Course Name:Business Intelligence and Analytics
Professor Name:Prof. Saji.K.Mathew
Department Name:Department of Management Studies
Institute Name:Indian Institute of Technology Madras
Week:11
Lecture:42

ANN FOR FINANCIAL TIME SERIES MODELLING | BI&A | Prof. Saji K Mathew

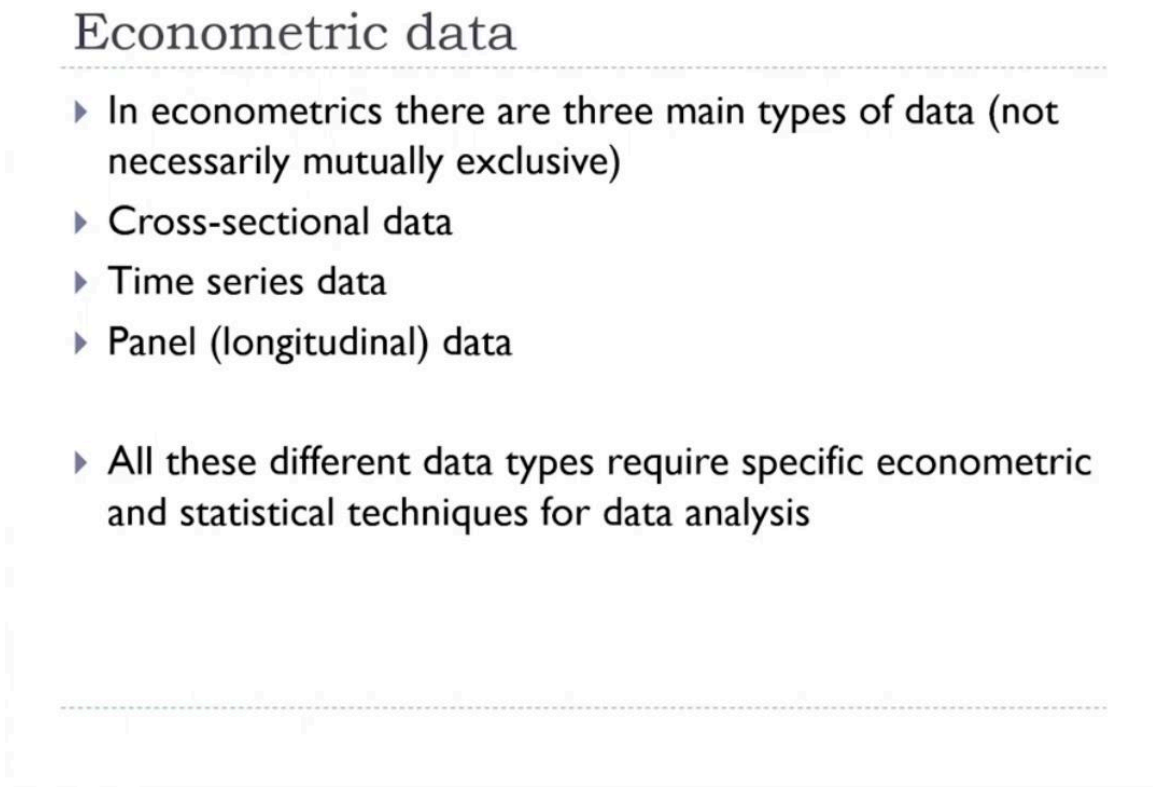
Hello, and welcome back. In the last session we learnt certain fundamentals of neural networks, artificial neural networks. The idea was to give a very basic foundation of what is a neuron and how a neuron gets translated into an artificial neuron. And then when neurons are connected together in the form of a layered network, how that particular arrangement is good for learning or you can create algorithm to make that network learn or you could train that neural network. So there is something called the structure of the network or the typology and that is the design of a neural network and what we have touched upon is one kind of neural network topology known as the feed forward network.

I only looked at feed forward network because that is useful for our project which is basically to design a neural network to learn the patterns in a complex time series, a historical time series. And then we also looked at algorithms that are used for training neural networks. So we looked at perceptron training rule and the limitation of it and then we came to category of training algorithms known as back propagation algorithms or BP not blood pressure, but back propagate, not black, back propagation algorithms, BP algorithms as they are called. And so we learnt about gradient descent algorithm which is an algorithm that is used to compute weight vectors or differential weight vectors, to adjust the weights constantly as you input data tuple by tuple to a predesigned neural network.

