Contents

Introduction to Time Series

Introduction

Welcome to the module on ‘**Time Series**’. In the last module, you learnt how to use **tree models**, such as decision trees and random forests, to solve classification problems in an intuitive manner. You also learnt how to use tree models to solve regression problems.

In this module

Now in this module, you will learn about time series. Often, you come across **time stamped data**. Time stamped data is basically a sequence of data that has time values attached to the sequence of values, such as

Time Stamp	Value
1	2.4
2	3.1
3	5
4	4.5
5	7.2
6	6.8

Time Stamped Data

Assume you have to **forecast** the value for timestamp 7. Now, based on the knowledge you gained in the previous modules, you may think, why not use regression for forecasting?

However, it is not enough to simply use regression to make the forecast. You have to do something more than that to make an accurate forecast. You will learn more about this in this module.

Let's listen to Raj to learn why forecasting is necessary in an industry setting.

In this session

In this session, you will be introduced to the idea of time series. The topics that will be covered are

- How does time series analysis **differ from regression analysis**?
- What are the **basic features** and **characteristics** of a time series?
- What are the **basic components** of a time series?

Downloads

This module on time series has integrated lecture notes for all the sessions combined, and it will complement and supplement your knowledge. You can download the lecture notes from the link given below, and refer to it as and when required.

Introduction to Time Series

Time Series vs Regression

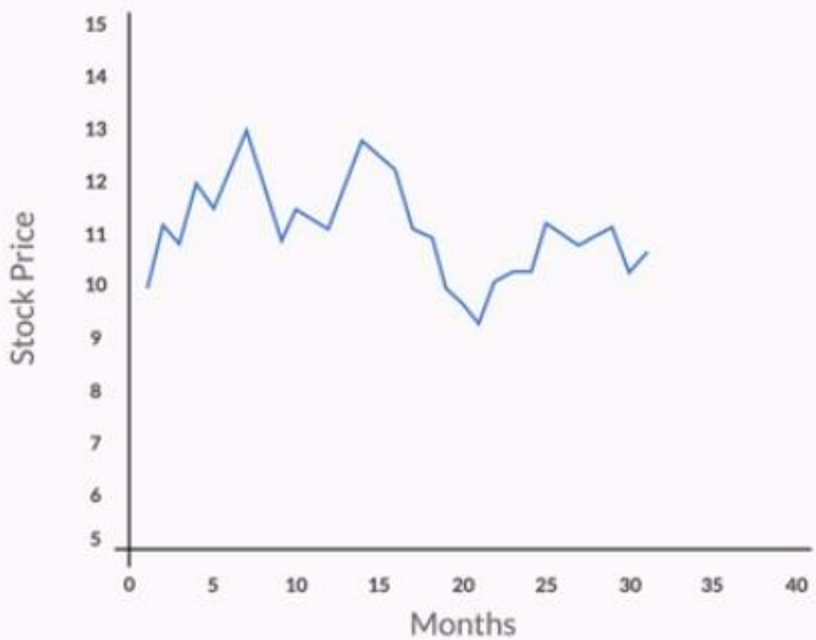
So, as you saw in the last section, while a time series analysis may look similar to simple regression, the results obtained by regression alone are not enough for making an accurate forecast. So, in this session, let's explore why this is the case; and if this is true, you'll learn what else needs to be done.

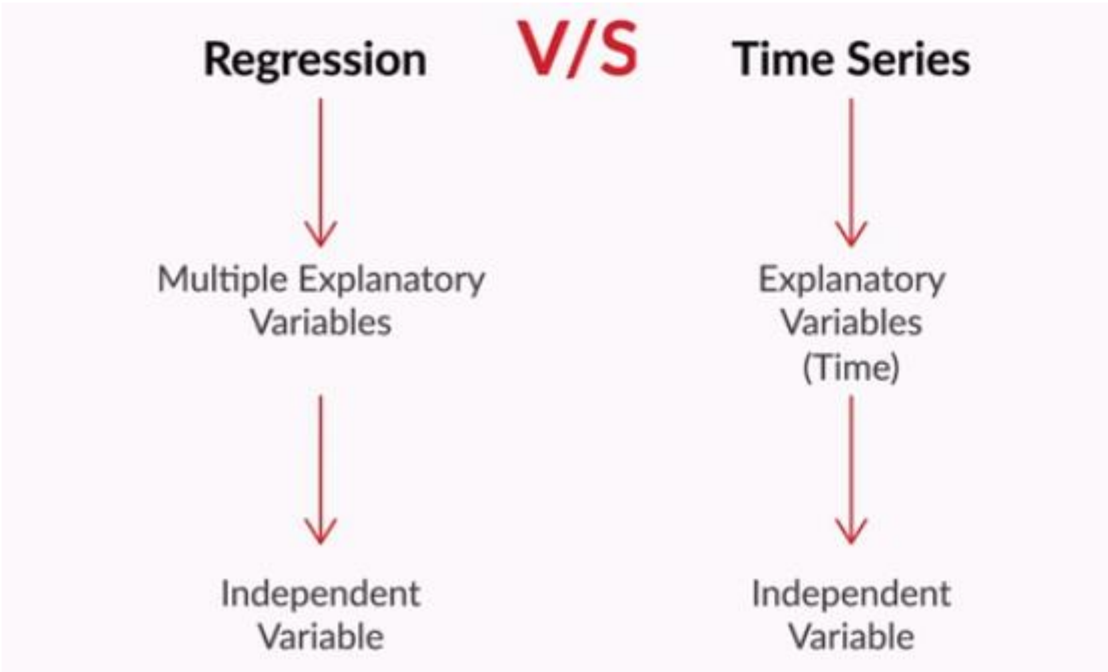
However, before getting into it, let's first understand the major differences between regression and time series, from a non-mathematical viewpoint.

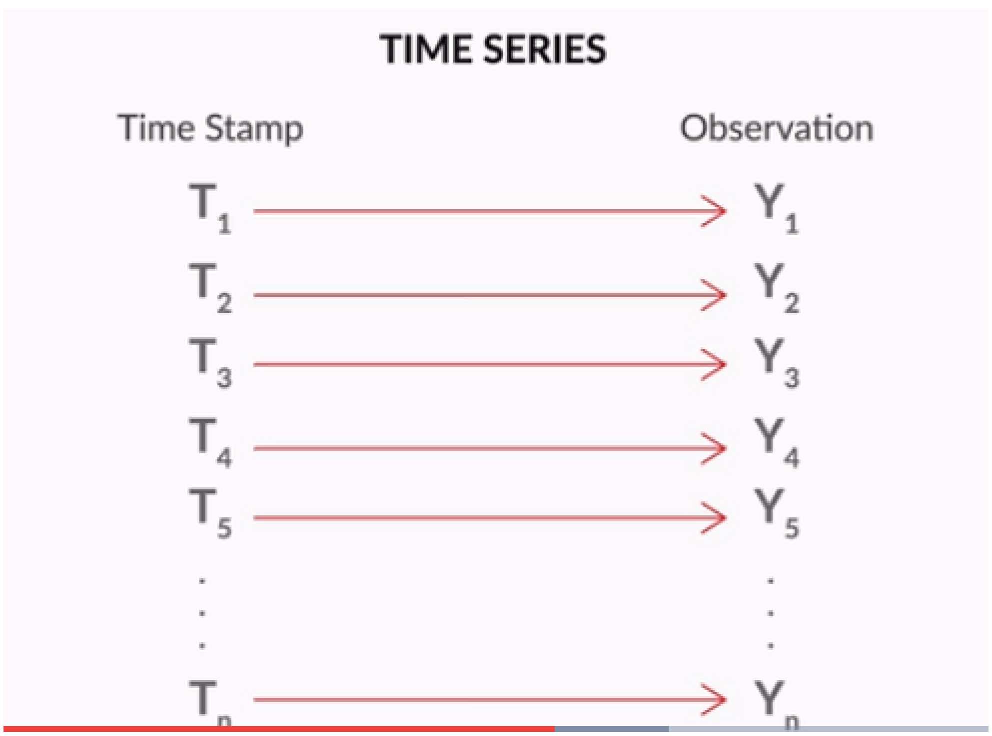
INVESTMENT DECISIONS:

1. Company performance
2. Company leadership
3. Balance sheet
4. Government policies
5. Macro-economic condition
6. RBI policies

TRADING USING TIME SERIES ANALYSIS



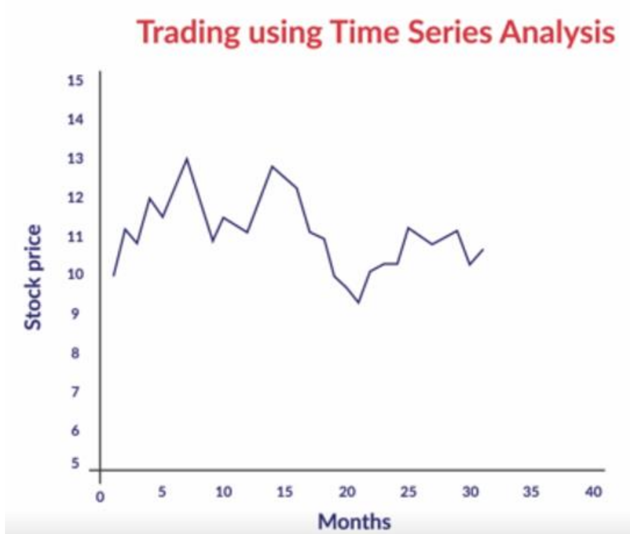




So, a time series is a series of **time stamped** values. In other words, it is a sequence of values with time values attached to it.

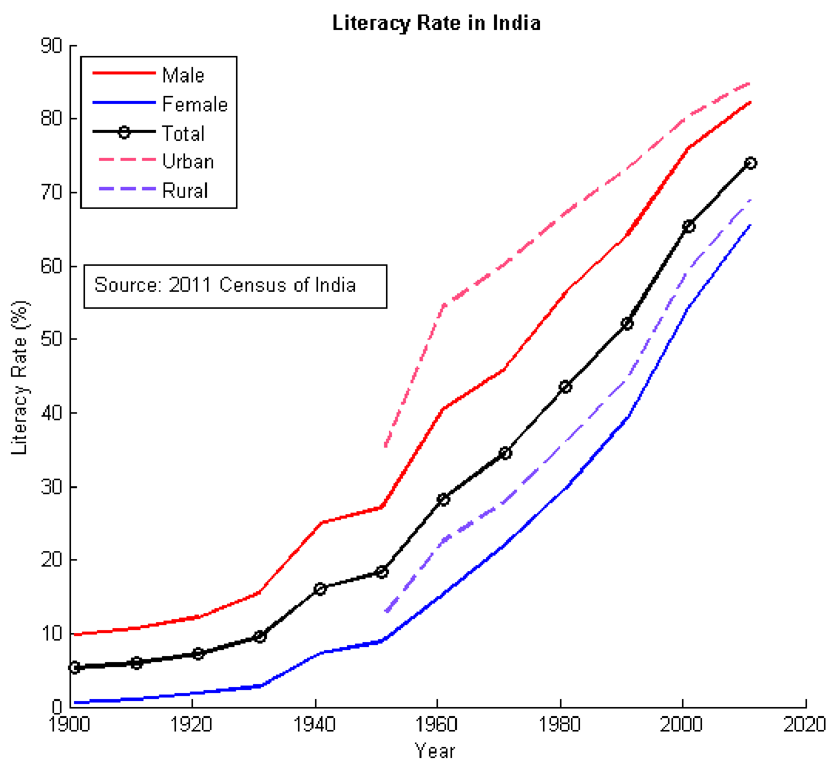
A few examples of such a series are

- Daily stock market figures



Daily Stock Market Figures

- Demographic/Development data (population, birth rate, infant mortality figures, literacy, per-capita income, school enrolment figures) by year



Literacy Rate in India (1900 - 2020)

Using time series analysis, you can **forecast**

- The value of the stock market index for a future month, or
- The value of the literacy rate for a future census

We will look into exactly how this is done, later in the module.

Now, you may think that you can just use regression or advanced regression to make a forecast, taking time as an independent variable. However, this will not work due to various reasons. One reason is that in a time series, the **sequence is important**. For example, let's take the data from before:

Time Stamp	Value
1	2.4
2	3.1
3	5
4	4.5
5	7.2
6	6.8

Original Time Stamped Data

Using regression or advanced regression, let's say you predict the value for timestamp 7. Now let's say you shuffle the data around like this:

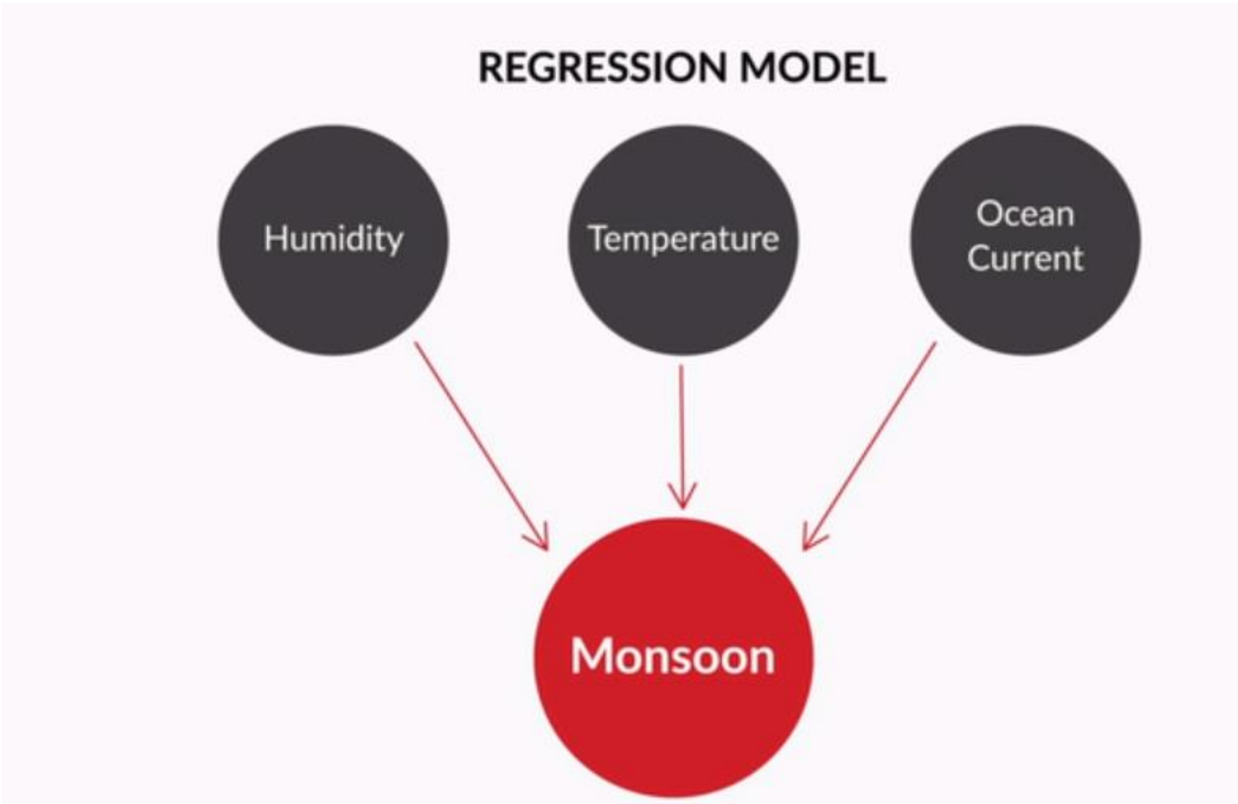
Time Stamp	Value
1	2.4
2	4.5
3	7.2
4	3.1
5	5
6	6.8

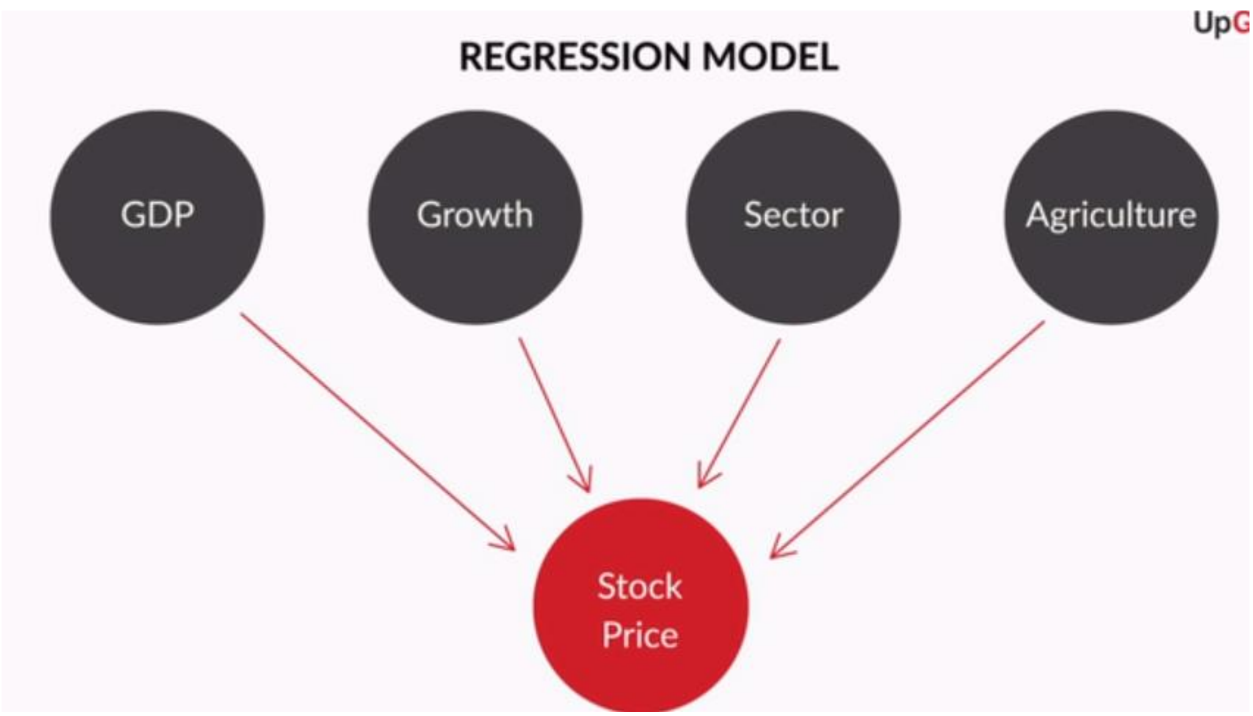
Shuffled Time Stamped Data

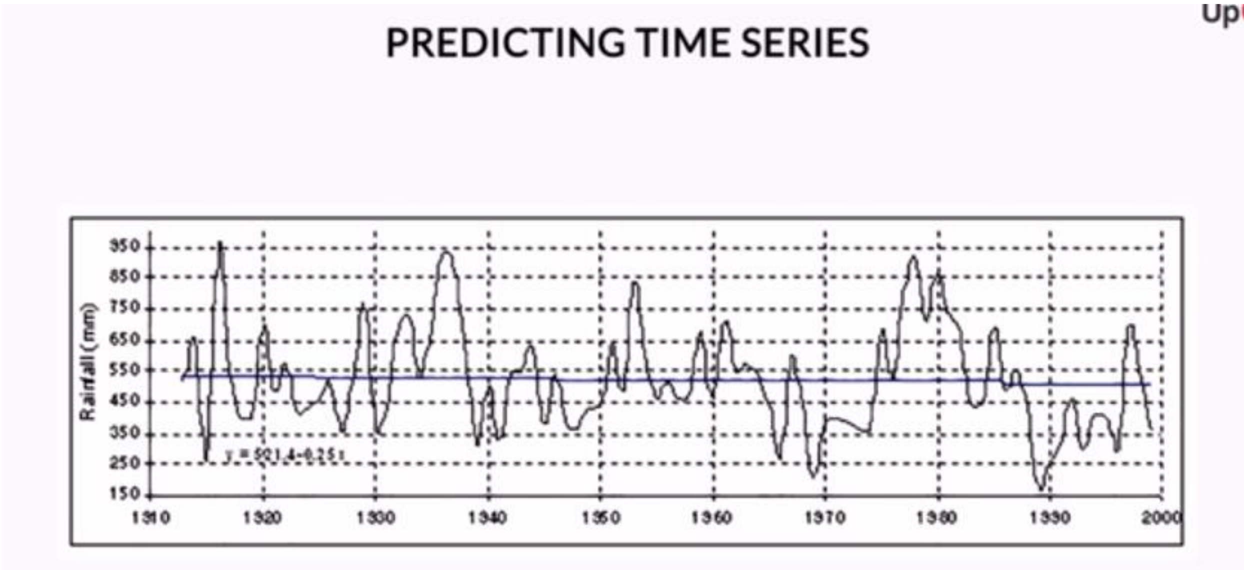
This data will also give you the same prediction for time stamp 7 if you use regression. However, a time series analysis will give you different forecasts for the original data and for the shuffled one.

Why does this happen? This happens because while forecasting using **time series**, your model predicts not only on the basis of the **values given**, but also on the basis of the **sequence** in which the values are given. Hence, the sequence is very important in a time series analysis and should not be played with.

Let's listen to Prof. Raghavan to understand what the second difference between time series analysis and advanced regression is.


REGRESSION MODEL
Humidity
Temperature
Ocean Current
Monsoon


UpG
REGRESSION MODEL
GDP
Growth
Sector
Agriculture
Stock Price



Questions: 1 / 1

Time Series vs Regression

Let's say that you are trying to predict the literacy rate of India in the 2021 census. Which of the following statements about this time series model is correct?

1 - Unlike regression models, the value of literacy at a particular timestamp (say 2001) may be correlated to the value of literacy at a previous timestamp (say 1981).

2 - Unlike regression models, we are not concerned with what the reasons are for a rise in literacy rate. We just want to find out what the literacy rate will be in 2021, based on the literacy rates of the past.

- Neither statement 1 nor 2 is correct
- Statement 1 is correct, but statement 2 is not
- Statement 2 is correct, but statement 1 is not
- **Both statement 1 and 2 are correct**

Feedback : *As you learnt earlier, the time series model will make forecasts for literacy rates in the future, based on the the literacy rates of the past, and the order they were in (that's why you can't randomise the order). Clearly, the past value of literacy affects the future value, and hence, is correlated.*

Also, as was taught recently, in a time series analysis you just try to predict future values, based on the past values and their order, you are not worried about the causal factors for rising literacy.

✓ Correct

So, the two most important **differences** between time series and regression are:

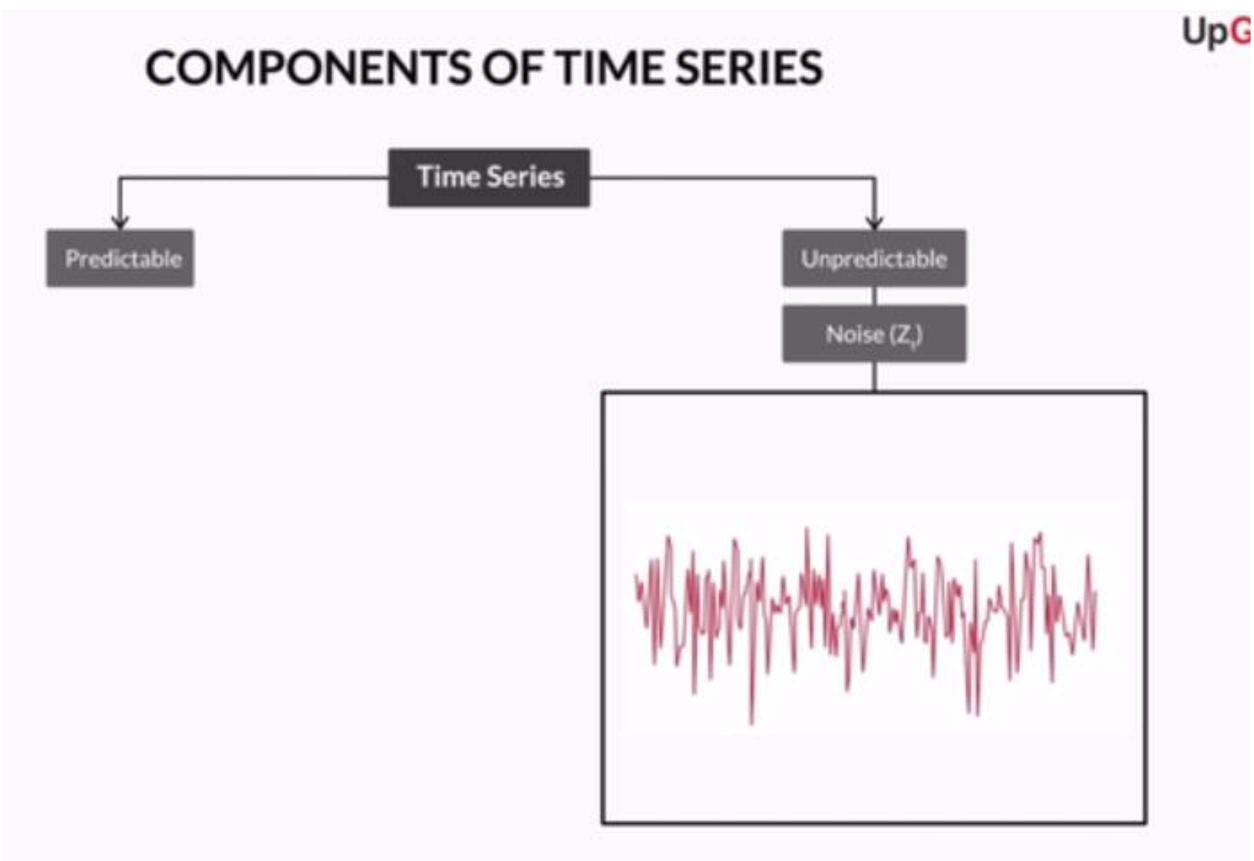
- Time series have a strong **temporal (time-based) dependence** — each of these data sets essentially consists of a series of time stamped observations, i.e., each observation is tied to a specific time instance. Thus, unlike regression, the order of the data is important in a time series.
- In a time series, you are **not concerned with the causal relationship** between the response and explanatory variable. The cause behind the changes in the response variable is very much a black box.

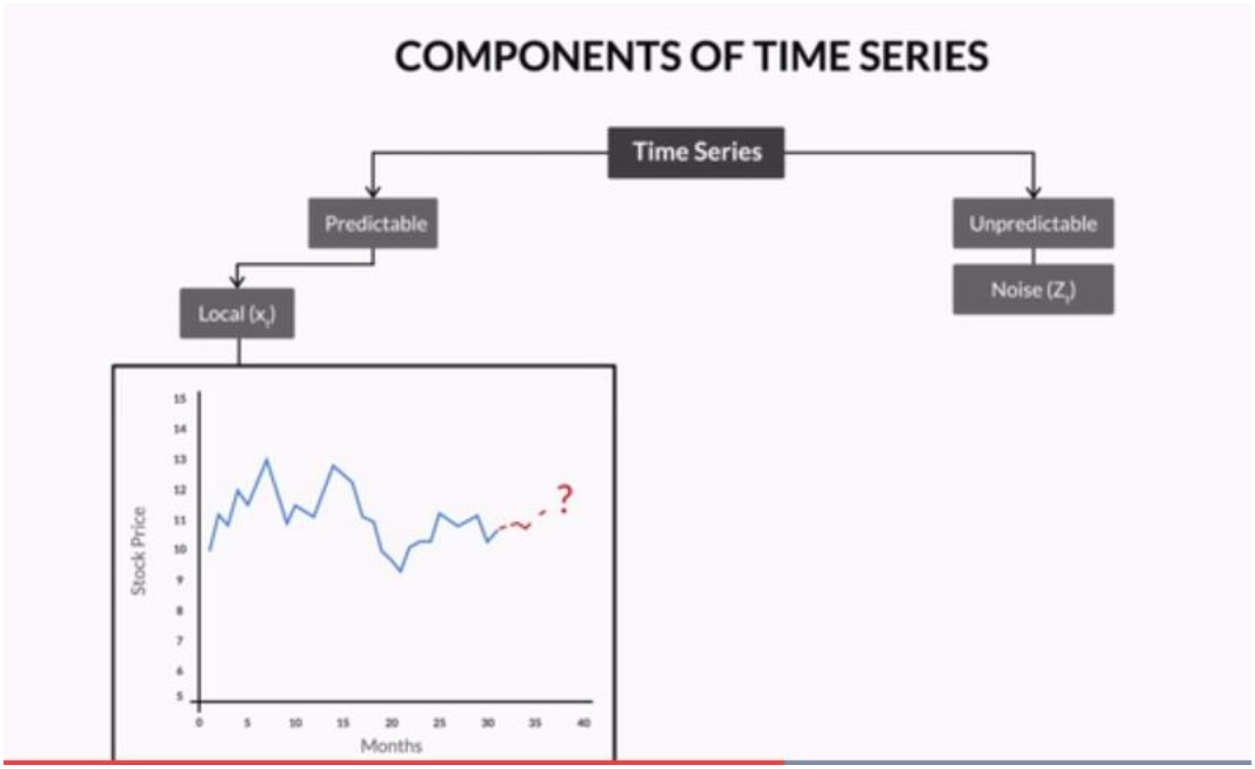
For example, let's say you want to predict what the value of the stock market index will be next month. You will not look at why the stock market index increases in value or if it's because of an increase in GDP or some other factor. You will only look at what the sequence of values was for the past months and predict for the next month, based on that sequence.

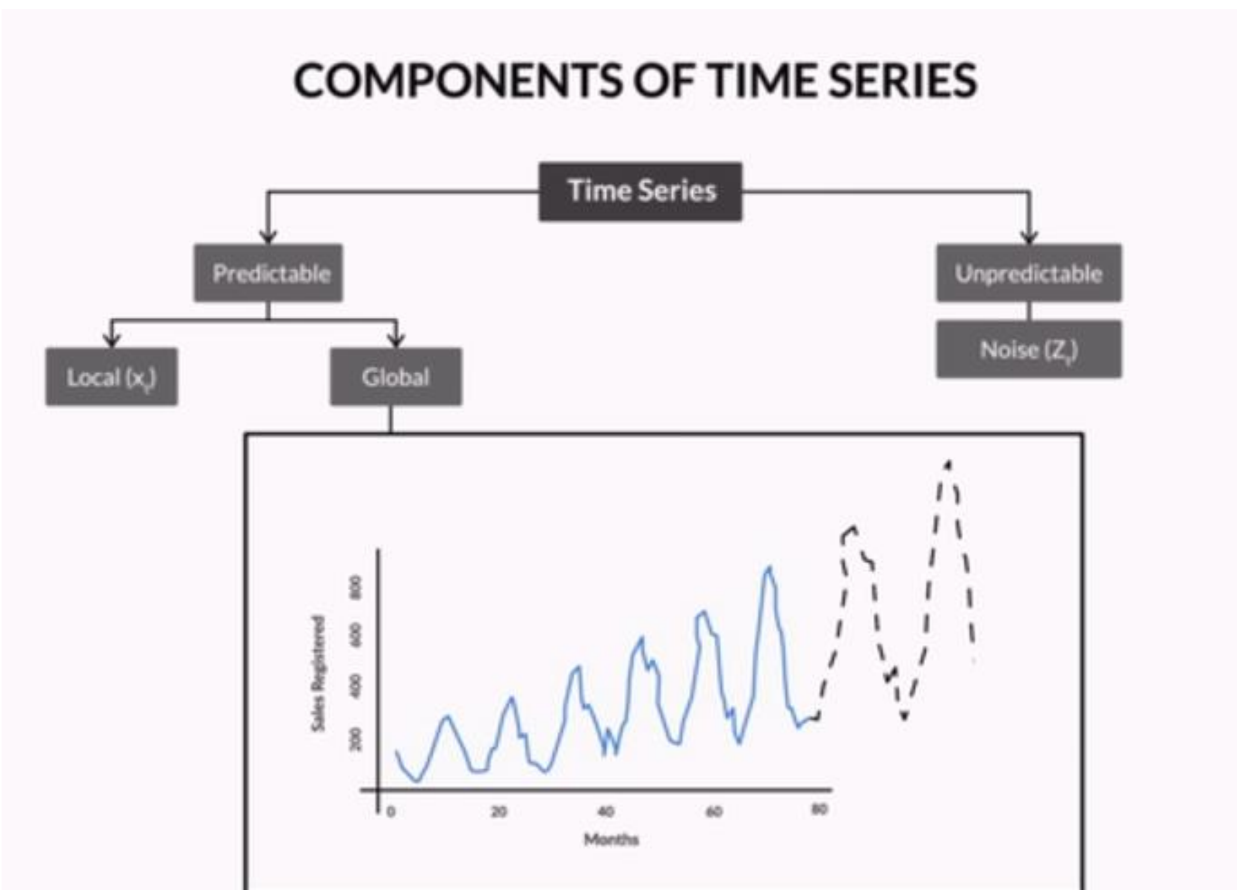
In the next segment, you will learn about some of the other characteristics of a time series; they will eventually help you analyse and forecast the values in any time series.

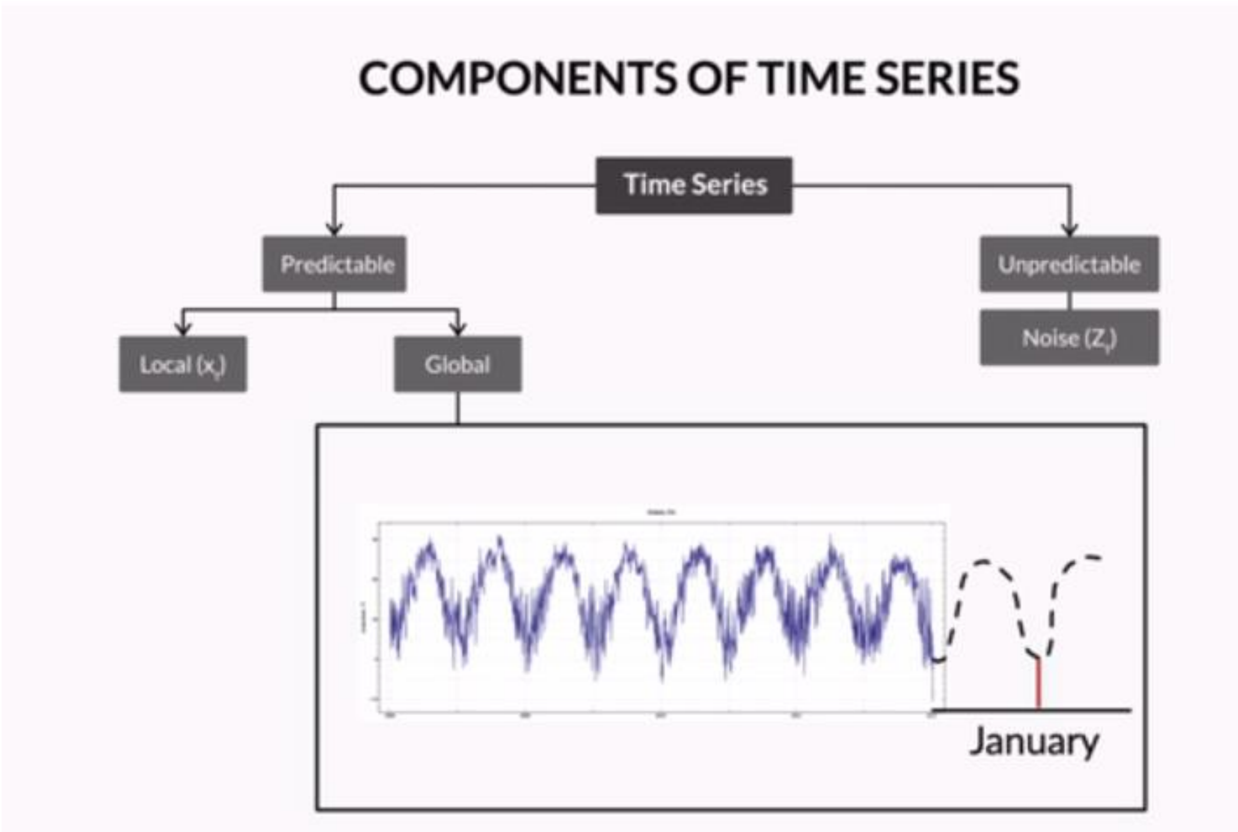
Components of Time Series - I

Before you make time series forecasts, it is important to first identify the different types of time series and their different components. This is because each type or component of a time series needs to be dealt with and analysed differently. Let's learn about this in detail.




COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Noise (Z_t)
Local (x_t)
Stock Price
Months
?
15 14 13 12 11 10 9 8 7 6 5
0 5 10 15 20 25 30 35 40


COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Local (x_t)
Global
Noise (Z_t)
Sales Registered
Months
200 400 600 800
0 20 40 60 80


COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Local (x_t)
Global
Noise (Z_t)
January

Questions: 1 / 1

Components of Time Series - I

Out the following events, which one is an example of a local pattern variation -

A) Due to a recent terrorist attack in Britain, the number of tourists visiting the city of Manchester has decreased, suddenly. Now, the number is returning back to normal, steadily.

B) Due to countries like India focusing on their respective tourism industries, the number of tourists visiting the British city of Manchester is on the decline.

