INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
NPTEL
NPTEL ONLINE CERTIFICATION COURSE
Business Analytics & Data Mining Modeling
Using R – Part II
Lecture-13
Understanding Time Series– Part II
With
Dr. Gaurav Dixit
Department of Management Studies
Indian Institute of Technology Roorkee

Business Analytics & Data Mining Modeling Using R - Part II

Lecture-13
Understanding Time Series-Part II

With
Dr. Gaurav Dixit
Department of Management Studies
Indian Institute of Technology Roorkee

Welcome to the course Business Analytics and Data Mining Modeling Using R – Part 2, so in previous lectures we started our discussion on time series forecasting and we started our this first topic that is understanding time series.

So we talked about time series forecasting and its application in for different industries, then we also understood the difference between time series and cross sectional data, and from that data point of view how you know time series forecasting and analysis in general is different from cross sectional analysis, so all those points were also discussed in the previous lecture itself.

Then we talked about the technological advances and how that have impacted time series data collection and availability of time series in our data sets. Then the next point that we started our

discussion on rows, choice of time scale, so we talked about two important factors that are important in terms of deciding which time scale should be used, so first one is as per the forecast requirements as we talked about, if it is for retail funds and you know if we want to forecast sales then typically the planning and execution is geared at towards you know, year on year basis, then probably we need to forecast the sales for next year, if the planning and execution is dependent, you know it's everything is in the firm its gear towards you know monthly, planning and monthly execution then probably we need sales number for each month, so depending on those requirements also, those requirement will also in a way forces to use a particular time scale, however we can also look at other factor, for example level and sources of noise, because from the analytics point of view it will have a great role in this because the forecasting power is influenced depending on the level and source of noise as we talked about in the previous lecture that if we you know aggregate data to a course or level, then probably some of the sources of noise might cancel out, so that might be helpful in terms of forecasting the time series.

So let's start our discussion on the next point, so that is about when multiple time series are to be forecasted, so there are different approaches that could be used when this is the scenario, so typically if we talk about the business context there most of the industry, most of the firms they have you know they face this kind of scenario, for example they might have many products, so the companies might be selling many products, and the requirement would be to forecast you know multiple products at the same time, depending on the kind of products that they are, and sometimes sale of one product might actually effect the influence the sale of other products, so keeping that in mind also sometime the requirement is to forecast the multiple time series, so

Understanding Time Series

- Choice of time scale
  - As per the forecast requirements
  - Level and sources of noise
    - Impact forecasting power
    - Aggregation of data to a coarser level can be used to average out some sources of noise
- When multiple time series are to be forecasted
  - Forecast each series individually or
  - Use multivariate time series model

one approach could be forecast each series individually, or the second approaches using multivariatetime series models, so these are the two approaches, so let's discuss few more points about these two, so one is as we talked about single series forecasting where we are

Understanding Time Series

Single-series forecasting

- Preferred approach is to forecast each series individually
  - Due to simplicity
- Doesn't account for the possible relationships between series

Multivariate time series model

- Require statistical expertise for estimation and maintenance
- Directly model the cross correlations between series
- Impose restrictive assumptions on data and cross-series structure
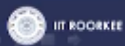
looking to forecast each series individually, so when this kind of approach is going to be suitable, we'll discuss this and we'll also compare this with the other approach that is multivariate time series modeling, so if we talk about single series forecasting or specifically you know forecasting a time series individually, this is the preferred approach and the main reason is its simplicity, because we can easily apply our, and it is expertise in terms of building model forecasting model for a you know single time series, it is simpler and you know then each series can you know model for each of those series and can be done individually and then the results can be interpret as per the requirement.

However one problem with this kind of approaches that it doesn't account for the possible relationships between series, so this is one problem because we are building forecasting model for each series individually, so therefore of course in that process itself we are not accounting for any interrelationships that could be there between those series, so as I talked about some products you know we might have the forecasting requirement, you know, for some products and you know sale of one product might influence the sale of other products, so the time series data that we might have for those products, they might have certain relationships as well, so if we are looking to follow this approach, single series forecasting approach so this might not be accounted, however at the same time if these products don't have you know such relationships and the time series that are there, they are typically independent then probably single series forecasting is the best approach that we could have because of its simplicity.