So that is what we have learnt about neural networks. So today our plan is to apply neural network to a problem that is very familiar to students of business and management and those who are investors and those who are generally aware of something known as stock markets. So we are going to a problem area which is not something that we discussed so far, most of our problems were, you know taken from sales, marketing and this type of context. So today we are going to capital markets market context to get a time series or a data set and then design a neural network and then use a training algorithm with its hyper parameters to train and then see how the

model performs. So that is what we are going to do first.

So before I go and use the neural network, it is important to have again some foundations about the financial time series in stock markets. And let me actually give you a broad overview of it and what are the techniques or methods currently used for modeling such time series and then we would apply neural networks. So that we are aware of the problem well and then, before we apply it. So students have often times asked me what are the different types of data or, well it is not about data unit or data type like the categorical data, ordinal, interval, metric data and so on, that is something that most students are aware of, but data sets, different types of data that is collected for the purpose of data analysis. So in econometrics, which econometrics mean the data analysis techniques generally used by economists.



So they have developed their own tools and techniques and literature available on econometric analysis of data, both macro and micro data analysis. So I am drawing from the literature related to econometric data analysis. So in econometrics, they talk about three types of data, one is the cross sectional data, other is time series data and then the panel data. So you must have heard about it in different context. So cross sectional data,

as you are aware is the type of data that we used so far, meaning data collected at the same time from different points.

So if there are say 1000 stores, you collect data from a sample of those 1000 stores at the same time. So it should not be different times, you know in the last session on clustering we discussed this, then time itself can add variability to the data, therefore cross sectional data means data collected at the same time.

Cross-sectional data

- A type of one-dimensional data set collected by observing many subjects (such as individuals, firms or countries/regions) at the same point of time, or without regarding the differences in time
- In general used to compare the differences among the subjects I Order does not matter
- Examples: Explaining people's wages by reference to their education level

Individual	Income	Marital Status	Educational Attainment
1	1500	single	university degree
2	2500	married	graduate degree
3	2000	separates	university degree

Then there is time series data. So there are examples here, so you can go through this slide, I will leave this with you and then time series data is the next type of data which is not data collected at the same time but data collected at fixed time intervals. Data collected at fixed time intervals. So the data source remains the same. It is the same data source, you know if you are collecting a person's heartbeat, it should be that person's heartbeat and of course, it beats at regular intervals. If it is not regular, then there is a problem. So that is an example of a time series or you collect data about temperature of Chennai every day and that is a time series but it has to be Chennai, it should not be other cities. So it becomes a time series of a phenomenon that you collect at regular fixed, regular intervals.

Panel data

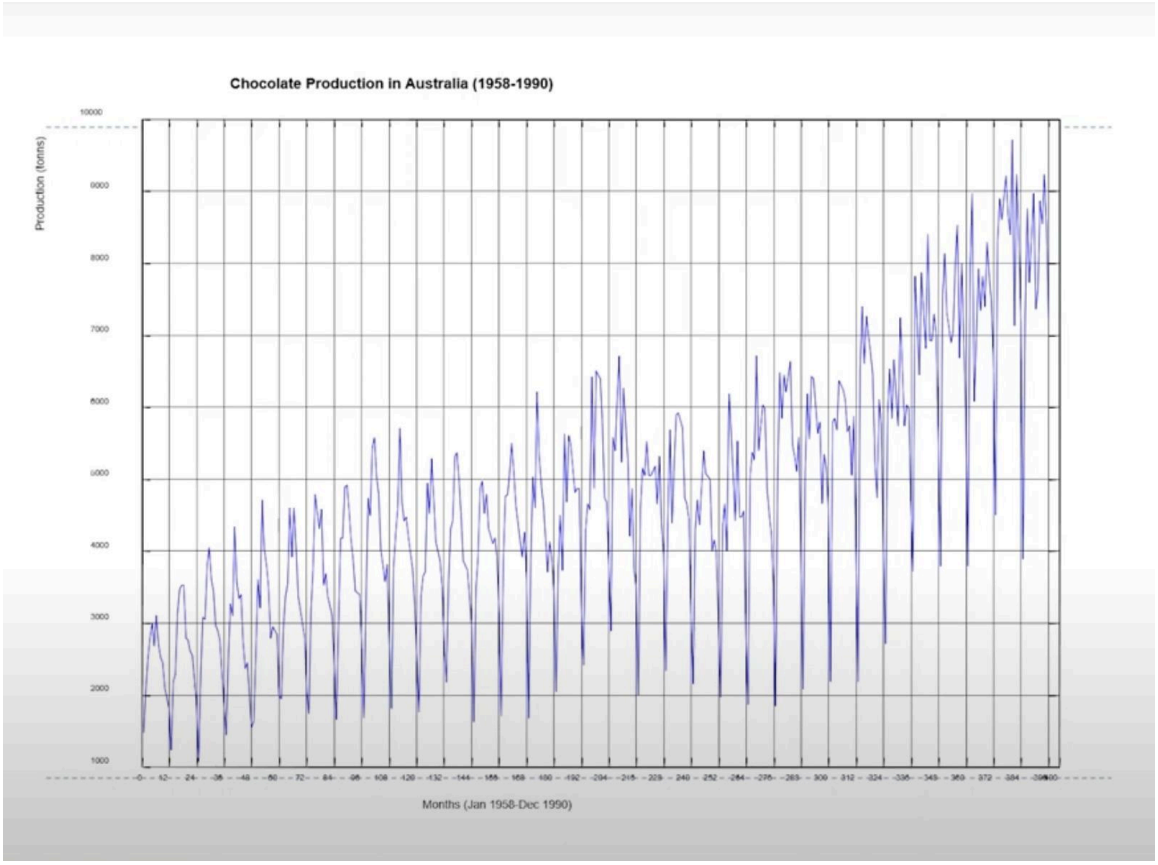
- Data that involve repeated observations of the same items over long periods of time
- Not necessarily cohort study
- Panel data (Longitudinal) studies track the same subject (people, countries, same set of stocks)
- Measurements are observed or taken on the same subjects repeatedly