Only A

Feedback : *Local pattern variations are those that show up in the short-term and balance out in the long-term, unlike global variations. Statement A looks like an example of a short-term change that will even out after a few days; then more people will start visiting the city again.*

Correct

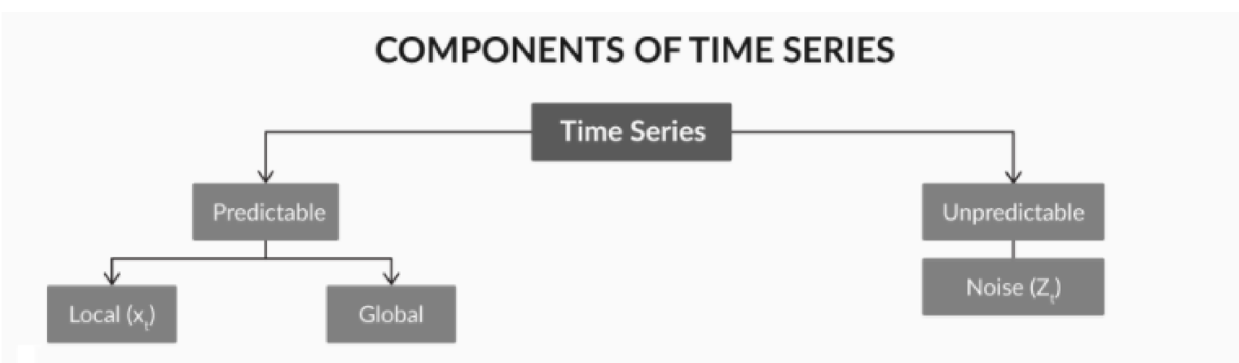
Only B

Both

Your answer is **Correct.**

Continue

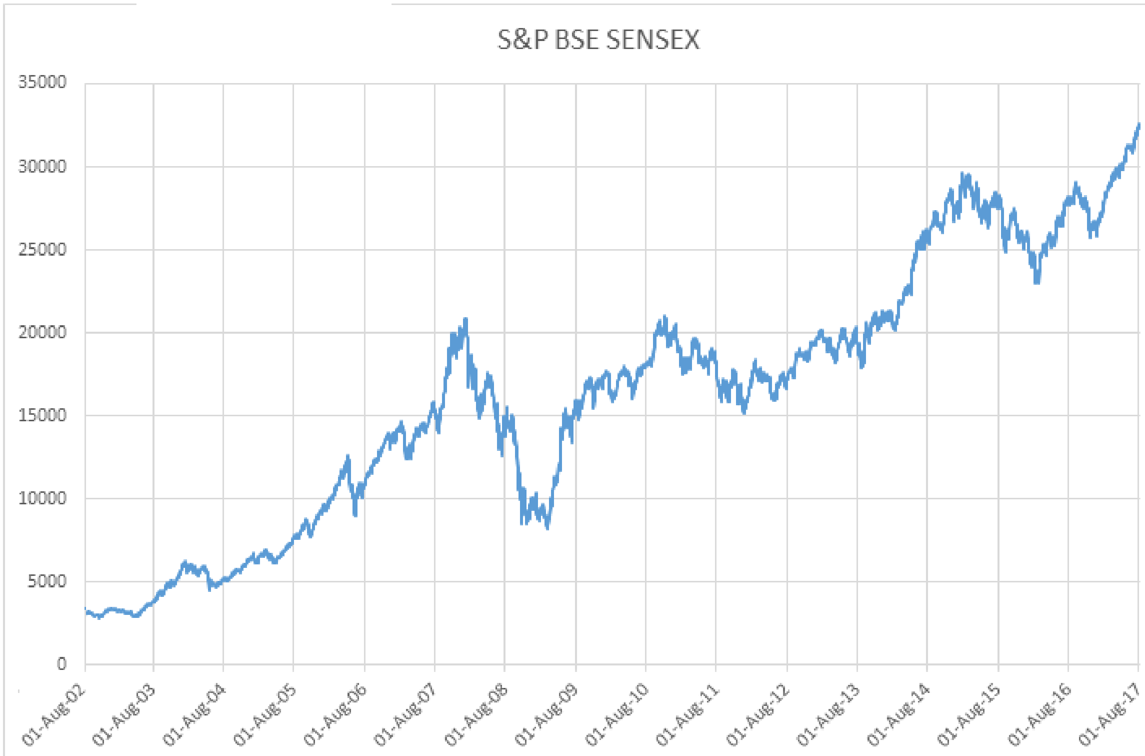
So far, you have learnt about the following components of a time series:



Components of a Time Series

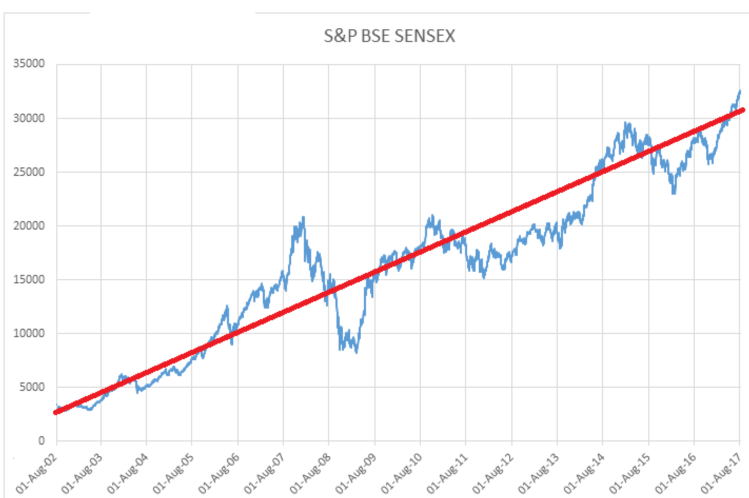
So you have the **globally predictable** part and the **locally predictable** part. You also have the **unpredictable part**, which is something you cannot predict.

To understand these better, let's look at the **BSE data** for the last 15 years.



BSE Sensex Data (Last 15 years)

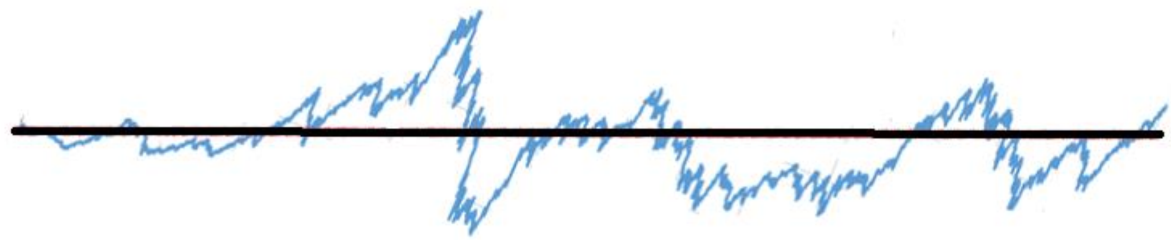
So in general, you can see that the value of the index has been increasing over the years. Hence, if you fit a **regression model** on to it, like you saw earlier, you will get a line like this:



Regression Fit on BSE Sensex Data

The red line here, the one that shows the regression fit, gives the **overall pattern** of the data. So overall, you expect the index to go up every year. Hence, this red line is indicative of the overall pattern of the data, or in other words, the global pattern.

But then, you see that there are a lot of **wiggles** in the data, on top of the simple regression fit. Let’s plot these wiggles, the ones that would be left after removing the regression fit.



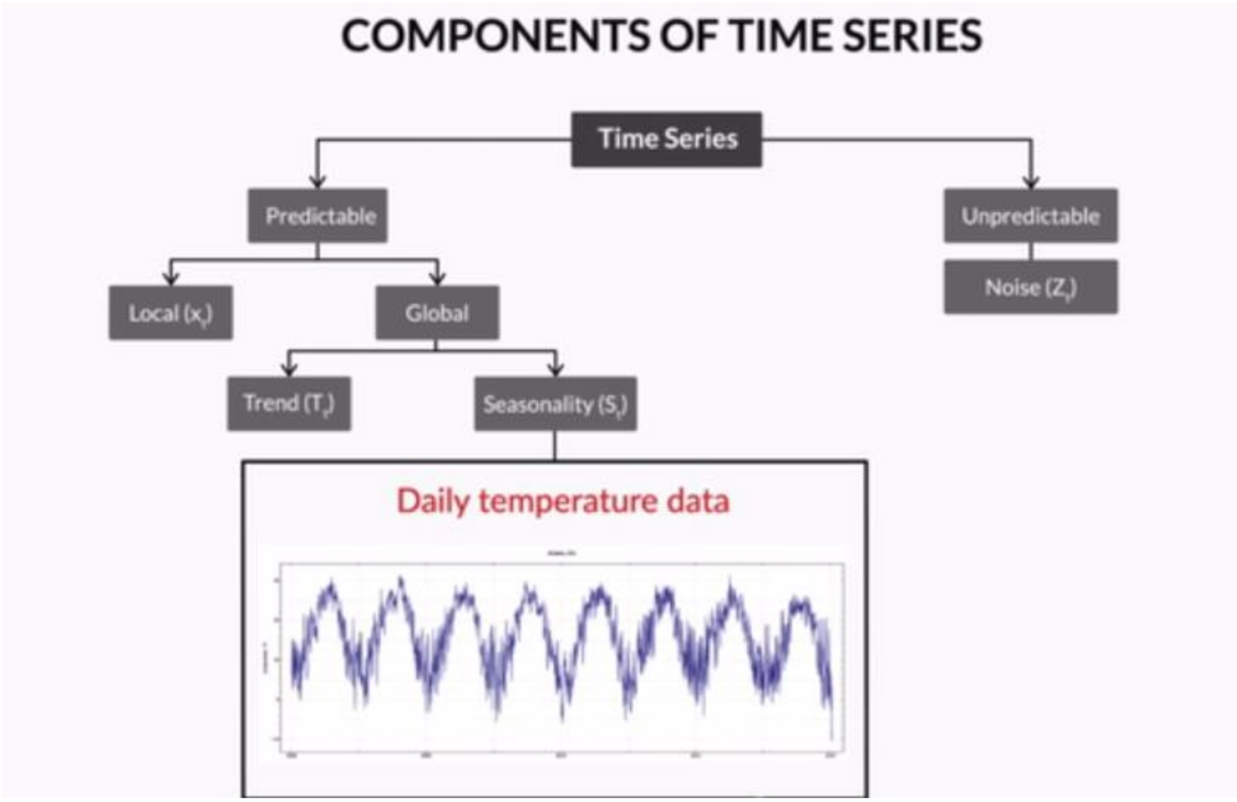
Wiggles in BSE Sensex Data

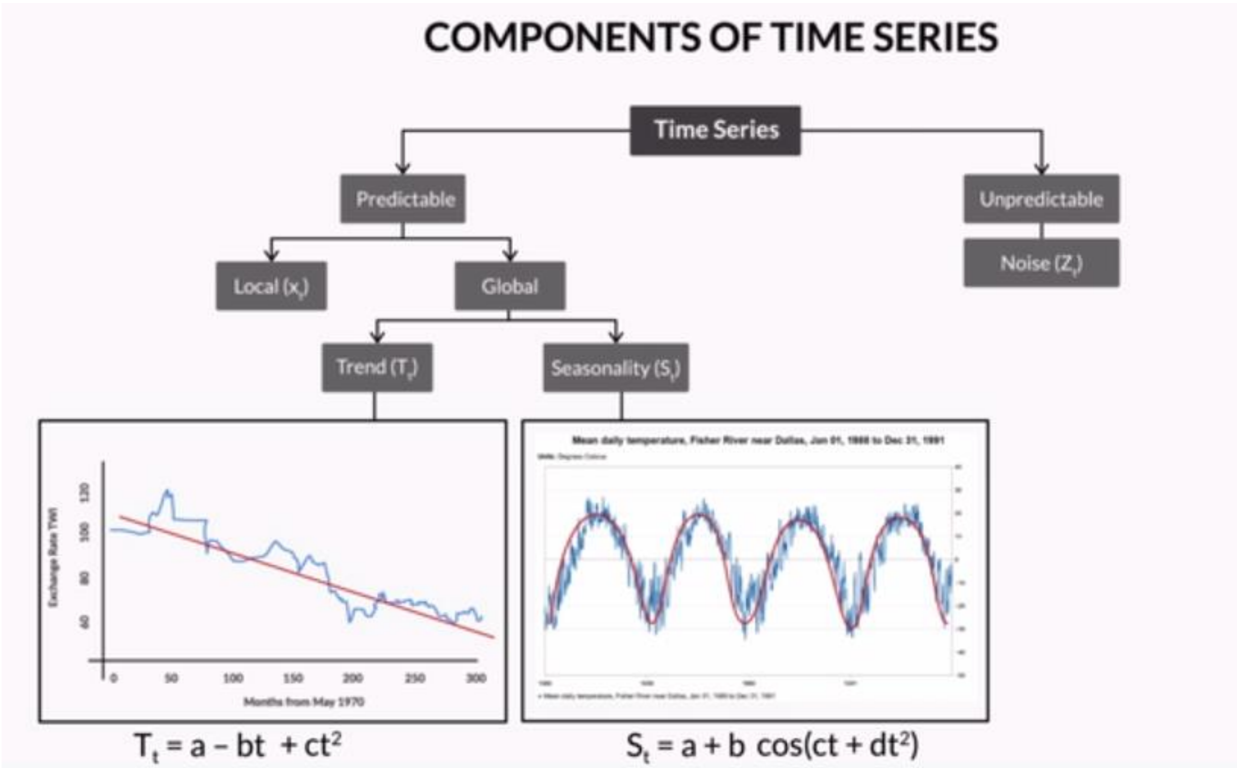
In a time series, you model this part of the data as well. This is the **locally predictable** part of the time series, or in other words, the local pattern in the time series. It is called local because the values here don’t show any long-term pattern. The values don’t increase regularly or repeat with a set frequency; so there is **no overall global pattern**. However, locally, there is some predictability: the value of the wiggle today depends somewhat on what the value of the wiggle was yesterday.

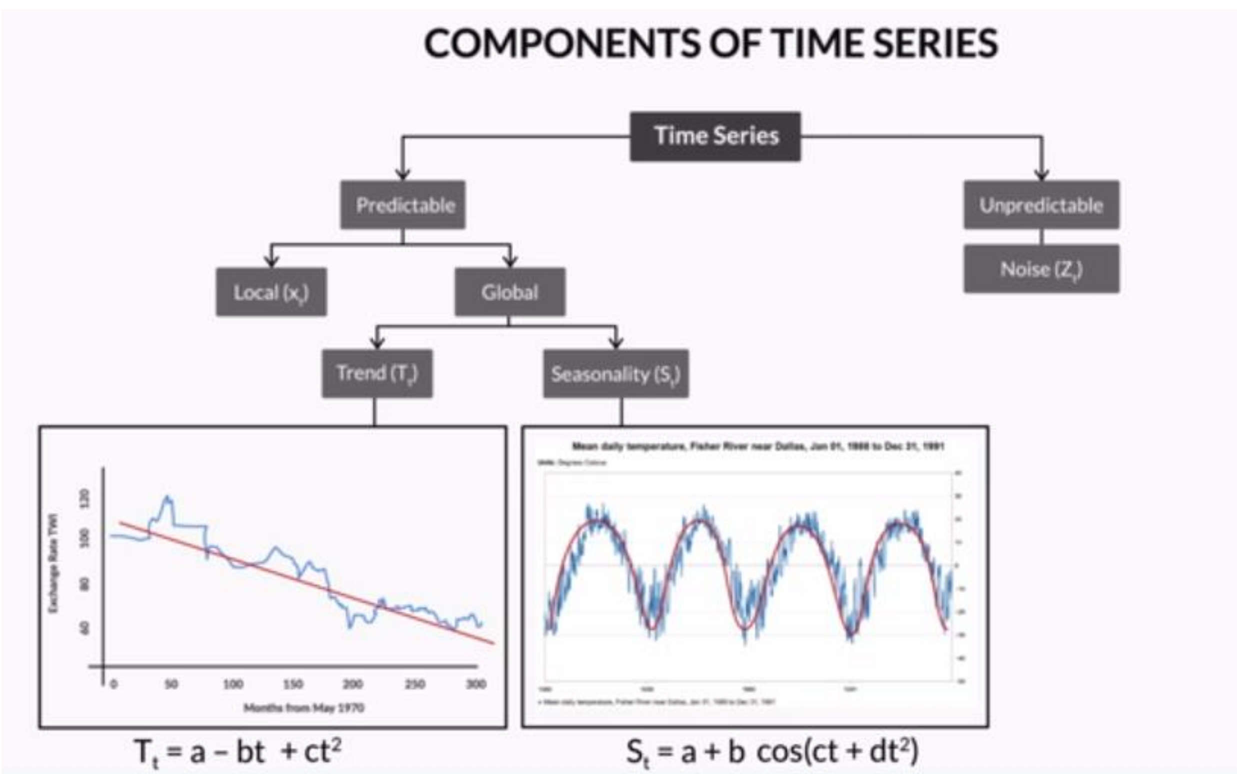
Look at it this way: because of the **overall global pattern**, you expect the stock market index to go up every year as India grows economically. So the underlying cause of the global pattern is **India’s long-term economic growth**.

However, other than the general overall growth, you do experience some **up-and-down variations** if you zoom in on a day-by-day or a week-by-week level. These variations happen mostly because of what the index value was the previous day. If the market starts to go up, you generally expect it to go up the next day too. If it has been going up for a long time, you expect it to decrease soon. The underlying causes of these local wiggles are the **market sentiment**, opinions of stock market experts, and various other factors that pop up in a sporadic fashion.

In short, there are three main components of a time series: **global, local and noise**. In the upcoming lecture, we will break down the global component further into two subcomponents: **trend** and **seasonality**.


COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Noise (Z_t)
Local (x_t)
Global
Trend (T_t)
Seasonality (S_t)
Daily temperature data


COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Noise (Z_t)
Local (x_t)
Global
Trend (T_t)
Seasonality (S_t)
Exchange Rate TWI
60 80 100 120
0 50 100 150 200 250 300
Months from May 1970
Mean daily temperature, Fisher River near Dallas, Jan 01, 1988 to Dec 31, 1991
$T_t = a - bt + ct^2$
$S_t = a + b\cos(ct + dt^2)$


COMPONENTS OF TIME SERIES
Time Series
Predictable
Unpredictable
Noise (Z_t)
Local (x_t)
Global
Trend (T_t)
Seasonality (S_t)
Exchange Rate TWI
60 80 100 120
0 50 100 150 200 250 300
Months from May 1970
Mean daily temperature, Fisher River near Dallas, Jan 01, 1988 to Dec 31, 1991
$T_t = a - bt + ct^2$
$S_t = a + b\cos(ct + dt^2)$

Questions: 1 / 1

Components of Time Series - I

A restaurant has been experiencing higher sales during the weekends, as compared to the weekdays. Select the correct component of the time series that explains the daily restaurant sales patterns for this restaurant.

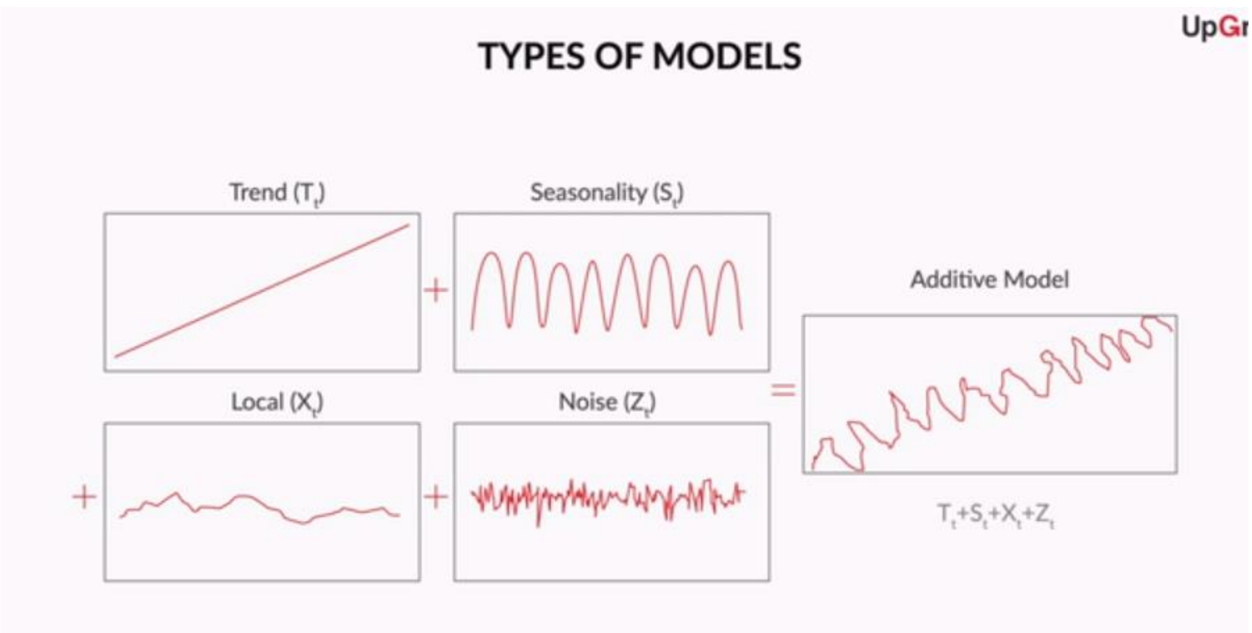
- Trend
- **Seasonality** ✓ Correct

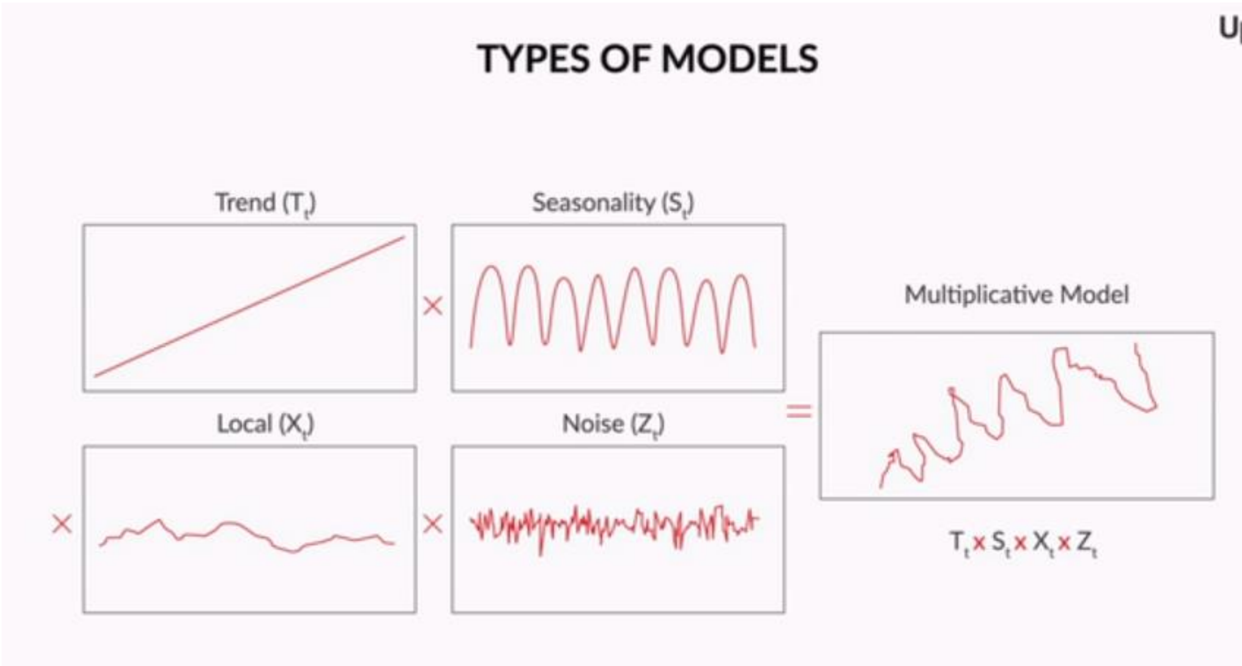
  Feedback : *When there are patterns in a data set that repeat over known, fixed periods of a time set, such patterns are known as seasonality. A season may refer to a time period as denoted by the calendar seasons, such as summer or winter, and commercial seasons, such as the holiday season. Companies that understand the seasonality of their business can time their inventories, staffing and other decisions to coincide with the expected seasonality of the associated activities.*
- Non predictable part - Noise (Zt)

Your answer is **Correct.**

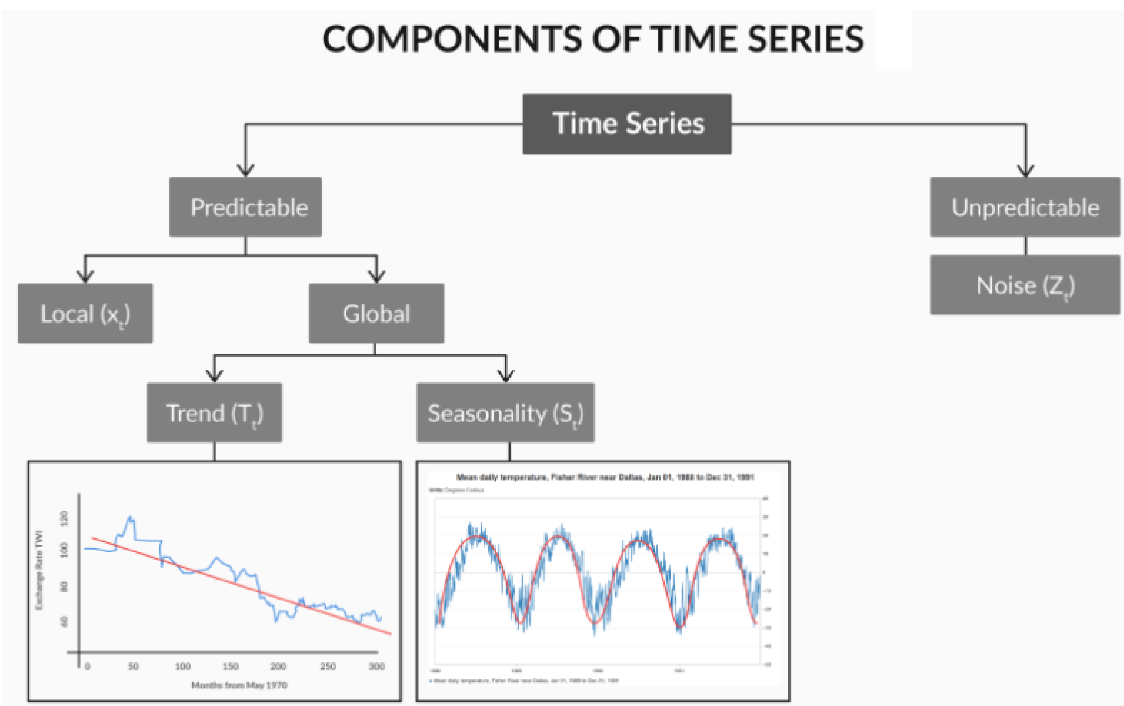
Continue

2:50 / 5:07 Medium 1x





So the globally predictable component of a time series can be of two types — **trend** and **seasonality**— as shown in this image:



Components of a Time Series

Here, the **trend** refers to any pattern that talks about the **overall increase or decrease** in the values, whereas **seasonality** refers to a **repeating pattern** of values seen in the data.

Trend and seasonality can also both appear in a time series. For example, let’s look at the **sales data**example from the lecture:

Time Series With Trend and Seasonality

Now, clearly this data has both types of patterns; it has a **trend**, as the overall pattern suggests an **increase in sales**. However, the sales are also **seasonal**, as a similar up-and-down pattern is **repeating** itself **every 12 months**.

A similar example would be the sales of an ecommerce start-up such as Flipkart. You would expect Flipkart’s sales to increase every year, but you would also expect a seasonal variation on top of that, i.e. sales will be higher in the months around Diwali and lower in the off-seasons.

However, a time series can also have only a trend or only a seasonality. For example, the stock market data you saw in the previous section has only a trend in it and no seasonality.

You also saw the different ways in which these components of the time series can be related to each other. Their combination can either be **additive** or **multiplicative**.

$$\text{Additive Model} - TS = T_t + S_t + X_t + Z_t$$

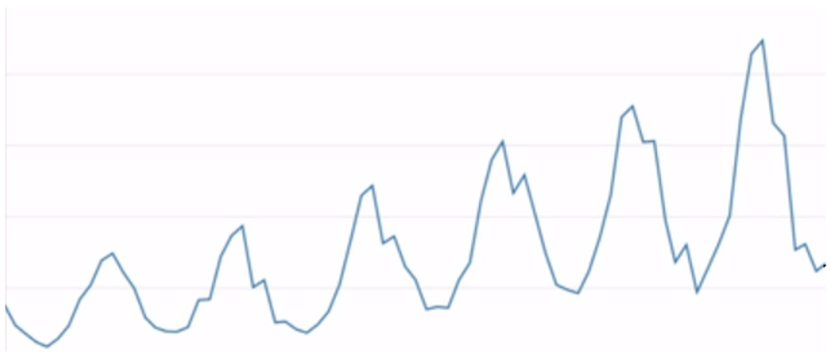
$$\text{Multiplicative Model} - TS = T_t * S_t * X_t * Z_t$$

Additive and Multiplicative Models

Questions: 1 / 1

Components of Time Series - I

Consider the graph given below, which represents a time series:



Which of the following models is the time series given above most likely to follow?

62%	Additive
38%	Multiplicative

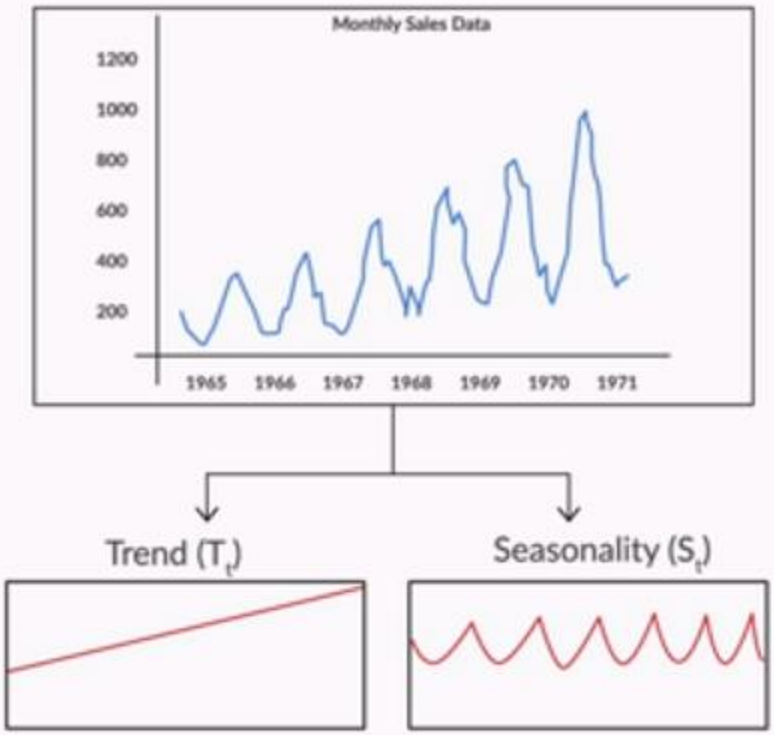
Only 159 people have taken this poll from your cohort. The results would be updated as we get more responses.

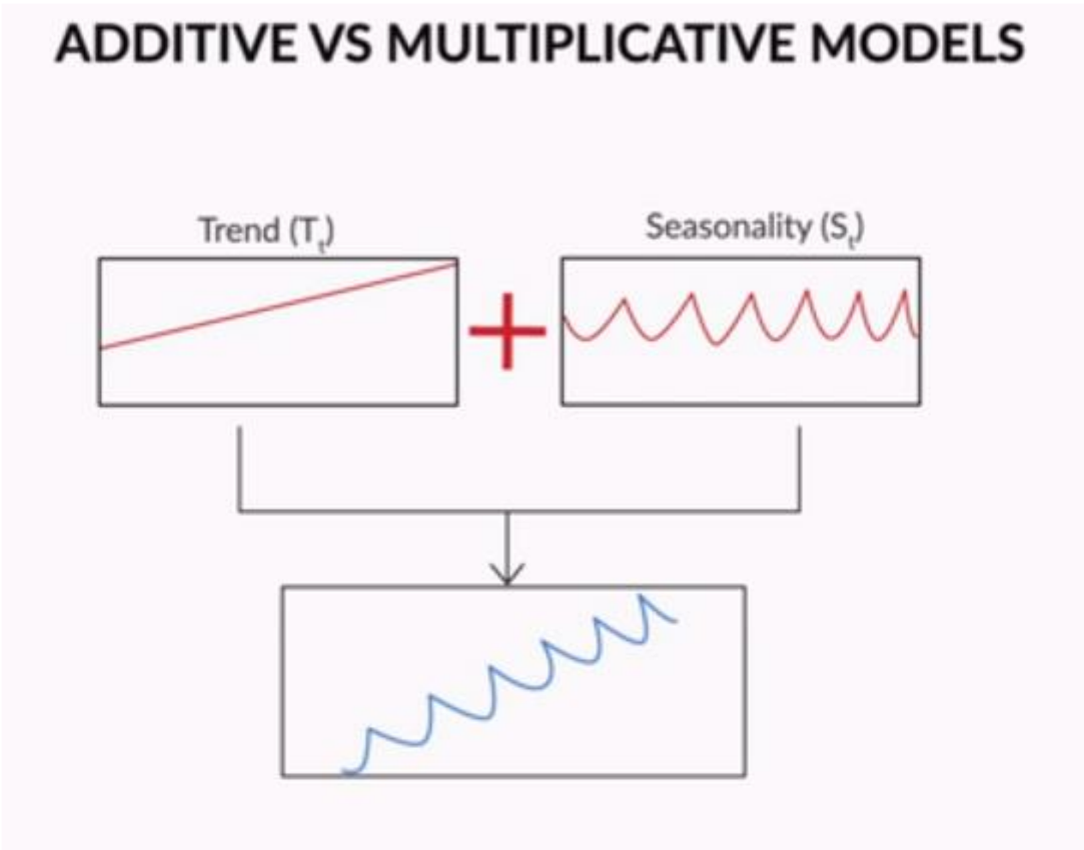
Components of Time Series - II

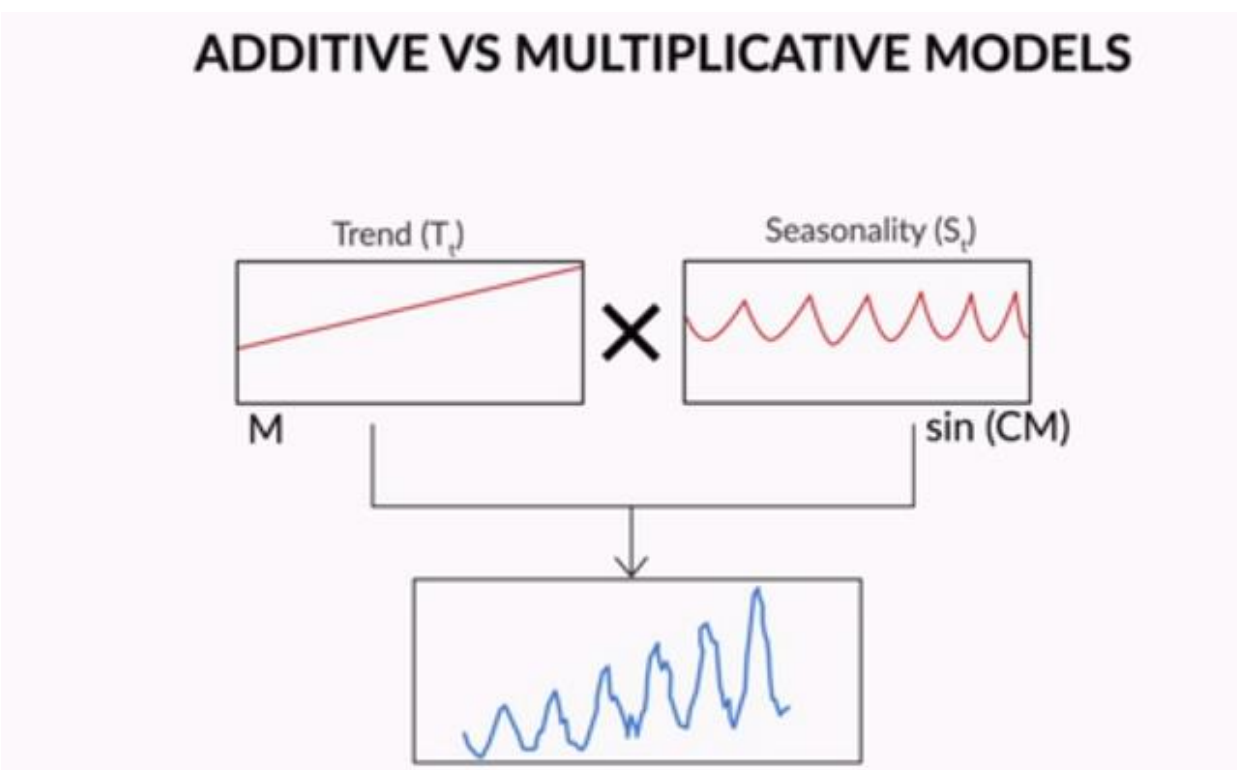
In the last session, you saw the different ways in which the components of a time series can be related to each other. These components may be combined additively or multiplicatively.

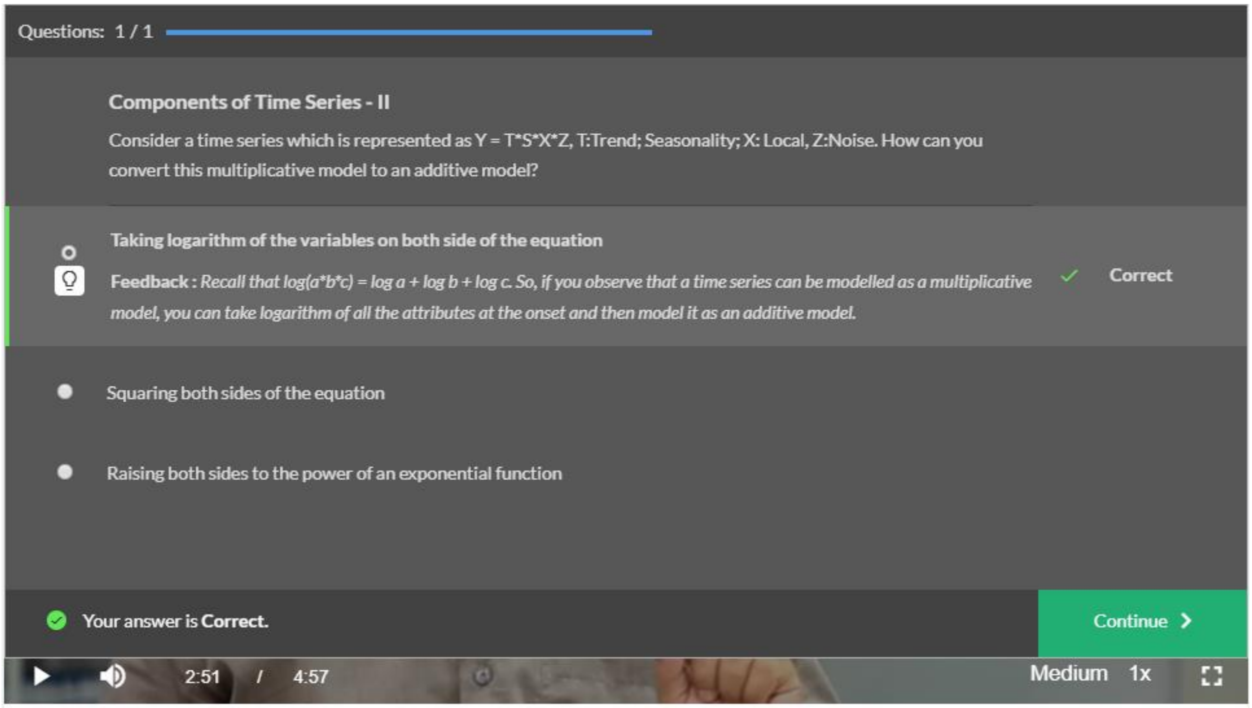
Let’s look at the difference between the two methods (**additive** and **multiplicative**), and how they are treated and identified.

ADDITIVE VS MULTIPLICATIVE MODELS




ADDITIVE VS MULTIPLICATIVE MODELS
Trend (T_t)
Seasonality (S_t)


ADDITIVE VS MULTIPLICATIVE MODELS
Trend (T_t)
Seasonality (S_t)
M
sin (CM)



As a broad rule of thumb,

- When the **magnitude** of the seasonal pattern in the data **increases** with an increase in data values, and decreases with a decrease in the data values, the **multiplicative model** may be a better choice.
- When the **magnitude** of the seasonal pattern in the data **does not directly correlate** with the value of the series, the **additive model** may be a better choice.

You typically work with additive models. An easy way to **transform** a multiplicative model to an additive model is to carry out the analysis on the **log** of the values in the time series.

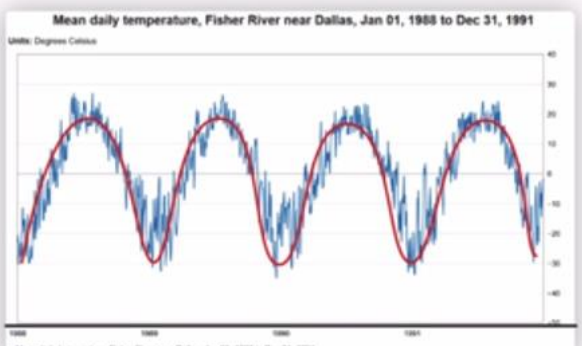
$$\text{Multiplicative Model} - TS = T_t * S_t * X_t * Z_t$$

$$\text{Log of Multiplicative Model} - \log(TS) = \log(T_t) + \log(S_t) + \log(X_t) + \log(Z_t)$$

Multiplicative to additive transformation

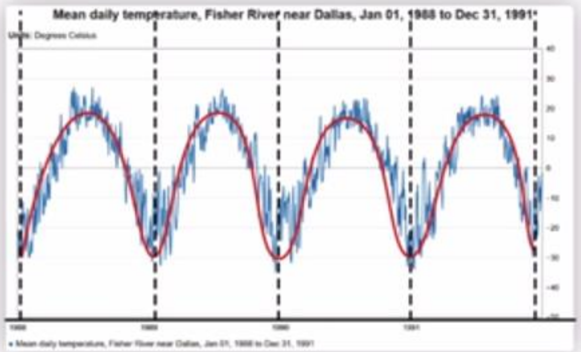
EXAMPLE OF CYCLICITY

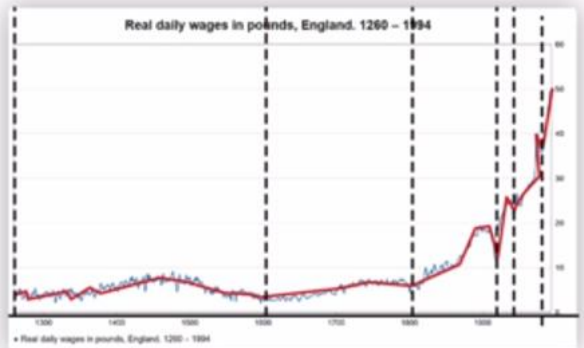
CYCLICITY VS SEASONALITY



UpGrad

CYCLICITY VS SEASONALITY





Questions: 1 / 1

Components of Time Series - II

Below are given two sets of information (A - C, I - IV) - examples and the components of time series. Select the correct option which maps the examples with the appropriate time series component.