So if we compare some of these points which we just discuss to the multivariate time it is smallest scenario, so in this first thing is we require a higher statistical expertise for estimation and maintenance because of the multivariate nature of this modeling approach, so the whole approach is because of this, you know the whole approach is slightly complex and in this advantage is that we can directly model the cross co-relations between series, so this multivariate time series model approach, it allows us to model this cross co-relations, however because of the complexity of this approach itself, the available techniques, available models that we have they typically imposed restrictive assumptions on data and cross series structure,

so that is one limitation as well, so these are the two main approaches that could be used, if we talk about other approaches, so one is the econometric modeling, so econometric modeling

Understanding Time Series

- ## Other approaches
  - ### Econometric modeling
    - Information from one or more series is used as inputs to model the series of interest
      - Typically, these inputs are based on causal relationships

  - ### Alternative approaches involving heuristics
    - Use associations between series of interest and external variables heuristically

could also be used for time series forecasting, so what typically happens is information for one or more series is used as inputs to model the series of you know interest, so in econometric modeling so when we have, when we are looking to forecast multiple series so essentially one or more series we can extract some important information, meaningful information and use that as input to model the series of interest, so typically these inputs are based on casual relationships so econometric modeling as because some of these models they have, they are theoretical basis, therefore you know usefulness of the relationship, meaningfulness of the relationship or the casual relationships that is typically analyzed and taken advantage, and in the time series of interest that is incorporated and forecasting is done.

Then there are alternative approaches involving heuristics which can also be used, so using some of the heuristic principle that we observe for our series of interest so we can use some of those associations between series of interest and the external variables heuristically, so these kind of approach is can also be done, because sometimes we, for example even for economic activity, for example if there is an observation that about the boom and bust cycles of economic, what economy or for that matter even for a countries economy, so if a particular you know external event happens before or after or you know indicating you know some sort of you know disturbance in the economic activity, you know as such then those kind of associations, those kind of external factors can also be used, so this is the approach where alternative approaches involving heuristics, these external factors, external variables, events can also be used to incorporate in time series forecasting, so this brings you know, this brings us to our next discussion point that is about time series analysis and time series forecasting, so as we have
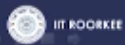
Understanding Time Series

Time series analysis

- Explanatory modeling

- Done to determine the time series components
  - Seasonal patterns, trends, relation to external factors etc.

Time series forecasting

- Predictive Modeling

- Done to forecast future values of the series
  - Using time series and other external information

discussed in the previous lecture that typically two kinds of task that we perform in time series context, so one is analysis which is mainly descript in nature for explanatory purposes, and then the second one is the forecasting.

So now we are going to discuss some of the differences between what we call time series analysis and time series forecasting, so if we look at time series analysis as I have talked about, this is mainly explanatory modeling, if we look at time series forecasting it is predictive modeling, because in time series forecasting we are looking to forecast future values of a particular series, so this is mainly predictive in nature, if we look at time series analysis we are looking to understand the time series as such and therefore it is you know primarily explanatory modeling, so as we can see in the next point as well, time series analysis is done to determine the time series components, so when we are looking to understand a particular time series you know it would be better if we are able to find out the time series components, so seasonal patterns, trends, relations through external factors, so some of these kind of things if we are able to determine so it is going to be easier for us to analyze the time series, so this is what typically happens in a time series analysis.

If we look at the you know time series forecasting it is done to forecast future values of the series as I talked about, so we use the time series data itself and other external information that is available, so in time series forecasting as the name suggest also that we are looking to forecast future values of the series using the series data and also external information, however in time series analysis main focus is understanding the time series and determining the time series components is one part of that exercise, so let's move forward.
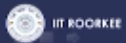
Understanding Time Series

Time series analysis

- In the modeling process, structural methods are preferred
  - To either explain associations or causality
- Retrospective in nature
  - Can use future values

Time series forecasting

- Data-driven methods are preferred
  - Despite their black-box approach
- Prospective in nature
  - Can't use future values

So few other competitive points about these two, so in time series analysis in the modeling process typically structural methods are preferred, so this is as we have talked about to either explain associations or causality, so there the typically statistical method and structural methods so they are preferred because of the nature of the you know analysis that is done, so if we look at the time series forecasting typically data driven methods are preferred, and this is despite they are black box approach because we are looking for our task is to be able to you know forecast the future values of the series, so because of this you know data driven method or typically considered to be better in doing this kind of task, so they are preferred and despite their black box approach, however as I said in time series analysis because the modeling is for explanatory reasons and therefore structural method and statistical methods they are preferred, and we are looking to understand the associations and causality.