Company	Year	Profit
1	2000	1.2 billion
1	2001	1.3 billion
1	2002	2 billion
2	2000	0.2 billion
2	2001	0.3 billion
2	2002	0.1 billion
3	2000	3 billion
3	2001	3.5 billion
3	2002	4 billion

So time is having a regularity here and the third type of data is the panel data where of course again, it is a longitudinal data that is data collected with respect to time but from multiple points. So or in other words, there are multiple variables in a panel data. It is not one source but multiple sources or you collect data about different variables at a given frequency and that becomes panel data. Good question. Panel data is longitudinal.

That is this data is collected at fixed intervals over a period of time, whereas cross sectional data collected from different points at the same time. So time brings variability to the panel data. But it is multiple time series, you can understand it that way. It is multiple time series.

So our interest today is to model time series data. That is what we are looking at today. So an example of a time series is given in the chart and this is chocolate production in Australia from 1958 to 1990. It is old data, does not matter this is econometric data but it is country level data, you can see that it is micro level data. So this is a good example of an econometric time series where the x axis or what varies, as the x variable is time. You can see that the x variable is time here.



And so this is showing year wise data and it also depicts year wise variations in the data. So you can see that this is whole one year. So within the year, you can see how chocolate production is varying. So it is speaking at some point and then there is some pattern of the chocolate production. It must be tied to the consumption pattern but you can see that how this is varying.

And generally it is said that a time series has four components. I hope you have studied this, four constituents of a time series. You can add them up, you can multiply based on that there are different time series, classical time series models. So what are the constituents of a time series? What is that? Level, seasonality. Seasonality, good, seasonality is what you can actually see in the encircled portion of the graph because it is not the same throughout the year.

There is a seasonal variation of the production of chocolate. So that is seasonality. So within an year, there is a variation. A seasonality is about within year seasonal variations of the data. So that is seasonality.

What is the other component? So one out of four answered, there are three more

answers. So there is a trend. I can actually fit a trend line to this data. I can fit a linear regression to this data. What would be my x variable? When I fit a trend line to a time series? Time.

You code time and time itself is a variable and that is the distinction of a time series or panel data etc. Time is a variable. So there is a trend line. There is a seasonality. What are the other variations you can observe? Two more.

These are fundamentals of time series before we go on to model it. Level. Level. What? Maybe I am not understanding your language. It could be equivalence, but I am not familiar with that language.

Extrapolation. Extrapolation and interpolation that is about, that is in modeling. So I am just asking about the constituents of a time series. There is seasonality. There is trend.