A) Death rate decreased due to advance in science

B) A fire in a factory delaying production for some weeks.

C) The sale of air condition increases during summer

I) Trend

II) Seasonality

III) Cyclicity

IV) Noise

- A = I, B = IV, C = III

- **A = I, B = IV, C = II** ✓ Correct

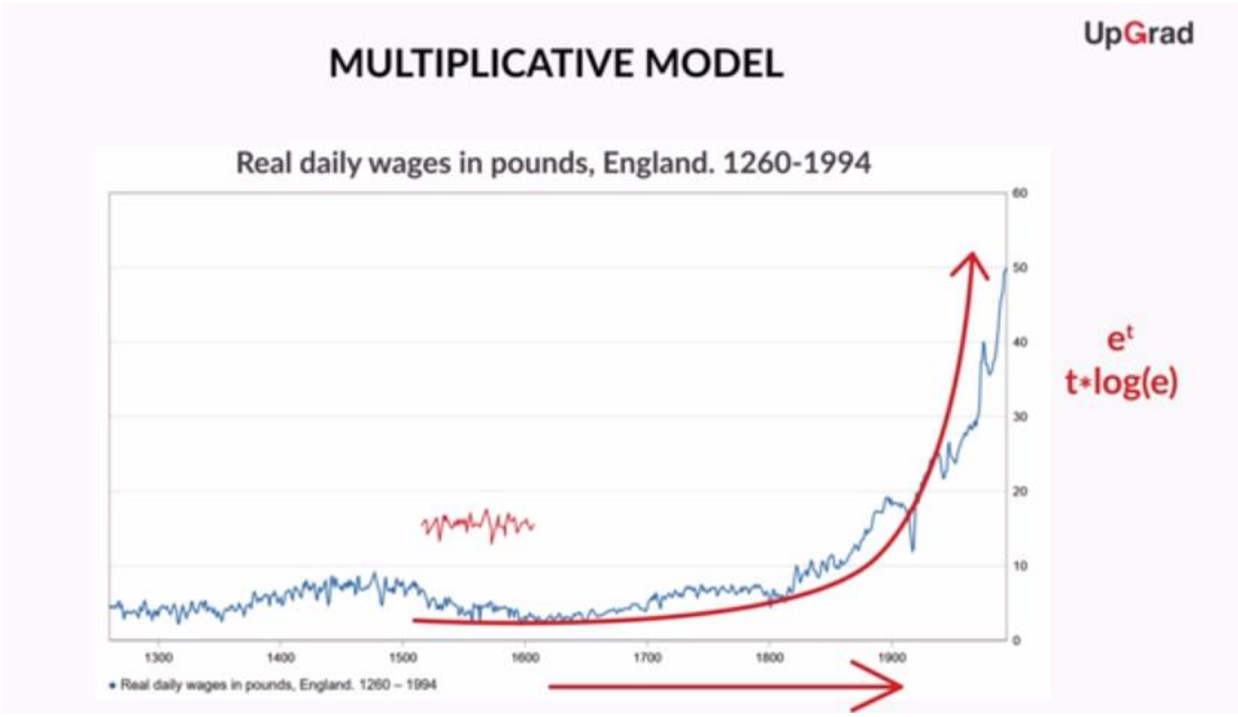
  Feedback : *Death rate decreased due to advance in science: It is a pattern which has taken place over time gradually and hence a trend. A fire in a factory delaying production for some weeks: It is an unpredicted event causing some unpredictability in the data and hence it is noise. The sale of air condition increases during summer: It is a periodic event occurring every summer.*

- A = IV, B = I, C = III

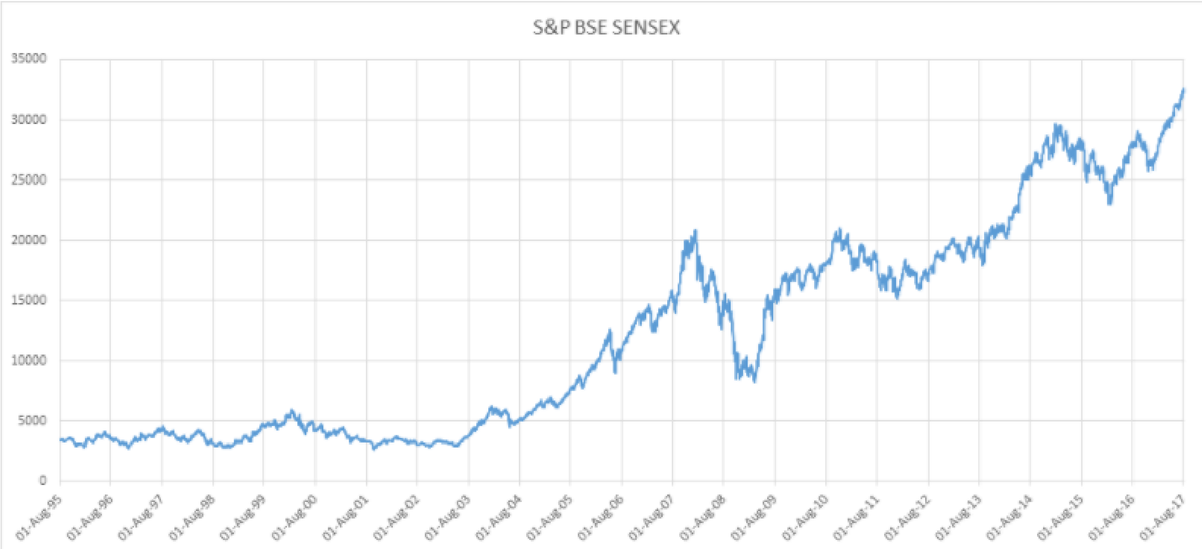
- A = IV, B = I, C = II

Your answer is **Correct.**

Continue ›



Let’s revisit the stock market data example from an earlier session. However, this time, let’s look at the **BSE data** for the last 22 years, instead of the last 15 years.



BSE Sensex Data (Last 22 Years)

Clearly, you can see that the data not only has a **trend**, but it also has some **cyclicity**. Clearly, the BSE index goes through some up-and-down **cycles**; it can possibly be related to economic depressions, booms, etc. However, the **period** for these up-and-down cycles is **not fixed**, and this is where cyclicity is different from seasonality. In general, you will see such cyclicity in most macroeconomic contexts, provided you have the data for a sufficient number of years.

We will not go into the modelling of cyclicity in this course, as it is too complex and will not be needed in many time series analyses.

Additional Links

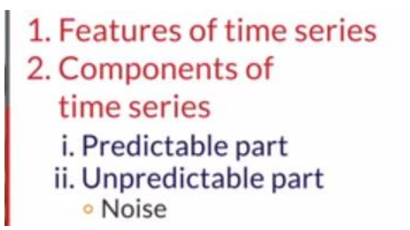
You can view the four data sets referred to in the session from the following links. You can play around by changing the time period and chart types in the plots.

- Monthly sales of company X, Jan ’65 – May ’71, C. Cahtfield
- Mean daily temperature, Fisher River near Dallas, Jan 01, 1988 to Dec 31, 1991
- Exchange Rate TWI, May 1970 – Aug 1995
- Real daily wages in pounds, England. 1260 – 1994

Summary

This was a brief introductory session on time series. Let’s revisit some of the concepts that we covered in this session:

- How does time series analysis **differ from regression analysis**?
- What are the **basic characteristics** of a time series?
- What are the **different components** of a time series?



1. Time series concept, Features of time series and Time series Vs Regression
2. Components of Time Series: Predictable and Unpredictable
    2.1. Predictable
        2.1.1. Local – Auto-regression
        2.2.2. Global - Trend, Cyclic and Seasonality
    2.2 Unpredictable - Noise
3. Types of models: Additive Vs Multiplicative

Separating Time Series Components in R

The solution file for the comprehension question on the previous page can be accessed using the link given below:

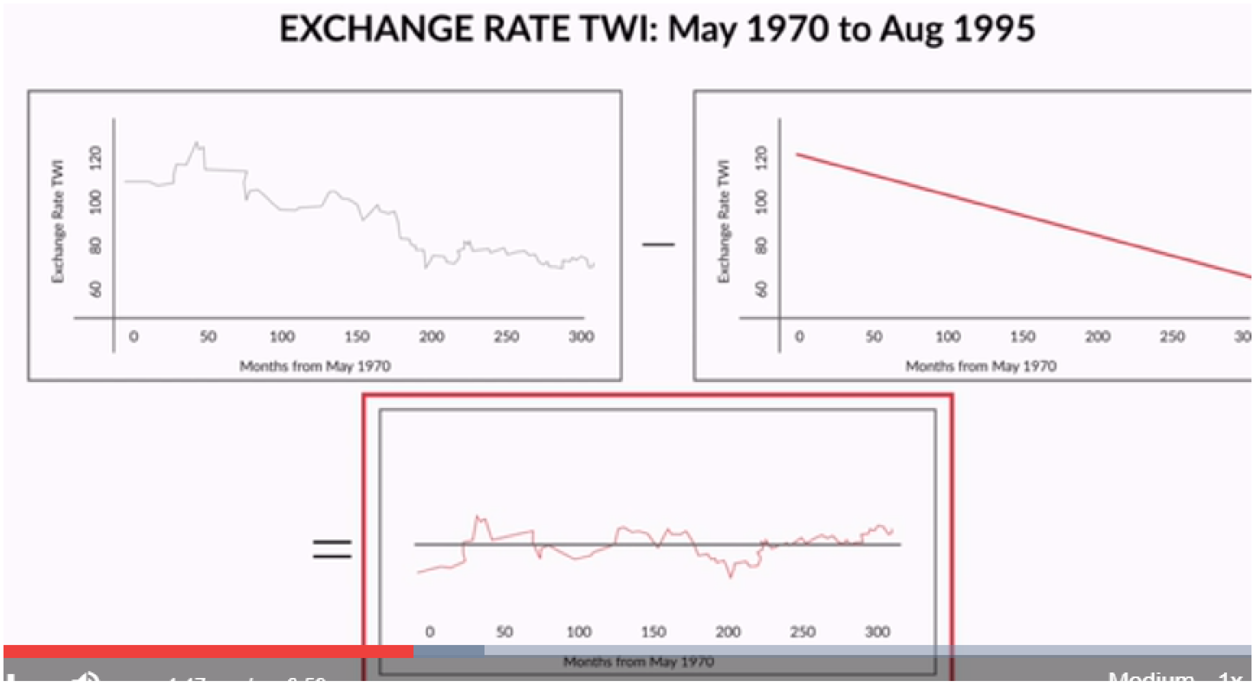
Separating Time Series Components (Comprehension)

file_download**Download**

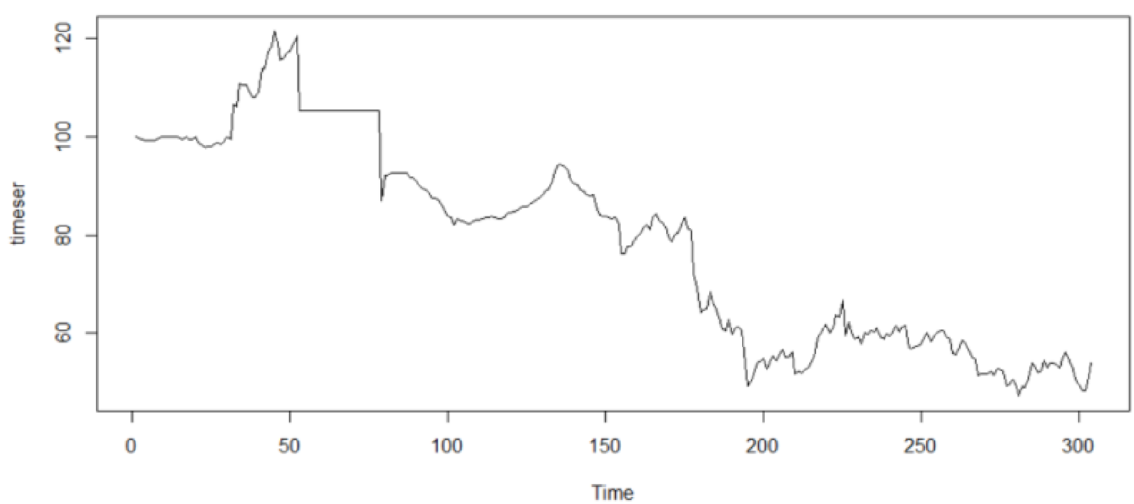
Now, let’s look at how Prof. Raghavan took the same data set and separated its various components.

If you want to follow along with what the professor is teaching, you can refer to the following R code file:

Separating Time Series Components (Prof. Raghavan)



So when you started off with this time series, the exchange rate TWI was from 1970 to 1985.



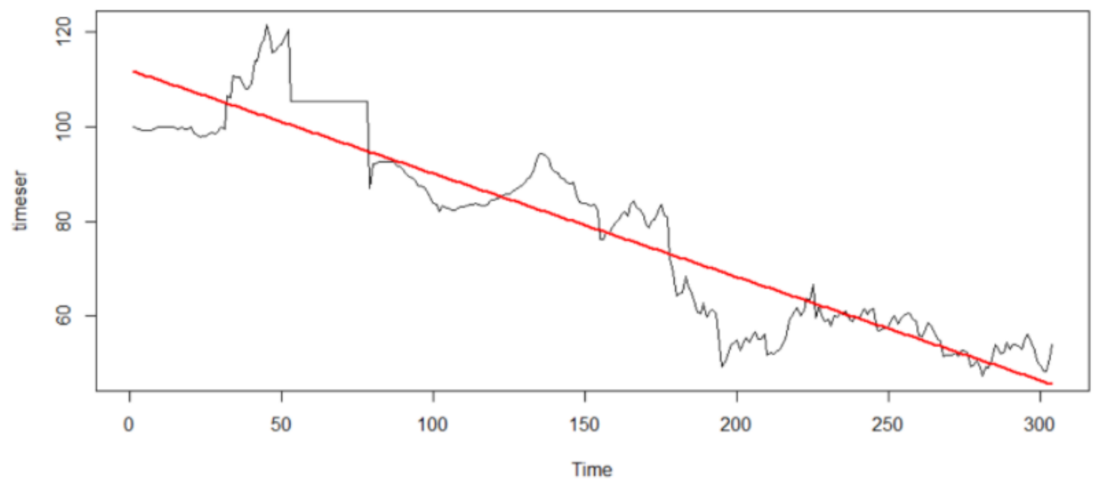
Exchange Rate TWI (1970-1985)

Now, after **fitting a regression line** on it, you found that the **global trend** in this series can be represented by the equation $Tt=111.82-0.22tTt=111.82-0.22t$. This means that, on average, the exchange rate decreases by 0.22 every month.

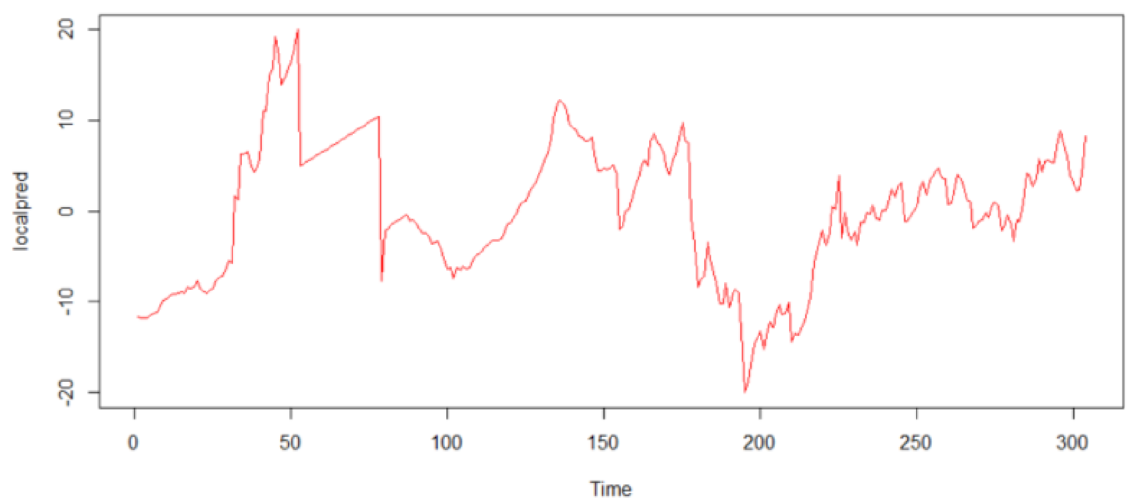
By the way, you only modelled the trend in this series, because as discussed earlier, this time series only has a trend. It does not have any seasonal component. If a seasonal component was present, you would have modelled that too, by fitting a sin or a cos function.

Now, after **fitting a regression line** on it, you found that the **global trend** in this series can be represented by the equation Tt=111.82−0.22tTt=111.82−0.22t. This means that, on average, the exchange rate decreases by 0.22 every month.

By the way, you only modelled the trend in this series, because as discussed earlier, this time series only has a trend. It does not have any seasonal component. If a seasonal component was present, you would have modelled that too, by fitting a sin or a cos function.



Regression Fit on Exchange Rate Data

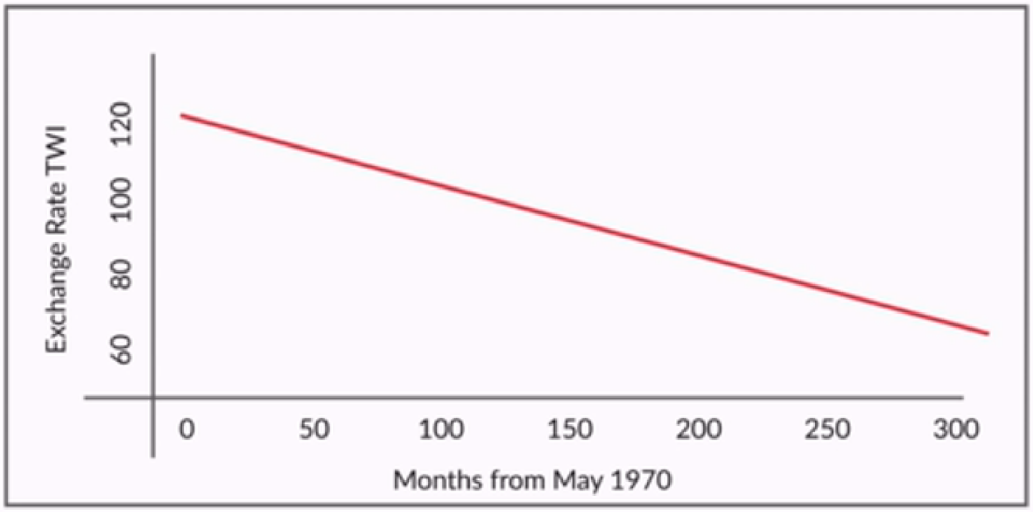


Local Variations in Exchange Rate

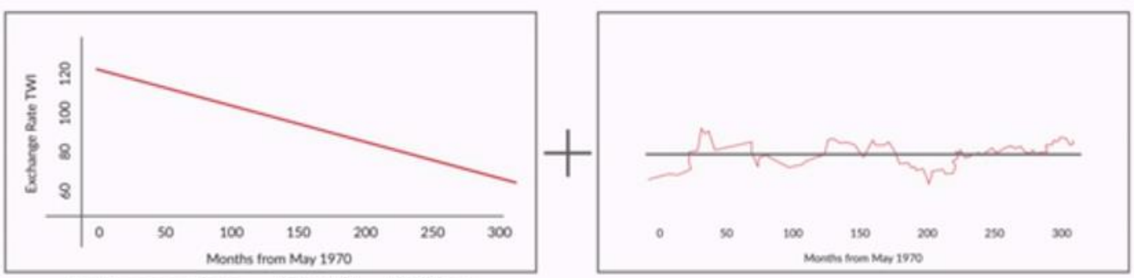
These are the variations that are **dependent on the immediate past**. For example, if the exchange rate has been falling too steeply lately, you expect it to increase soon and become consistent with the global prediction.

Global trend:

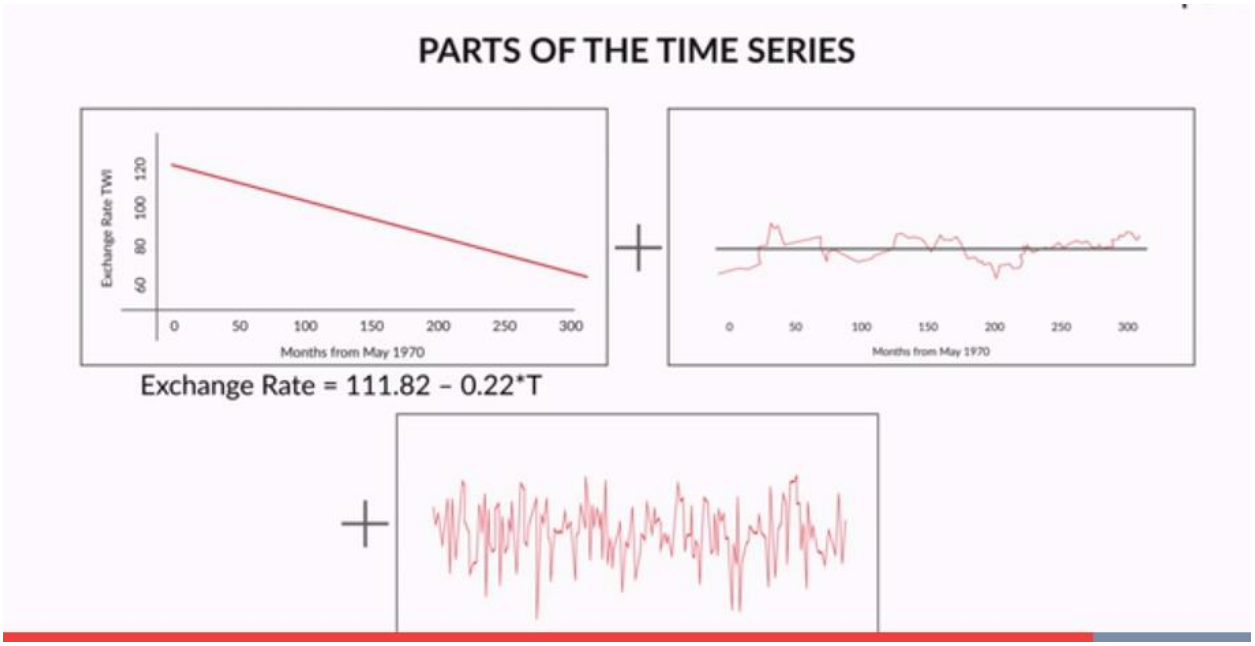
PARTS OF THE TIME SERIES



Exchange Rate = 111.82 – 0.22*T



Exchange Rate = 111.82 – 0.22*T



Questions: 1 / 1

Separating Time Series Components in R

Assuming that the modelling is done correctly, the variable localpred will contain (the name of the variable doesn't imply anything; it is only suggestive)

- Locally predictable part of the model

- **Locally predictable part of the model + Noise**

  Feedback : *The total time series is the sum of the variables globalpred and localpred. Since the globally predictable part of the time series has been modelled using the variable globalpred, the variable localpred will contain both the locally predictable part of the time series and the unpredictable noise.* ✓ Correct

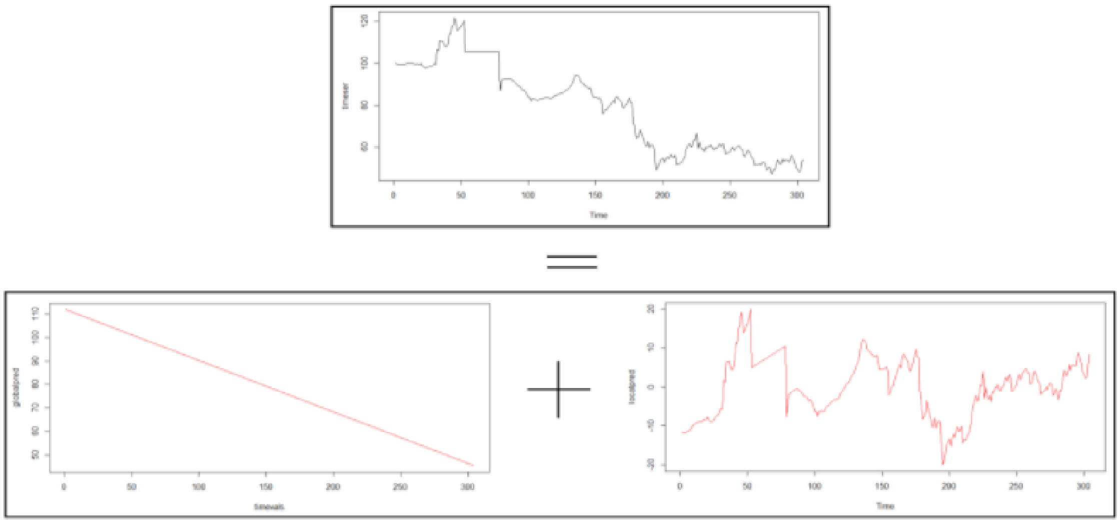
- **Globally predictable part of the model + Locally predictable part of the model**

  Feedback : *If the modelling has been done correctly, i.e, if the variable globalpred was modelled correctly, then the globally predictable part should not be present in the data frame localpred, which was obtained by subtracting the globalpred from the original time series.* × Incorrect

- Globally predictable part of the model + Noise

Your answer is **Incorrect.**

So, the time series you have (Exchange Rate) can be broken down into two parts:



Two Parts of a Time Series

Here, the left graph tells you about the global trend of the data, while the right graph is what is left after you subtract the global trend from the series.

So the series on the right, when **modelled**, gives you the **locally predictable part** of the model. But how is this locally predictable part modelled? You'll learn this in the next session.

Also, assuming that everything has been modelled correctly, if you subtract your prediction (global + local) from the original values, you will get white noise. Now what is white noise? How can you identify if a given time series is white noise? These are also questions we will address in the next session.

Time Series Modelling (Stationarity and ARMA Processes)

Steps for Modelling a Time Series

Introduction

Welcome to the second session on 'Time Series Analysis'.

In the previous session, you gained a basic understanding of what a time series is and learnt how it **differs from a normal regression** analysis. In this session, you will dive deeper into time series modelling, which will eventually help you in making a time series forecast.

In this session

Some of the concepts that you will learn in this session are

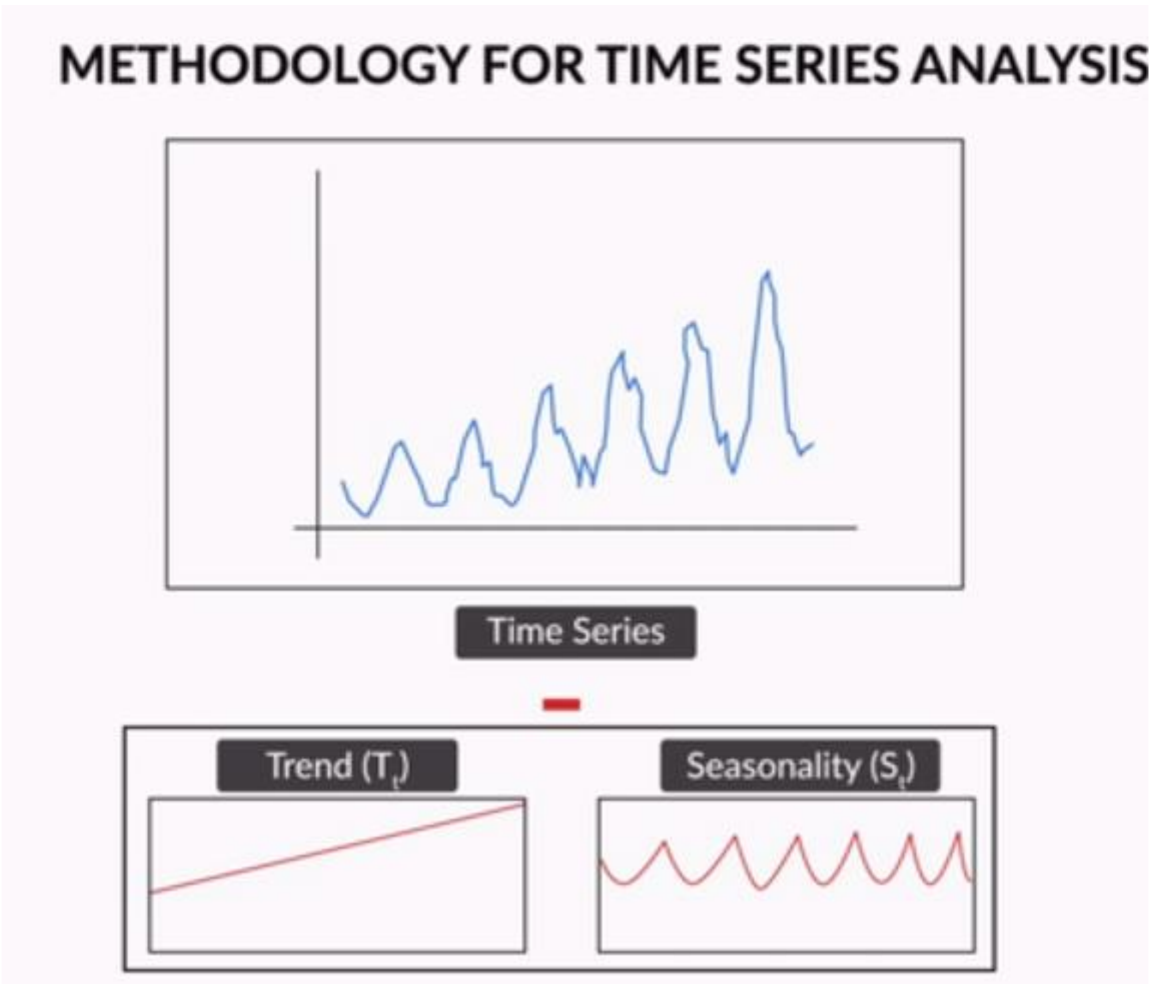
- Stationarity
- White noise
- Autocorrelation function

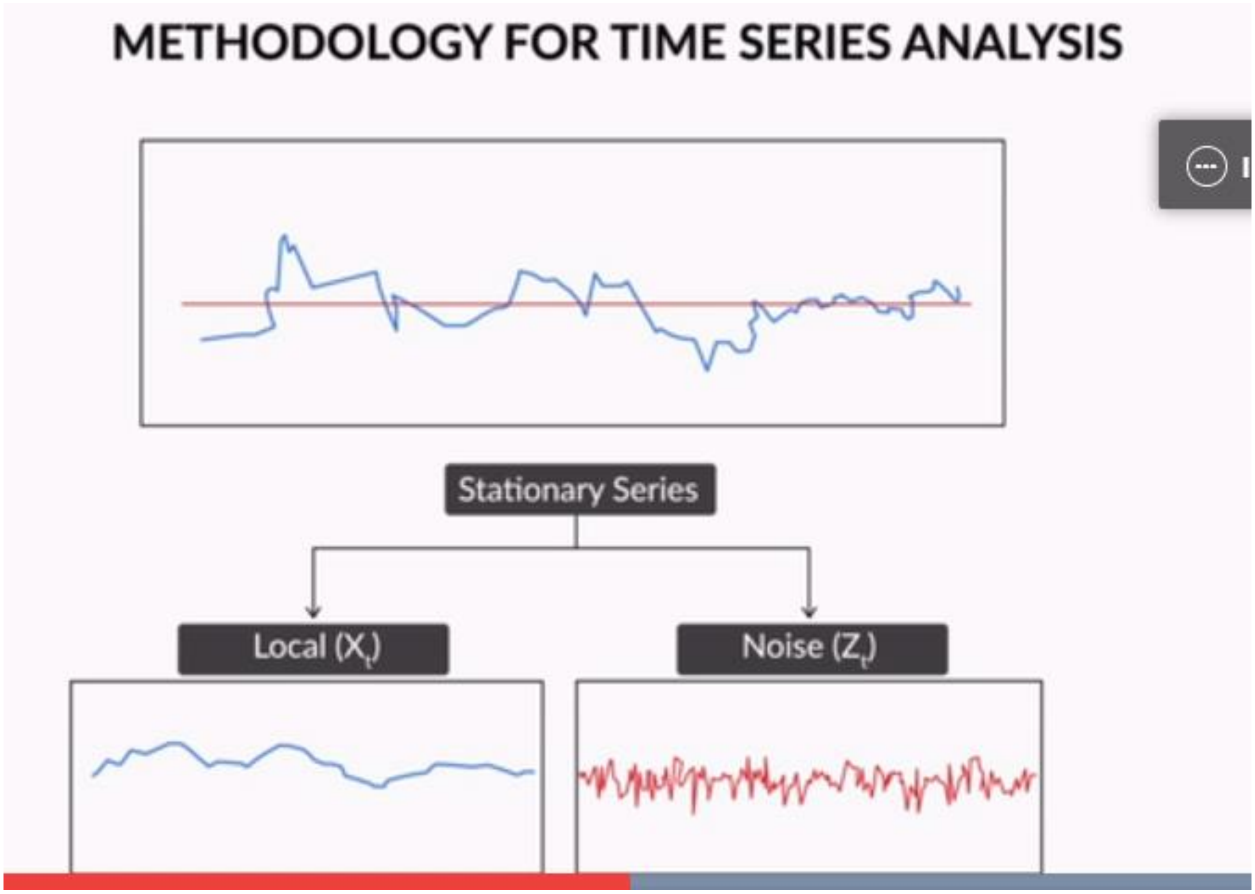
- Partial autocorrelation function
- AR model
- MA model
- ARMA model

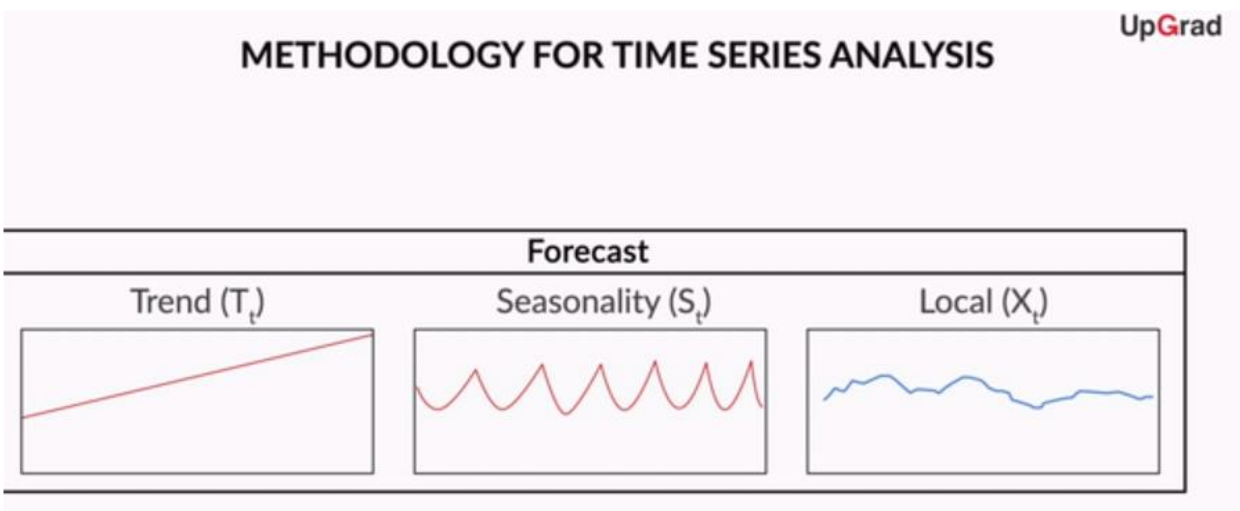
Downloads

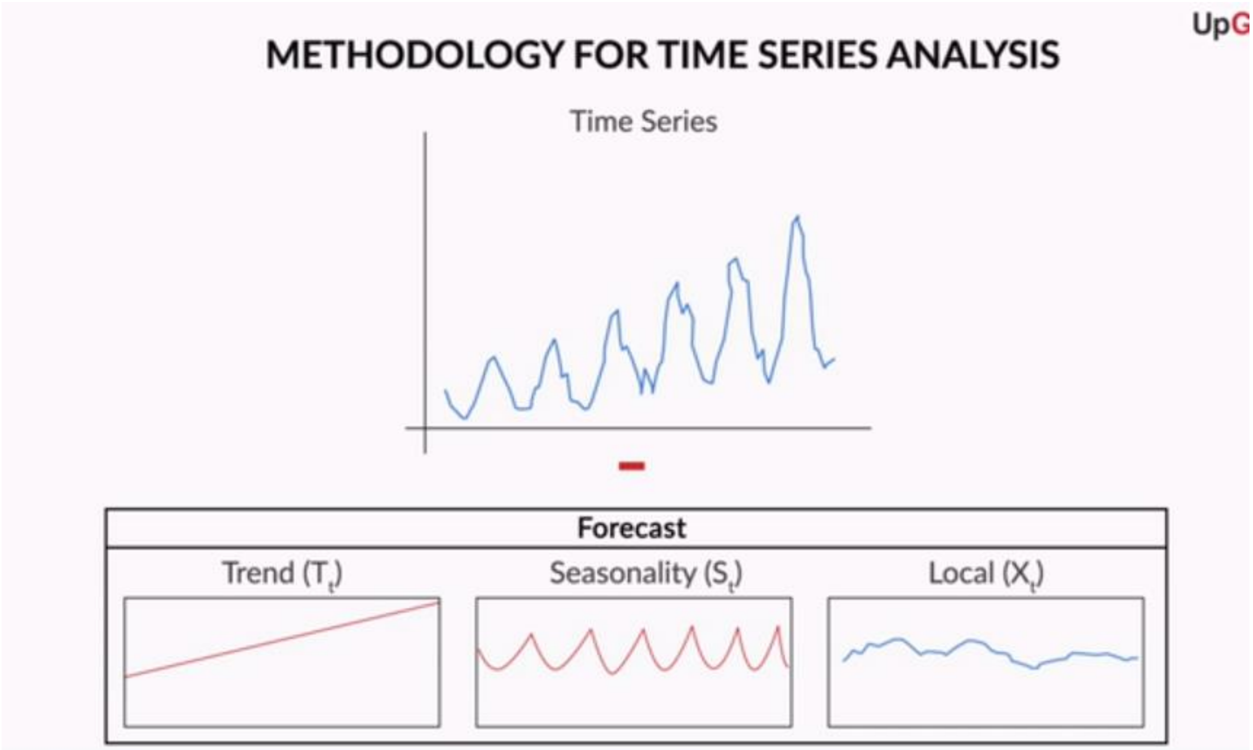
This module on time series has integrated lecture notes for all the sessions combined, and it will complement and supplement your learning. You can download the lecture notes from the link given below, and you can refer to it as and when required.