Time series analysis is also retrospective in nature, so when we are looking to explain something so that particular event or that particular thing has happened, so with that consideration we are looking to analyze it, so therefore even the future values can also be used in doing this analysis because we are not looking to you know predict or forecast future values, neither we are looking to understand a particular time series, therefore mainly it is retrospective in nature and we can also use future values because we are looking to understand a series, we are looking to understand the association and causality and different components, but when we talk about the time series forecasting it is prospective in nature and we cannot use future values because the main task itself is to be able to forecast future values of the series, so this is also one important difference.

Understanding Time Series

- Forecasting Methods
  - Regression-based forecasting methods

  - Smoothing methods

  - Data mining methods such as neural networks
    - Typically used in analyzing cross-sectional data
    - Used in time series forecasting as well
      - Specially to incorporate external information into the forecasts

So this was about the time series analysis and time series forecasting, so let's move forward, so the next important discussion point is the kind of forecasting methods that we are going to use, so as I said the major discussion in this particular module is going to be around time series forecasting, so next we are going to discuss the types of forecasting methods that are popular in this particular task, time series forecasting task, so first one is regression based forecasting methods, so in this the you know regression based models are typically formulated and then they are used for forecasting time series, so as we have talked about for this module that this particular aspect regression based forecasting method we are going to cover in the coming you know, coming lectures.

The second method is smoothing methods, so this is, though regression based forecasting that we just talked about, this is mainly statistical you know, these are statistical based approaches, statistical based methods, but if we talk about the smoothing methods they are typically data driven methods so we'll discuss these methods also in coming lectures, so these are the two popular and common approaches for forecasting time series, one is regression based more of a statistical based approach, the second one is smoothing methods which is more of a data driven approach.

Then of course there are data mining methods which are sometimes found to be quite useful in forecasting time series, so data mining methods such as neural networks they have also been used time series, and nowadays you know many applications so they are looking to incorporate neural network in their time series forecasting task, time series forecasting analytics quite often and few points about these methods that these data mining as we have discussed in you know previous modules, in this course and previous course as well that typically we used data mining methods to analyze cross sectional data, however we can use them in time series forecasting as well and this is quite a regular thing specifically if we are looking to incorporate external information into the forecast, so neural network method and other data mining methods are you know more preferred if we are looking to you know incorporate, if we are looking to include other external information into the forecast then these kind of methods are also being used for time series forecasting, so overall if we look at these are the three approaches regression based

forecasting you know methods which are typically you know statistical in nature, then we look at the smoothing methods which are typically data driven, then we have the data mining methods such as neural networks.

Understanding Time Series

- Forecasting Methods
  - Combining forecasting methods
    - Multi-level forecasting
      - In the first level, one method is used to forecast future values of the series
      - In the second level, another method is used to forecast the residual series of the first model and then the results are used it to correct the first level forecasts
      - Main idea is to benefit from the specific strengths of different methods to capture different aspects of the series
        » The approach can be used in analyzing the cross-sectional data as well

Let's move forward, so within forecasting methods there is another approach that can be tried out, so which is combining forecasting methods, so this is also something that can be done, so either we can go for regression based method the statistical kind of method or smoothing methods or data mining methods, however we can also use some of the you know, some of this methods from this different approaches and use this approach combining forecasting method approach, so there are two types of you know approaches within this combining forecasting methods, so first one is what we call multilevel forecasting and the second one is we ensemble approach, so we'll first start out with the this first approach that is multilevel forecasting.

So if we talk about the multilevel forecasting, since we are discussing combining you know methods for forecasting time series, so the name of this approach itself multilevel forecasting suggest that we are going to apply different methods at different levels, so in the first level we use one method to forecast future values of the series and in the second method we use the another technique to forecast the residual series of the first model that we have done in the first level, and then the results are used to correct the first level forecast, so we select one method based on you know certain criteria or certain preference or you know strength of that particular method to forecast future values of the series, so this happens in the first level. And in the second level we can use another method which can now you know the residual series that we have after the first level that we can again forecast using the second method and the reserves than can be used to correct the first level forecast, so in this way multilevel forecasting approach can also be implemented to forecast a particular time series.

So what is the main idea behind this particular approach, multilevel forecasting approach, let's discuss this, so main idea about, main idea of this particular approach is to benefit from the specific strengths of different methods to capture different aspects of the series, right, so different methods could be suitable to analyze or to be able to model you know different aspects of a particular time series, so we are looking to, through this approach we are looking to use

those strength, so first level if a particular method is suitable for you know forecasting you know certain aspect of a time series which is the prominent one given the data of that series then probably we'll use that method there, and then the residual series is going to be there which will have the remaining aspect of the time series which could again be you know forecasted using the second method given the strength of that method, so in that sense we are trying to combine the strength of two particular methods and doing a multilevel forecasting so that we are able to improve our forecast.

