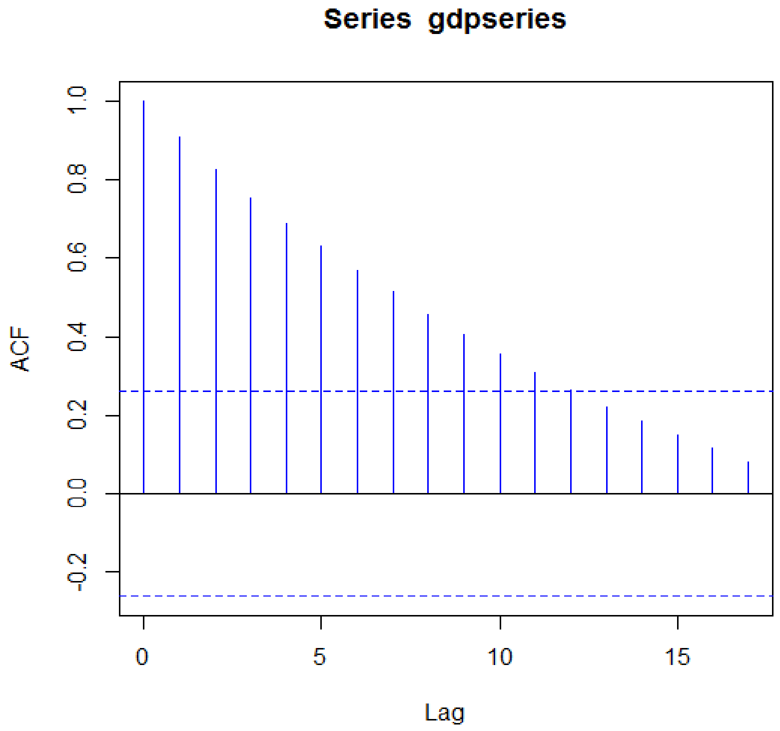
Time Series Analysis in R
Correlogram for Difference Time Series