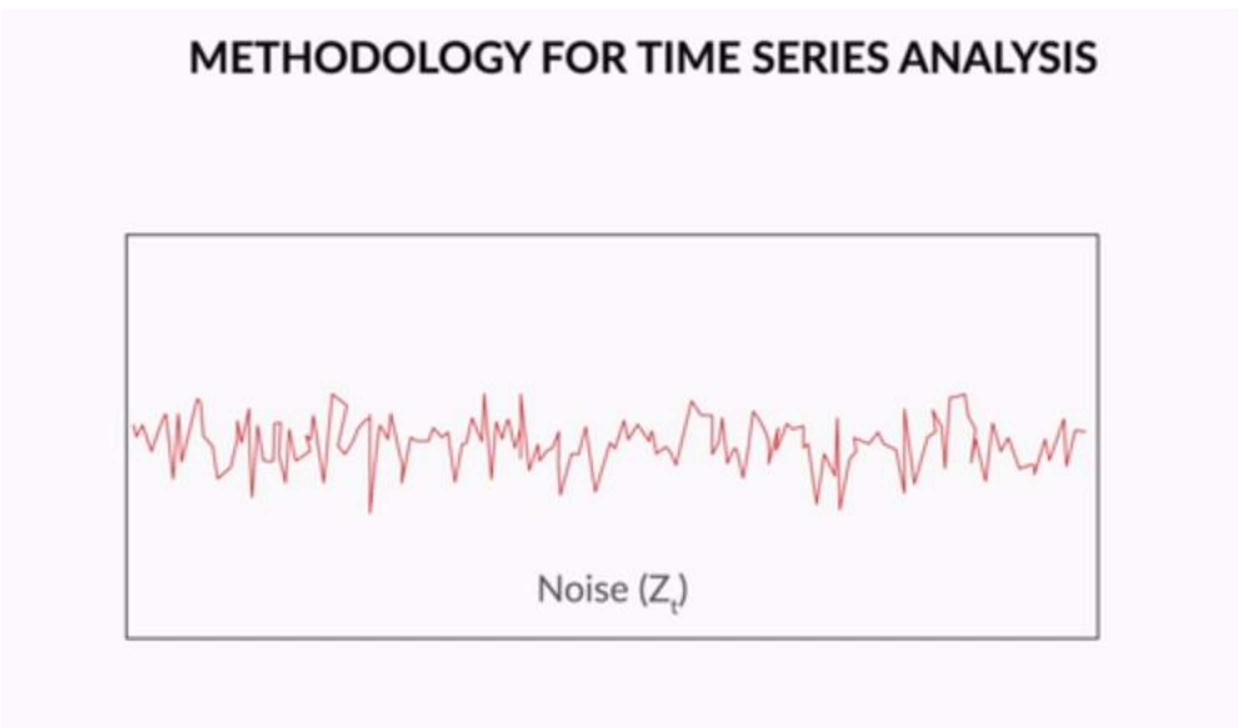
Let's quickly recap the overall methodology for modelling a time series.




METHODOLOGY FOR TIME SERIES ANALYSIS
Stationary Series
Local (X_t)
Noise (Z_t)


UpGrad
METHODOLOGY FOR TIME SERIES ANALYSIS
Forecast
Trend (T_t)
Seasonality (S_t)
Local (X_t)


UpG
METHODOLOGY FOR TIME SERIES ANALYSIS
Time Series
Forecast
Trend (T_t)
Seasonality (S_t)
Local (X_t)


METHODOLOGY FOR TIME SERIES ANALYSIS
Noise (Z_t)

Different steps of the time series modelling process can be summarised

In general, the different steps of the time series modelling process can be summarised as follows:

1. Visualise the time series.
2. Recognise the trend and seasonal component.
3. Apply regression to model the trend and the seasonality.
4. Remove the trend and seasonal component from the series. What remains is the stationary part: a combination of the autoregressive, and white noise.
5. Model this stationary time series.
6. Combine the forecast of this model with the trend and seasonal component.
7. Find the residual series by subtracting the forecasted value from the actual observed value.
8. Check if the residual series is pure white noise.

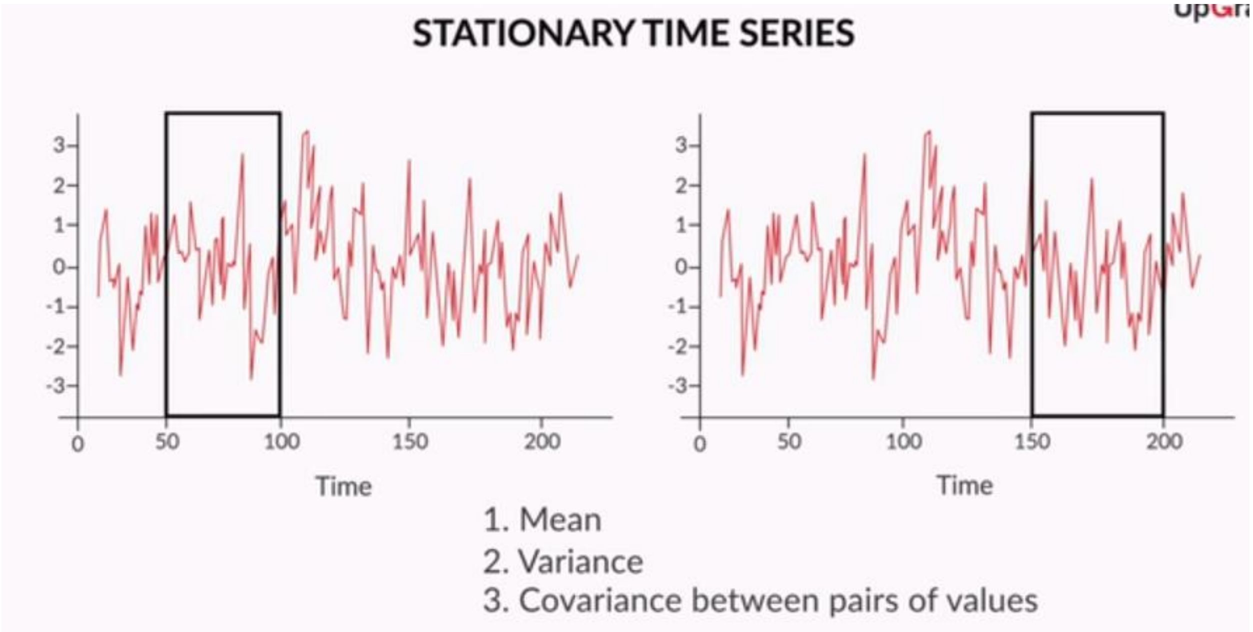
As Prof. Raghavan said, the local part can be modelled only if the series is "**stationary**". But what is a stationary series? You'll learn more about this in the next section.

Stationarity

So now that you've modelled the globally predictable part using regression, you have to model the locally predictable part.

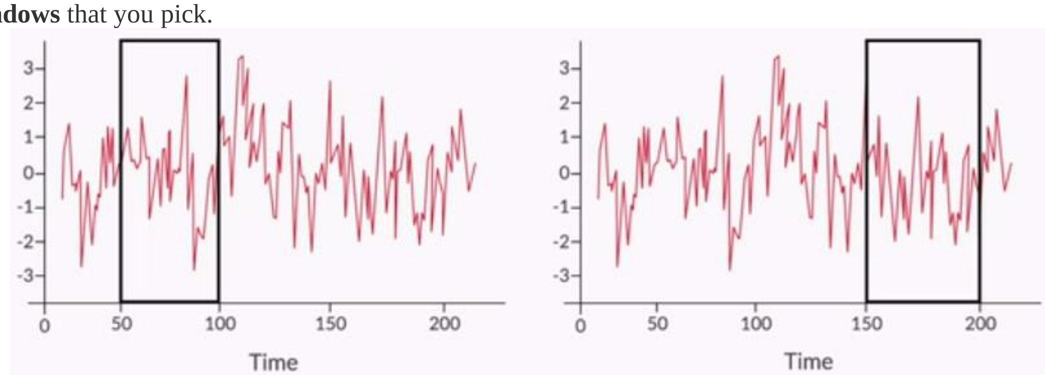
However, before modelling the locally predictable part, you have to ensure that it is **stationary** or at least **weakly stationary**; otherwise, it cannot be modelled.

So let's look at what a stationary time series is and how it can be identified.



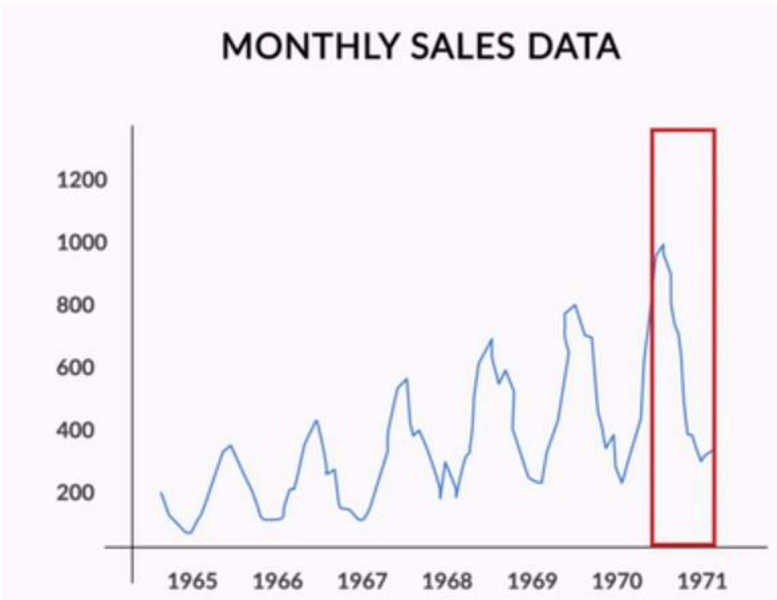
So as the professor said, if a time series is stationary, its **statistical properties** will be the **same throughout the series**, irrespective of the time at which you observe them.

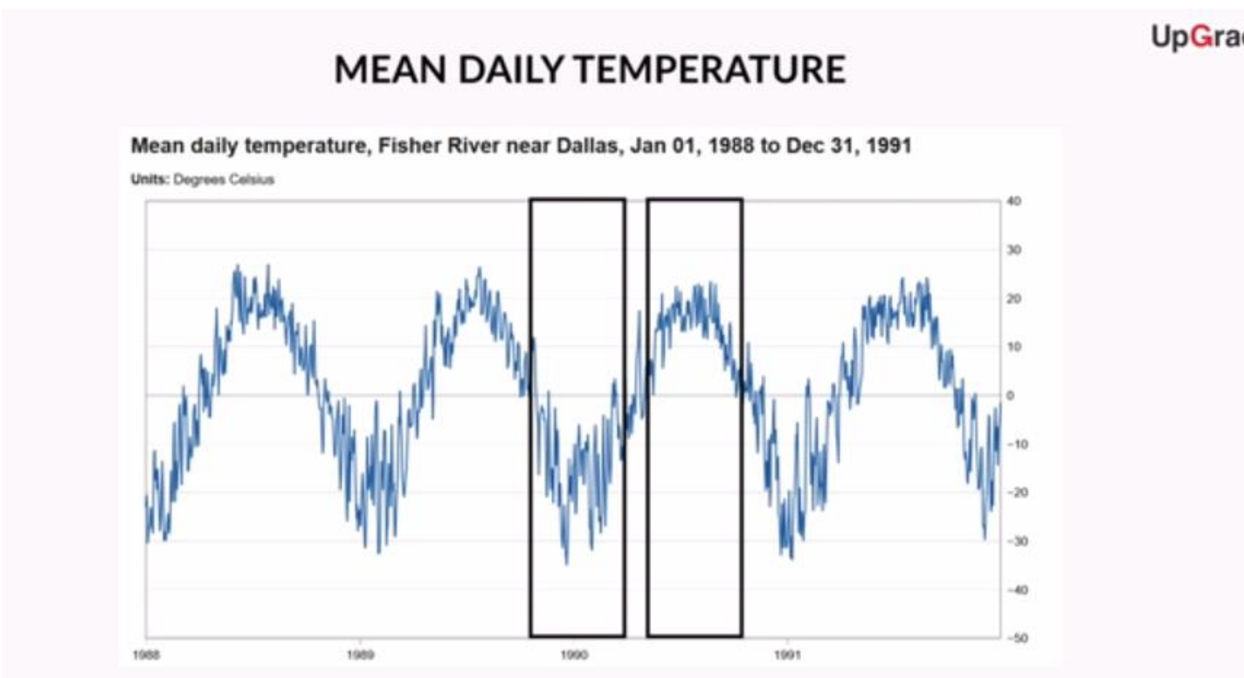
In other words, for a stationary time series, properties such as **mean, variance, etc.** will be the **same for any two time windows** that you pick.

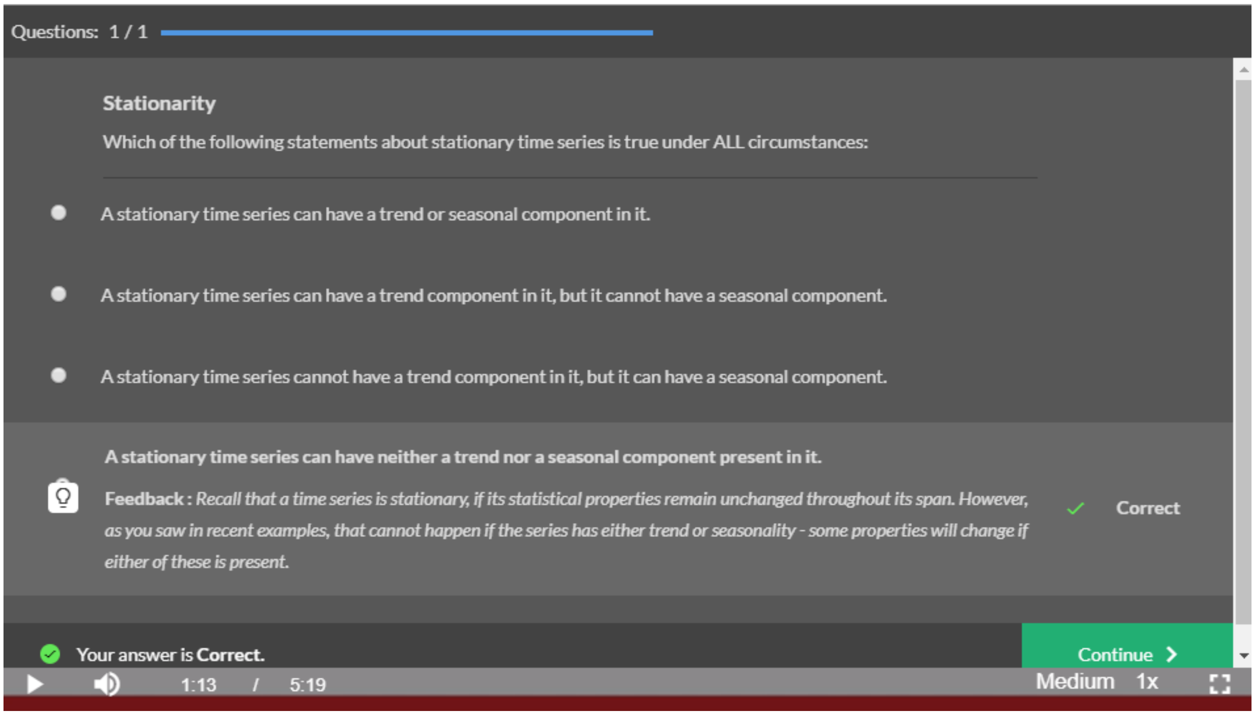


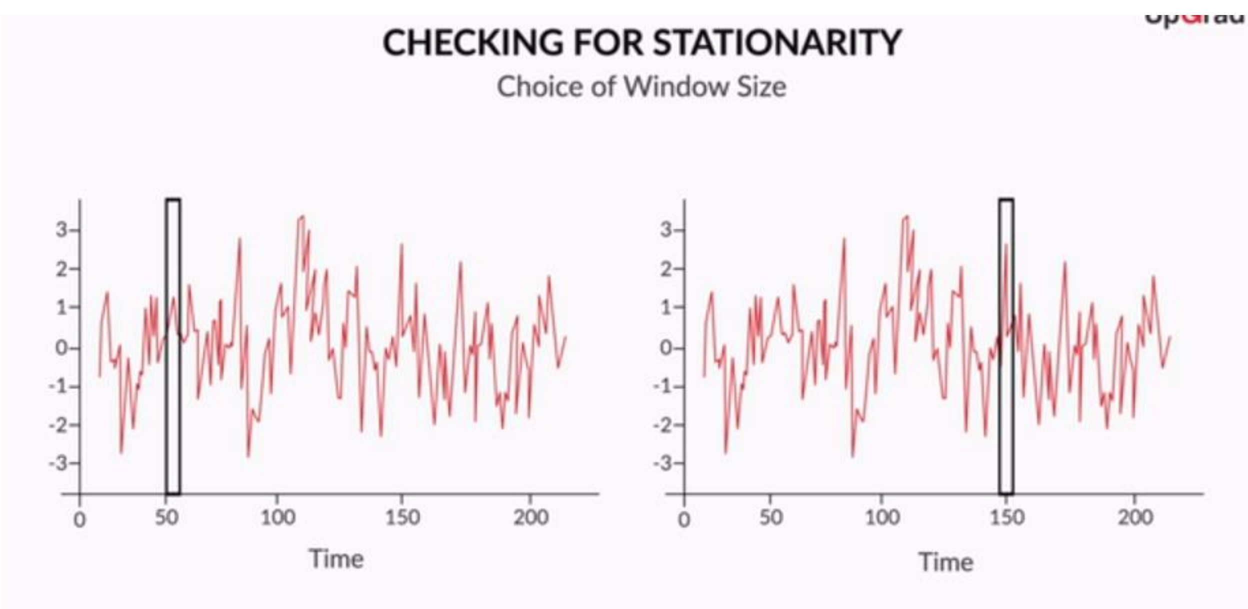
Stationary Time Series

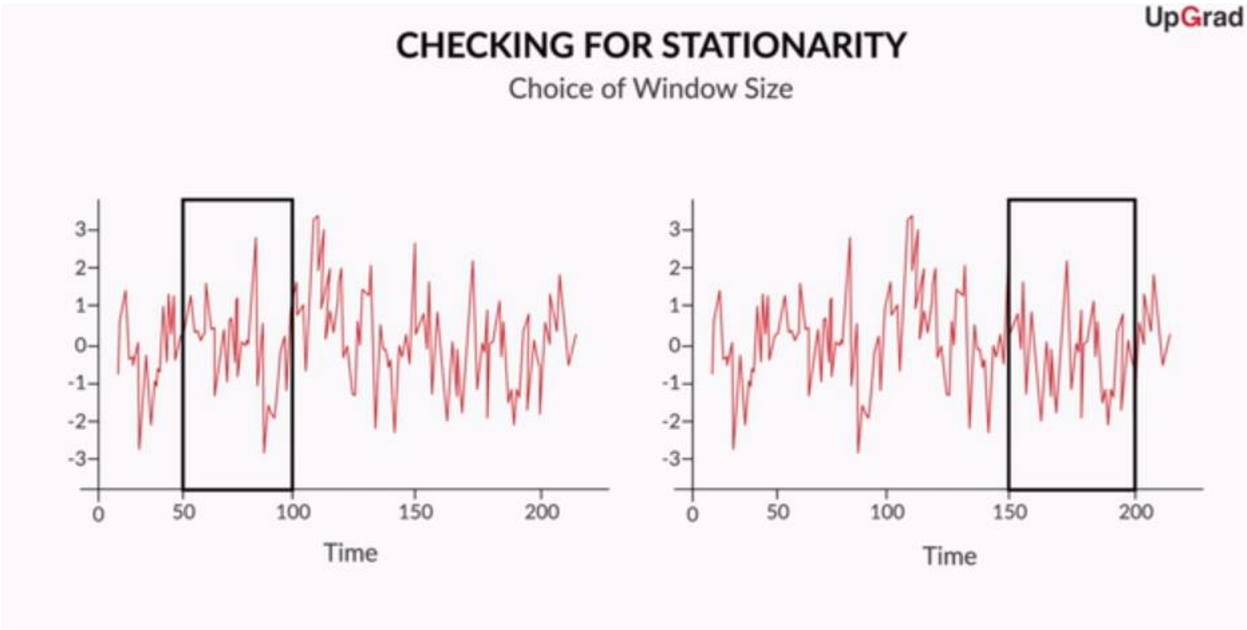
Let’s now listen to Prof. Raghavan as he delves deeper into the notion of stationarity, with more examples.


MONTHLY SALES DATA
1200
1000
800
600
400
200
1965
1966
1967
1968
1969
1970
1971


MEAN DAILY TEMPERATURE
UpGrad
Mean daily temperature, Fisher River near Dallas, Jan 01, 1988 to Dec 31, 1991
Units: Degrees Celsius
40
30
20
10
0
-10
-20
-30
-40
-50
1988
1989
1990
1991


Questions: 1 / 1
Stationarity
Which of the following statements about stationary time series is true under ALL circumstances:
A stationary time series can have a trend or seasonal component in it.
A stationary time series can have a trend component in it, but it cannot have a seasonal component.
A stationary time series cannot have a trend component in it, but it can have a seasonal component.
A stationary time series can have neither a trend nor a seasonal component present in it.
Feedback : Recall that a time series is stationary, if its statistical properties remain unchanged throughout its span. However, as you saw in recent examples, that cannot happen if the series has either trend or seasonality - some properties will change if either of these is present.
Correct
Your answer is Correct.
Continue
1:13 / 5:19
Medium 1x

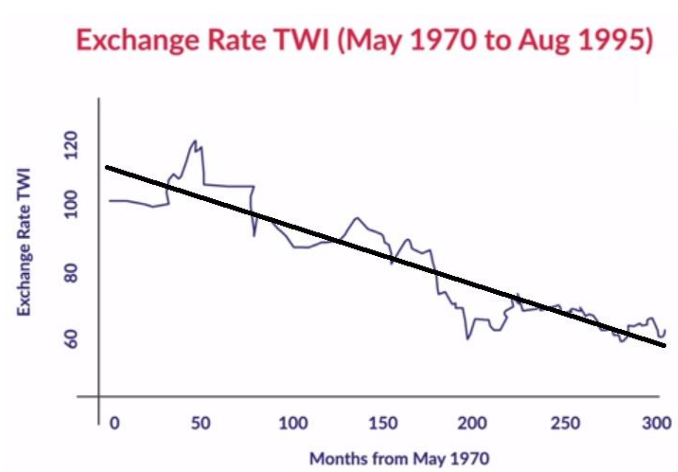

CHECKING FOR STATIONARITY
Choice of Window Size
3
2
1
0
-1
-2
-3
0
50
100
150
200
Time
3
2
1
0
-1
-2
-3
0
50
100
150
200
Time



In general, a stationary time series will have **no long-term predictable patterns** such as trends or seasonality. Time plots will show the series to roughly have a **horizontal trend** with the **constant variance**.

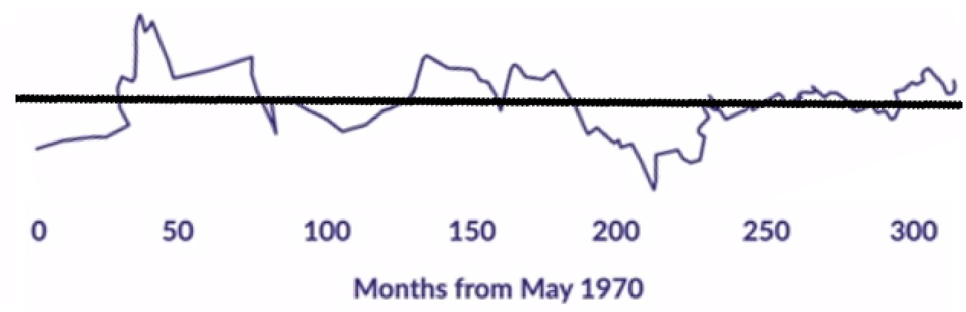
So why is stationarity so important? To understand this, let's go back to the exchange rate TWI example. If you remember, you were trying to predict what the exchange rate's value would be in the near future.

For that, you modelled the global trend of the data using linear regression, which is shown here using the black line.



Modelling Global Trend

Then, after removing the global pattern from the data, you got the following series:



Series Left After Removing Global Trend

The regression model should capture all the variations caused in the exchange rate due to the trend and seasonality. When you model the residual series — the one you get after subtracting the regressed series from the original one — there should be **no trend or seasonality left** in it. Using the residual series, you only want to model the local patterns, i.e. the dependence of values on past values. You don't want any global patterns to seep in there.

But how do you model this stationary (de-trended and de-seasonalised) series? We will look into this in later sections.

Questions: 1 / 2

Stationarity

The image given below shows the graph of a time series.

Is this series stationary?

Yes

No

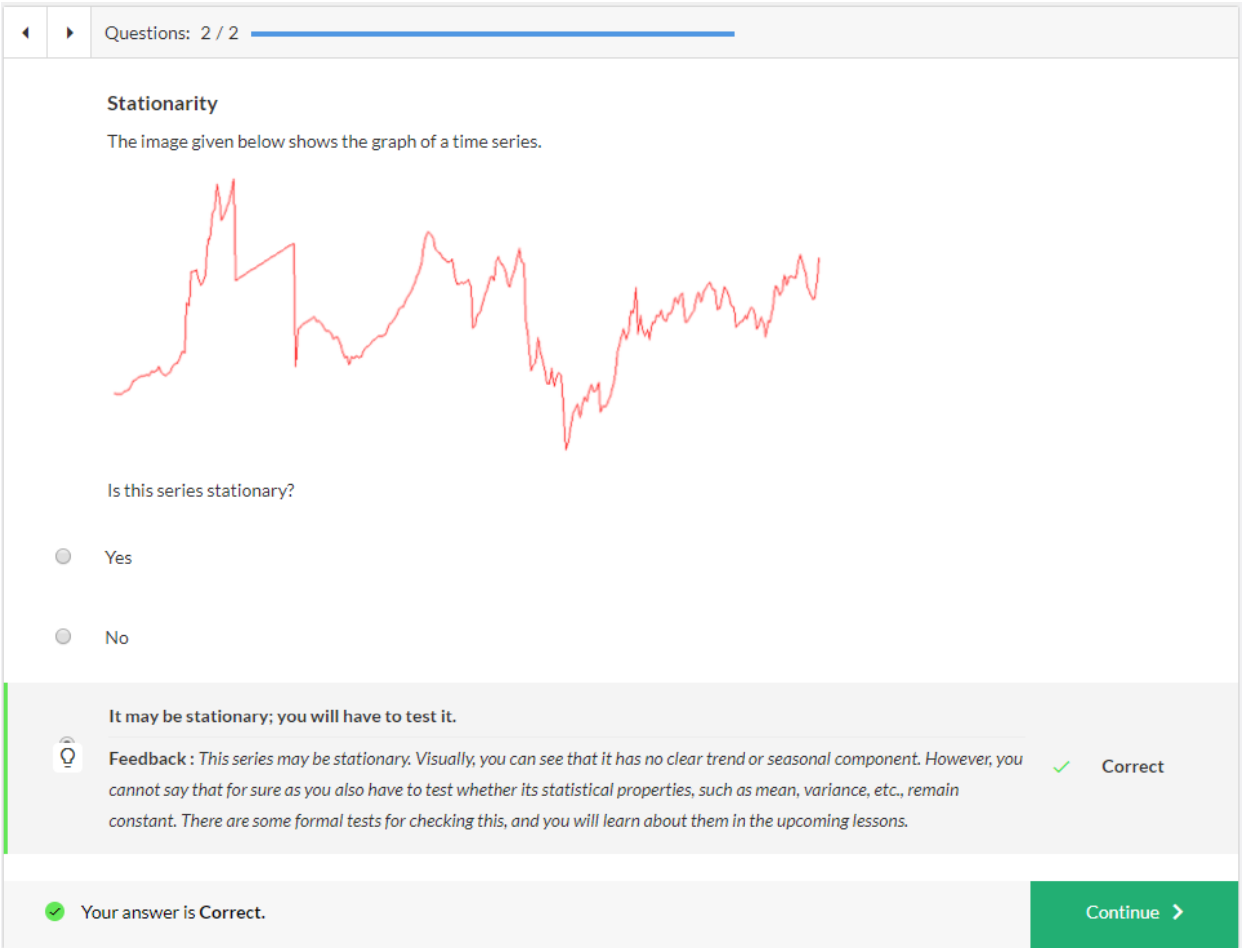
Feedback : *This series cannot be stationary. Visually, you can see that it has a very clear trend component. Stationary time series however, are not supposed to have trend or seasonal components.*

Correct

It may be stationary; you will have to test it.

Your answer is **Correct.**

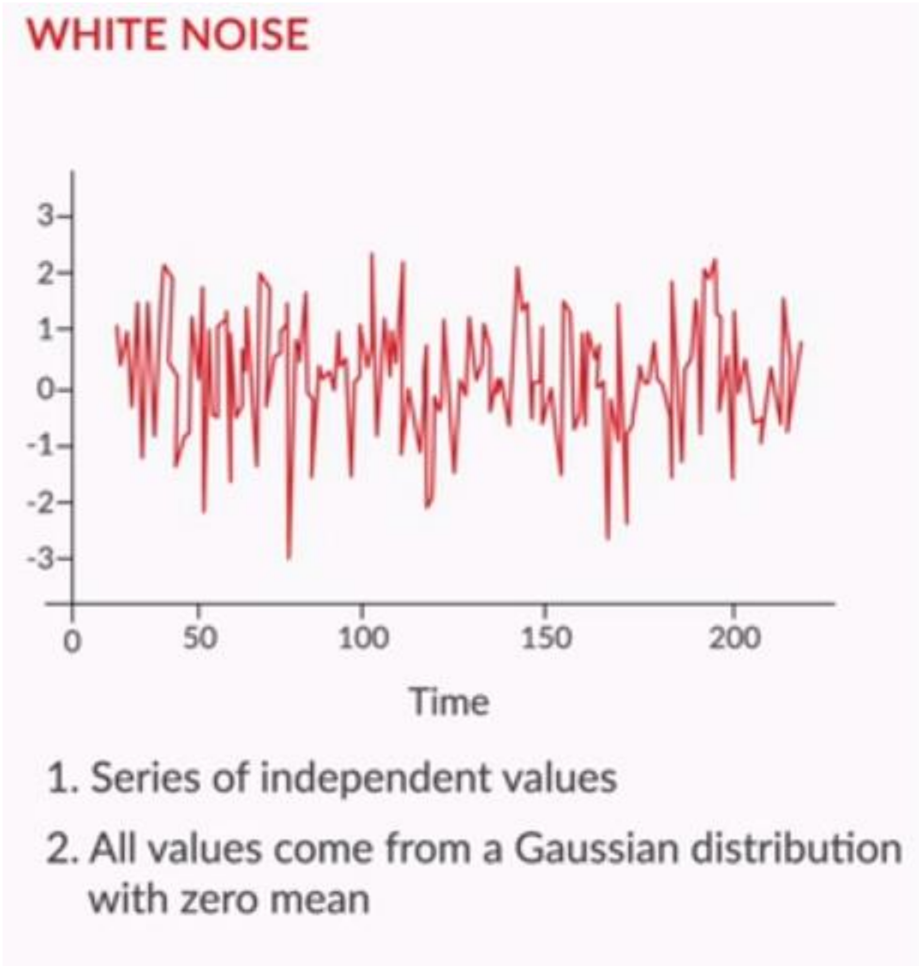
Continue



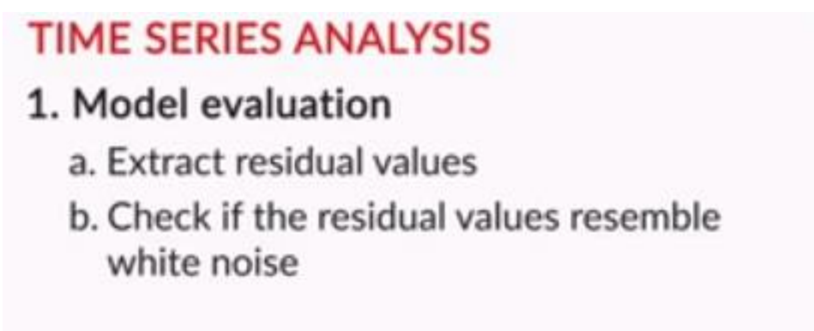
White Noise

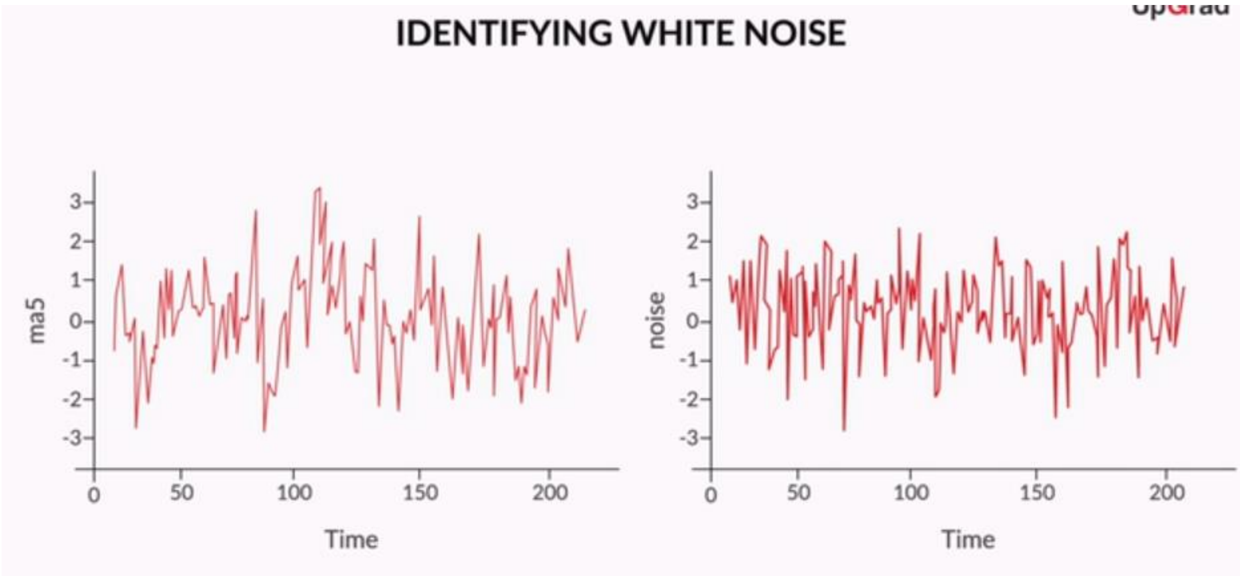
So far, you learnt that a time series has three main components: the **global pattern** (which captures both the trend and seasonality), the **local** (predictable) pattern which should be stationary, and the (unpredictable) **noise**.

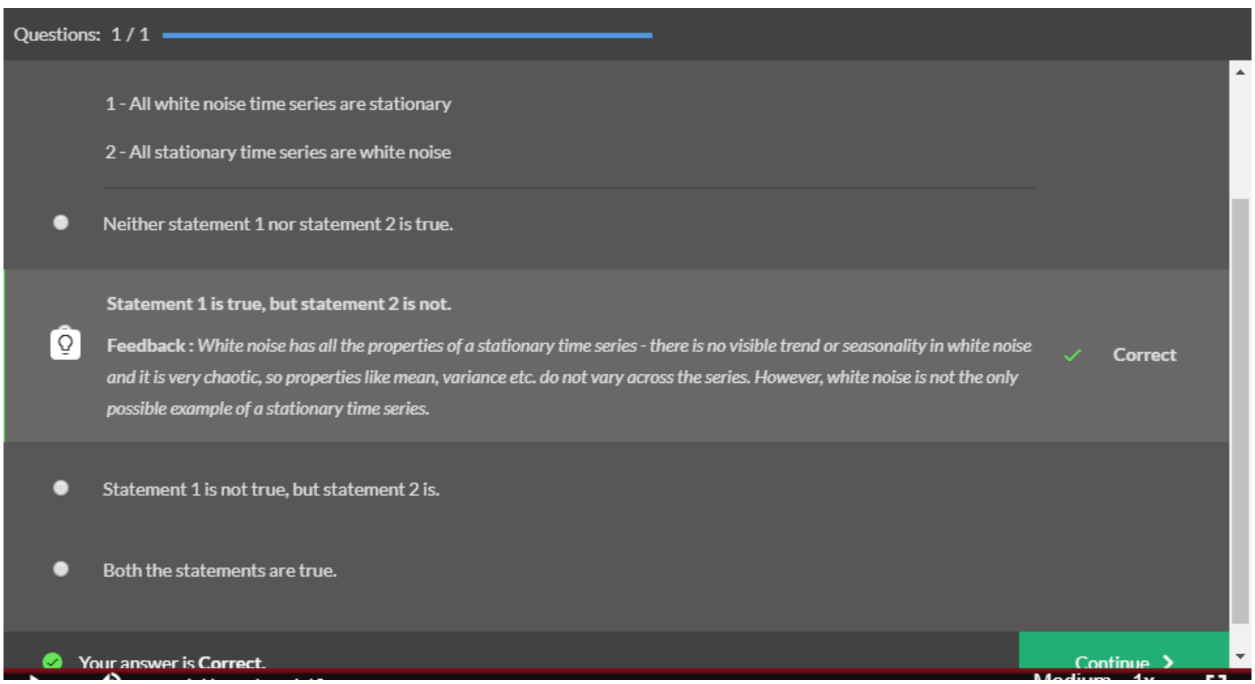
Let's understand noise in detail. Remember that noise (or **white noise**) is what remains when all the predictable parts of a time series have been modelled and extracted from it. It is a set of **independent and uncorrelated** values. If you plot white noise over time, it will look something like this:



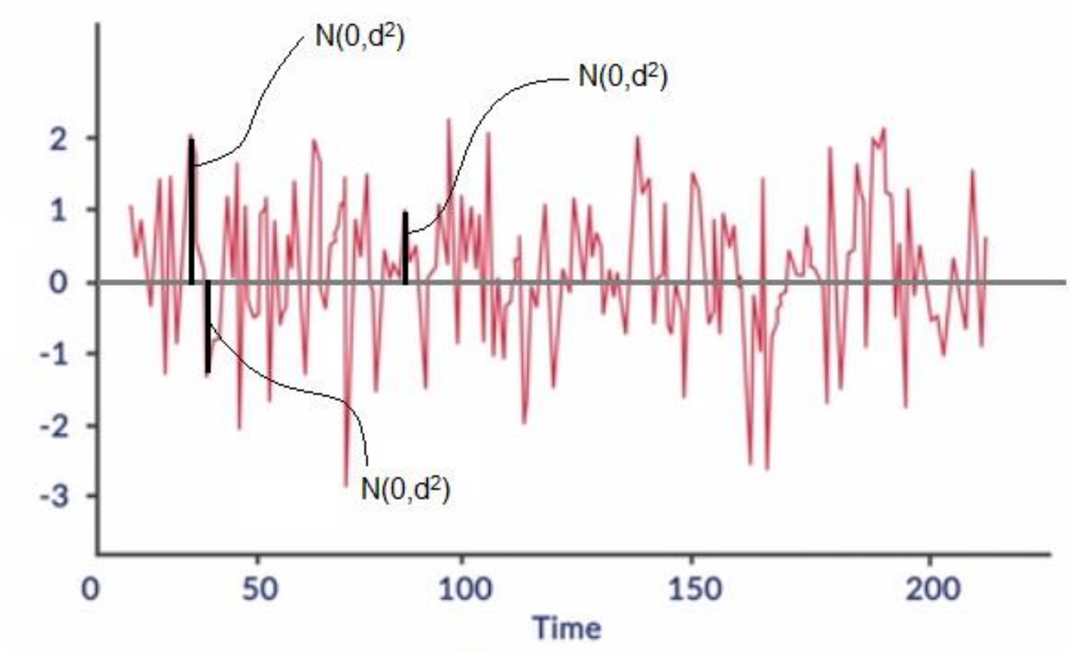
Notice that there are no identifiable trend, seasonal or cyclical components. So a white noise series is basically an example of a stationary series. Let's learn more about it.