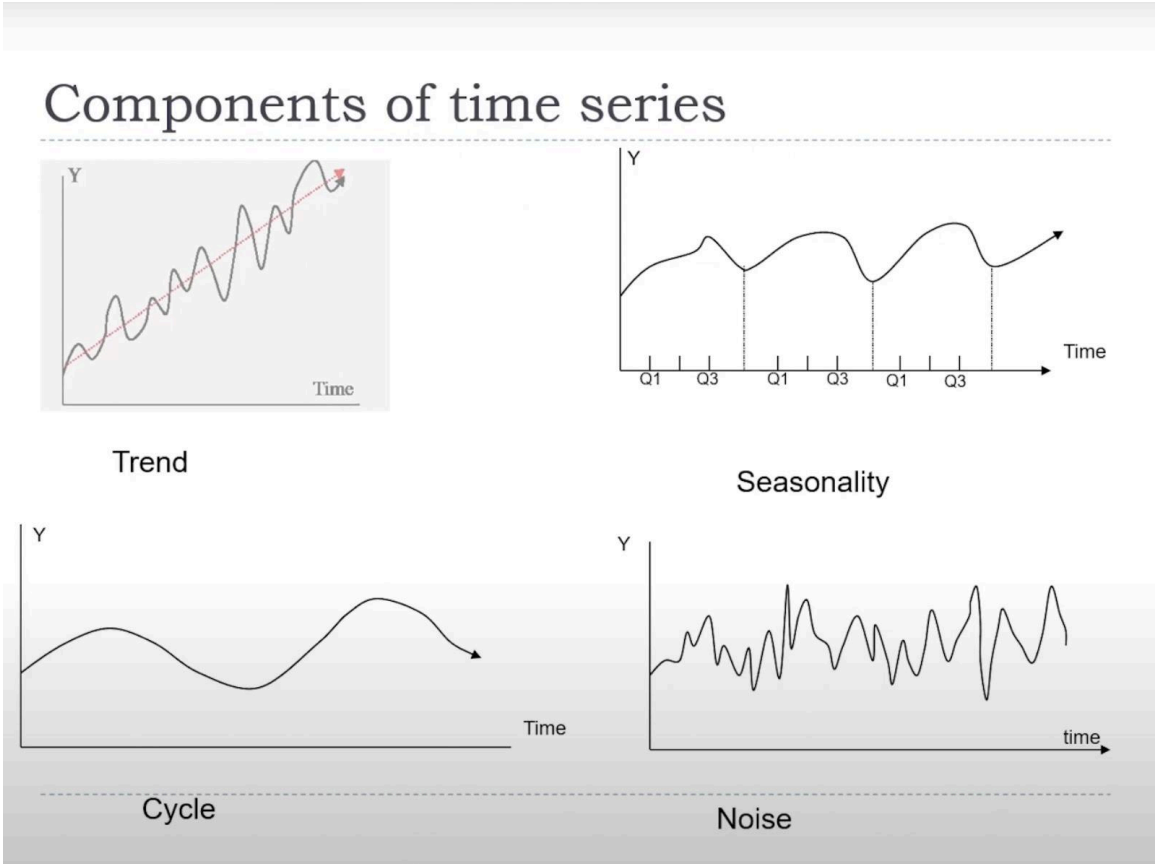
There are two more. Very good. Cyclicity. There is cyclicity. Economic cycles which affects the production and what is the difference between seasonality and cyclicity? Seasonality is within an year. Season changes within an year and therefore, production changes within an year.

Cyclicity is. It follows. We may not be able to actually observe it clearly. I am just drawing an approximate line. Over five, six years, there are economic cycles and economic cycles influence production and consumption. And that pattern is also inside of this time series. That is another pattern hidden within this time series.

So if you are trying to model the time series, you can think of the constituents. There is a trend. There is a seasonality. There is also cyclicity of datum. With cyclicity, the period would be five to six years, which follows the economic cycle.

Last component. Stationarity. Stationarity is a characteristic of the data. I am happy you are familiar with time series modeling. That is common sense. Everything is. So you have actually talked about all regularities within the data.

Isn't there an irregularity which cannot be modeled, which is the randomness. There is a random component which does not follow any regularity and that is the fourth component. So those four components, if you can assemble together, that can actually model an economic time series like this. An econometric, not economic time series, econometric time series like this. So this slide depicts the four components.