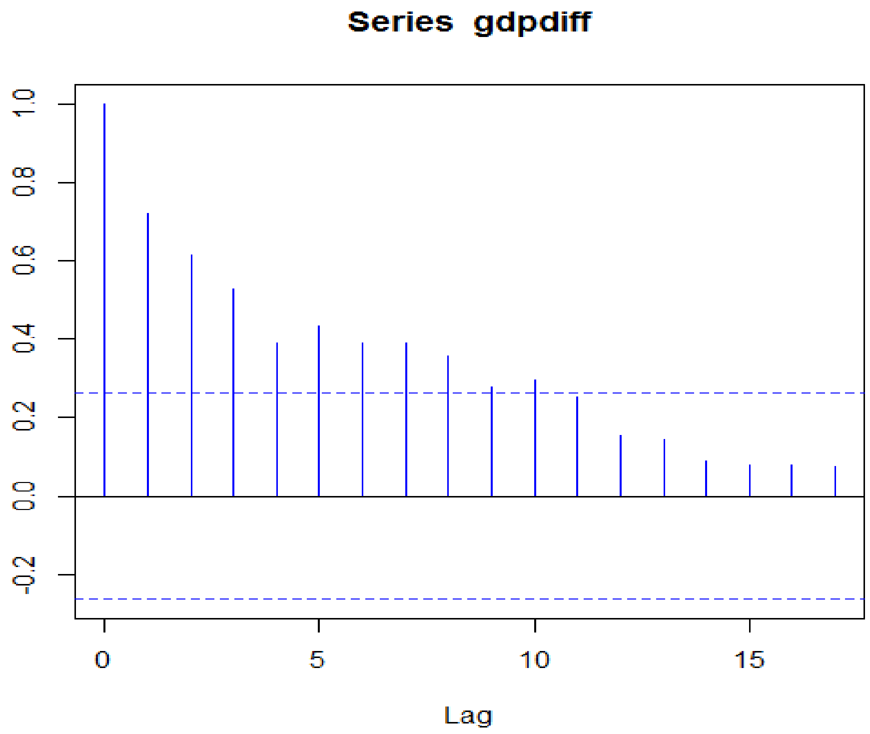
```
acf(gdpdiff,col="blue")
```



Slow decay.
Sign of non-stationarity

Time Series Analysis in R
Correlograms



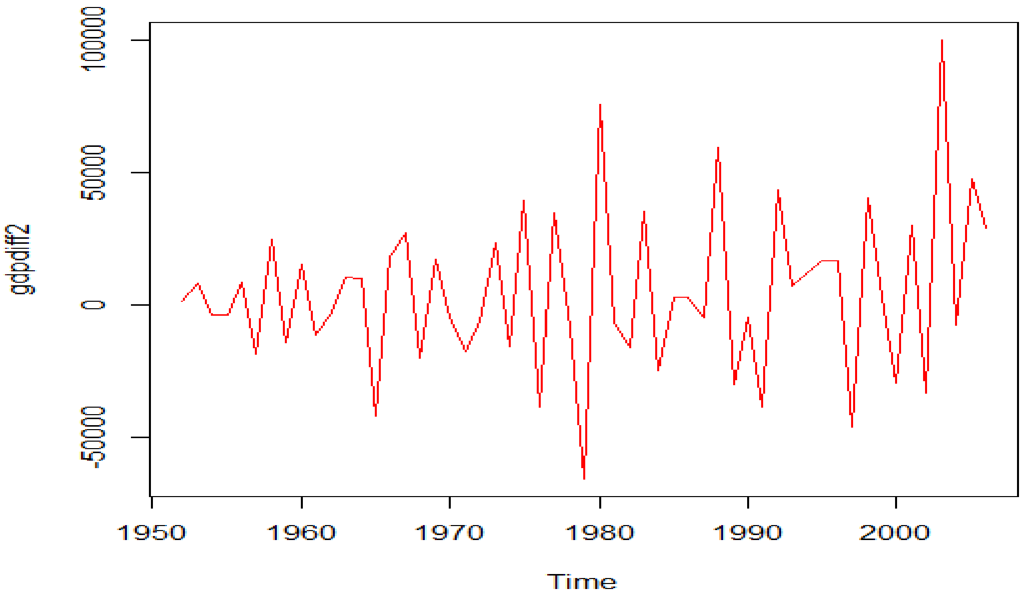


Stationarity is not achieved with first difference

DATA SCIENCE INSTITUTE

How Many Times Should Time Series Be Differenced to make Stationary?

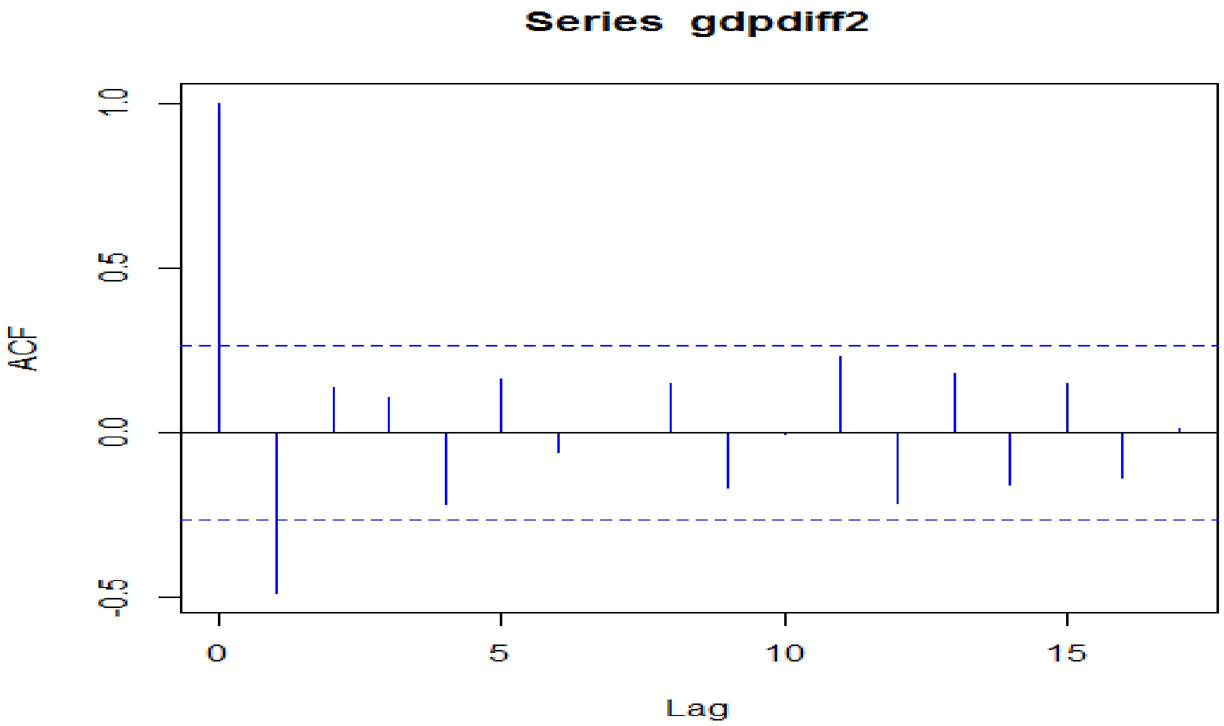
```
install.packages("forecast")
library(forecast)
ndiffs(gdpseries)   # gives 2
gdpdiff2<-diff(gdpseries,differences=2)
plot(gdpdiff2,col="red")
```



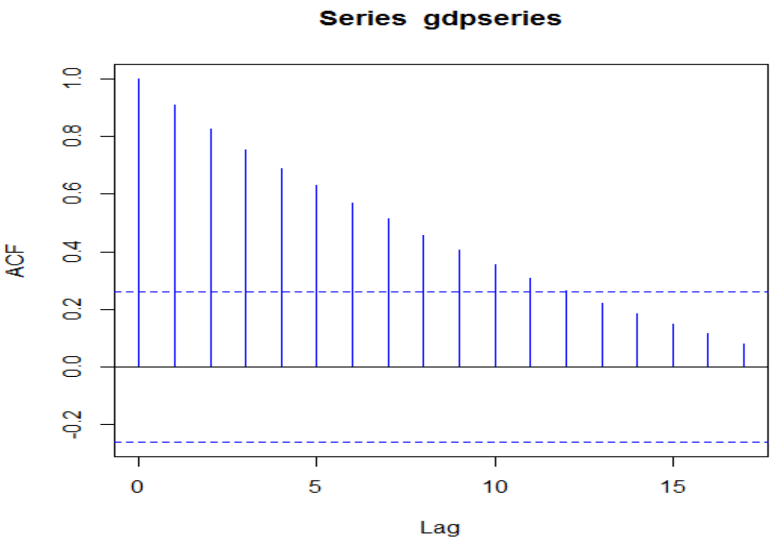
DATA SCIENCE INSTITUTE

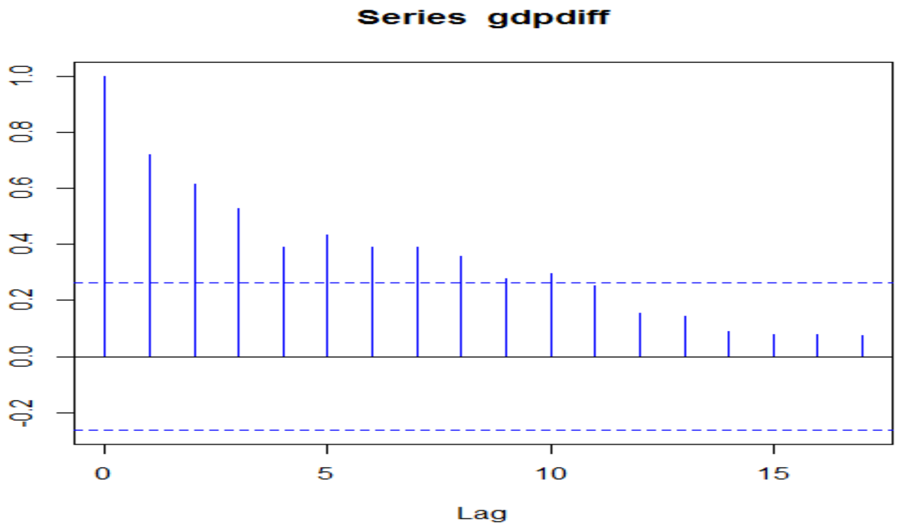
Time Series Analysis in R
Correlogram for Difference2 Time Series

```
acf(gdpdiff2,col="blue")
```

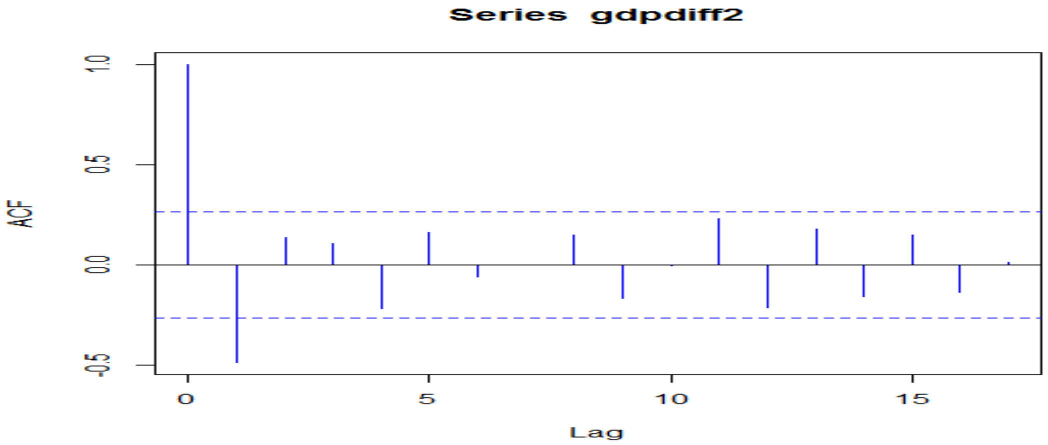


Time Series Analysis in R
Correlograms





Variai
Cumu
e
Propo



)0

Stationarity is achieved with second order difference

DATA SCIENCE
INSTITUTE

Dickey Fuller (DF) test

- Consider X_t (t=1,2,3,...) is a time series of the form

$$X_t=\rho X_{t-1}+U_t. \quad ....(1)$$

- If $\rho=1$ then X_t becomes a random walk.

- We assume that $U_t \sim IID(0, \sigma^2)$, i.e U_t is a white noise.

- So we are interested in testing for $\rho=1$.

DATA SCIENCE INSTITUTE

D-F (Unit Root) Test

- H_0: $\rho=1$ (Non stationary) , H_1: $\rho<1$(stationary)

$$(X_t-X_{t-1}) = \Delta X_t=(\rho-1)X_{t-1}+U_t \ldots\ldots\ldots\ldots\text{from}(1)$$

- H_0: $\rho^*=0$, H_1: $\rho^*<0$, $\rho^*=(\rho-1)$.
- Test statistic is $(\rho^*/SE(\rho^*))$.
- Test statistic follows DF distribution under null.
- Reject H_0 if tcal < DF table value.

Time Series Analysis in R
Dickey Fuller Test

```
library(urca)
df<-ur.df(gdpseries,lag=0)
summary(df)
```

Value of test-statistic is: 19.2745

Critical values for test statistics:

	1pct	5pct	10pct
tau1	-2.6	-1.95	-1.61

Inference: Time series is non-stationary. Value of test statistic is greater than 5% critical value.

Time Series Analysis in R
Dickey Fuller Test

```
library(urca)
df<-ur.df(gdpdiff2,lag=0)
summary(df)
```

```
Value of test-statistic is:  -11.9083

Critical values for test statistics:

      1pct   5pct  10pct
tau1  -2.6  -1.95  -1.61
```

Inference: Time series is stationary. Value of test statistic is less than 5% critical value.

THANK YOU!!

DATA SCIENCE
INSTITUTE