So **white noise** is basically a series of values that are all **independent**. Typically, for noise, you can assume that all these values come from a **Gaussian distribution** with a **zero mean**.



White noise

Remember the methodology for modelling a times series. At the end of the analysis, you will need to know if the series you have is pure noise or not. However, since the graphs for stationary series and white noise look so similar, it may actually be a little difficult to tell the two apart.
more_horizIncomplete

TESTS FOR PURE NOISE

1. Histogram Plot

2. Q-Q Plot

As you saw above, only performing a visual inspection of the time series plot will not help you confirm if the series is white noise. This is because any stationary series would resemble a pure white noise series.

Hence, you will have to perform some **concrete tests** to check whether the series is noise, or if it is just a stationary series. As you learnt earlier, if the series is white noise, the values in it will belong to a **normal distribution**. Hence, you can test if the series' values belong to a normal distribution or not. For this, you've learnt two tests:

- Histogram test
- Q-Q plot test

You can download the R-codes used by Prof. Raghavan, for the testing of white noise, from the link given below.

White noise test:

Other than the tests mentioned above, i.e. the histogram test and the q-q plot test, there are a few more popular tests that can be used to understand whether a series is white noise or not. Some of them are listed here:

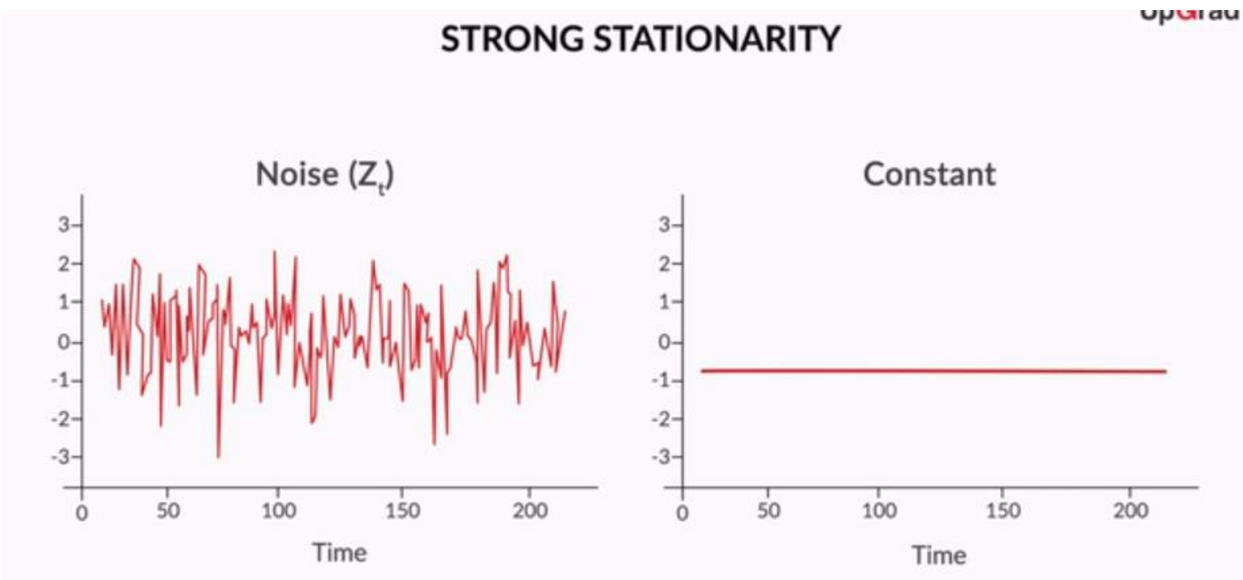
1. Ljung-Box (Portmanteau) test
2. Turning point test
3. Difference sign test
4. Runs test
5. Rank test
6. ACF and PACF

Due to time constraints, we have not covered all these tests in this course. We already covered histogram plots and Q-Q plots, and we will cover ACF and PACF plots in a subsequent section. However, if you wish to go through the other five tests, you can go through sections 4.2.2 to 4.2.6 of the lecture notes.

Identifying stationary time series - ACF and PACF

As discussed earlier, there are some specialised tests that can help you distinguish white noise from an ordinary stationary series. These tests are necessary because visually, a stationary time series does not really look very different from white noise.

One such test is called an **autocorrelation function (ACF)**. Let's listen to Prof. Raghavan as he explains what ACF is.

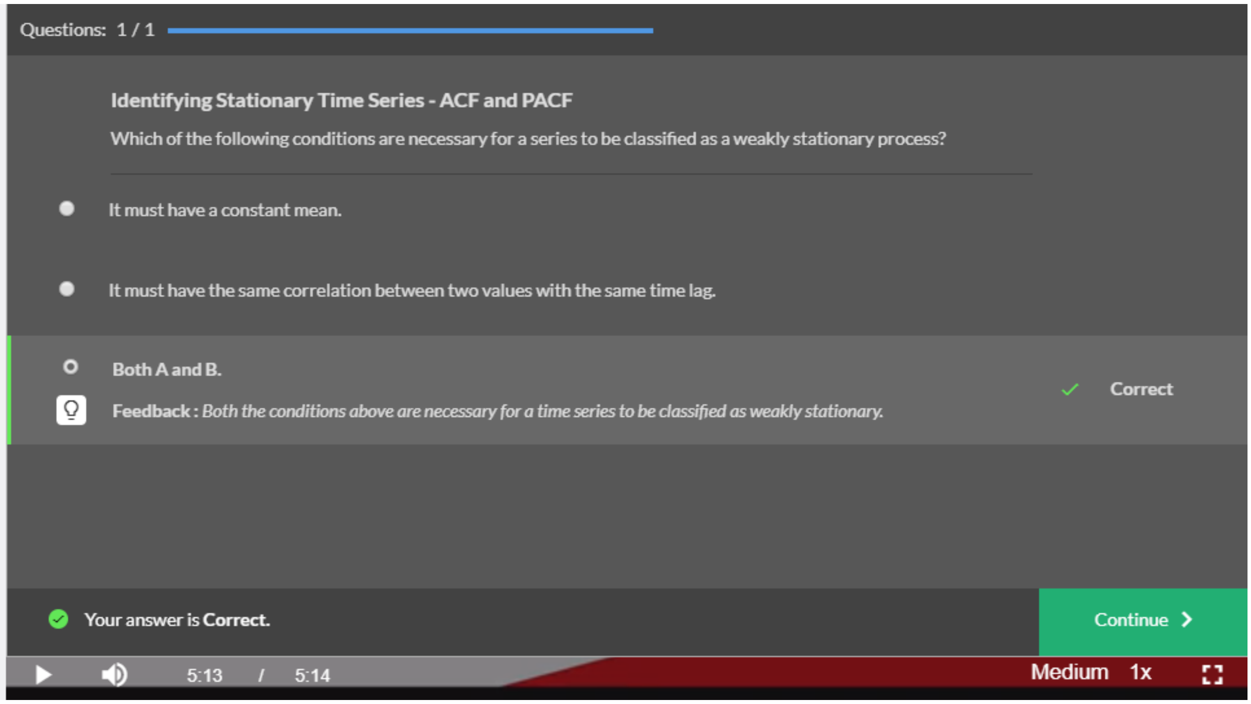


AUTOCORRELATION

Timestamp	Values
T_1	V_1
T_2	V_2
T_3	V_3
T_4	V_4
T_5	V_5
T_6	V_6
T_7	V_7
T_8	V_8
T_9	V_9
T_{10}	V_{10}
T_{11}	V_{11}
T_{12}	V_{12}
T_{13}	V_{13}
T_{14}	V_{14}
T_{15}	V_{15}

LAGGED TIME SERIES

Timestamp	Observation	Lag 1	Lag 2
T1	42.98000		
T2	52.27116	42.98000	
T3	49.61729	52.27116	42.98000
T4	39.33107	49.61729	52.27116
T5	49.71779	39.33107	49.61729
T6	49.48074	49.71779	39.33107
T7	46.20758	49.48074	49.71779
T8	47.63529	46.20758	49.48074
T9	55.92837	47.63529	46.20758
T10	49.63925	55.92837	47.63529
T11	46.29526	49.63925	55.92837
T12	41.91689	46.29526	49.63925
T13	43.14725	41.91689	46.29526
T14	52.29256	43.14725	41.91689
T15	52.83307	52.29256	43.14725
T16	47.52686	52.83307	52.29256
T17	49.60583	47.52686	52.83307
T18	52.79283	49.60583	47.52686
T19	50.82943	52.79283	49.60583



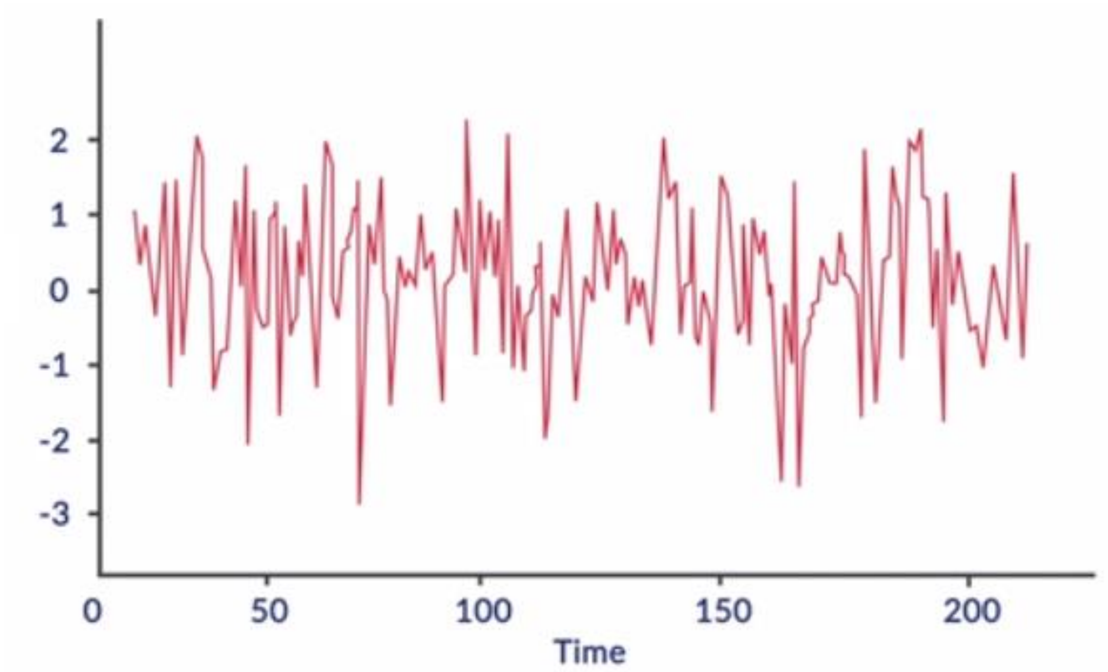
Now let's look at strong and weak stationarity again.

Strong stationarity: If a series is stationary, then the shape of the time series plot will look identical, irrespective of the time, t, at which you start collecting/observing the data. The time series 'shifted' to the right or left makes little difference to the shape of the plot.

Weak stationarity: If a series is stationary, then the pairwise relationships are preserved. In other words, the time series 'shifted' to the right or left makes a fixed difference to the shape of the plot.

The only two series that are stationary in the strongest sense are
1. White noise
2. Constant function

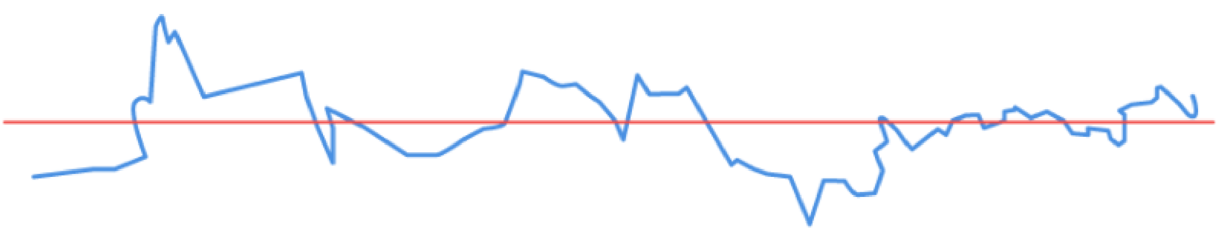
Remember the exchange rate time series example. At the end of the entire modelling exercise, when you have modelled both the globally and locally predictable parts of the time series, you will have to check whether the final residue is white noise or not.



Final Residue (Original Series - Global Predictions - Local Predictions)

In other words, you will have to check whether the **final residual series** is **strongly stationary** or not.

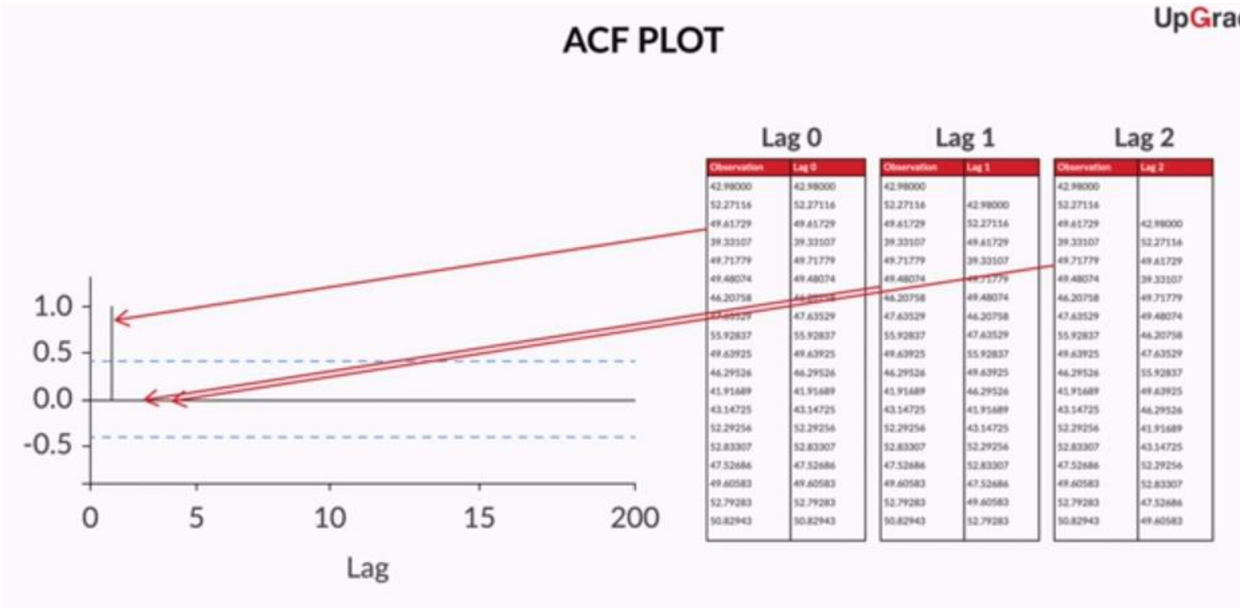
Also, in the intermediate step, you will check whether the series obtained after subtracting the global predictability from the original values is stationary or not. Actually, here, you check for **weak stationarity** and not strong stationarity.

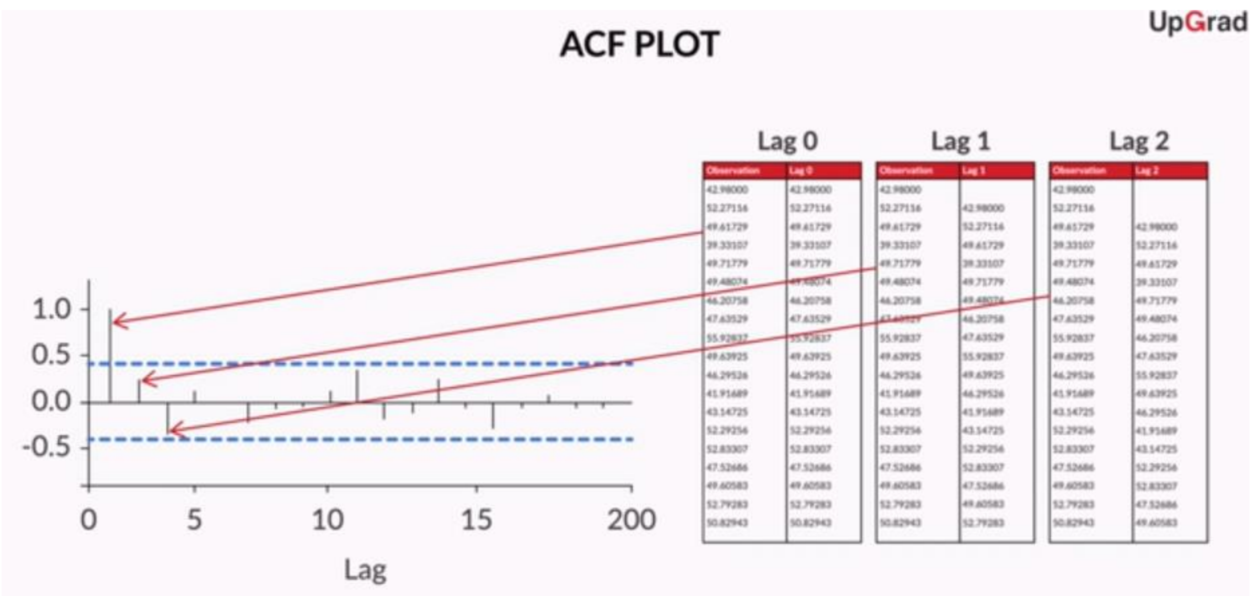


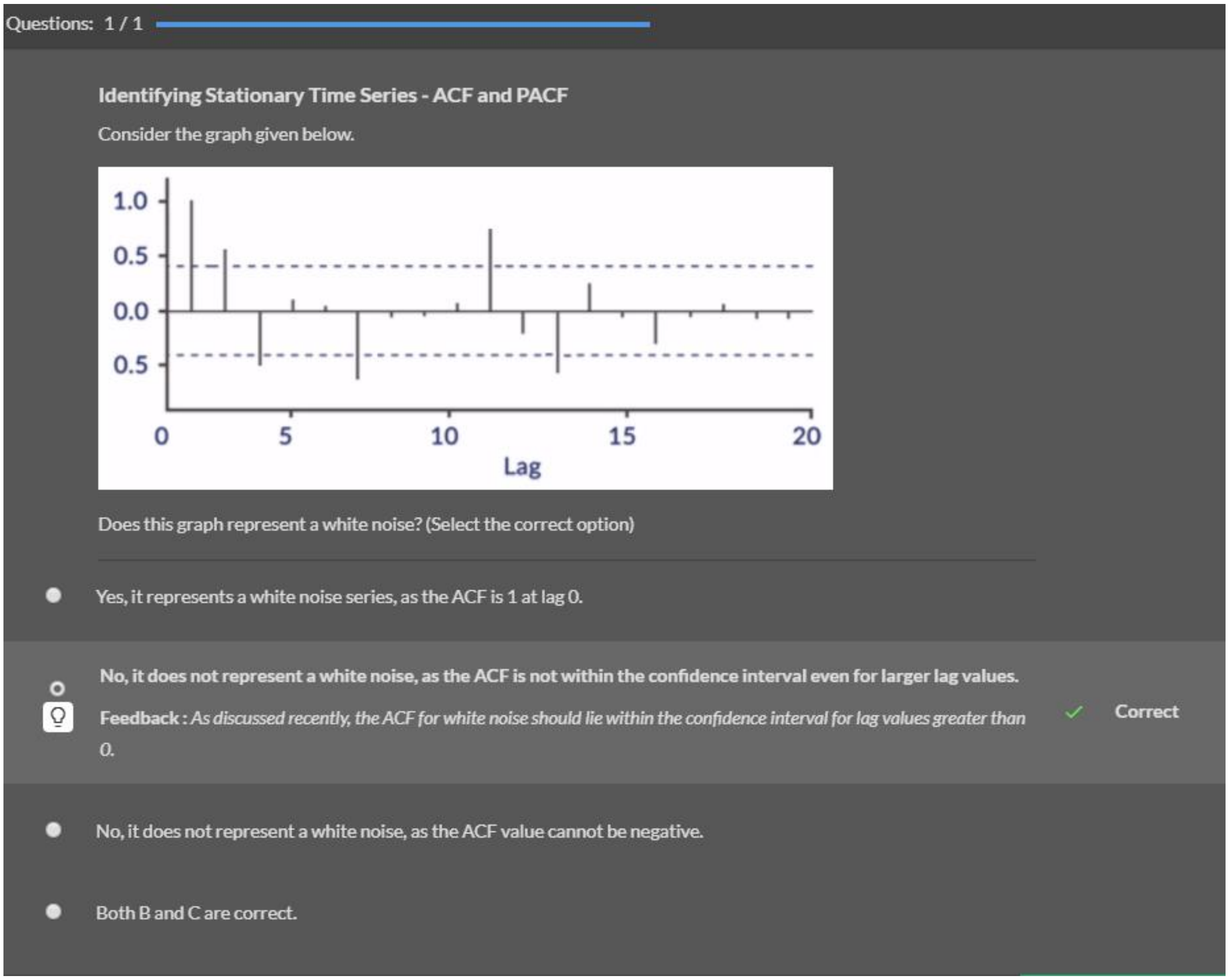
Intermediate Residue (Original Series - Global Predictions)

Think about it: if the series is strongly stationary, then that means the exchange rate's values are actually not dependent on past values. Whereas, if the series is weakly stationary, then it means that the exchange rate's values are dependent on previous exchange rate values, and they are dependent on them in a fixed way. There is no long-term trend or seasonality, but you do have some local predictability in the exchange rate data. The values of the past are dependent on the future, and you now model this relationship.

The **ACF** can actually help you **identify** and **distinguish** between **strong** and **weak stationarity**. First, let's look at how it can help you identify strong stationarity, i.e. the presence of white noise.


ACF PLOT
UpGrad
Lag 0
Lag 1
Lag 2
Observation
1.0
0.5
0.0
-0.5
0
5
10
15
200
Lag


ACF PLOT
UpGrad
Lag 0
Lag 1
Lag 2
Observation
1.0
0.5
0.0
-0.5
0
5
10
15
200
Lag



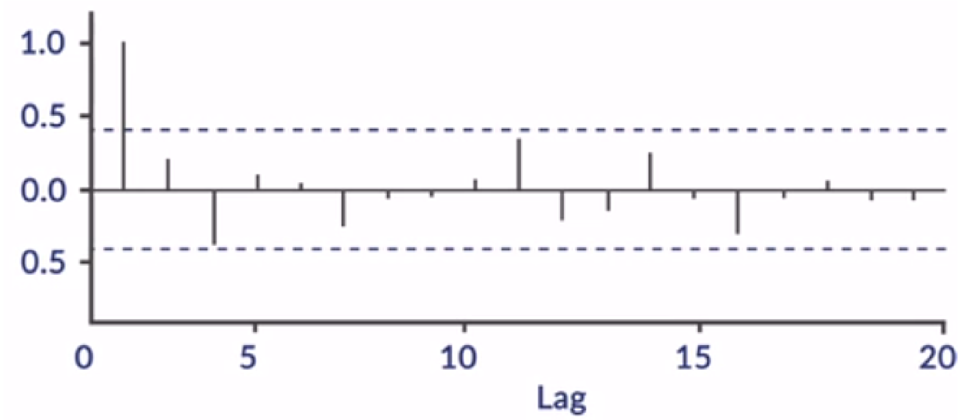
You can download the R-codes, for generating the ACF and PACF plot, from the link given below.

ACF & PACF

file_download**Download**

So for white noise, the autocorrelation function (**ACF**) is **zero** for all **non-zero lags**. In other words, past values are not correlated with present values at all, making the series a sequence of independent, uncorrelated values.

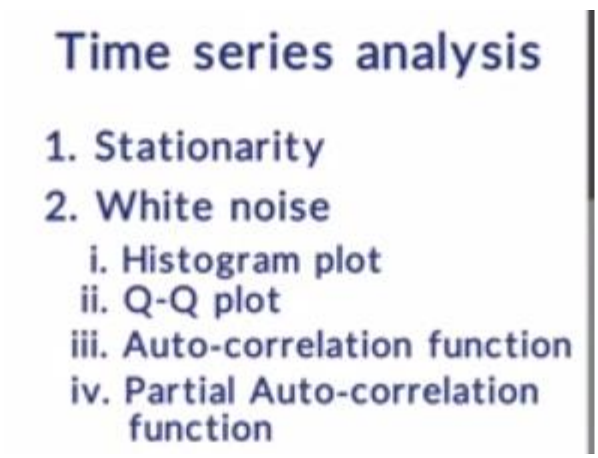
However, in reality, the value of correlation will not exactly be equal to zero. Hence, using hypothesis testing, you can check if it is significantly different from zero.

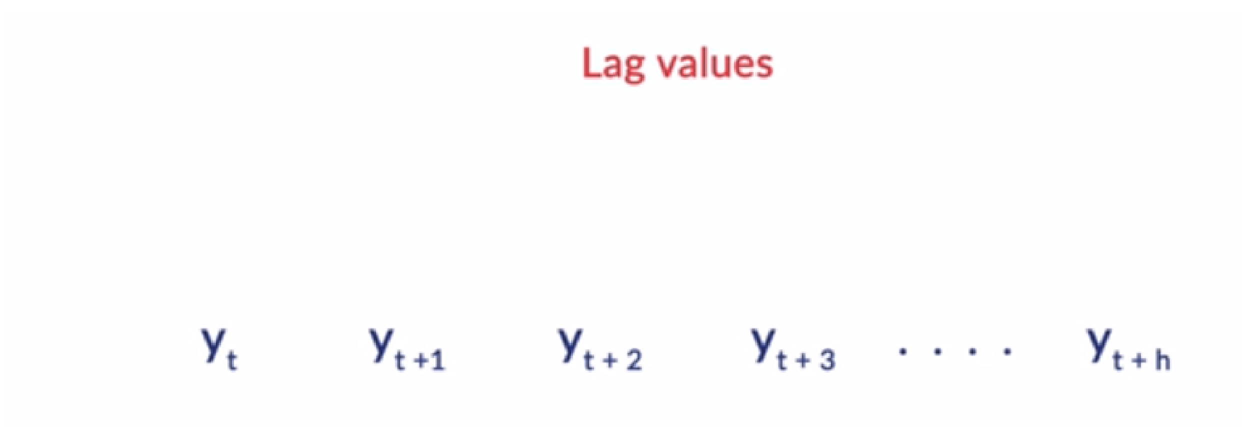


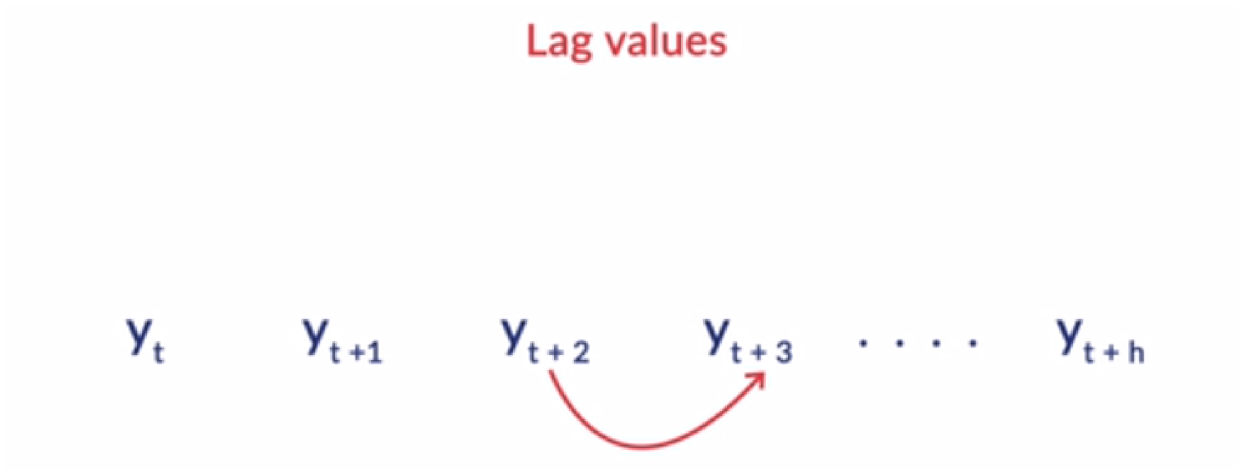
ACF Plot for White Noise

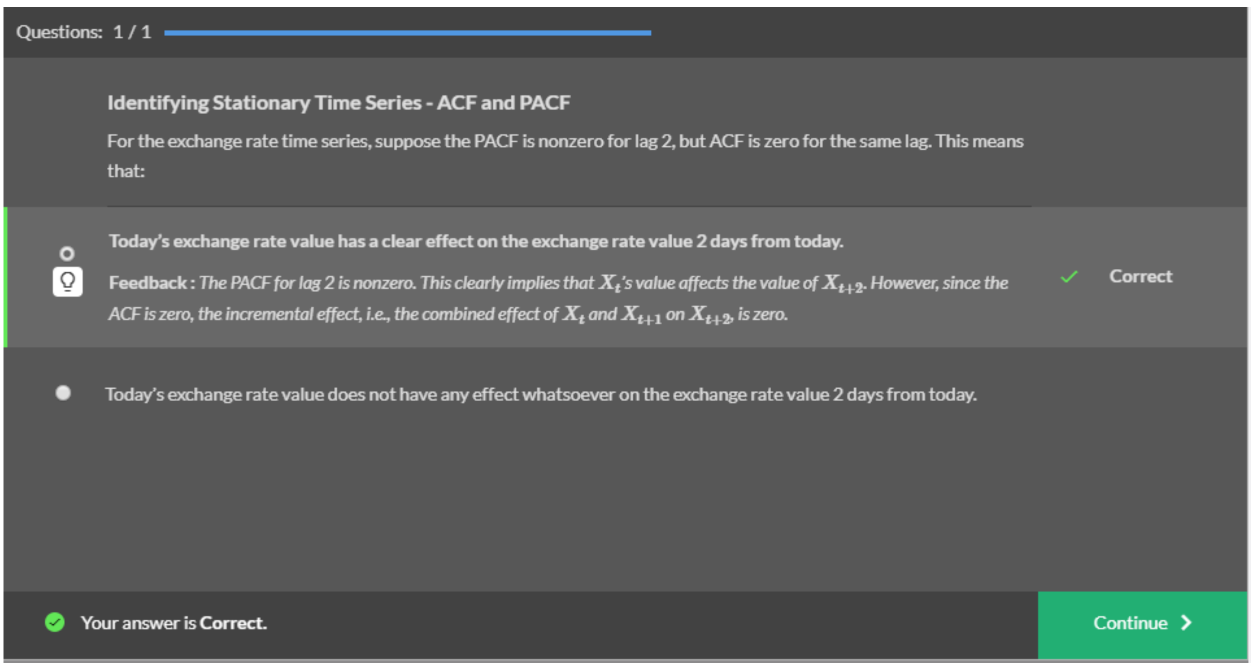
So if the correlation value is between the blue lines (that signify the upper and lower limits of the confidence interval), you can say that it isn't significantly different from zero; and hence, you can take it to be zero.

However, the autocorrelation function is not the only way to differentiate between white noise and a simple stationary series. You also have the partial autocorrelation function (PACF) method.
more_horizIncomplete.









Notice that the ACF looks at the correlation between values at timestamps that are 'h' unit steps away — say, XtXt and Xt+hXt+h. However, this correlation is a sum result of several '**partial**' **correlations**between Xt+hXt+h and other XiXis (such as Xt+1,Xt+2,Xt+3,Xt+1,Xt+2,Xt+3, …. up to Xt+h−1Xt+h−1). It may be useful to try '**isolating**' the **direct correlation** between just Xt+hXt+h and XtXt, without the influence of any of the intermediate values. This is how **PACF** functions.

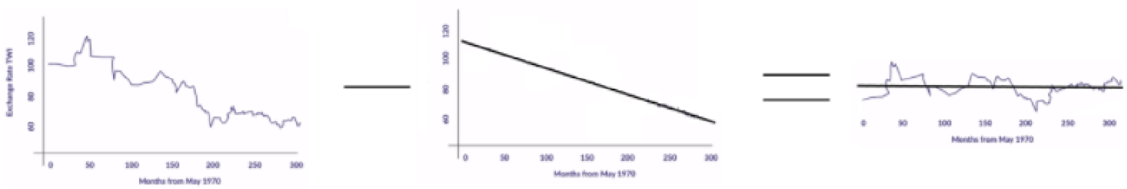
Now let's say you want to understand how the exchange rate in month 12 is related to the exchange rate in month 8. Here, when the ACF defines their relationship for you, it is actually the incremental effect, i.e. the effect of months 8, 9, 10, and 11 on month 12. However, PACF simply isolates the effect of month 8 out, and it tells you that the exchange rate in month 8 affects the exchange rate in month 12.

For a more elaborate and mathematical understanding of these concepts, please refer to section 4.1 of the lecture notes.

Modelling Stationary Time Series - I

By now, you have a basic understanding of white noise and the different tests you can carry out to confirm if a series given to you is, in fact, white noise.

Now, let's go back to the locally predictable pattern, i.e the weakly stationary time series. Remember the exchange rate example:



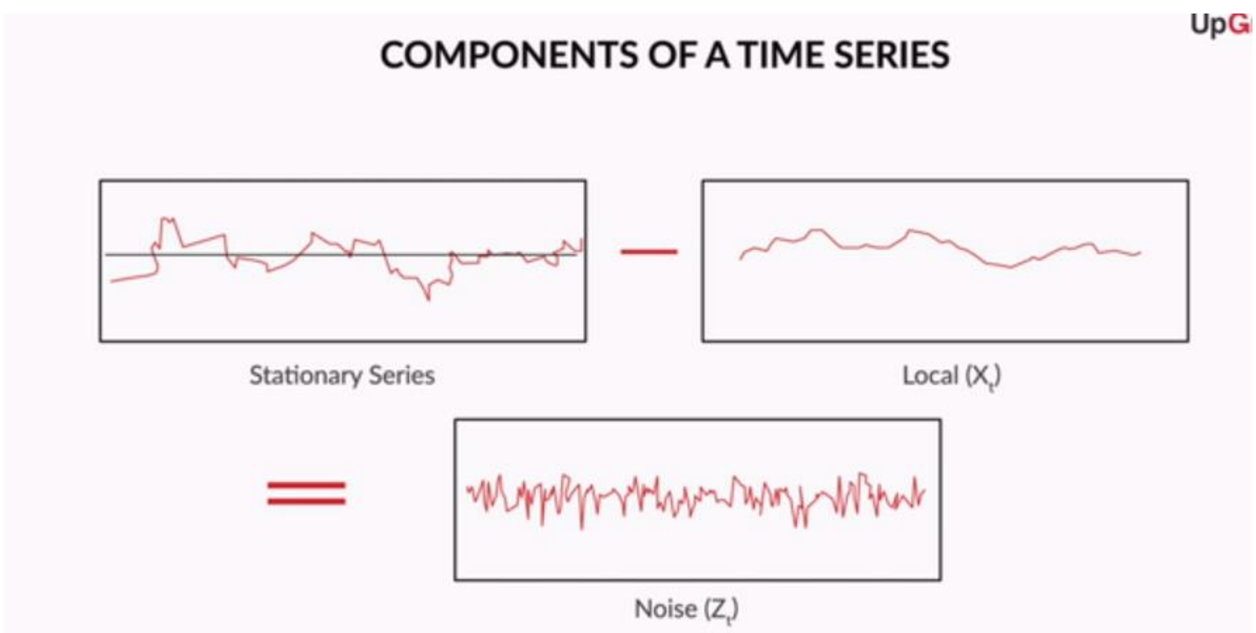
Intermediate Residue (Original Series - Global Predictions)

Once you model the global pattern of the data and subtract it from the original series, you should be left with a time series that has no trend and seasonality in it, i.e. a time series that is stationary. Now if it is stationary, it could be weakly stationary or strongly stationary. **Weak stationarity** would imply that the values of this time series **depend on past values** in a fixed manner. So there is **local predictability** of some kind that is present.

This raises two questions:

1. How can you **identify** if a given time series is weakly stationary or not?
2. If a time series is weakly stationary, how do you **find its equation**?

First, let's look at how the second question, i.e. modelling the weakly stationary series and finding the correct equation, is addressed.



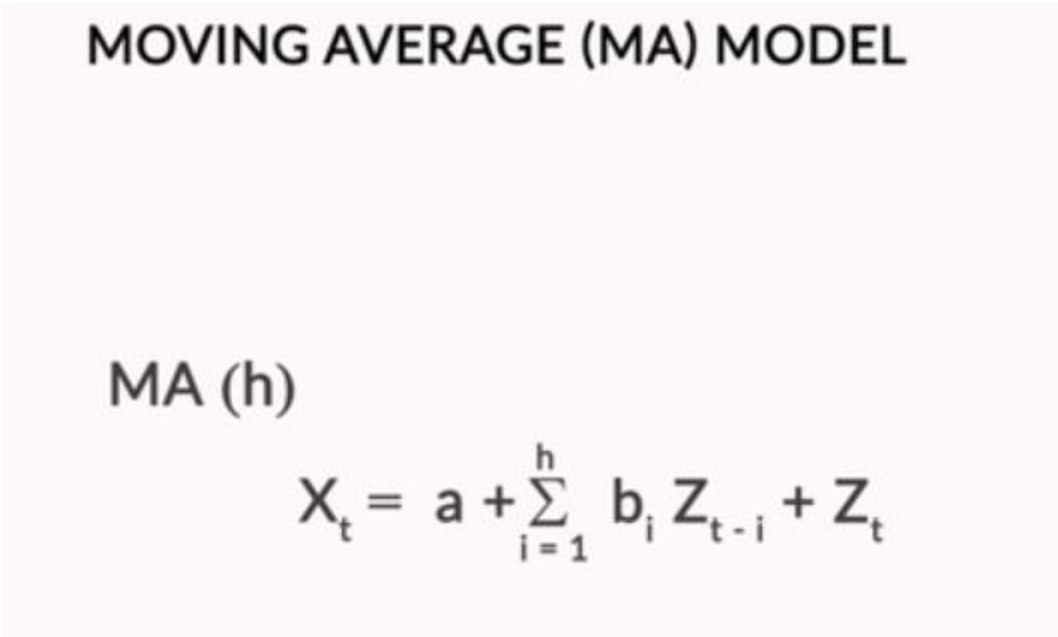
AUTOREGRESSIVE (AR) MODEL

AR $\boxed{(h)}$

$$X_t = a + \sum_{i=1}^{h} b_i X_{t-i} + Z_t$$

Time series analysis

1. Stationarity
2. White noise
   i. Histogram plot
   ii. Q-Q plot
   iii. Auto-correlation function
   iv. Partial Auto-correlation function
3. AR(h) model
4. MA(h) model



We have two basic types of time series models:

Autoregressive (AR) model

An autoregressive time series is one where the value at time 't' depends on the values at times (t-1)....(t-h) superimposed on a white noise term. You define an autoregressive time series AR(h) of the order 'h' as a series $\{X_t\}_0^T$ where $[X_t = \mu + \sum_{i=1}^{h} \alpha_i X_{t-i} + Z_t, \quad Z_t = \mathcal{W}(0, \sigma^2)]$ for some constants μ and $\alpha_i, 1 \leq i \leq h$. The coefficient α_i represents the influence (weight) of the value of the time series 'i' steps in the past, on the current value.

Moving average (MA) model

A moving average time series is one where the influence of the noise at some time step 't' carries over to the value at t+1, or possibly up to t+h for some fixed 'h'. Formally, a moving average time series $\{X_t\}$ of the order 'h' (denoted MA(h)) is $[X_t = \mu + Y_t + \sum_{i=1}^{h} \alpha_i Y_{t-i}]$, where $\{Y_t\} = \mathcal{W}(0, \sigma^2)$ for some constants μ and $\alpha_i, 1 \leq i \leq h$. Note that this gives you a process that is centred around the constant value μ . The value at time 't' in an MA(h) process is, therefore, the noise at the current time, t, superimposed on the cumulative weighted influence of the noise at 'h' previous timestamps.