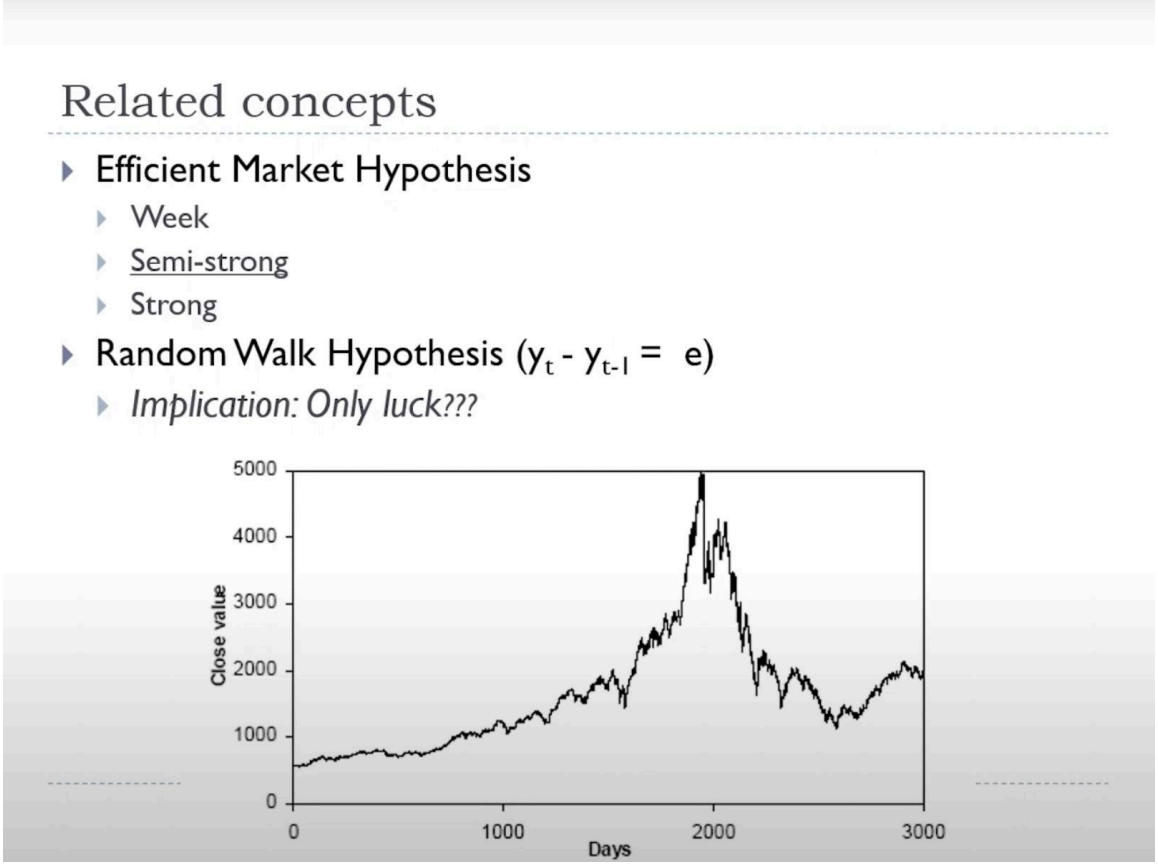


Trend, seasonality, cycle and noise or irregularity or randomness within the data. So classical time series models, if you look at statistical literature, you will find separate modeling techniques for trend, for seasonality, for moving averages and then you actually bring these components together multiplicatively or additively and there are models in statistics for modeling time series. So the ARIMA models are widely used in modeling of econometric time series like this. So that is a classical or statistical approach. That is the statistical way of modeling a time series and particularly an econometric time series.

So this is well established but we are moving on from here to neural networks, suggesting that we are using another type of technique for modeling time series. That is our project. So since we are going to draw a time series from stock market, it is important to know some basic concepts about stock market time series. Some of you may be familiar with it because you invest in stock markets.

I do not. Related literature, very broadly there are two common theories that you must be familiar with. One is the EMH, efficient market hypothesis which categorizes every stock market into three categories. Weak, semi strong and strong forms of stock market based on how efficiently information flows within the market, within the actors of the

market. So weak form of a market is that the flow of information is very weak. So for example, the information about an asset or an organization is not efficiently flowing into the investors.



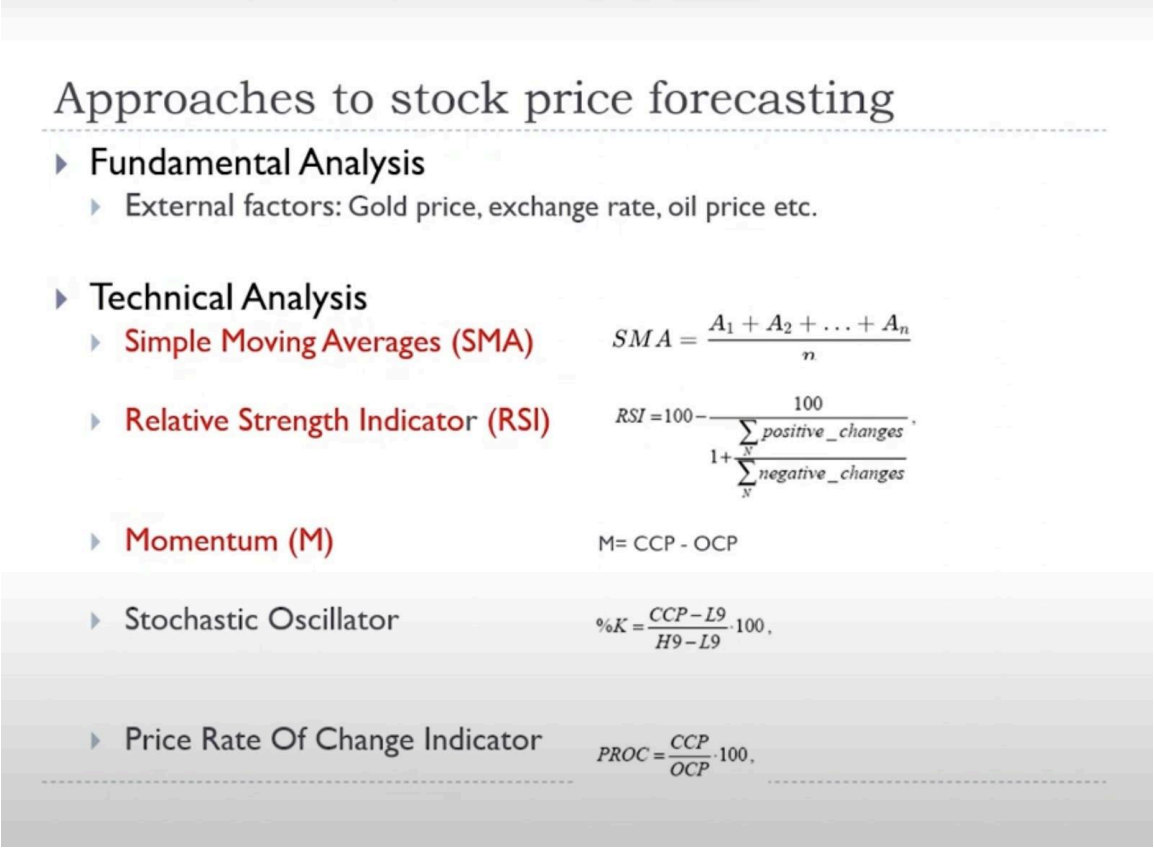
Semi strong is something between weak and strong. Strong form of market is where the information flow is very efficient. That is one aspect of the stock market. Other is the random walk hypothesis. What does it mean? That stock markets behave very randomly.

In that case, there is no point in trying to model a stock market time series because it is like a drunkard man. Once you are drunk, where is your next step going to be? It is not something that you can predict. So the random walk hypothesis would suggest that stock prices move in a random way and therefore, there is no regularity in the time series and therefore, it is not a good idea to model it, because you are trying to model randomness. So that is another argument or so y_t, so today's price minus yesterday's price is a noise.

So there is no regularity there. So that is the argument. Do you agree with this? In that case, we can stop the class right away because we should not go forward if random walk hypothesis is 100 percent true. So all attempts to model a time series or be it a time series itself or an index of the stock price is based on the assumption that there is some regular component. There is irregularity and also there is some regularity or, like suggesting today's price is reflected in yesterday's price. So today's price will be related to

yesterday's price or day before yesterday's price.

It is actually dependent on a certain indicators of the series itself and that is assumption for modeling a time series. And if there is no regularity, how can people make money consistently in stock market? A Warren Buffet could not be existing in the stock market, if there is no pattern in the time series. Of course, not time series, of course if you talk to him, if you have a lunch with him, you know how much to pay, but he will not recommend time series modeling or chart analysis. He looks at the fundamental analysis. But let us move on with the assumption that there is pattern in time series like this and approaches to stock price forecasting if you read literature, it is a very broad summary.



There are two approaches to stock price forecasting. One is the fundamental analysis, other is the technical analysis and I am hoping you can understand it easily. The fundamental analysis suggests that the movement of a stock price series is influenced by the fundamentals of the economy. It is an economy that drives a price series. So it could be exchange rates, it could be gold price, all the environmental factors actually influence a time series.