If we look at the overall process of this particular approach, multilevel forecasting, this multilevel you know multilevel analysis can also be applied on cross sectional data as well, so the different methods that we have talked about in one module in the previous course that is supervise learning, and the other module in this course itself that is unsupervised learning methods, so there also for certain you know types of task we can again apply two methods which are, which have different strength and capture the different aspects of cross sectional data, so this approach can also be applied for analyzing cross sectional data in unsupervised and supervised learning you know methods, so let's move forward.

Understanding Time Series

- Forecasting Methods
  - Combining forecasting methods
    - Ensemble approach
      - Multiple methods are used to forecast future values of the series
      - Results are averaged out to produce the final forecasts
      - Main idea is to produce more robust, accurate, and reliable forecasts

So talk about the second approach within the combining forecasting methods, so the second approach can be called the ensemble approach, so what we do in this particular approach, so multiple methods are used to forecast future values of the series, now in the second step results are averaged out to produce the final forecast, so multiple methods so different method that we think are going to be good enough to forecast a particular time series, so we are going to apply those methods to make their own forecast, then those results or those forecast are going to be averaged out in you know in some fashion and the final forecast is going to be produced, so what is going to be the advantage? What is going to be the benefit? So the main idea behind this ensemble approach is to produce more robust, accurate and reliable forecast, if we are applying different methods then they would be you know as we have talked about in the previous approach also, so they will have their own strength and weaknesses, but if we you know use their forecast and averaged it out so they might you know complement in terms of you know forecast accuracy and you know in terms of averaging out the, canceling out of noise, different

noise level and sources, so therefore main idea is to produce more you know robust, accurate, and reliable forecast and if we are able to select methods which are going to complement each other and the way we do the averaging of results we can actually improve our final forecast, so these are the two approaches within combining forecasting methods, so overall we have discussed you know four types of main approaches, one is regression based forecasting, the second one is smoothing methods, the third one is applying data mining methods like neural network and the fourth one is combining forecasting methods where we have again talked about two main sub approaches, the first one being the multilevel forecasting, and the second one that we just discussed is the ensemble approach.

So these are the overall forecasting methods that are you know common and popular for time series forecasting, so let's move forward. So as we have discussed before in this lecture that when we talked about the comparison between time series analysis and time series forecasting that, in the time series analysis determining the time series components is also an important part because we are looking to understand the time series, so however some of those aspects are important even for time series forecasting, so let's discuss time series components, so

Understanding Time Series

- Time series components
    - Identifying time series components is typically the first step before applying forecasting methods
    - Four components
        - Level, trend , seasonality
            - Characterize the underlying series
        - Noise
    - Level
        - Average values of the series across periods

identifying time series components is typically the first step before applying forecasting methods, because identification of these time series components is typically going to help in the next few steps of time series forecasting, so typically we do this you know identification process of time series components and whatever insights that we get through this process by understanding different components that is incorporated in the next few steps that we use in our forecasting.

So what are these time series components? So typically four components are there, so first the we have grouped together, so these are level, trend and seasonality, so these three components, the level of time series, the trend that is there in the time series, and the seasonality part of time series they in a sense characterize the underlying series that is there, so these three components are more important for us to identify because they characterize the underlying series and when we talk about the you know forecasting time series it is these components if we are able to

analyze and implement and model in some fashion then they are going to give us a good enough forecast.

Then the fourth component is actually the noise which is something you know that would be the left out, so something that we cannot you know explain or account for, so these are the four components, first three are the you know, you know we can say the structural components, the level, trend and seasonality, the another one is the noise, so let's discuss these components one by one, so what we mean by level? So level is average values of the series across you know periods, so what are the you know average values you know for the entire time series and if we can visualize the time series in you know some sub time periods so how those average values are being taken for those sub time periods, so in a sense that would give us the you know idea about the level of the series, so which you know which values are actually being taken by the series at an average overall average level, so that is what we mean by level, so for example if we look at the, if we follow the news that is there in the, for the you know stock market, the Sensex index or the Nifty index, so if we look at those numbers you know people will, in the news channel and other media they will talk about that Sensex has reached up to this number 9,000, 10,000, 8,000, so when we you know when we anchor our discussion around a particular certain you know number 10,000, 8,000 or 9,000 so essentially we are talking about the level of you know that particular series, so because the values are going to be around that particular level, so in a sense that gives us a sense about where the, that particular, you know where that particular series, the kind of values that particular series is taking, so that gives us an idea about the average values.
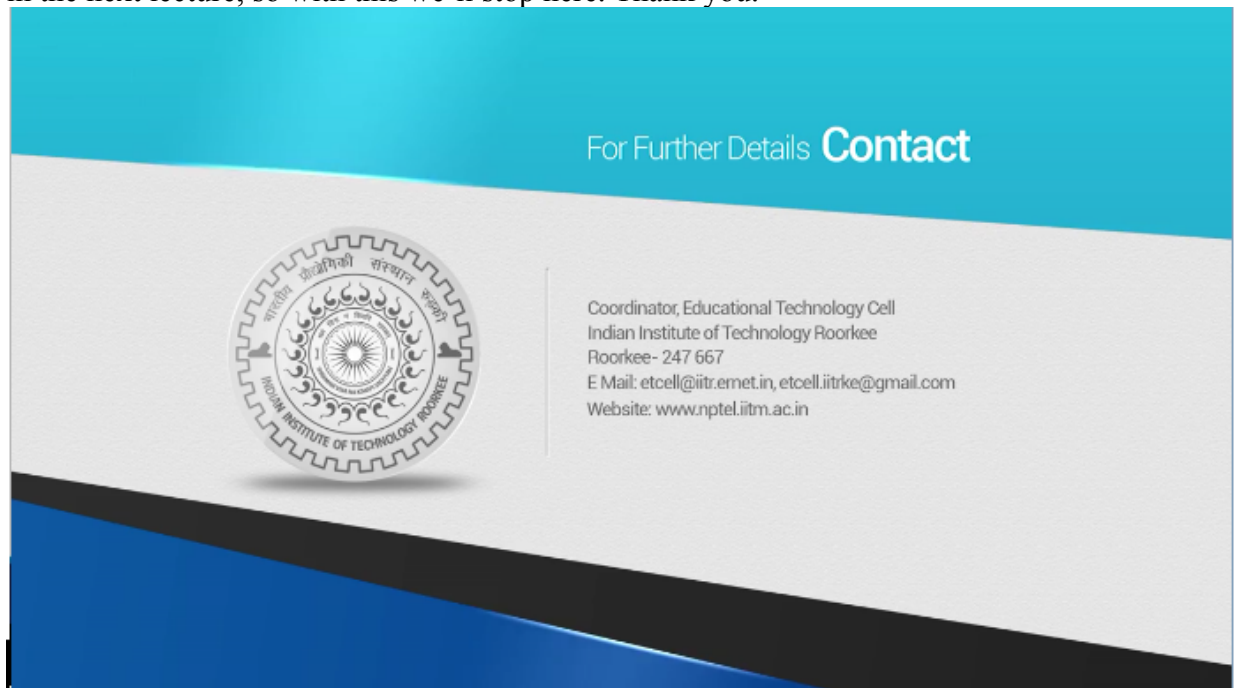
Understanding Time Series

- Time series components
  - Trend
    - Pattern of change in the series from one period to next
  - Seasonality
    - A short-term cyclical behavior observed several times in the given series
  - Noise
    - Random variation due to unaccounted causes
    - Always present in a series to some extent

Now the second component is trend, so what we mean by this? So pattern of change in the series from one period to next, so typically if we look at the you know a particular time series even if we, for example if we look at the you know stock market performance, share market performance and if we look at the BSE index and how you know what has been the pattern in the series from one period to the next, so this pattern actually reflects the presence of a trend in the series and so typically for example if there is a linear trend so on a longer you know time horizon on a longer time period we would see that the particular time series is increasing a

linearly in fashion, so there are going to be swings and cycles in a shorter time period, however if we look at take a long term prospective then we would see you know, we might be seeing a linearly increasing trend, for example India's GDP numbers if we you know plot them, if we take the GDP numbers from you know, 80's to till now and if we plot them then we would see that you know, we might see a linear line there, so linear trend might be visible there, so overall trend would capture that you know change in the series from one period to next year, so that means year on year how the series has been changing, what has been the trend, whether the trend has been linear or it could be polynomial, it could be exponential, it could be cubic, so if whether there is any structural in the way the series is changing so that is something that we identify as trend.

So there are few more components for example seasonality and noise, so we'll talk about them in the next lecture, so with this we'll stop here. Thank you.

Produced by
Mohan Raj.S
Graphics
Binoy V.P
Web Team
Dr. Nibedita Bisoyi
Neetesh Kumar
Jitender Kumar
Vivek Kumar
Dharamveer Singh
Gaurav Kumar
An educational Technology cell
IIT Roorkee Production
WANT TO SEE MORE LIKE THIS
SUBSCRIBE