Time series analysis

1. Stationarity
2. White noise
   i. Histogram plot
   ii. Q-Q plot
   iii. Auto-correlation function
   iv. Partial Auto-correlation function
3. AR(h) model
4. MA(h) model
5. ARMA(p,q) model

AUTO - REGRESSIVE - MOVING - AVERAGE (ARMA) MODEL

ARMA (p,q)

$$X_t = a + \sum_{i=1}^{L=p} b_i X_{t-i} + \sum_{i=1}^{i=q} c_i Z_{t-i} + Z_t$$

Having looked at MA(q) and AR(p) separately, you can now combine the two into a composite ARMA(p,q) model.

Autoregressive - Moving average (ARMA) model

A time series that exhibits the characteristics of an AR(p) and/or an MA(q) process can be modelled using an ARMA(p,q) model. Formally, you define an ARMA(p,q) process as a series $\{X_t\}$, where

$$[X_t = \mu + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i Z_{t-i} + Z_t].$$

Questions: 1 / 2

Modelling Stationary Time Series - I

Consider the following time series model:

$$Y_t = 0.3 + 0.5Z_{t-1} - 0.4Z_{t-2} + Z_t$$

This model can be represented as

- AR(2)
- MA(2)

  Feedback : *Notice that the model has a lag of the order of 2 and is dependent on error terms in the previous two timestamps. So it is an MA(2) model.*

  ✓ Correct
- ARMA(2,1)
- ARMA (1,2)

Your answer is **Correct.**

Continue

Modelling Stationary Time Series - II

In this segment, you will learn to identify the most **optimal values of p and q** corresponding to the AR(p) and MA(q) process from the ACF and PACF plots.

ACF & PACF PLOTS FOR ARMA PROCESSES

Process	ACF	PACF
AR (p)	Infinite waning exponential or sinusoidal tail	PAC = 0 for h > p
MA (q)	AC = 0 for h > q	Infinite waning exponential or sinusoidal tail
ARMA (p,q)	Like AR (p) for h > q	Like MA (q) for h > p

Questions: 2 / 2

Modelling Stationary Time Series - I

Consider the following time series model:

$$Y_t = 0.3 + 0.8Y_{t-1} + 0.5Z_{t-1} - 0.4Z_{t-2} + Z_t$$

This model can be represented as

- AR(2)
- MA(2)
- **ARMA(1,2)** — Correct

  Feedback : *Notice that the model has a lag of the order of 2 for the error term and a lag of the order of 1 for the value of the variable in the previous timestamp. So it is an ARMA(1,2) model.*
- ARMA (2,1)

Your answer is **Correct.**

Continue

Modelling Stationary Time Series - II

In this segment, you will learn to identify the most optimal values of p and q corresponding to the AR(p) and MA(q) process from the ACF and PACF plots.

ACF & PACF PLOTS FOR ARMA PROCESSES

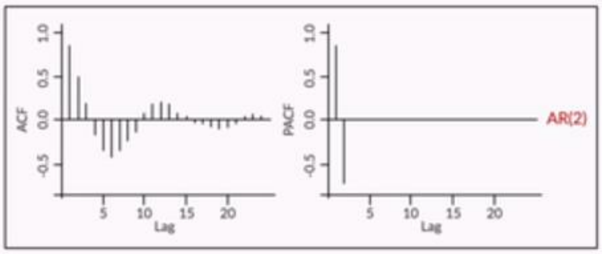
Process	ACF	PACF
AR (p)	Infinite waning exponential or sinusoidal tail	PAC = 0 for h > p
MA (q)	AC = 0 for h > q	Infinite waning exponential or sinusoidal tail
ARMA (p,q)	Like AR (p) for h > q	Like MA (q) for h > p

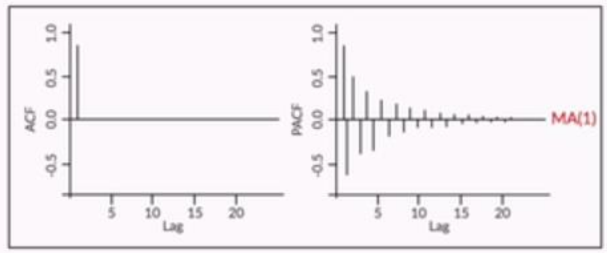
UpGrad

ACF & PACF PLOTS FOR ARMA PROCESSES

Incomplete

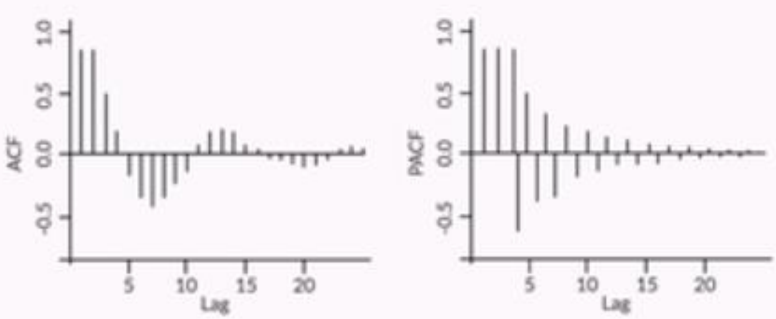
Process	ACF	PACF
AR (p)	Infinite waning exponential or sinusoidal tail	PAC = 0 for h > p
MA (q)	AC = 0 for h > q	Infinite waning exponential or sinusoidal tail
ARMA (p,q)	Like AR (p) for h > q	Like MA (q) for h > p



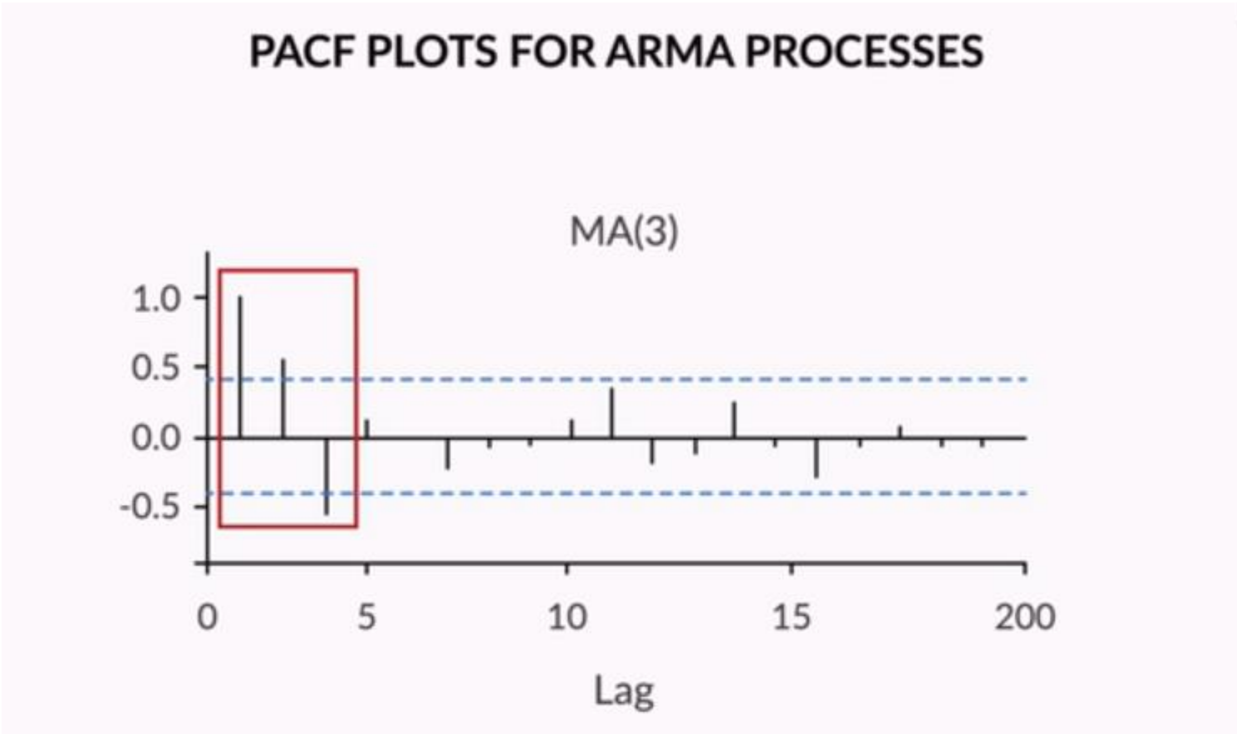


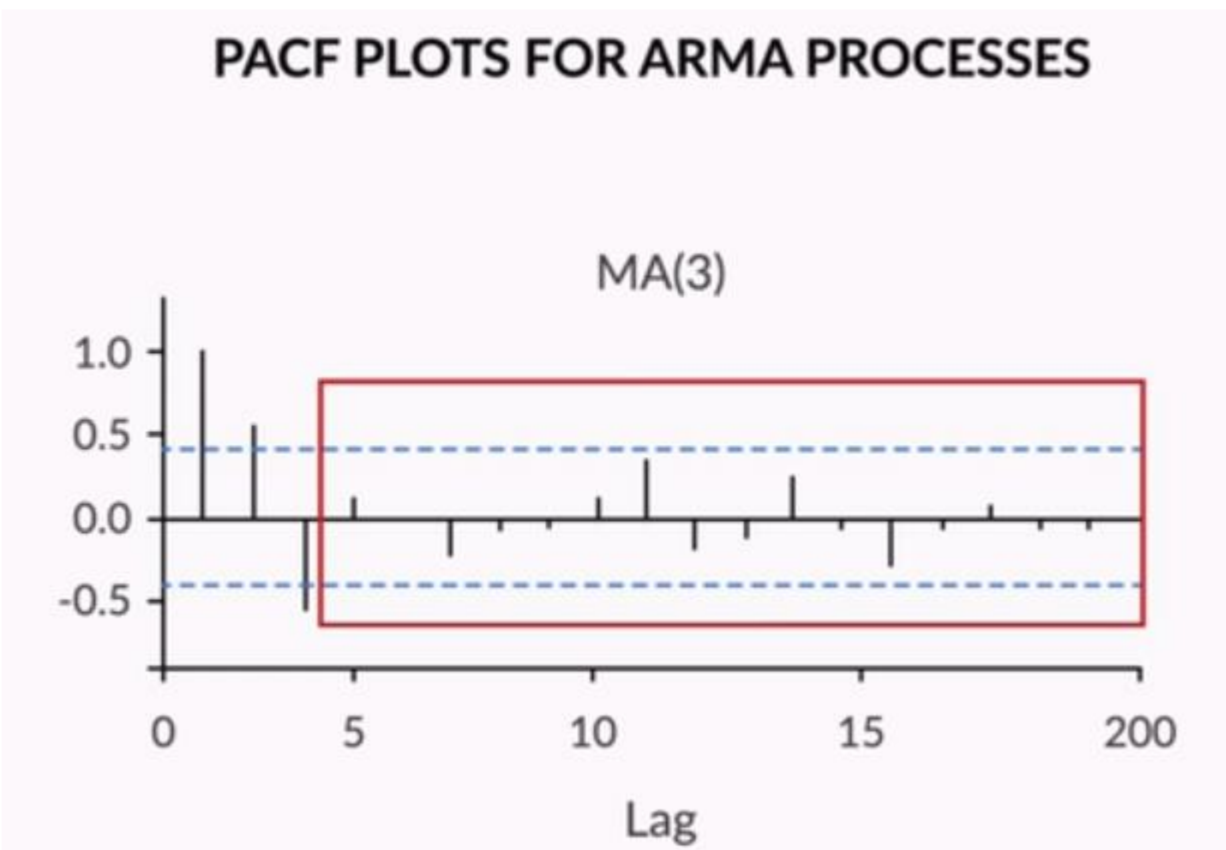
ACF & PACF PLOTS FOR ARMA PROCESSES

Process	ACF	PACF
AR (p)	Infinite waning exponential or sinusoidal tail	PAC = 0 for h > p
MA (q)	AC = 0 for h > q	Infinite waning exponential or sinusoidal tail
ARMA (p,q)	Like AR (p) for h > q	Like MA (q) for h > p



ARMA (2,1)


PACF PLOTS FOR ARMA PROCESSES
MA(3)
1.0
0.5
0.0
-0.5
0
5
10
15
200
Lag


PACF PLOTS FOR ARMA PROCESSES
MA(3)
1.0
0.5
0.0
-0.5
0
5
10
15
200
Lag

Therefore, you look for the cutoff value in the **PACF** plot for the most optimal p in the **AR(p)** model, and the **ACF** plot for q in the **MA(q)** process. For the ARMA(p,q) process, you need to find the cutoff lag values from both the ACF and the PACF plots. This can be summarised in the following table:

Theoretical Characteristics of the ACF and PACF plots for ARMA (p,q) Processes

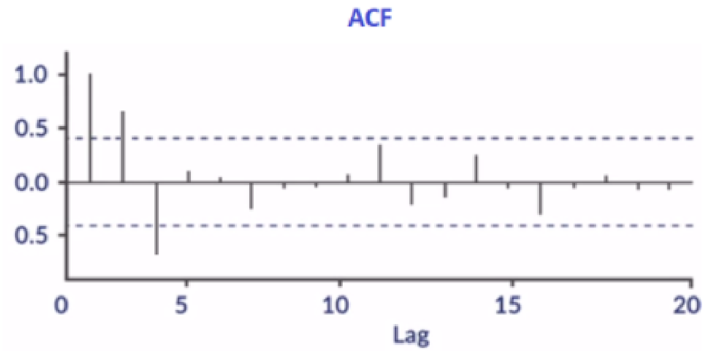
Process	ACF	PACF
AR (p)	Infinite waning exponential or sinusoidal tail	PAC = 0 for h > p
MA (q)	AC = 0 for h > q	Infinite waning exponetial or sinusoidal tail
ARMA (p,q)	Like AR (p) for h > q	Like MA (q) for h > p

Finding Optimal p & q

Questions: 1 / 2

Modelling Stationary Time Series - II

Consider the ACF plot below.



Which model will the given time series most likely follow?

- MA(5)

  Feedback : *Check the lag for which there is a drastic change in the absolute value of the ACF. Also, try to recall if the ACF is used to check the AR or MA model.*

  ✕ Incorrect

- MA(3)

  Feedback : *The picture in this question has three significant ACF peaks; after these, the ACF value lies well within the confidence interval. Hence, this information is consistent with an MA(3) process.*

  ✓ Correct

- AR(3)
- AR(5)

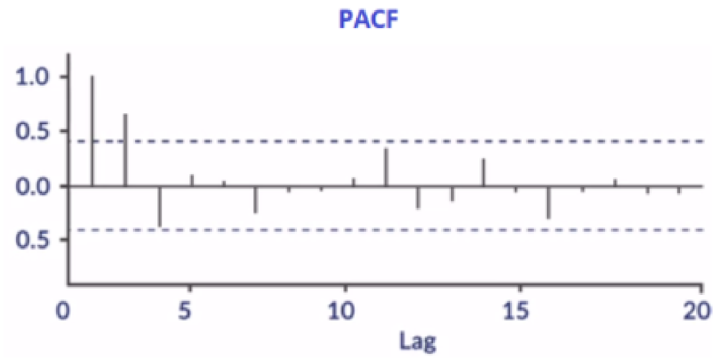
Your answer is **Incorrect.**

Continue

Questions: 2 / 2

Modelling Stationary Time Series - II

Consider the PACF plot below.



Which model will the given time series most likely follow?

- MA(2)
- MA(5)
- AR(2)

  Feedback : *The picture in this question has two significant PACF peaks (the fact that the second one is negative is irrelevant and simply indicates that the signs of the coefficients in the model are different from one another) in which the PACF value lies well within the confidence interval. Hence, this information is consistent with an AR(2) process.*

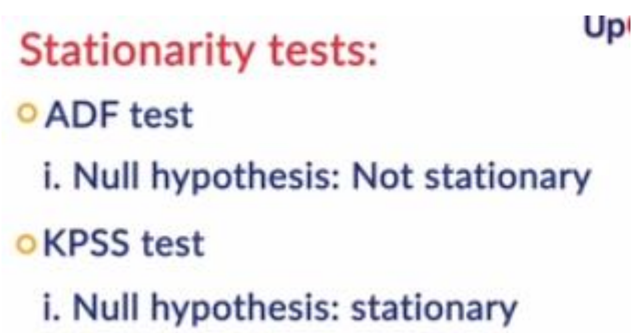
  ✓ Correct
- AR(5)

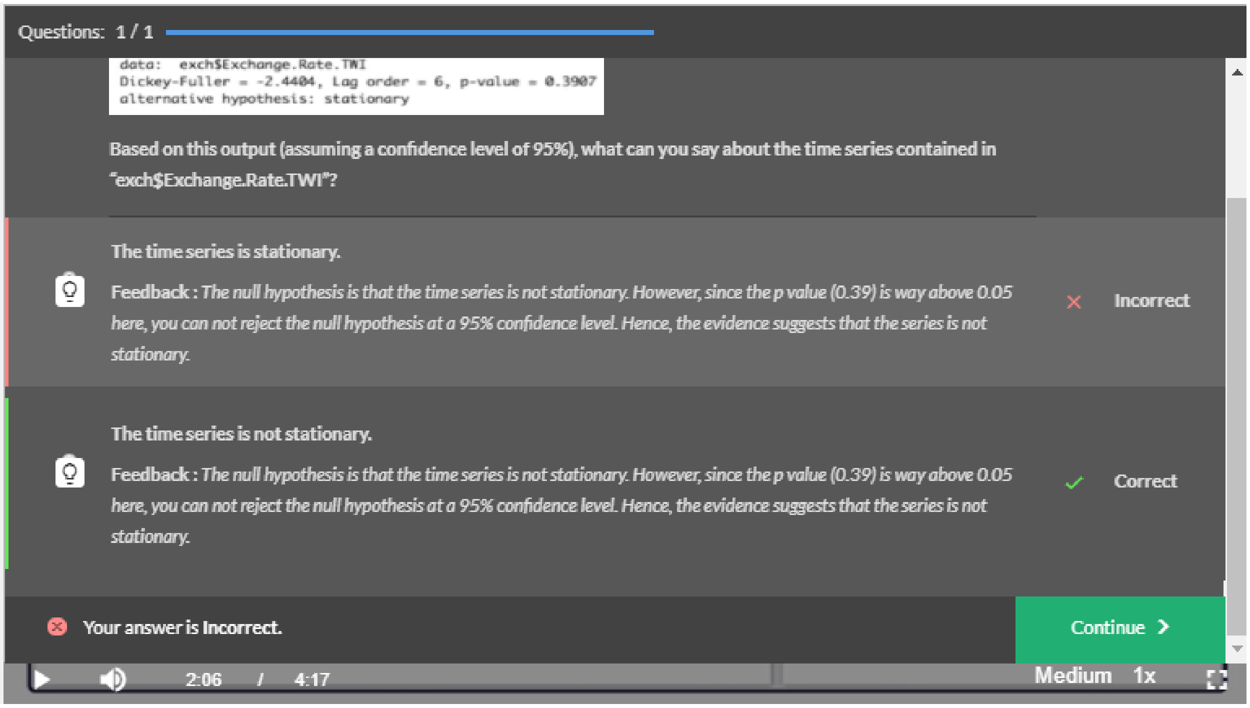
Your answer is **Correct.**

Continue ›

There are also some formal tests to check stationarity. But remember, these tests actually test for strong stationarity.

So let’s find out what these tests are.





So as you saw, there are two formal tests for checking the stationarity of a time series, both with opposite null and alternate hypotheses.

- **ADF test:** Null hypothesis assumes that the series is not stationary
- **KPSS test:** Null hypothesis assumes that the series is stationary

Here's a summary of all the tests you saw so far:

1. Tests for **normality (white noise/QQ plot)**: Tests for strong stationarity (white noise)
2. **ACF/PACF**:
    1. Test for **strong stationarity (white noise)**: ACF/PACF should not be significantly different from zero for non-zero lags
    2. Test for **weak stationarity (local predictability):** Check ACF/PACF plot patterns for patterns similar to the AR, MA, and ARMA processes
3. **Formal tests (ADF, KPSS):** Tests for strong stationarity (white noise)

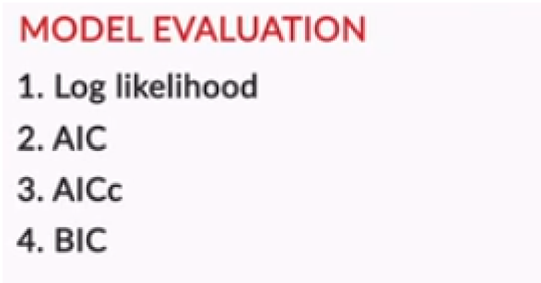
Finding the Best-Fit ARMA Model

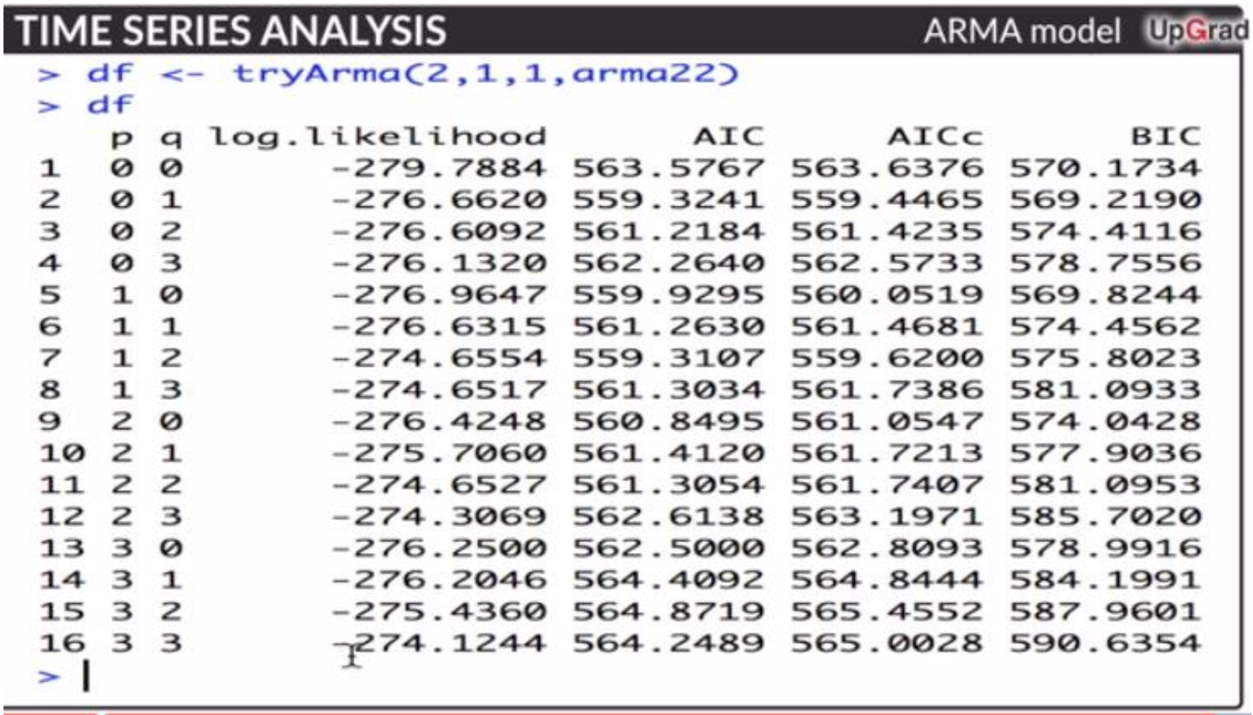
Let's suppose, from the ACF and PACF plot you concluded, that the model is most likely to follow an AR(2) or MA(3) process. But, is there any other way to optimise the value of p and q? Let's hear about this from Prof. Raghavan.

You can download the R-code file for this segment from the link given, and you can code along in R.

So what you did is, you simulated an ARMA process. Now, you're going to try different ARMA (p,q) combinations and compare them, based on parameters such as **log likelihood, AIC, AICc, and BIC**.

Based on these comparisons, you'll then decide on the best-fit p and q.






```
> df <- tryArma(2,1,1,arma22)
> df
```

	p	q	log.likelihood	AIC	AICc	BIC
1	0	0	-279.7884	563.5767	563.6376	570.1734
2	0	1	-276.6620	559.3241	559.4465	569.2190
3	0	2	-276.6092	561.2184	561.4235	574.4116
4	0	3	-276.1320	562.2640	562.5733	578.7556
5	1	0	-276.9647	559.9295	560.0519	569.8244
6	1	1	-276.6315	561.2630	561.4681	574.4562
7	1	2	-274.6554	559.3107	559.6200	575.8023
8	1	3	-274.6517	561.3034	561.7386	581.0933
9	2	0	-276.4248	560.8495	561.0547	574.0428
10	2	1	-275.7060	561.4120	561.7213	577.9036
11	2	2	-274.6527	561.3054	561.7407	581.0953
12	2	3	-274.3069	562.6138	563.1971	585.7020
13	3	0	-276.2500	562.5000	562.8093	578.9916
14	3	1	-276.2046	564.4092	564.8444	584.1991
15	3	2	-275.4360	564.8719	565.4552	587.9601
16	3	3	-274.1244	564.2489	565.0028	590.6354

A brief explanation of each of these measures is given below:

- **Log likelihood:** This refers to the log of the likelihood (probability) of the given data set being generated by the chosen model. The higher the likelihood, the better the fit of the model to the data. Note that under the Gaussian assumption, the model that maximises the likelihood also minimises the cumulative square error. Since probabilities have to be < 1, the log likelihoods are all negative.
- **Akaike information criterion (AIC):** This is a theoretical measure of the information in the model. It also takes into account the model complexity. In this case, as the order (p+q) of the model increases, the model becomes more complex. As you saw in the module on regression, in general, the more complex the model, the more prone it is to overfitting. Therefore, you try to pick a model that has an AIC that's as small as possible.
- **AIC corrected (AICc):** This is the AIC measure corrected for the size of the data set, i.e. in this case, the length of the time series 'n'.
- **Bayesian information criterion (BIC):** This is a measure that's similar to the AICc; however, it's arrived at using Bayesian methods. Again, the lower the BIC measure, the better the fit of the model.

Note that in all these measures, the common pattern is

- The higher the likelihood, the lower the measure, and
- The higher the model complexity, the higher the measure

These measures are not always entirely consistent with each other, and some amount of judgement (and trial) is required to pick the right model. In the example above, you could pick either the (0,2) model or the (1,1) model. Both have the same cumulative order, but order 2 for MA seems worse (longer range dependence and hence, more complex) than order 1 in both the AR and MA. So you finally pick (1,1).

Model Evaluation - MAPE

Until now, you learnt how to model the globally predictable part (regression) and the locally predictable part (ARMA fitting). Now you're in a position to make a prediction for the time series by combining both the globally predictable and the locally predictable part.

However, once you’ve made the predictions, you will naturally need to evaluate the model. For that, one measure that is used widely in the industry is the MAPE (mean absolute percentage error).
more_horizIncomplete

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

So the MAPE (mean absolute percentage error) is given by

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|.$$

It is **similar to R-squared but still slightly different.** Some pros and cons of using the MAPE are

1. R-squared punishes big deviations very strictly compared to the MAPE.
2. The MAPE will not perform well if one of the actual data points is equal to zero.
3. It is difficult to interpret as a percentage.
4. The MAPE is known to favour models that consistently predict lower values. This may introduce bias.

Questions: 1 / 1

Model Evaluation - MAPE

The demand for a product for the last six months (t=1 to 6) has been 15, 15, 17, 18, 20, and 19. The manager wants to predict the demand for this time series using the following simple linear trend equation: d = 12 + 2t. What is the MAPE for the given data set?

- 10.2%
- **11.0%**

  Feedback : *Remember that the MAPE is calculated as follows: the difference between the actual and predicted value of the dependent variable. The result is divided by the actual value. The absolute value in this calculation is summed up for every forecasted point in time, and it is divided by the number of data points or 'n'. Multiplying by 100 makes it a percentage error. On substituting the values for the given case, you will get the MAPE as 11.1%.*

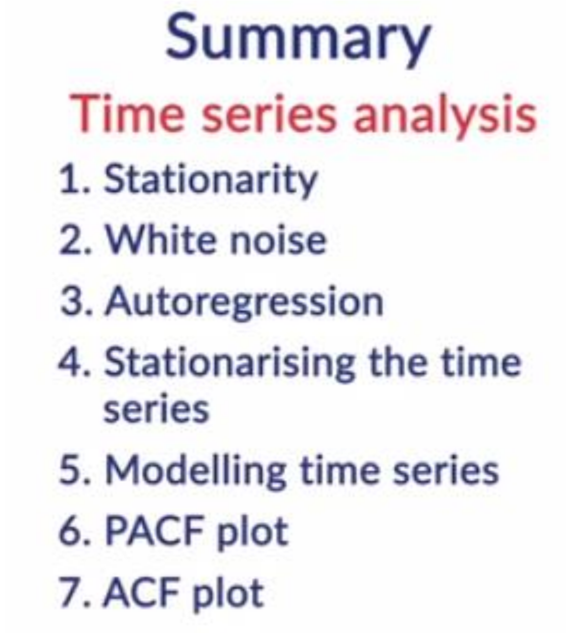
  ✓ Correct
- 2.6%
- 5.3%

Your answer is **Correct.**

Summary

You covered a lot of conceptual ground in this session. Let's revisit some of those topics:

1. Stationarity
2. The two components of a stationary time series: white noise and the autoregressive part
3. ACF and PACF functions and their plot
4. Making a time series stationary
5. Modelling a stationary time series:
    1. AR model
    2. MA model
    3. ARMA model
6. Evaluating a time series model



ARIMA Modelling and Time Series Smoothing

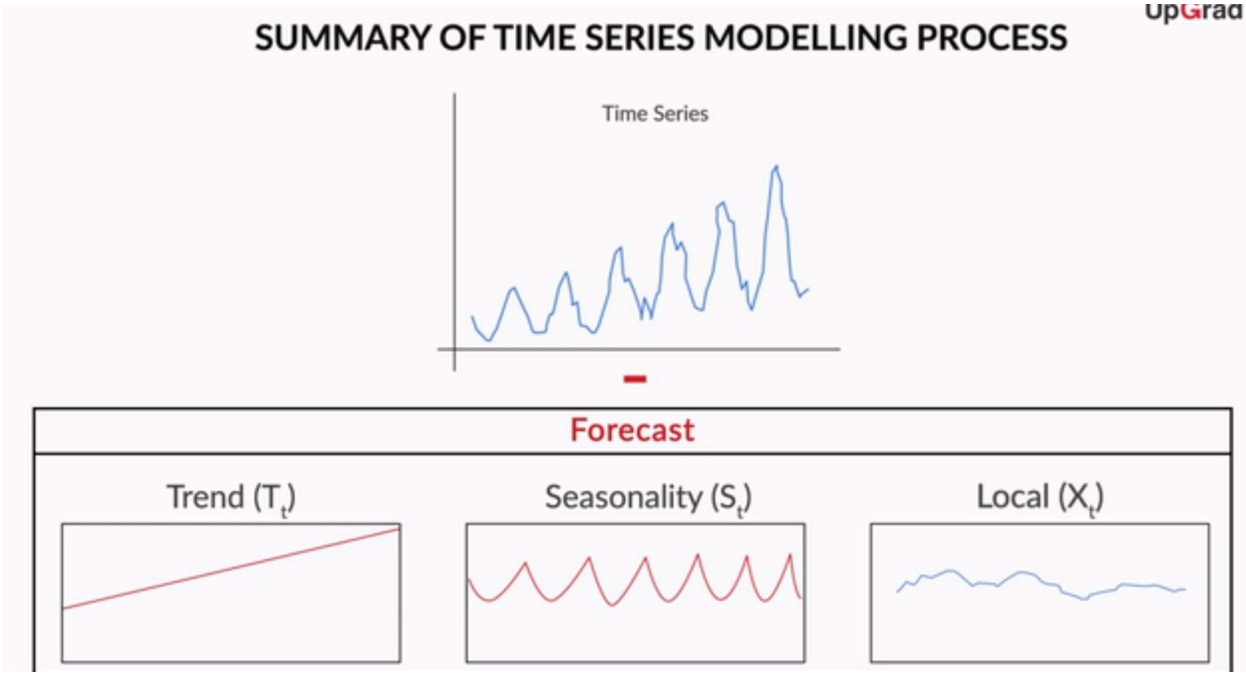
Introduction

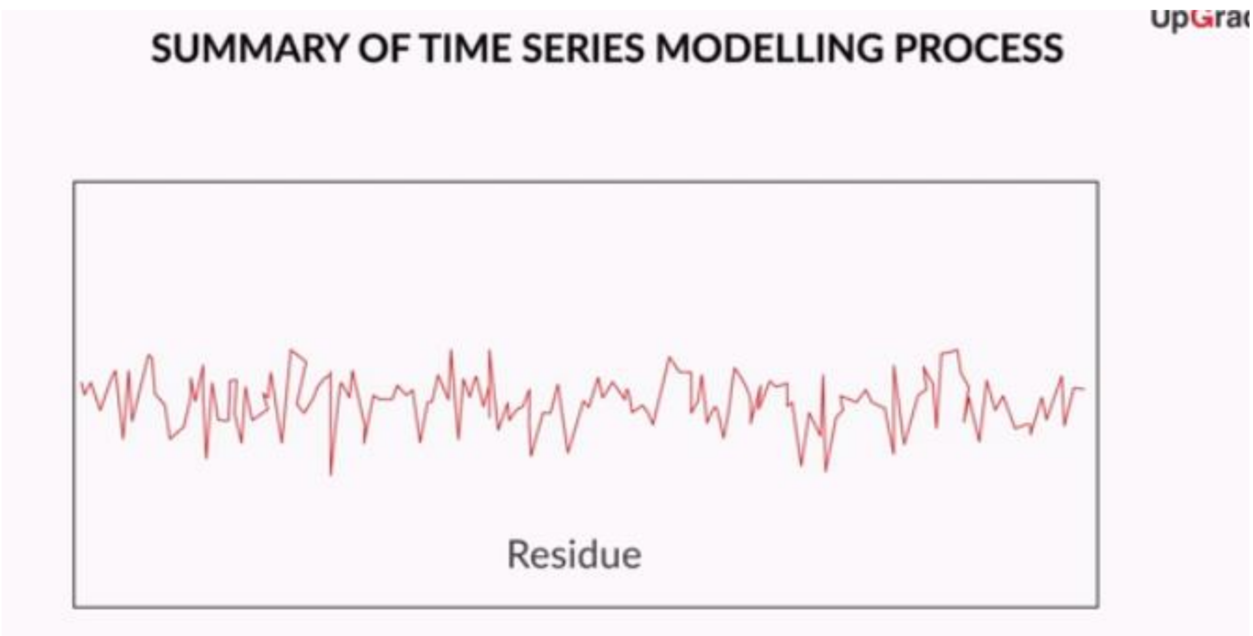
Welcome to the third session on 'Time Series Analysis'.

In the previous sessions, you learnt the basics of time series. You also learnt how to model and evaluate a time series by making it stationary. To make the time series stationary, you removed its trend and seasonality by modelling them. Then, you could model the remaining (weakly) stationary portion using ARMA processes.

In this session

You will look at another approach that can be used to make the time series stationary. You can use the in-built **ARIMA function**, which uses the method of **differencing** to make the data stationary, before modelling it as an ARMA process.


upGrad
SUMMARY OF TIME SERIES MODELLING PROCESS
Time Series
Forecast
Trend (T_t)
Seasonality (S_t)
Local (X_t)


upGrad
SUMMARY OF TIME SERIES MODELLING PROCESS
Residue

Questions: 1 / 1

original series. We'll call this subtracted series, the residual series.

If the residual series is white noise, then the modelling part has been done correctly. However, if the residual series is not white noise, then that means that:

- The global pattern has not been identified correctly. The local pattern though, has been identified and modelled correctly.
- The local pattern has not been identified correctly. The global pattern though, has been identified and modelled correctly.
- Either the local or the global pattern has not been identified correctly. We cannot yet tell which one that is.

Feedback : *If the residual series is not white noise, it is not unpredictable. This clearly means that there are some predictable patterns that are still left in the time series. Had the modelling part been done correctly, all the predictive patterns would have been captured in earlier steps, in either the global pattern or the local pattern.*

✓ Correct

Your answer is Correct.

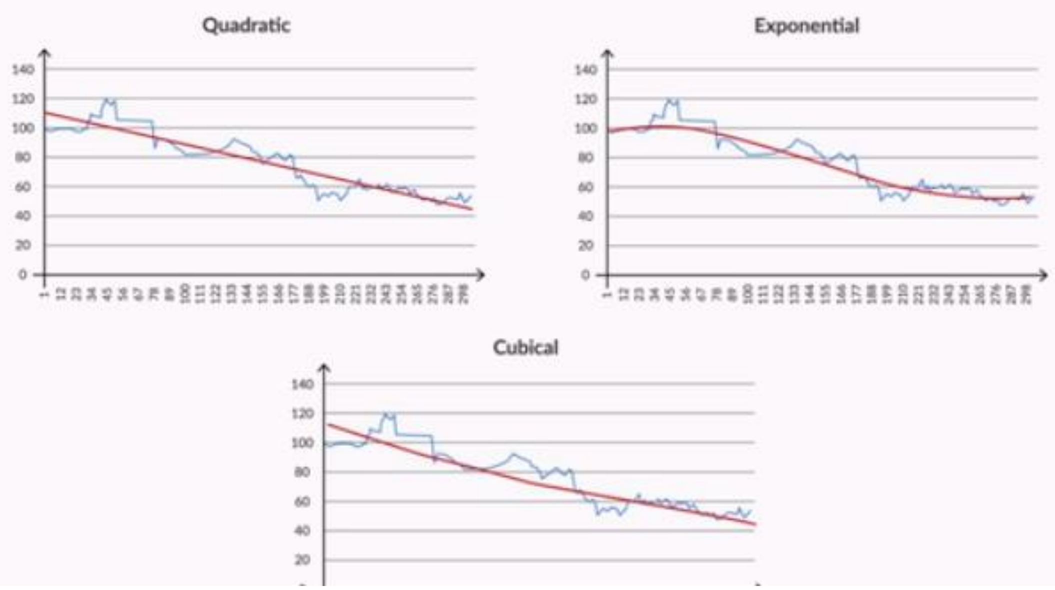
Continue

1:51 / 4:20 Medium 1x

TIME SERIES MODELLING

1. Classical decomposition
2. ARIMA

CHOICE OF FUNCTION FOR MODELLING TREND



In previous sessions, you saw how a time series can be made stationary by removing its trend and seasonality. The stationary series that you would be left with would then be modeled using ARMA processes. That was the classical decomposition approach for modeling the time series.
Now, you will look at another approach that can be used for making the time series stationary. This approach is called ARIMA modeling. It makes the time series stationary through a particular algorithm called differencing, after which, it can once again be modeled as an ARMA process.
So let's look at what the algorithm for differencing is.

Differencing

Differencing

$$x_1, x_2, \text{-------------} \; x_n$$

$$y_1, y_2, \text{-------------} \; y_{n-1}$$

$$y_2 = x_2 - x_1 \, , \; x_1 = y_1$$

$$y_3 = x_3 - x_2$$

$$y_4 = x_4 - x_3$$

Questions: 1 / 1

Time Series Differencing - I

Here, you have a time series that is to be differenced in order to make it stationary.