So how you move, depends on where the vehicle takes you. So look at the vehicle and predict the time series. That is the approach in fundamental analysis. You are sitting in a car, so look at the movement of the car instead of looking at your movement. So that is sort of a concept. That is fundamental analysis.

In technical analysis, instead of looking at where the car or cycle is moving, you look at the person, you look at the series itself, look at the characteristics of the series and then derive certain indicators from the series itself and use them as predictors of the future of the movement of the time series. So that is technical analysis. So in this particular session, we are going to look at technical analysis, not fundamental analysis. So in technical analysis, if you go to any website or any portal which provides you the time series of stocks, you will also generally see the technical analysis done in the form of certain charts.

So those are basic analysis they do and provide to investors, particularly for regular daily investors or daily traders and those indicators are number one, the moving average of the time series. There is something called MA, moving average or simple moving average. You must have learned about different types of moving averages. One is simple moving average, then exponential moving average and so on. But simple moving average is very insightful, in the sense the data series does not look at that days or that points data, but it actually moves as an average of the nearby points, suggesting that it is the average that matters because there is some information in the average. So you try to chart the average which gives you information about what is the average movement of the series.

So you know the formula for a moving average and you also know that typically an odd number is used as a denominator for moving average and if you are using a odd number say 3, so data point 1, 2, 3 and 4 and if you start a moving average calculation, the moving average will start from the third point, meaning that you lost two points information, the starting information will be lost. You can actually substitute it, but if you just go by the moving average calculation there is a loss of data in the initial period. You are familiar with it. There is another indicator widely used for stock prices, that is the relative strength indicator or RSI, relative strength indicator. Does it ring a bell for any investors here? This is something, some people recommend to look at how the prices are moving or how relatively the positive movements and negative movements are happening.

So RSI is a useful indicator to generate buy signal and sell signal. Buy signal and sell signal. So what do you think looking at the formula? This is the formula. Closely look at this formula. If the positive movements are more than the negative movements, what would be the value of RSI? Will it move towards 100 or will it move towards 0? If positive movements are much higher than negative movements, will RSI move up or will

RSI come down? Look at the formula. It is 100 minus 100 divided by number of positive changes, 1 plus number of positive changes divided by number of negative changes.

So if positive changes are more, RSI moves up. RSI will move up and it is suggested that when RSI crosses 80, it is good to buy or it is good to sell. The asset has been increased, its price has been going up, going up, going up, it reached 80. So positive movements are very strong. Is it buy signal or sell signal? It is the right time to sell because anything that moves up, it is overvalued probably, it may come down soon. So it generates a buy signal, sorry, sell signal and when RSI is going down and down, of course this is not the only indicator you should look at.

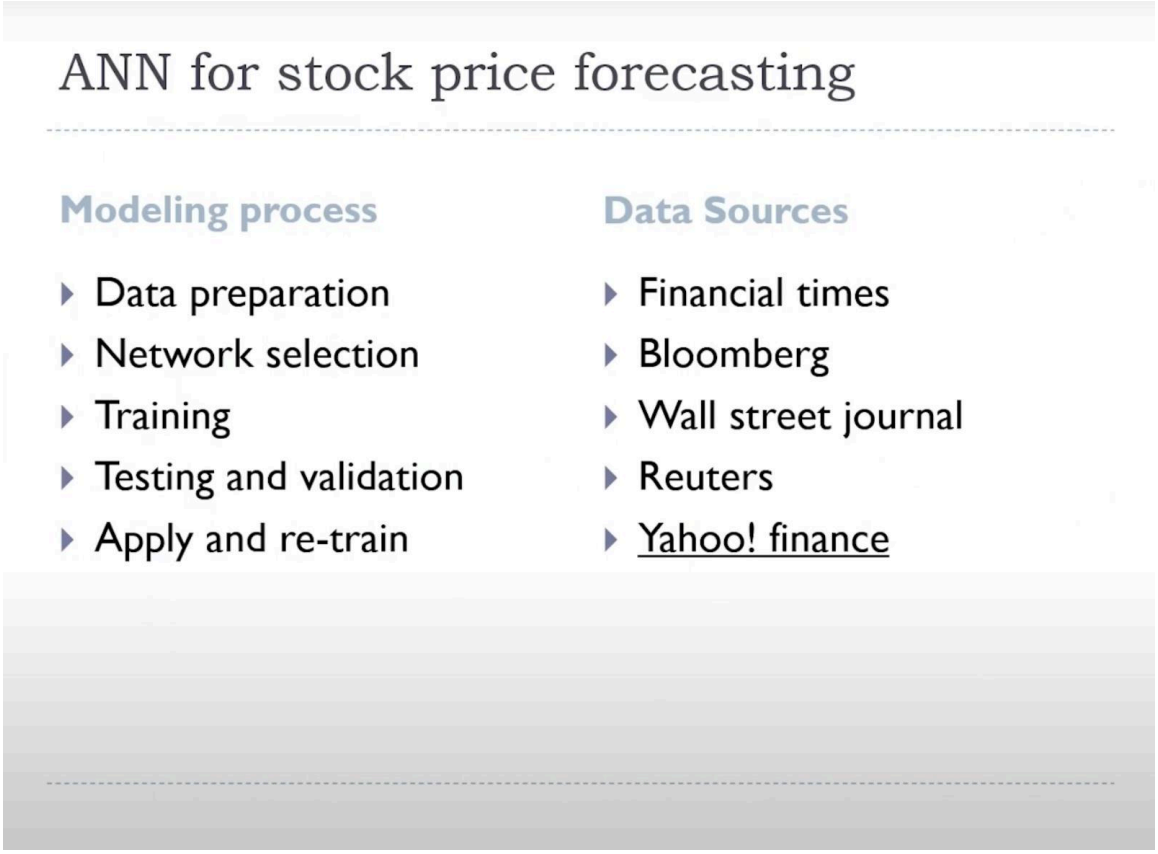
You should look at other indicators as well, especially fundamental aspects, but everything else remaining the same if RSI is going down beyond 20, it is a buy signal. That is what literature says. Then there is something called momentum. So in all these indicators, you can see what is interesting for a stock price is how the series is moving, not the absolute value of a time series at a point in time, but how it is movement.

The movement is more interesting or useful than the values itself. And therefore, there are indicators like momentum which is current closing price minus old closing price CCP-OCP and there is stochastic oscillator again CCP minus lowest in nine periods divided by highest in nine periods minus lowest in nine periods into 100 and then price rate of change indicator which is current closing price minus old closing price with respect to a period. So you can see that all these are indicators of the movement of a time series than that series itself. And you may wonder why you have these different indicators because all these indicators are derived. They can also be called derived indicators.

These are derived from the original series. Original series is the actual data point, but then you create or derive multiple indicators from the same series. And the purpose is, each of these indicators gives you a different sense or a different insight about how the prices are moving. And that is why you look at the multiple indicators. For example, momentum versus RSI. You get different insights about the movement pattern versus whether you should buy your sell etc.

So in our modeling today we are going to use a time series and then we are going to derive the ones, the indicators which are given in red. So we will write some script and get these values calculated and then use them as variables, use them as separate indicators in our modeling. And that is based on literature in this area which suggests that in ANN modeling, instead of using the time series alone, if you add more derived

indicators then the training is more efficient. That is what empirical evidence shows, empirical research shows.



Empirical means based on observations. And that is the basis for using the different indicators from a time series like the stock price series for training ANN. Now our aim is to apply ANN for stock price forecasting. So this is a very fundamental level modeling and those who want to apply it, can extend it to specific requirements. So I only model the stock price. So someone may not be interested in what is the stock price tomorrow, but you may more be interested in the movement pattern.

Our aim is first to sort of do this first and see how well modeling approach like the ANN works for stock prices. And I must tell you it is widely used for stock markets, ANN is widely used for stock markets. So what are the steps involved? We know this is basically analytics process. It is the same thing, but translated specifically for ANN.

Data preparation is the starting point. Then we have to design a network. We discussed it yesterday, a feed forward network. And then we have to choose a training algorithm and input data and use the algorithm to train the network using the data or fit the data

into the network and use a training algorithm for that purpose. And then we have to test and validate the models. And of course, you can retrain when you have more data.

So that is a continual process. So time series data is generated every day. So you can train it for up to data today and after some time, you can train it with updated data. So that is retraining. So what is the data source? There are multiple sources of stock price data. Some you have to pay and some are free. For example, I think some of, some faculty members in finance, they have a subscription to Bloomberg.

So Bloomberg is a paid access and it is very credible. So therefore, Bloomberg is widely used as a data source for econometric data. But I am going to use Yahoo Finance portal for stock prices and they have actually freely given access to stock prices and NASDAQ listed companies stock prices are available in Yahoo Finance. That is what I am going to use in this particular exercise.