Time stamp	Original Series (x)	Differenced Series (y)
1	0.14	0.14
2	0.28	-
3	0.41	-
4	0.57	-

The values of the difference time series for time stamps 2, 3 and 4 are:

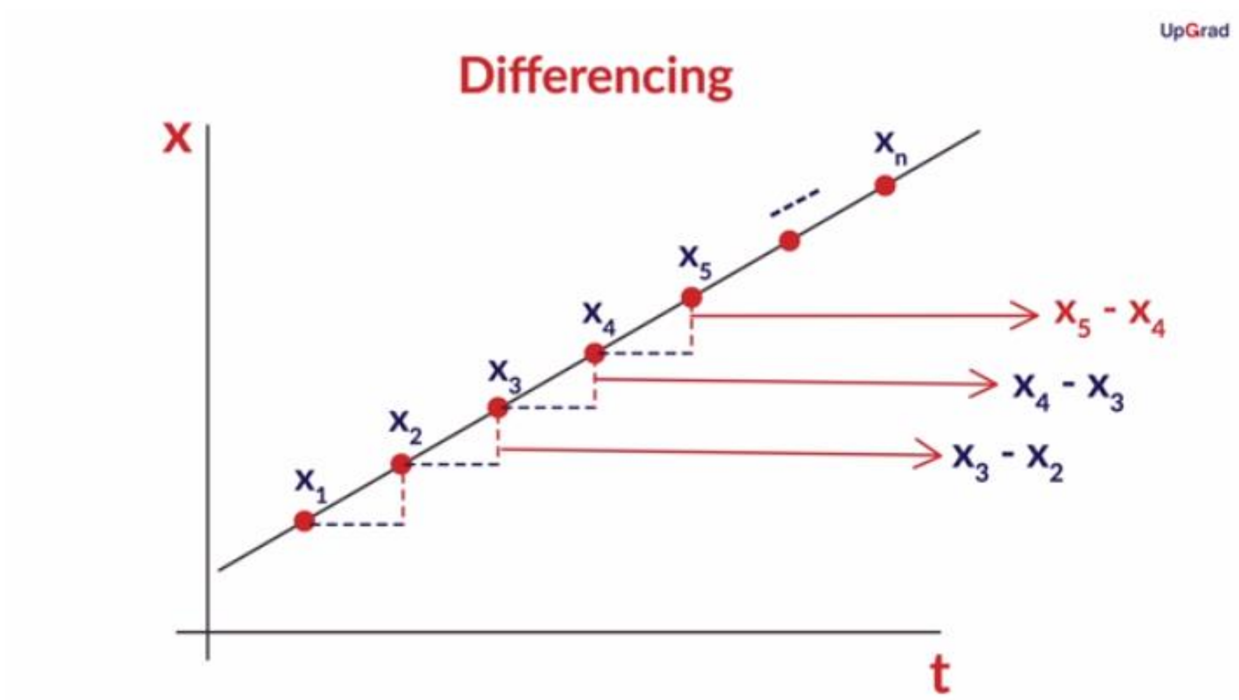
- $y_2 = 0.14, y_3 = 0.13, y_4 = 0.16$ ✓ Correct

  Feedback : *The second term of the differenced series, i.e. y_2 is given by $x_2 - x_1$. Also, $y_3 = x_3 - x_2 and y_4 = x_4 - x_3$. Putting in these values, you will get $y_2 = 0.14, y_3 = 0.13, y_4 = 0.16$.*

- $y_2 = 0.13, y_3 = 0.16, y_4 =$ not enough information
- $y_2 = 0.13, y_3 = 0.16, y_4 = 0$

Your answer is Correct.

Continue >



So, differencing is another method that can be used to make your model stationary. For example, suppose you get a stationary series after one level of differencing. When you model this differenced series, it may turn out to be an ARMA (2,2) process.

But one question that could arise is, how do you use this model to make predictions? Well, first you predict for the differenced series. In the example above, the differenced series is an ARMA (2,2) process. Once you've predicted the values for the differenced series, you then get the original series' predictions by 'reverse differencing' the time series.

Questions: 1 / 2

Time Series Differencing - I

Suppose you are using a time series analysis to forecast the exchange rate.

The time series given to you for this analysis, when differenced once, becomes a stationary series. This stationary series can be modelled as an ARMA (2,2) process.

Now, your predictions for this time series will be

Timestamp	Original Series Predictions (x)	Differenced Series Predictions (y)
1		0.08
2		0.05
3		0.02
4		-0.09

- x_1 = cannot be found, $x_2 = 0.13, x_3 = 0.15, x_4 = 0.06$
- x_1 = cannot be found, $x_2 = 0.08, x_3 = 0.13, x_4 = 0.15$
- $x_1 = 0.08, x_2 = 0.13, x_3 = 0.15, x_4 = 0.06$

  Feedback : *The first term,* $x1$*, will simply be equal to* y_1*. It will give you* $x_1 = 0.08$*. For the second term, you can use* $x_2 - x_1 = y_2$*, which gives you* $x_2 = 0.08 + 0.05 = 0.13$*. Similarly, you can find* $x_3(0.15)$ *and* $x_4(0.06)$*.*

  Correct
- $x_1 = 0.13, x_2 = 0.15, x_3 = 0.09, x_4$ = not enough information

Your answer is **Correct.**

Continue

Questions: 2 / 2

Time Series Differencing - I

Suppose the time series becomes stationary after two differencing steps, not one.

Now your prediction for timestamp 2 will be

Timestamp	Original Series Predictions (x)	Level 1 Differenced Series Predictions (y)	Level 2 Differenced Series Predictions (y')
1			0.08
2			0.05
3			0.02
4			-0.09

{Here, the level 1 differenced series means the series obtained after differencing the original series, and the level 2 differenced series means the series obtained after differencing the level 1 differencing series.}

- 0.23
- **0.21**

  Feedback : *For the level 1 series, y_1 is simply equal to y'_1, i.e. 0.08. For y_2, you can use the equation $y_2 - y_1 = y'_2$. This will give you $y_2 = y_1 + y'_2$, i.e. 0.13. Similarly, you can find that x_1 = 0.08 and x_2 = 0.08 + 0.13 = 0.21.*

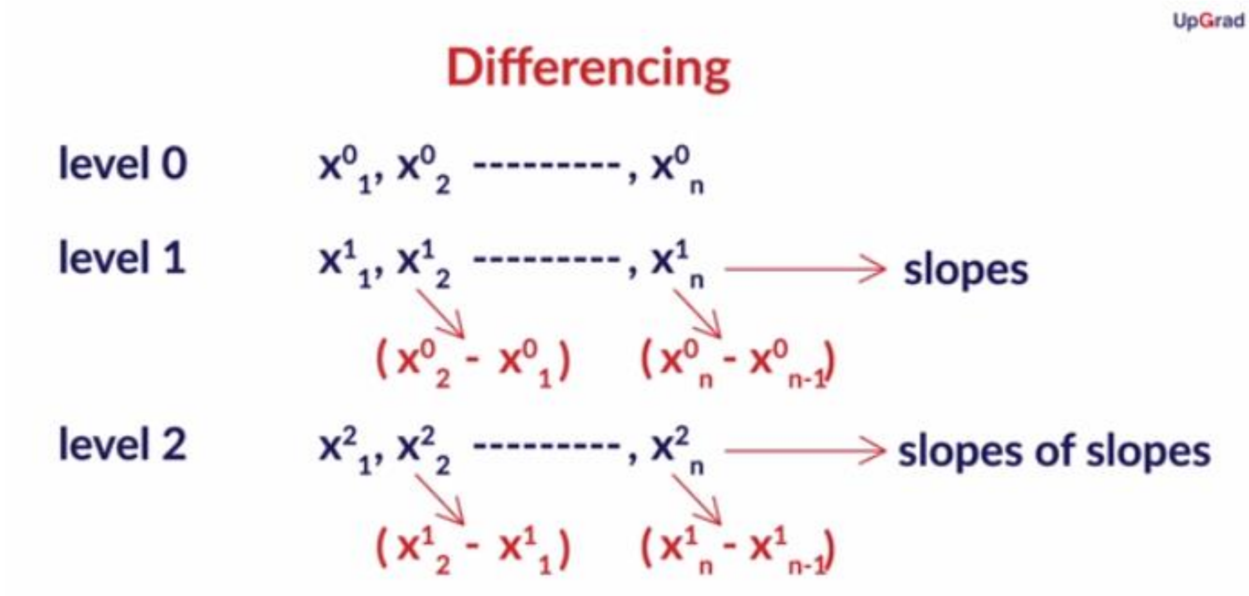
  Correct
- 0.19
- 0.17

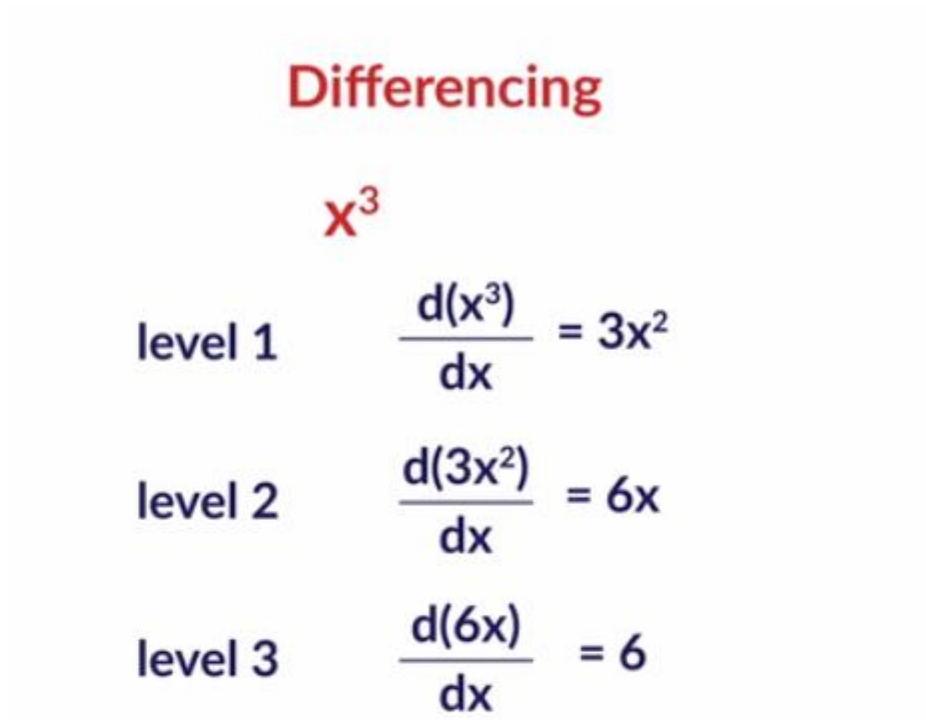
Your answer is **Correct.**

Continue

Time Series Differencing - II

So far, you've briefly seen how differencing can be used as an alternative method to make time series forecasts. However, an important assumption that arises during time series differencing is that differencing the time series a few times will result in it becoming stationary. Why would that happen? What if it doesn't happen? What do you do then?





So that's a broad intuition behind why differencing works. In a way, you can say that differencing is like differentiating. So a few levels of differencing will give you a de-trended (weakly) stationary series.

However, because of the same reason, you can say that this method of differencing will only work efficiently for series that can be approximated to low degree polynomials, such as

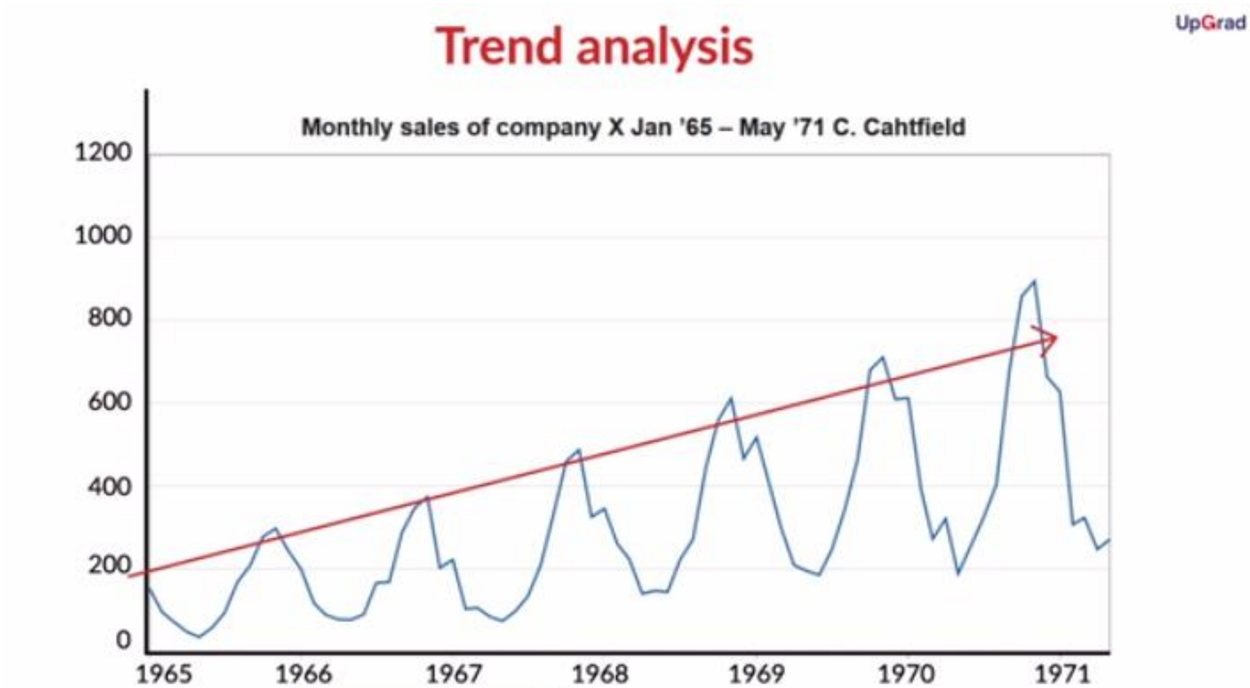
ax2+bx+c,ax3+bx2+cx+dax2+bx+c,ax3+bx2+cx+d, etc. A sinusoidal series, for example, will not effectively become constant after two or three levels of differencing, and hence, cannot be modelled effectively using differencing.

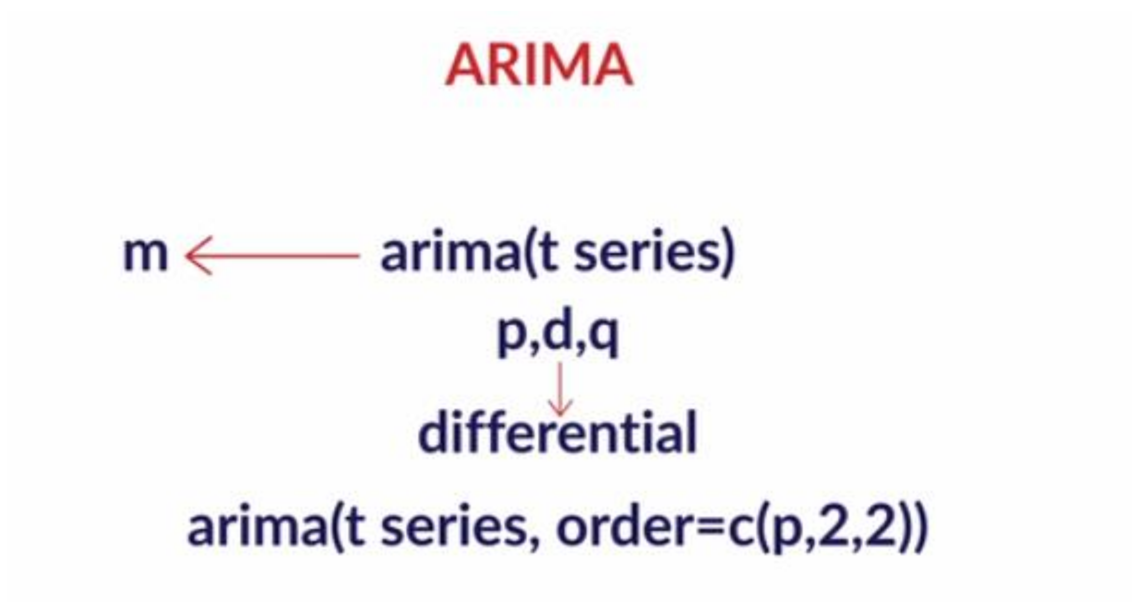
ARIMA - Difference the series. If the differenced series is stationary, model it as an ARMA process.

Classical Decomposition - Remove the trend and seasonality of the time series by modelling them. If the de-trended and de-seasonalised series is stationary, model it as an ARMA process.

Let’s now listen to Prof. Raghavan as he goes through the pros and cons of each of these methodologies.

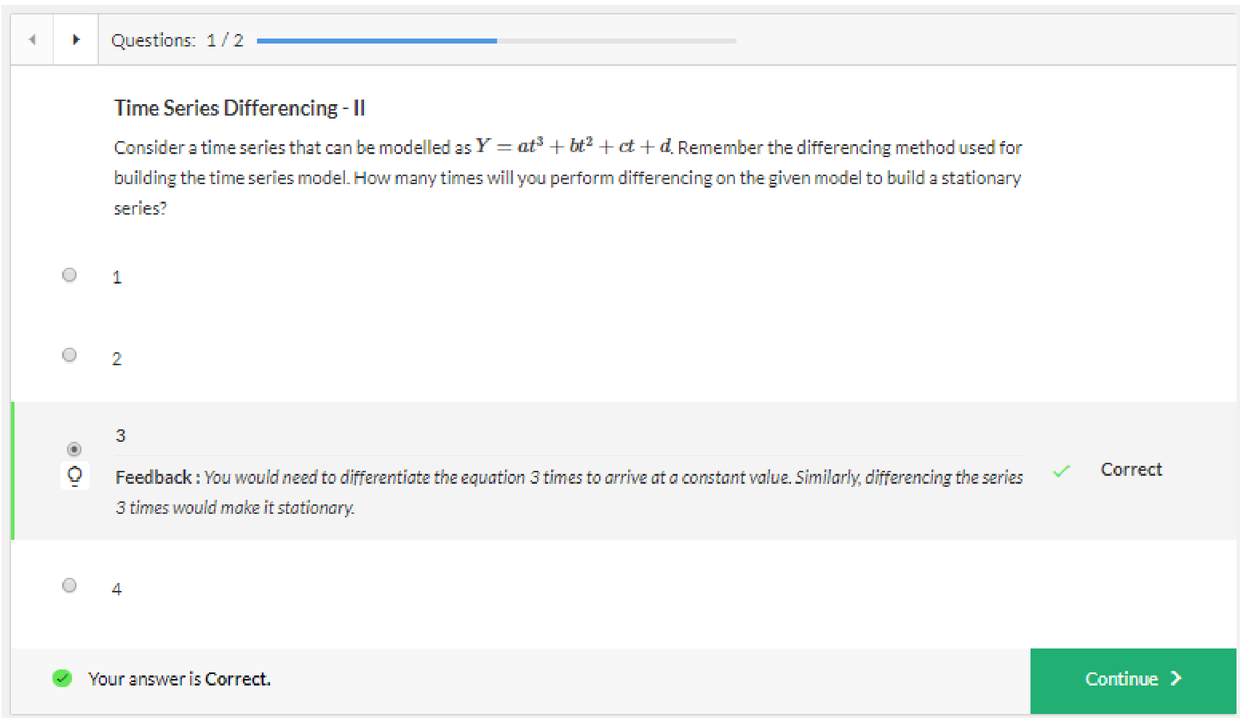
So now you know two methods to forecast a time series:





Let's revisit the pros and cons of the two methods discussed above:

- **Classical Decomposition**: The obvious downside of this methodology is that you need to manually 'guess' a good trend and seasonal pattern in the data. This is often not such a big issue, though. On visualisation, most series you encounter in practice reveal such trends rather quickly.
- **Differencing**: This method can easily be automated, and that removes the onus of guessing an appropriate trend/seasonal pattern from the analyst. An important drawback of the differencing approach is that though the differencing eventually does remove non-stationarity and produce a good fit, it is hard to visualise the trend and seasonality implicit in the data. An explicit visualisation of such trends and seasonality can be critical for several applications, which is missed out in this approach.



Questions: 2 / 2

Time Series Differencing - II

Let's say that you are analysing a time series called T. This time series reduces to a stationary time series T' after 3 differencing steps. Suppose the stationary time series T', that you get after these 3 differencing steps, is an ARMA (1,2) process.

What kind of model does the original time series, T, follow?

- ARIMA (1, 2, 3)
- **ARIMA (1, 3, 2)**

  Feedback : *Remember that an ARIMA model is defined by the parameters p, d and q, where d represents the number of differencing steps it takes to make the time series stationary, and p and q represent the order of the AR and MA components of the stationary differenced series. Hence, this is an ARIMA (1, 3, 2) model.*

  ✓ Correct
- ARIMA (3, 1, 2)
- ARIMA (3, 2, 1)

Your answer is **Correct.**

Continue

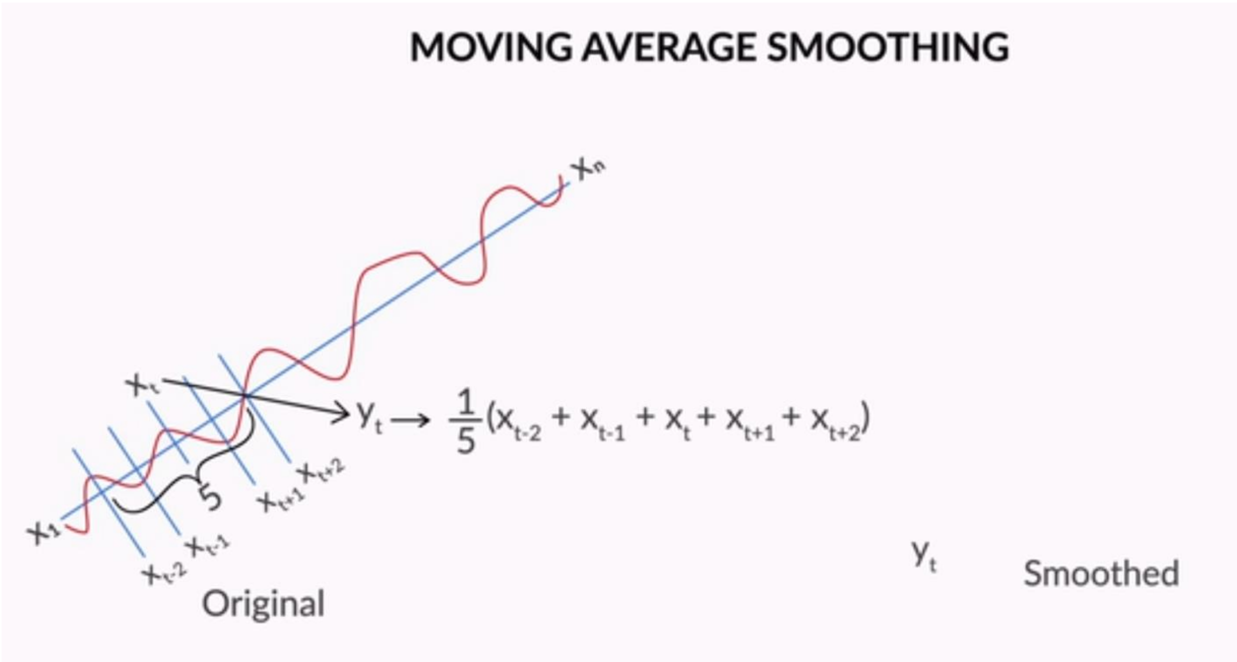
Time Series Smoothing

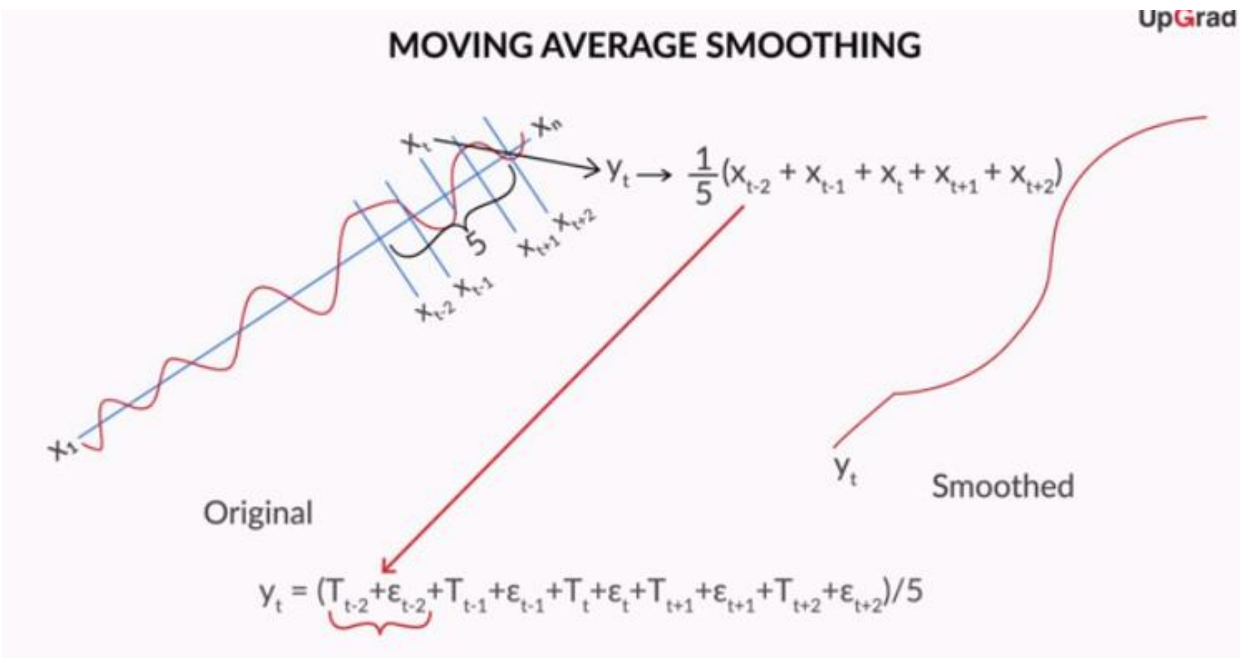
You would have noticed, from the examples of time series shown earlier, that visually 'spotting' a trend or seasonality is made unclear by the local 'spikiness' of the data; the 'spikiness' could be due to noise or a very strong autoregressive behaviour. From earlier examples, you have also seen that visually autoregressive data is hard to distinguish from pure noisy data.

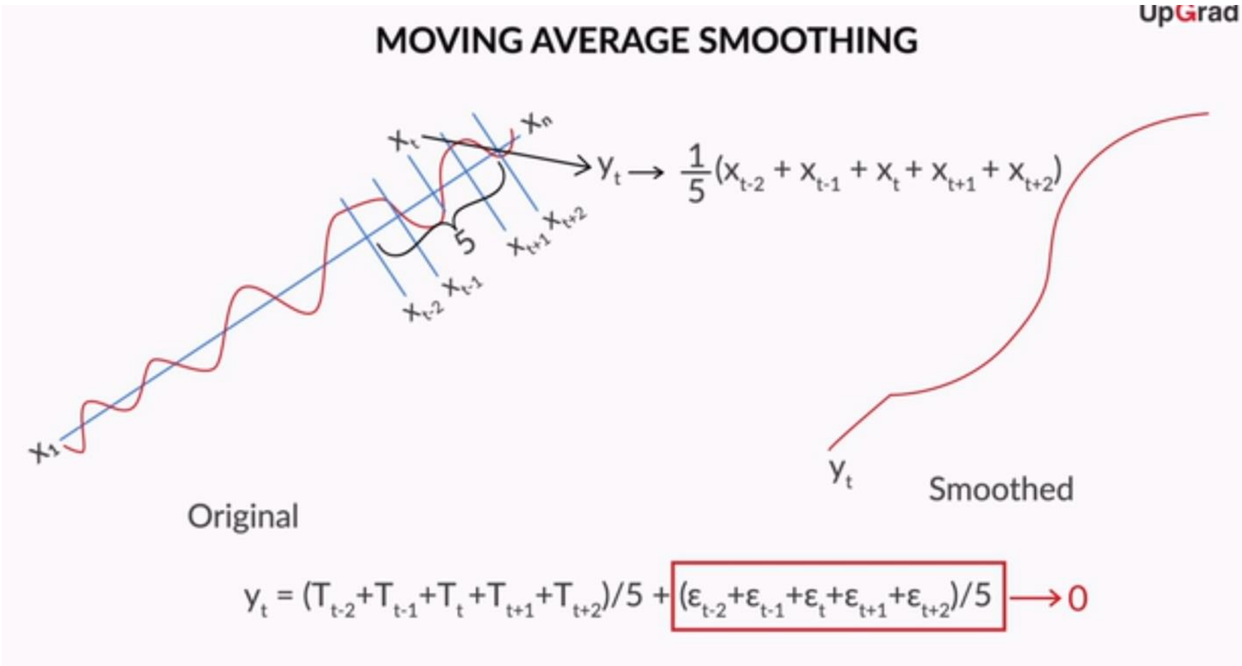
Smoothing is the process of making the curve smoother by 'averaging' out the noise to make the trend and seasonality more apparent. We will explore a few common ways of smoothing, in this segment.

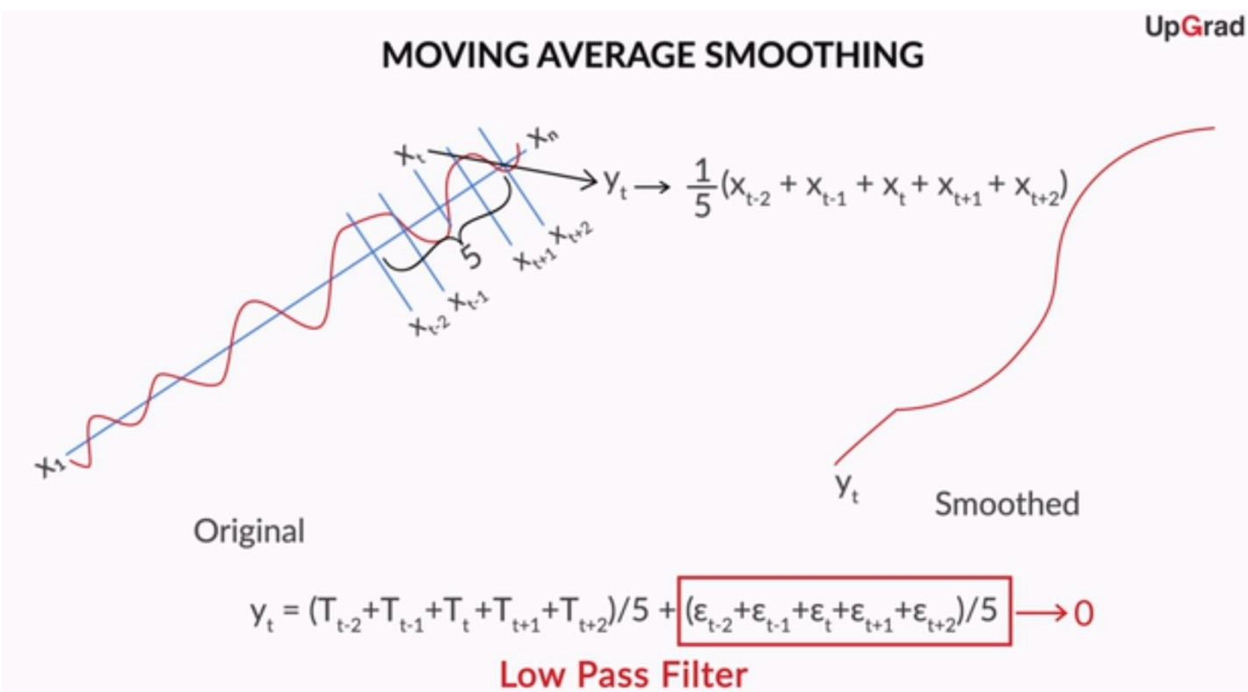
SMOOTHING

1. Getting rid of the noise


MOVING AVERAGE SMOOTHING
x_n
x_t
$y_t \rightarrow \frac{1}{5}(x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})$
5
x_1
x_{t-2} x_{t-1} x_{t+1} x_{t+2}
y_t
Smoothed
Original


UpGrad
MOVING AVERAGE SMOOTHING
x_t
x_n
$y_t \rightarrow \frac{1}{5}(x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})$
5
x_{t-2} x_{t-1} x_{t+1} x_{t+2}
x_1
y_t
Smoothed
Original
$y_t = (\underbrace{T_{t-2}+\varepsilon_{t-2}}+T_{t-1}+\varepsilon_{t-1}+T_t+\varepsilon_t+T_{t+1}+\varepsilon_{t+1}+T_{t+2}+\varepsilon_{t+2})/5$





So, a **moving average smoothing** algorithm with a window size of say, 5, would replace the original value x_t with y_t, where $y_t = (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})/5$.

Questions: 1 / 1

Time Series Smoothing

The following table contains the sale of a product in a department store for the first 6 months of the last year.

Month	Sale (in 100s)
Jan	36
Feb	45
Mar	81
Apr	90
May	108
Jun	144

Consider a moving average smoothing method with a filter size of 3. Use the MA method as discussed in the lecture (i.e. consider the average to be achieved in both directions). Find the smoothed value of sales for March, April and May.

- 72, 54, 93
- 72 ,66, 57
- **72, 93, 114**

  Feedback : *For April, the 3-month moving average smoothed value will be an average of (81 + 90 + 108)/3 = 93. Similarly, the smoothed value can be calculated for other months.*

  Correct
- 57, 66, 72

Your answer is **Correct.**

Time series analysis

1. Differencing
2. Classical decomposition
   i. Moving average smoothing
   ii. Exponential smoothing

Exponential Smoothing

Smoothing techniques

$$Y_t = \alpha . X_t + (1 - \alpha)\, Y_{t-1}, \quad Y_1 = X_1$$

Smoothing techniques

$$Y_t = \alpha . X_t + (1 - \alpha)\, Y_{t-1}, \quad Y_1 = X_1$$

$$\alpha = 0.3$$

$$(X_t) \longrightarrow Y_t$$

$$Y_{t-1}$$

Note: Higher the values of alpha, less the smoothing and vice – versa

$$Y_t = \alpha.X_t + (1 - \alpha)\,Y_{t-1}, \quad Y_1 = X_1$$

$$\alpha = 0.3$$

$$(X_t) \longrightarrow Y_t$$

$$Y_{t-1}$$

$$\begin{aligned} Y_t &= \alpha.X_t + (1 - \alpha)\,Y_{t-1} \\ &= \alpha.X_t + (1 - \alpha)(\alpha.X_{t-1} + (1 - \alpha)Y_{t-2}) \\ &= \alpha.X_t + \alpha(1 - \alpha).X_{t-1} + (1 - \alpha)^2 Y_{t-2} \\ &= (1 - \alpha)^{t-1}X_t + \sum_{i=0}^{t-2} \alpha(1 - \alpha)^i.X_{t-1} \end{aligned}$$

The **exponential smoothing** technique is the most popular way of smoothing. In this scheme, the current value is replaced by a weighted sum of the current value and the previous smoothed value. Exponential smoothing takes the following form:

$[Y_t = \alpha.X_t + (1-\alpha)Y_{t-1}, \quad Y_1 = X_1]$ for some parameter $0 < \alpha < 1$

'Unrolling' the recursion in the expression above, you get

$$\begin{aligned} Y_t &= \alpha.X_t + (1-\alpha)Y_{t-1} \\ &= \alpha.X_t + (1-\alpha)\left(\alpha.X_{t-1} + (1-\alpha)Y_{t-2}\right) \\ &= \alpha.X_t + \alpha(1-\alpha).X_{t-1} + (1-\alpha)^2 Y_{t-2} \\ &= (1-\alpha)^{t-1}X_1 + \sum_{i=0}^{t-2} \alpha(1-\alpha)^i.X_{t-i} \end{aligned}$$

Questions: 1 / 1

Time Series Smoothing

The following table contains the sale of a product in a department store for the first 6 months of the last year.

Month	Sale (in 100s)
Jan	36
Feb	45
Mar	81
Apr	90
May	108
Jun	144

Now, calculate the smoothed value for April, May and June, using the exponential smoothing method. Take α=0.8 and Y_{Mar}= 80.

88, 104, 136

Feedback : *Remember that the formula for exponential smoothing is* $Y_t = \alpha * X_t + (1-\alpha) * Y_{t-1}$, *where* Y_t *is the smoothed value at the time, t, and* X_t *is the actual value at the time, t. For the calculation of the smoothed value of sales for March, substitute the values in the equation. So,* $Y_M ar = 0.8 * 90 + (1-0.8) * 80 = 88$. *Similarly, you can find the other values.*

✓ Correct

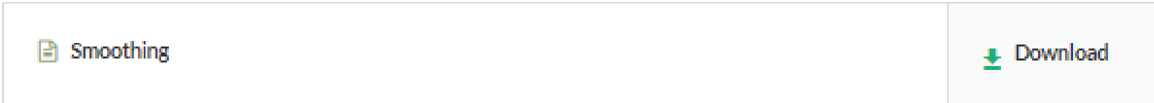
- 136, 104, 80
- 72, 66, 57
- 54, 72, 93

Your answer is **Correct.**

Smoothing Time Series Using R

Now let’s see how the smoothing techniques you just learnt can be implemented in R.

The R-codes used for the demonstration can be downloaded from the link below.



So in R, you can implement the **moving average** smoothing method using the command **filter()**, and you can implement **exponential** smoothing using the command **HoltWinters()**.

Also, you saw how the value of α affects the level of smoothing. Small values of α result in higher levels of smoothing (this may result in 'distorting' the original time series), and large values (close to 1) will not result in any smoothing at all.

End-to-End Analysis - Classical Decomposition

Please refer other doc in the folder “End-to-End Analysis - Classical Decomposition”

Summary

This brings us to the end of this session and the module on time series analysis. Some of the concepts that we covered in this session are

- Making a time series stationary through differencing
- Additive and multiplicative models
- Smoothing techniques
- End-to-end analysis of a time series

Forecasting using Smoothing Techniques

Introduction

Welcome to the fourth session on 'Time Series Analysis'.

In the previous sessions, you looked at two methods commonly used for analysing time series, i.e. classical decomposition and ARIMA modelling.

In this session

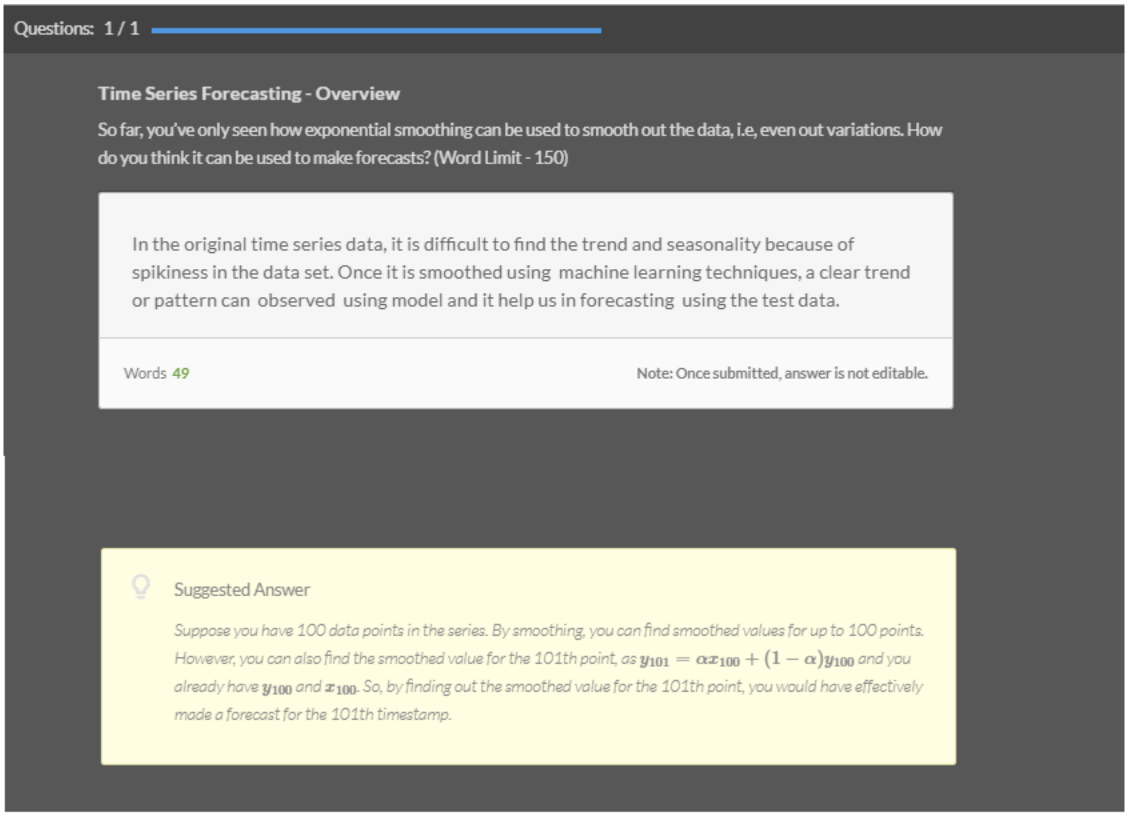
In this session, you will learn how smoothing techniques can be used to make forecasts for a time series. The methods you will study are

- **Average methods**
    - Simple average method
    - Simple moving average method
    - Weighted moving average method
- **Exponential smoothing methods**
    - Simple exponential smoothing
    - Holt's model
    - Holt-Winters model (Optional)

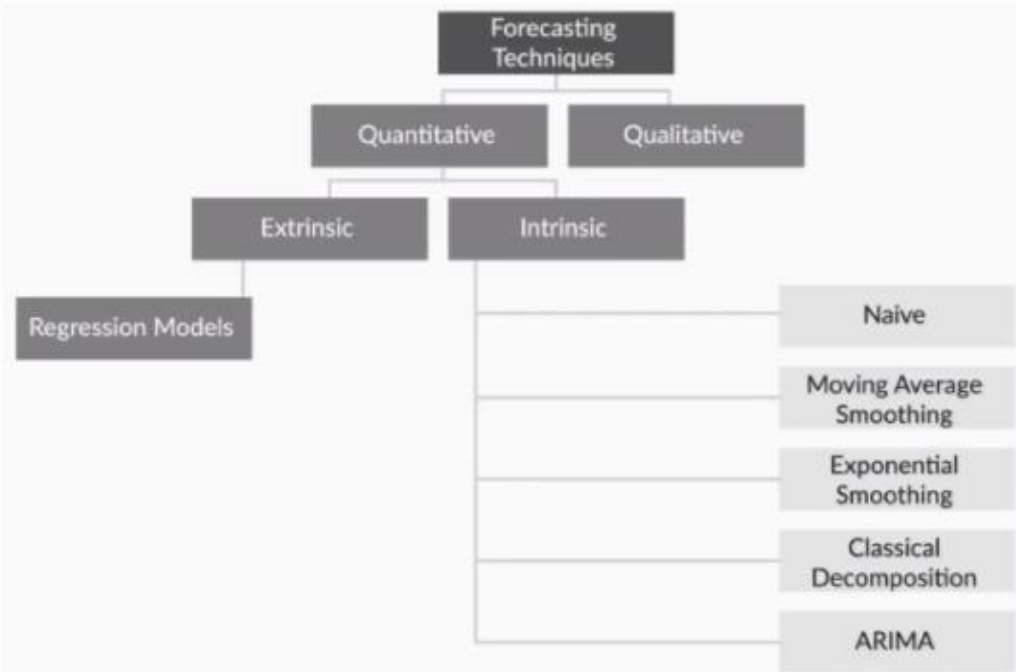
Time Series Forecasting - A Quick Overview

So far, you've learnt two methods for making forecasts: classical decomposition and ARIMA. As Raj will explain in the upcoming video, there are some more methods that can be used to make forecasts.

So let's listen as he tells us about these other methods of forecasting and when which of them should be used.



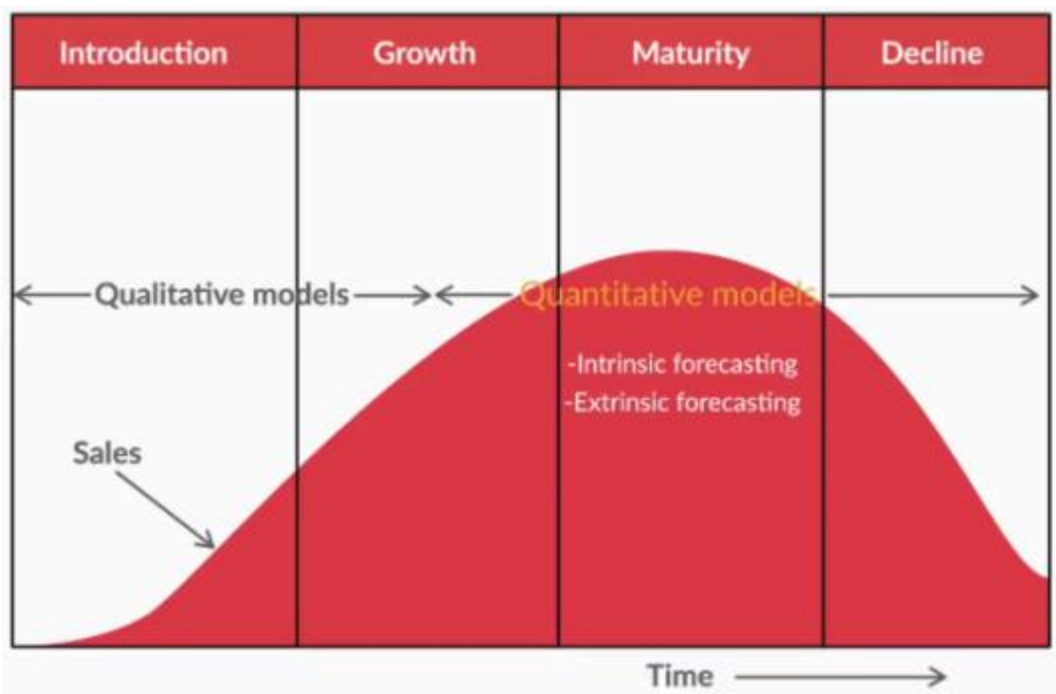
So all the techniques used for forecasting can be classified as follows:



Forecasting Techniques

Suppose you are an analyst who is trying to predict the sales for a given model of a television set, for the year 2018.

Which technique should you use to make this prediction? Actually, the most suitable technique for the forecast will vary depending on the type of data you have and on where you are in the product cycle.

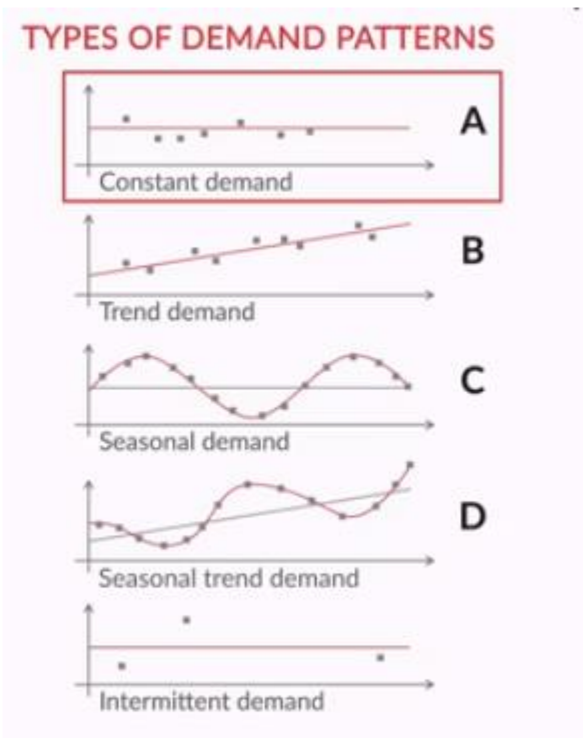


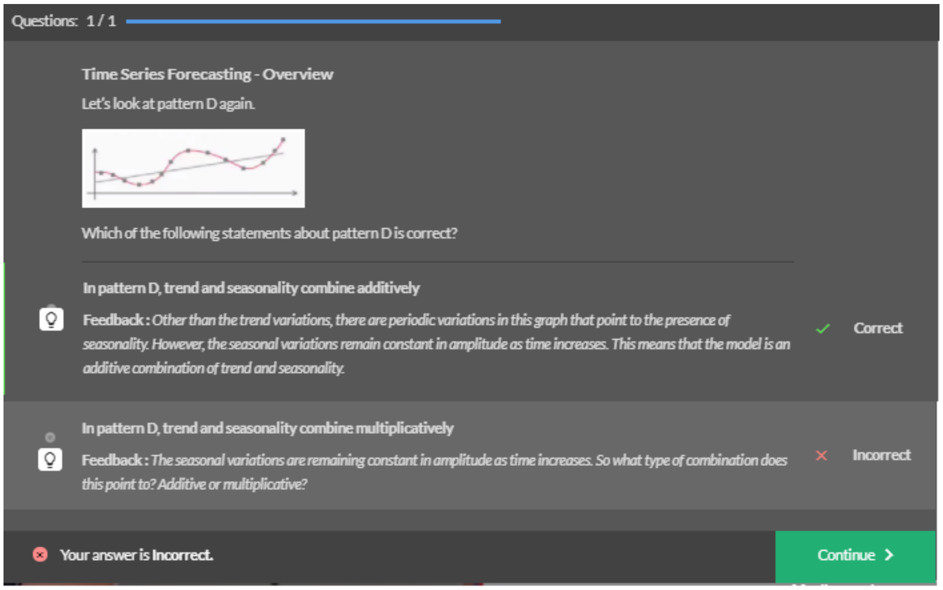
Product Cycle of a Typical Product

If you have data regarding the sales for the past years, i.e. 2014, 2015, 2016 and 2017, you can go for an **intrinsic method** of forecasting, such as ARIMA, classical decomposition, etc.; these use past sales data to predict future sales. However, in order to be able to do this, you will need some significant past data, and this will only be possible in the later stages of the product cycle.

If you don't have the sales data for many past years, you can use other data such as competitors' sales levels, market size, number of new features, etc., to make a decent forecast. In this case, you will be using regression, i.e. the **extrinsic method** of forecasting.

Now suppose you are stuck in the worst case, i.e. you don't have any past sales data, and you don't have any other data either (market size, etc.). In this case, you will simply have to use a **qualitative model**. In other words, you will just have to take the opinions of some experts who will provide guesses on what the sales will be. Such a situation will arrive early in the product cycle, especially if the product is very new, and neither you nor your competitor has ever sold it before.





So, that provides a quick review of what you learnt about time series so far.

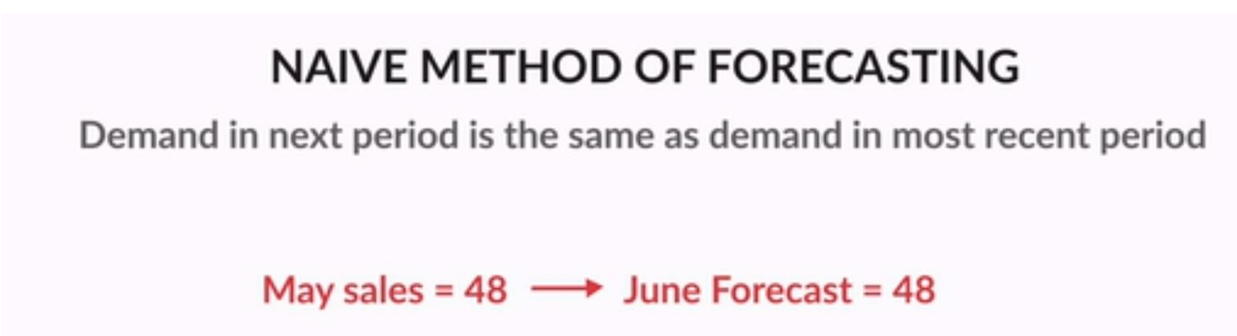
For the rest of this session, we'll explore some intrinsic methods of forecasting (i.e. naive, moving average, exponential smoothing, etc.), in detail.

Naive and Average Methods for Time Series Forecasting

Now let's explore the intrinsic methods of forecasting in detail. The various methods listed for time series forecasting were

1. Naive method
2. Average method(s)
3. Exponential smoothing method(s)
4. Classical decomposition method
5. ARIMA modelling

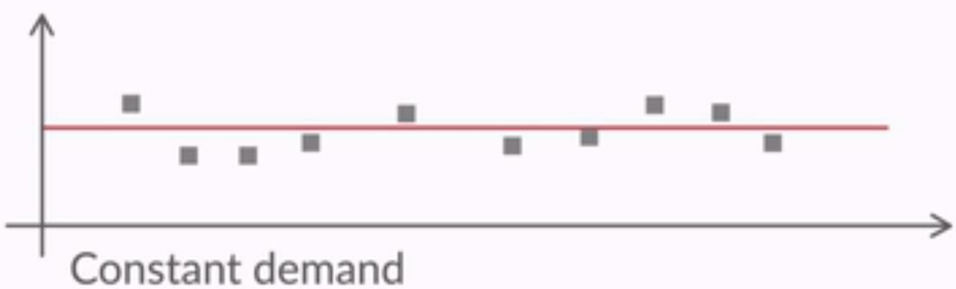
You've already been through classical decomposition and ARIMA modelling. Now let's explore the other methods and see how they can be used to make forecasts.



AVERAGE METHODS OF FORECASTING

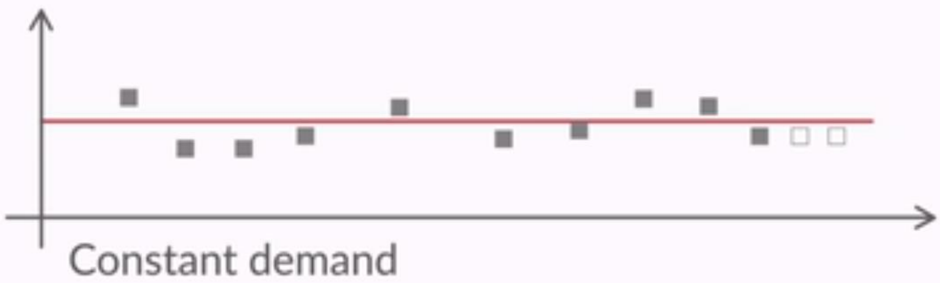
1. Simple Average (SA)
2. Simple Moving Average (SMA)
3. Weighted Moving Average (WMA)

SIMPLE AVERAGE FORECASTING



$$\text{Forecast for } 11^{th} \text{ month} = \frac{\text{Sum of last 10 months' demands}}{10}$$

SIMPLE AVERAGE FORECASTING



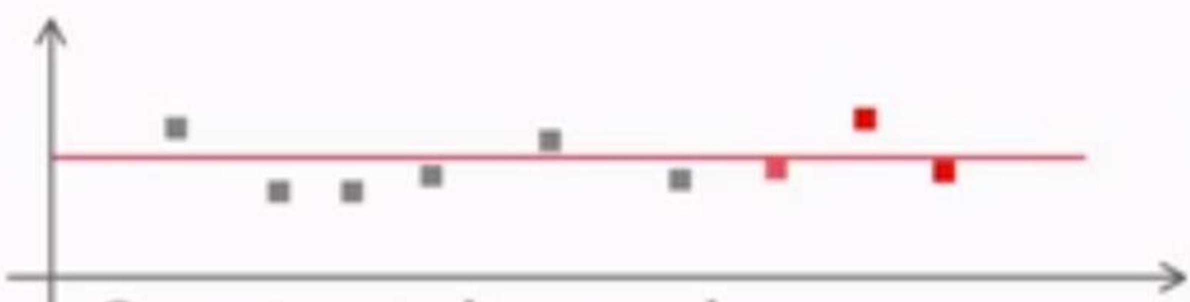
$$\text{Forecast for } (N+1)^{th} \text{ month} = \frac{\text{Sum of last N months' demands}}{N}$$

WHY SIMPLE MOVING AVERAGE?

SIMPLE MOVING AVERAGE SMOOTHING

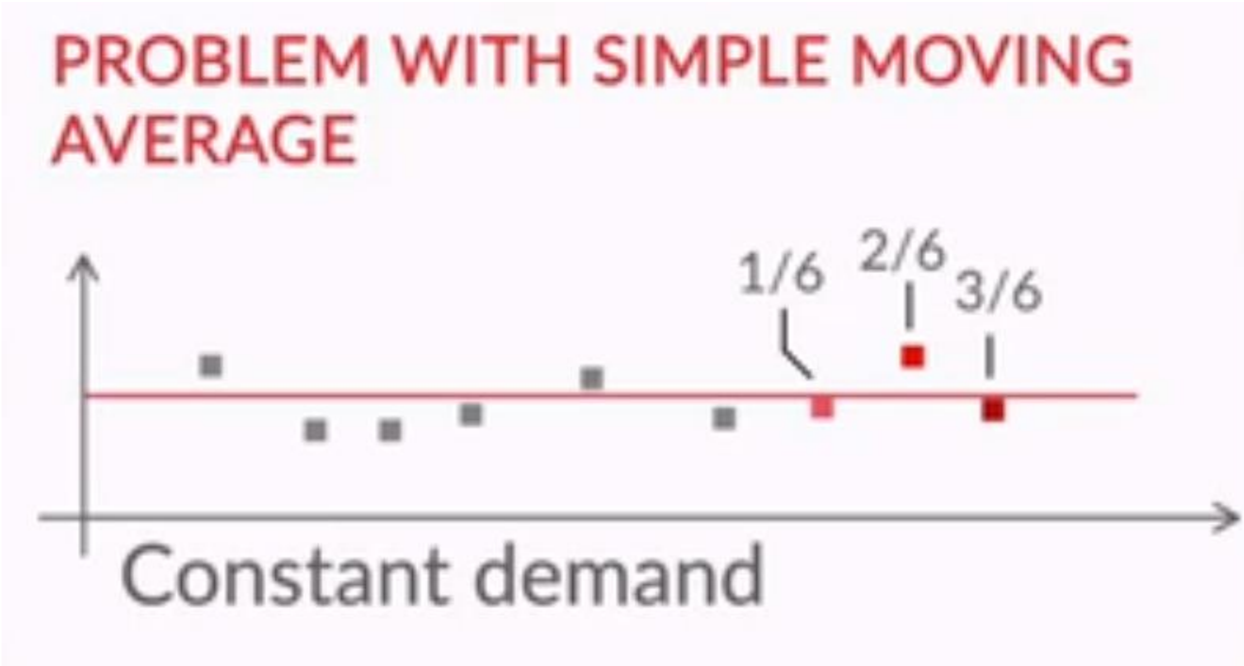
Month	Sales (000)	Moving Average Forecast (n=3)
1	4	NA
2	6	NA
3	5	NA
4	?	(4+6+5)/3 = 5
5	?	5
6	?	5

PROBLEM WITH SIMPLE MOVING AVERAGE



Constant demand

Problem: same weight is given to all the past data. It makes more sense to give more weight to the most recent data than the older data.



SIMPLE MOVING AVERAGE SMOOTHING

Weights: 3/6, 2/6, 1/6

Month	Sales (000)	Movie Average Forecast (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3	31/6 = 5.167
5	7	25/6 = 4.167
6		

$F_5 = 6*1/6 + 5*2/6 + 3*3/6 = 25/6$

Raj said 5.167, but he meant 4.167

SIMPLE MOVING AVERAGE SMOOTHING

Weights: 3/6, 2/6, 1/6

Month	Sales (000)	Movie Average Forecast (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3	31/6 = 5.167
5	7	25/6 = 4.167
6		32/6 = 5.333

$F_5 = 6*1/6 + 5*2/6 + 3*3/6 = 25/6$

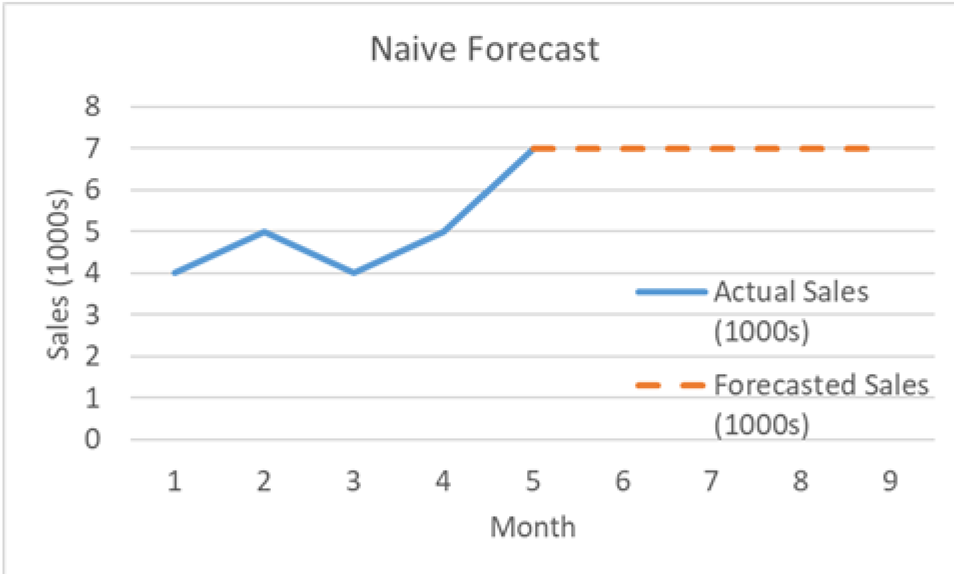
Rai said 5.167, but he meant 4.167

So, the various methods you just learnt for making predictions are

1. **Naive method**
    - Forecast = Last month's sales
2. **Average methods**
    - Simple average method
        - Forecast = Average of all past months' sales
    - Moving average method
        - Forecast = Average of last 3 (or 4, or n) months' sales
    - Weighted average method
        - Forecast = Weighted average of last 3 (or 4, or n) months' sales

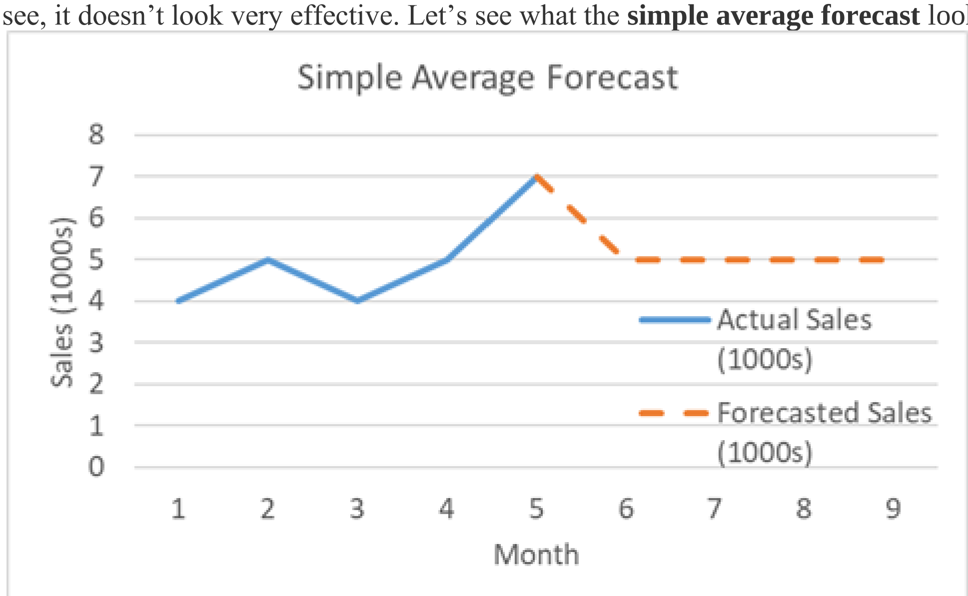
In order to understand these better, let's look at a sales example similar to the one provided in the lecture. Let's say that you have the sales data for a bookstore for the last five months, and you want to forecast what the sales level will be for the sixth month.

Now let's look at the naive method again. A forecast made using the naive method will look like this:



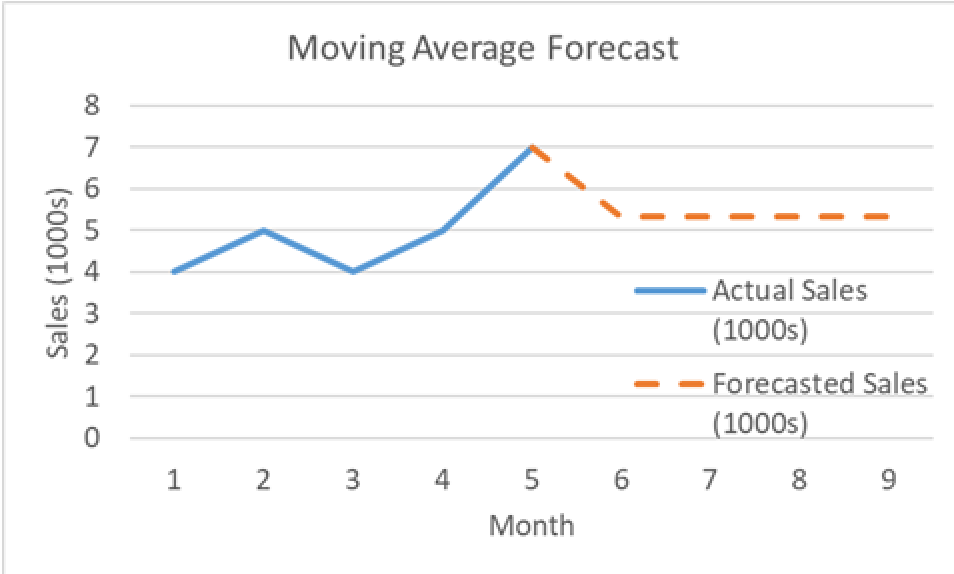
Naive Method Forecast

As you can see, it doesn't look very effective. Let's see what the **simple average forecast** looks like:



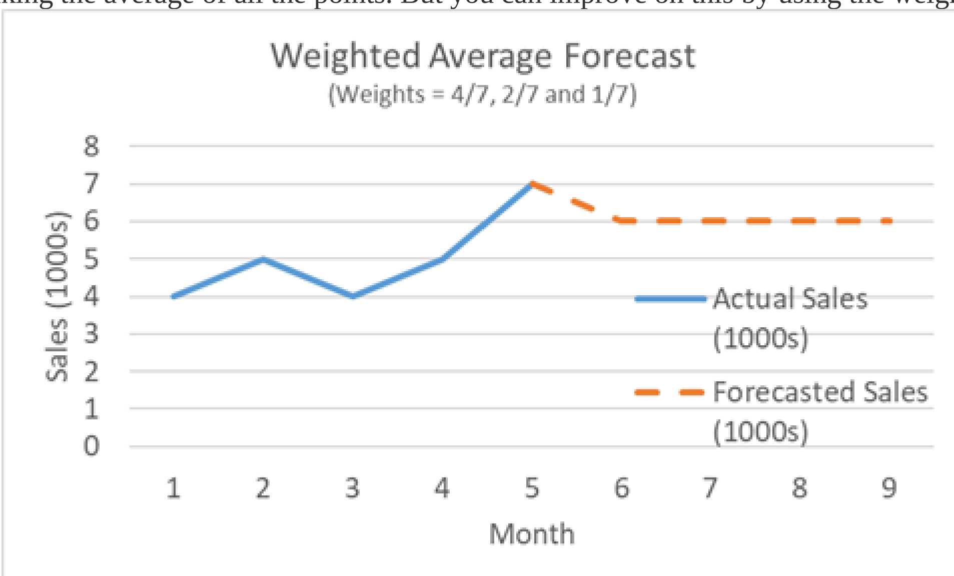
Simple Average Forecast

This looks better since it is more in line with the overall trend of values in the series. However, since the recent data points were higher for this series, a good forecast should take that into account. To do that, you can use the moving average forecast:



Moving Average Forecast

The moving average forecast is slightly higher than the simple average one, and hence, looks closer to the training data than a simple average does. This is because it takes the average of only the recent points, instead of taking the average of all the points. But you can improve on this by using the weighted average:



Weighted Average Forecast

The weighted average improves on the moving average by giving more weight to the most recent value (4/7), instead of to the not-so-recent values (2/7 and 1/7).

Questions: 1 / 1

Average Methods of Forecasting

You can make the forecast for the 6th month easily since you have the data for the past 5 months. Now suppose you are trying to make a forecast for the 7th month. This can be a problem because you don't have any past data for the 6th month. Which of the following strategies will you then apply to make a forecast?

81% Take the forecast for the 6th month as the past data, i.e. data for the 6th month, and forecast for the 7th month.

19% Simply forecast that the 7th month's sales = 6th month's sales since the data to make a proper forecast is not available.

Only 67 people have taken this poll from your cohort. The results would be updated as we get more responses.

So, in order to forecast for the 7th month, you can employ either of these two strategies:

1. Take the forecast for the 6th month as the past data, and forecast for the 7th month.
2. Simply forecast that the 7th month’s sales = 6th month’s sales since the data is insufficient for making a proper forecast.

Both are acceptable approaches. You used the second approach here, but that's done just to be safe.

So now that you’ve gone through the naive method and the average methods (methods that utilise the average) for forecasting, in the next session, you will go through the exponential smoothing methods used for forecasting in time series.

Exponential Smoothing Methods for Time Series Forecasting

So now you’ll learn about exponential smoothing methods for forecasting. But before you do, let’s quickly revise what exponential smoothing is.

EXPONENTIAL SMOOTHING

$F_{t+1} = \alpha^* d_t + (1-\alpha)^* F_t$

where d_t is last data point and F_t is forecast for that period

$0 \le \alpha \le 1$

$F_7 = \alpha^* d_6 + (1-\alpha)^* F_6$

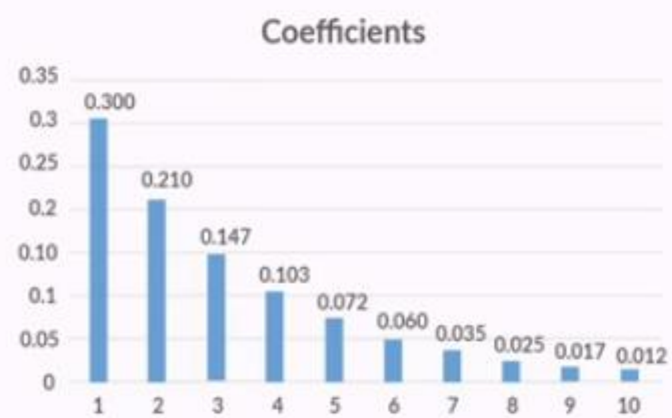
$F_7 = \alpha^* d_6 + (1-\alpha)^*(\alpha^* d_5 + (1-\alpha)^* F_5)$

$= \alpha^* d_6 + \alpha^*(1-\alpha)^* d_5 + (1-\alpha)^2 F_5$

$F_7 = \alpha^* d_6 + \alpha^*(1-\alpha)^* d_5 + \alpha^*(1-\alpha)^{2*} d_4 + \alpha^*(1-\alpha)^{3*} d_3 + \dots\dots + \alpha^*(1-\alpha)^{5*} d_1 + (1-\alpha)^6 F_1$

$F_{t+1} = \alpha^* d_t + \alpha^*(1-\alpha)^* d_{t-1} + \alpha^*(1-\alpha)^{2*} d_{t-2} + \alpha^*(1-\alpha)^{3*} d_{t-3} + \dots\dots + \alpha^*(1-\alpha)^{t-1*} d_1 + (1-\alpha)^{t*} F_1$

EXPONENTIAL SMOOTHING



So, to summarise, exponential smoothing is basically a weighted average smoothing model with a slightly different equation:

$$F_{t+1} = \alpha d_t + (1-\alpha)F_t$$

For example, let's take a model with = 0.3. Now, say you are trying to make a forecast for the fourth timestamp. The forecast would be

$F_4 = (0.3)d_3 + (0.7)F_3$

$F_4 = (0.3)d_3 + (0.21)d_2 + (0.49)F_2$

$F_4 = (0.3)d_3 + (0.21)d_2 + (0.147)d_1 + (0.343)F_1.$

The advantages this model offers are

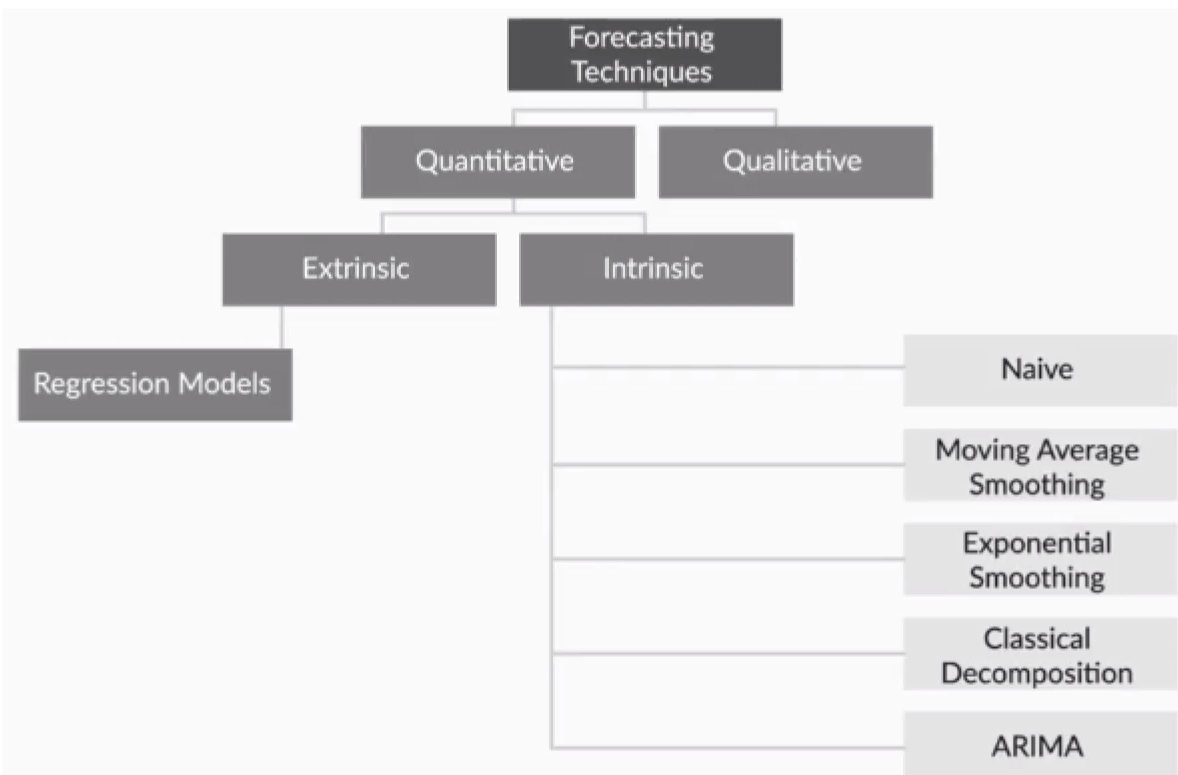
1. The forecast is influenced by all the past values.
2. Weightage given to recent data is much more than to past data.

Forecasting Using Simple Exponential Smoothing

Please see the folder – 'Forecasting Using Simple Exponential Smoothing'

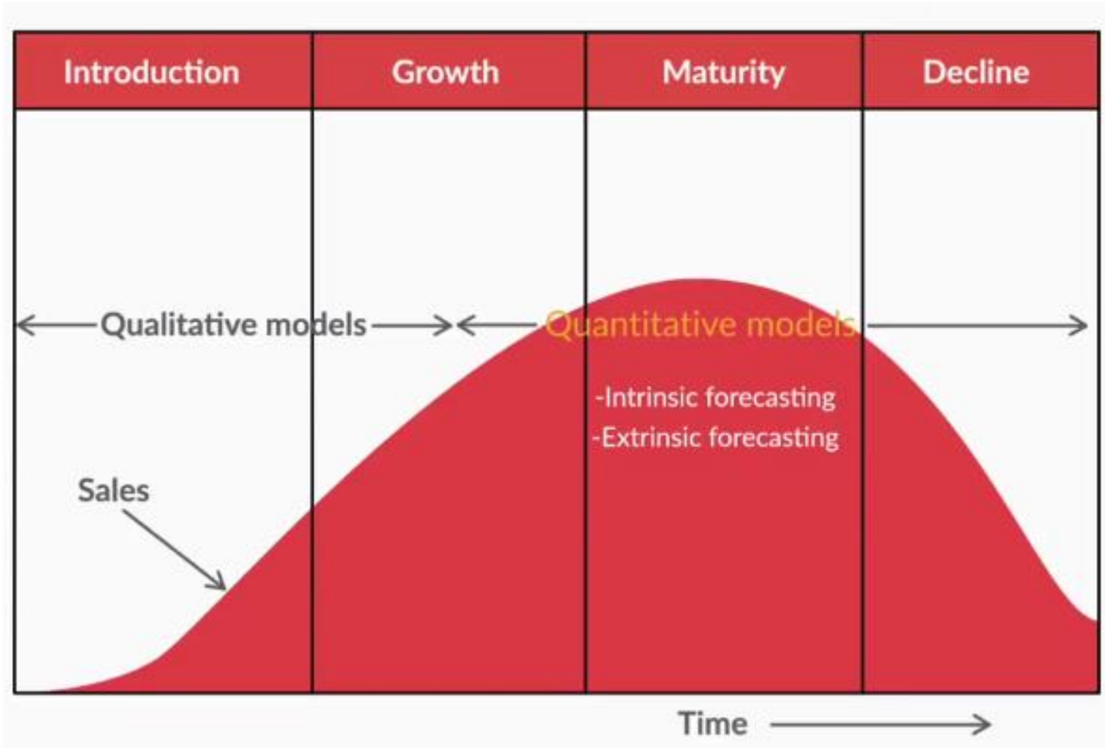
Summary

In this session, you started by getting an overview of the various techniques that are used to make a forecast. The different techniques used for this purpose are as shown below:



Forecasting Techniques - An Overview

Then, you saw when in the product cycle each of these techniques should be used.



Forecasting Techniques According to Uses in Product Cycle

Quite simply, the **qualitative models** are used when there is very **little data** to make a forecast. This usually happens early on in a product's lifecycle. In an **intermediate stage**, you would have information about the product, such as the competitor's pricing, etc., but you would still not have enough past data. In that case, you would use **extrinsic forecasting** to make a forecast. However, once you have **enough past data**, you can use **intrinsic methods** of forecasting, such as ARIMA or classical decomposition, to make a pretty good forecast.

Then you saw how smoothing techniques can actually be used to make forecasts. The different techniques that can be used to make forecasts are

1. **Naive method**
   - Forecast = Last month’s sales
2. **Average methods**
   - **Simple average method**
     - Forecast = Average of all past months’ sales
   - **Moving average method**
     - Forecast = Average of last 3 (or 4, or n) months’ sales
   - **Weighted average method**
     - Forecast = Weighted average of last 3 (or 4, or n) months’ sales
3. **Exponential smoothing methods**
   - **Simple exponential smoothing method**
     - Exponential smoothing used to forecast the level
   - **Holt's model**
     - Exponential smoothing used to forecast the level and trend
   - **Holt-Winters model** (optional)
     - Exponential smoothing used to forecast the level, trend and seasonality

The equations for Holt's Model are

$$S_t = \alpha d_t + (1-\alpha)F_t$$
$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}$$
$$F_{t+1} = S_t + b_t,$$

where S_t is the **level estimate**, b_t is the **trend estimate,** and F_t refers to the **overall forecast**.

Since Holt's model estimates both the level and the trend, it can give you better estimates for data that has a trend in it. For your own sales data example, you saw how Holt's model was able to capture the trend, but exponential smoothing wasn't.

Holt's Model vs Simple Exponential Smoothing

Lastly, for data that has a **level, trend, and seasonality**, you can use the **Holt-Winters Model**. Due to complexity and time constraints, this model was not covered in the main module. However, you can learn more about it in the optional session of this module.

You can download the lecture notes for this session (Time Series Forecasting Using Smoothing Techniques) from the link given below: