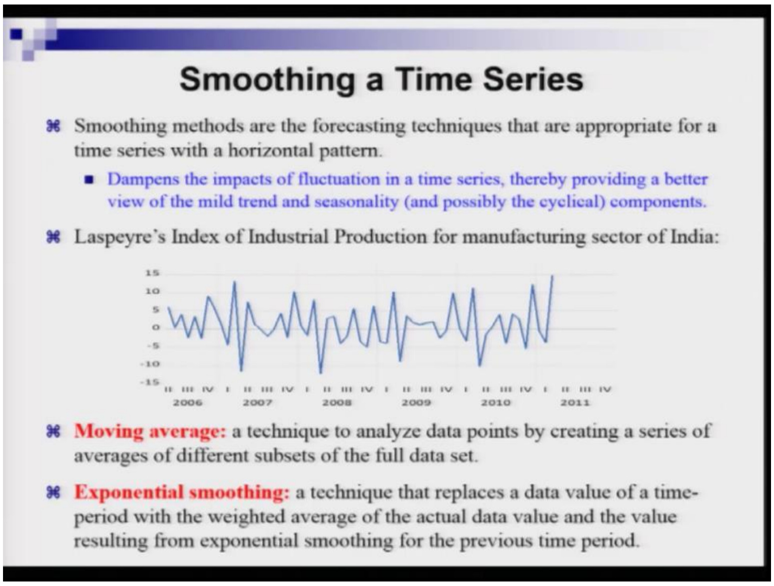
Applied Statistics and Econometrics
Professor Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur
Lecture-24
Classical Time Serries Analysis (Part-II)

Hello friends. Welcome back to the lecture series on Applied Statistics and Econometrics. So, we have been discussing classical time series analysis. We are going to continue with the discussion, and we are going to end that discussion in this lecture only. So, before we jump to the theoretical discussions and illustrations case studies, let us have a look at the agenda items for today.

(Refer Slide Time: 00:42)

So, we are going to start with the discussion on smoothing techniques, and we are going to cover two smoothing techniques. One is the popular one, which is called moving average method or it has got other names like rolling or running average method as well. And second, we are going to discuss exponential smoothing, which is a special case of moving average method. And then we are going to briefly talk about the issue of forecasting and then we are going to end today's discussion by having a very brief idea about forecasting accuracy.

(Refer Slide Time: 01:20)



So, what are smoothing methods? These are the forecasting techniques that are appropriate for a time series with horizontal pattern. So, let us know, look at an example of a time series variable, which is showing horizontal pattern. So, now I request you to concentrate on the graph that is placed in the middle of the slide. Here I am portraying the values of Laspeyre's Index of Industrial Production for our manufacturing sector of India.

So here you see, I have plotted 6 year's quarterly data for this Laspeyre's Index of Industrial Production for manufacturing sector, and you see there is zigzag around that number 0, which is basically at the center of the data, and you can see that there is no prominent long run trend in the data.

So, you cannot draw straight line either with a positively sloped one or a negatively sloped one. Here the fluctuations are all around the value 0. So, you here you can see that the Laspeyre IIP for manufacturing sector for this particular time period, of course, is showing a horizontal pattern.

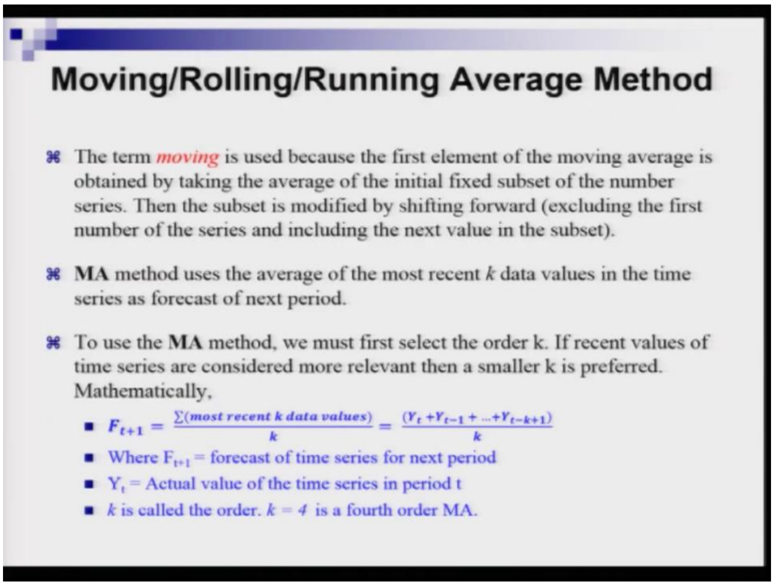
So, if we have such kind of time series data, then what does a smoothing technique do? So, a smoothing technique actually dampens this in our regular fluctuations around the constant mean. And what is the advantage of that? So, basically it helps us to disentangle the mild train defect is there is any and the seasonality effect or component that is embedded in the time series data.

So here we have two types of smoothing techniques in this course. So, we are going to start with the brief definitions for each one of them. And we are going to start with moving average first, because this is the most popular one. So, moving average, which is abbreviated as MA is a smoothing technique that is used to analyze data points by creating a series of averages of different subsets of the full data set.

And in this course, I am going to show you ample number of examples, how to compute moving average, so please wait for a moment. And then the second one is exponential smoothing. This is a statistical technique that replaces a data value of a particular time period with the weighted average of the actual data value and the value that results from the exponential smoothing for the previous time period.

So, this is slightly complicated. I hope that, these ideas of moving average and exponential smoothing will be much clearer to you when we will discuss some new medical examples. I am going to tell you the exact steps, how can you compute moving average and exponential smoothing, so that we will discuss in today's lecture only.

(Refer Slide Time: 04:33)



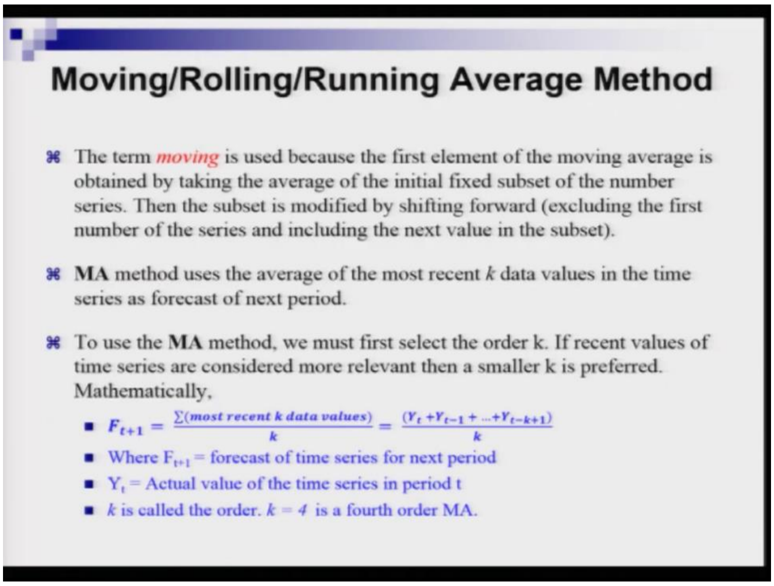
So, now why we are adding this term moving in front of the average method. So, we are adding the term moving because the first element of the moving average is obtained by taking the average of the initial fixed subset of the number of series. So, suppose we choose the first k number of observations in the time series then the subset is actually modified by shifting it forward.

So, what do we do there? We actually exclude the first number of the series and then include the next value in the main data set in that subset of k observations. And then again, we compute that verse so a new average value is obtained and this is the way the average number is constantly moving over time.

So, here in this slide, I am going to show you the steps or the formula, how to compute the moving average values or how to actually implement the moving average method. So, here we must first select the order of k. So, what do we mean by this order of k here? So, if we are dealing with say a monthly time series data set, then within a year there are 12 data points, and the seasonality, if it exists that will be basically reading from one month to the other.

So, you can have a 12-point moving average to reduce the fluctuations the seasonal fluctuations. But if you have a data set, which is a quarterly data set. So, if you have a dataset for say some number of years and for each year you observed the value of the time series variable for different quads, four quarters, then you can actually construct a four point moving average. So here for a monthly data set, the k is usually 12. And if you are dealing with the quarterly data set the usually, the value of k is 4.

(Refer Slide Time: 06:29)

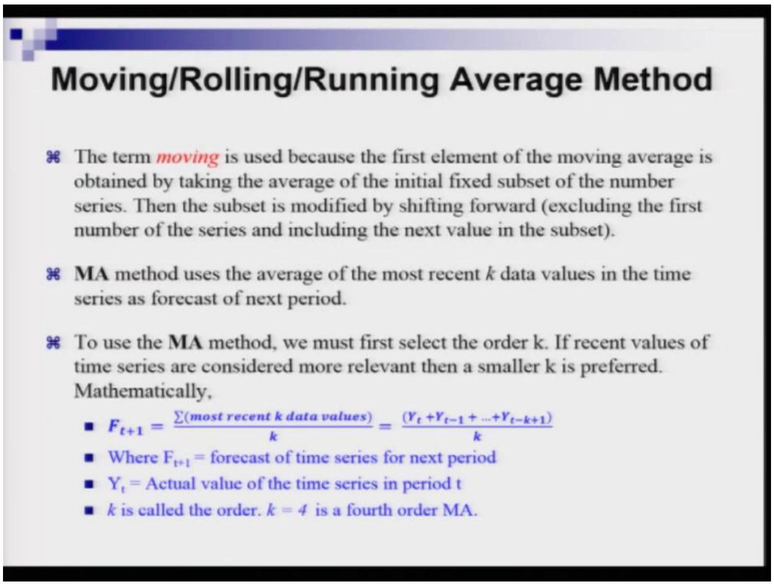


So now you concentrate on the last bullet point of this slide, where in the sub bullets, I am explaining the mathematical formula. And you look at the formula capital FT plus 1, so that is basically given by the sum of most recent k data values divided by the number of k. So, if you remember how I have tools in my case.

So, if I am dealing with the quarterly data set, then there are four quarters and then here k can take value 4, so I will basically start with first four values of the data set, and then I will compute an average and then I will, exclude the first observation from the dataset or that subset, and then I am going to include the very fifth observation in that time series data set.

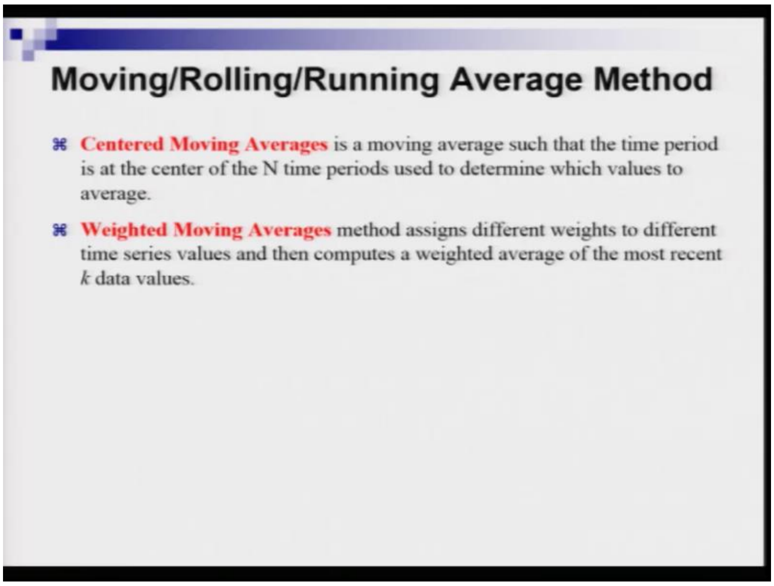
So, once we are done with the first average computation, we are going to exclude the first observation from that subset. So, I am going to exclude the quarter one of year one value, and then I am going to include quarter one of year two time series value in my subset and then again, I am going to compute that average and that will give me no another average value. So, this is the way I am going to proceed.

(Refer Slide Time: 07:53)



So here, I am showing you the formula where capital F t plus 1 means it is the forecast of time series for the next period. And Yts are basically the actual value of the time series in the period t and k is the order. So here it is important to note that there is a jargon. So, if you have chosen k to 4 for a quarterly data set then basically you are computing a fourth order, moving average. Now, the moving averages could be off mainly two types and they are called centered moving average and other one is called weighted moving average, so we are going to briefly have a look at these two items and then we are going to talk about a special case of moving average at length.

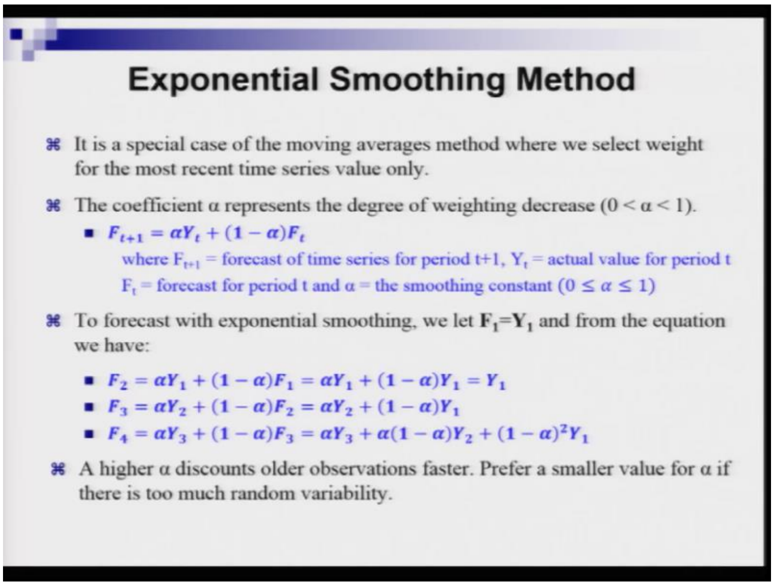
(Refer Slide Time: 08:45)



So, the centered moving average is a moving average method such that the time period is at the center of the end time periods use to determine which values to average. And again, later I am going to show you an example how to compute the centered moving average. And the weighted moving average method actually assigns different weights to the different time series values then computes a weighted average of the most recent k data values.

So, you see the simple moving average actually is not making use of any weight, whereas, weighted moving averages are making user weights. So, now we are going to consider the case of exponential smoothing which is a special case of the moving average method. It can also be seen, as a weighted average moving average method.

(Refer Slide Time: 09:43)



So, under the exponential smoothing method we select the weights for the most resent time series value only. And you see, we have to introduce a new concept here, and that is called the smoothing factor or sometimes it is called co-efficient of a degree of weighting decrease. So, this is generally presented by alpha in statistical literature and it is a fraction. So, alpha always takes value between 0 and 1. So, basically you are taking a convex combination of two different time series values.

So, let us revisit that formula for F t1, which is basically the forecast or fitted value of the time series variable for period t plus 1, where Yt is basically my actual observation for time period t. So now you see, as I was explaining you a couple of seconds before that I am going to talk about a weighted average or a convex combination of two different time series values. So, here alpha, is basically the smoothing factor weight or the coefficient of degree of weighting decrease whatever you want to call it, that alpha value is multiplied with Yt and then 1 minus alpha is applied to the value. Ft and Ft is really the forecast for time period t.

So, now you see the forecast for a time series point t plus 1 is basically weighted average of two different time series values or you can also say that it is a convex combination of two different times series values. One is of course, the value of the time series variable in the preceding period, but this Ft is basically the another variable to which you are applying this weight and that is basically the fitted value of the time period t which is the preceding period. So, to apply exponential smoothing for a particular time period you must have an idea about the fitted value of the preceding time period. So that is a big assumption that you must remember.

So, now to forecast with the exponential smoothing. Suppose I have a data set Y1 to Y capital T so they are a T data points. So now, as I was telling you that you have to have some idea about the Ft which is basically the forecasted or fitted value of the previous period. So, then you as you, that F1 is equal to Y1. And then you apply that equation to get the forecast or the fitted value for the second time period, which is t equal to small t equal to 2 and you have this expression F2.

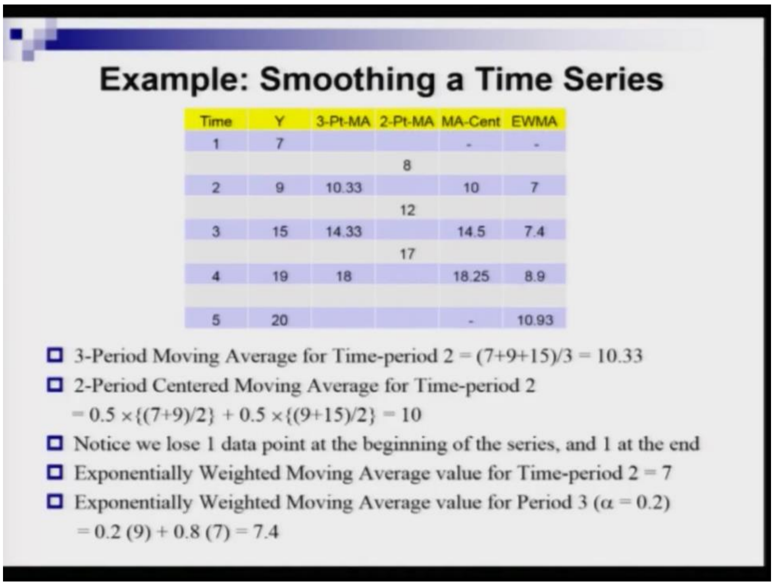
And you see, you get back actually the observation Y1 which is the actual value of the time series. Then you take Y1 as proxy for F2, and then you continue to calculate the fitted value or the forecasted value for the time period t equal to 3, and then you see you apply the value of Y1 in place of F2 and you get a much-complicated expression compared to the first one and you can continue like this.

So, here alpha plays a very big role. And how to choose a particular value of alpha. Because suppose if you have chosen an alpha value of say 0.02, then you will get a series of forecasted or fitted values, but if you have chosen alpha equal to 0.8, say, then you are going to get a very different set of forecasted value. So, how to actually choose the particular value of alpha. So, in statistics, literature, there is some guidance given how to choose alpha value.

So, a commonly used value for these constant smoothing factor alpha is given by 2 divided by capital T plus 1 where capital T is the number of observations in the time series data. So, note that, as a passing comment I should mention that a higher alpha discounts the older observations faster. And you should prefer a smaller value of alpha if there is too much random variability in that data. So, these are the guidance provided by statisticians in the literature. So, remember these as thumb rules.

So, now we are going to deal with a hypothetical data set. And I am going to show you how can you compute moving average and exponential moving average numbers for a hypothetical time series data set. I am only going to show you some examples. And I believe that these examples are going to help you much to understand the steps compute for moving average and exponential smoothing techniques.

(Refer Slide Time: 14:56)



So, here in this slide we are going to discuss an example of smoothing a time series. And here we have a hypothetical data set with five time periods, and the time series variable is Y. So, we see five values of Y. And here I am going to show you how one can compute the 3-point moving average and not 2-point moving average. And finally, we are going to show you how exponentially moving average numbers can also be calculated.

So here, if we start our discussion with three-period moving average then actually it is simple because if we say take time period 2 then we have to basically take three numbers, 7, 9, 15 and then we need to sum and then divide by 3 and then the resultant number is 10.33 and that is going to be the 3-point moving average for time period 2. Note that here there is no problem of centering because he calculated moving average is exactly mapped with the time period 2.

But what if we take k equal to 2 or this is a general problem I am going to discuss for any even value of k, the order of moving average. So, here to be the simplest possible case we are going to discuss the case of k equal to 2. So here, if we take the case of k equal to 2 then what will happen. So, to calculate the moving average for time period 2, we have to get first 2 values 7 and 9 and then we have to take the average, and you see the value of course is going to be 8, but where to place this newly created average.

Because you see, I cannot place this number 8 to either of these two time periods, 1 or 2. So actually this number 8 is placed somewhere in the middle of these two time periods, 1 and 2. So, that is why you see, I have created this gray cell where I do not have any time reference or if you wish you can put the time reference say 1.5, the midpoint of time periods 1 and 2, and then you can place these two-point moving average value 8 against this time 0.1 0.5.

But note that in the original data set, we do not have a time period 1.5. So, there is a problem of mapping between the two point or even point moving averages and the time periods given in your dataset. To resolve the issue, we have to now make use of this concept centered moving average.

So how do I do that? So, I have already calculated the moving average for the time periods 1 and 2. And next, what you have to do, you have to exclude the time period 1 and include the time period 3, but keep time period 2. So, now you are considering two values, 9 and 15. You take the average so that is going to be 12. And these newly created two-point moving average will, again be placed against the hypothetical time period, 2.5, which is exactly in between time periods, 2 and 3, which we observe in the data set.

So, now we have two, two-point moving average numbers, 8 and 12. Now, if you take mean of these two numbers you get 10, and then these newly created centered moving average number 10 will match to the time period two, and then you see the mapping between the moving average numbers and the original time period, and the original time period value is restored.

So, this is the way you can compute the moving average centered for other time periods as well. And I am showing you how to calculate the moving average centered values for time periods 3 and 4 as well. So, note an interesting thing here. If we are going to compute centered moving average, then we are going to lose one data point at the beginning of the series and the end of the series. So, that is why you see under the column of MA sent I have two dash at the time period 1, and again 1 at that time period 5. So, you lose out on these two time periods.

So, now we move on to the case of exponentially weighted moving average. So, if you remember, I told that for the first time period the fitted value you can, assume it to be the value of time period 1. So, the exponentially weighted moving average value for time period 2 will be the number 7 that you observed in the case of time period 1. But then how to calculate the EWMA for period 3, if you assume a alpha value of 0.2.

So, what will you do? So, in that case, you have to go back to the previous time period which is 2 and you have to figure out what is the actual observed value of a time series variable Y that is 9. You multiply the smoothing factor 0.2 with that actually observed number and then you multiply 1 minus alpha, which is 0.8 with the fitted value or the projected value for that time period.

And from the last bullet point only, we know that, that value is 7, so you then, after you, multiply you sum these two products and then you to get a number 7.4, which is the EWMA for period 3. Now, we are going to talk about a very important concept and that concept is called seasonal index. Now, why we must learn about seasonal index. Remember, in the last lecture I have told the entire objective of classical time series analysis is that it should help us in forecasting or projecting the future values of the time series.

So, suppose we have an initial data set where the time periods are 1 to capital T and we are interested to know what is the value at time periods say capital T plus 1, capital T plus 2, we do not have the values for capital T plus 1 and capital T plus 2, but from the historical data, which is with us for a first capital T time periods from there we want to draw some inference. We want to estimate the future values of the time series variable.

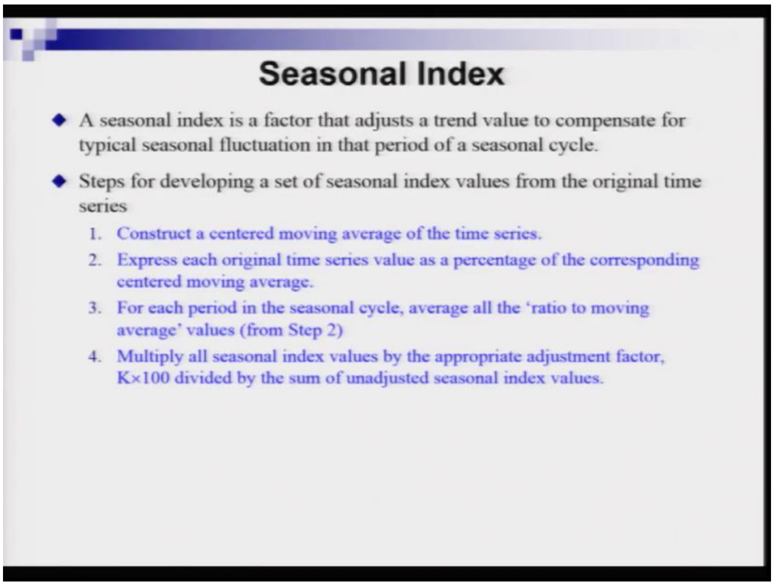
So, for that, you need to disentangle the different components of the time series variable or time series data namely train, seasonal component and then the cyclical component if there is any. And of course, there is a random component but here we are assuming for the sake of simplicity that our random component is very well behaved, it is purely random. And it is basically coming from a white noise already Gaussian distribution, so we do not have to worry about it.

So, there are these systematic factors that we need to take care of when we are trying to make projections about future. And we should do it step by step. In the last lecture we have discussed the case of linear trend. So, if there is a linear trend, if by plotting the time series variable you see that there is an upward or downward trend visible enough from the data then you just apply the OLS technique and get a trend equation estimated, and that will basically give you the fitted values of Y.

If you, get the OLS estimators for the population parameters and then plug the different values of time period in the equation, you will get the fitted values and that is basically the trend value, that is the trend component of your time series variable. So, we have studied that. Now, we are going to concentrate on two more cases. First of all, when we discuss that case of linear trend, I assume that the trend was visible, but trend not be visible. Trend maybe very mild. It is there, but it is not visible enough. If you are just looking at a graph printed on a piece of paper, so then basically what to do then you can apply the case of exponential smoothing.

And as I told you, that exponential smoothing is very useful when there is constant or regular fluctuation around constant means, so basically that is the horizontal pattern with which we started the discussion on smoothing. So then, you have to adopt a different technique, which is basically a smoothing technique.

(Refer Slide Time: 24:16)

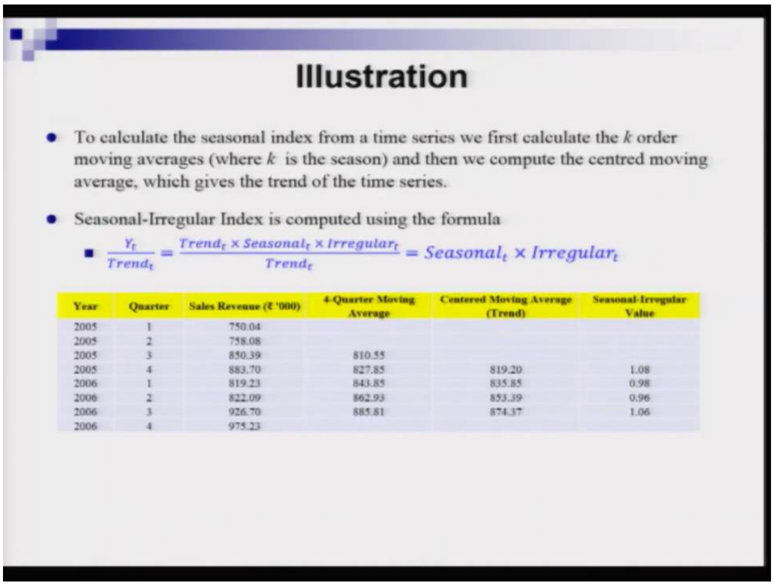


So, here we are going to now start the discussion about seasonal index, which helps us in making predictions or projections about the future values of time series. So, a seasonal index is basically a factor that adjusts a trend value to compensate for the typical seasonal fluctuation in that period of seasonal cycle.

Now, there are four steps for developing a set of seasonal index values from the original time series data. And I am going to just briefly discuss them. So first you to construct a centered moving average of the time series data, and again, the choice of k is arbitrary. So, if you are dealing with the quarterly time series data feel free to take a value of k equal to 4. If you are dealing with monthly time series data, then probably we will start with k equal to 12.

Then in the step two, you express each original time series value as a percentage of the corresponding centered moving average value. And in step three for each period in the seasonal cycle, average all the ratio two moving average values, which is basically coming from step two. And then finally in stage four, you have to multiply all seasonal index values by the appropriate adjustment factor k times 100 divided by the sum of un-adjusted seasonal index values. So, these steps may look a bit complicated as they are expressed in words, but in later slides hopefully this will be cleared when I am going to show you the actual use of seasonal index.

(Refer Slide Time: 26:03)



So, first, you need to calculate the seasonal index from a time series. And for that, we first calculate the k order moving average averages where case the number of season or quarters in the data. And here, of course, from the story you can say that we are dealing with small data set, so that is why we are making use of moving average method over the least squared principles.

And then we have to compute the centered moving average, which gives the trend of the time series. So, here the table at the bottom of the slide you are seeing some numbers given. So here suppose there is a fictional company who sales revenue data we have for two different years, 2005 and 2006, and the data is given for each quarter, so we have to tell eight observations.

So basically, we need to first compute the four-quarter moving average, so I am showing you the numbers for four quarter moving average figures. And you can now try your hand to apply the formula that you have learned from today's lecture, and apply that formula on the sales revenue figures that I am showing you in column three. And then you can get the four-quarter moving average numbers.

And that is basically the non-centered four-quarter moving average number. But you have to now find the centered moving average numbers and that will actually give you the trend.

Why? Because I told you that when you are calculating the non-centered moving average numbers, uh, there is a mismatch between the calculated moving average value and the time period. So, if you want to have a one-to-one mapping with the calculated moving average values with different time points in that case you have to use the centered moving average concept. So that is what we are doing here.

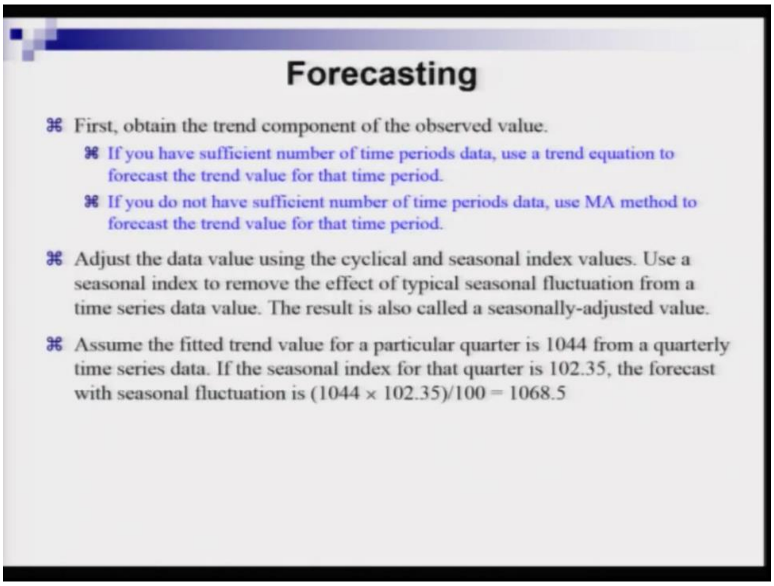
So here you see, I am also calculating the centered moving average and that is basically my trend observation. So that is basically giving me the trend values or the trend component of my time series data. And note that I can only open these trend values for the last quarter of 2005. And then, I can compute these rend values up to the third quarter of 2006.

So, the lesson is clear from here that, if you are interested to apply the centered moving average technique to find the trend component of short lived or small time series data set, then you will lose out initial time points and the last time period. For those time periods you cannot calculate the value of the centered moving average. Anyway.

So, suppose for those centered moving average trend values now what to do. We have to get the seasonal index, so that we can use the seasonal index to forecast the value for the time series variable. So how do we compute the seasonal or irregular index here? So here you need to divide the actual time series value Yt by the trend component.

And they you see, if I assume that I am working with a multiplicative time series model then trend will cancel out from numerator and the denominator, so then I will get this seasonal irregular index. So basically, this index now is basically a ratio of the actual time series observation by the fitted value of that trend component. So, you get this seasonal irregular index that is basically in the last column.

(Refer Slide Time: 30:23)



So, now let us link these seasonal index concept to the forecasting job. So, as I said, first, we have to obtain the trend component of the observed values from the time series variable. So, if you have sufficient number of time periods data then you can use the OLS technique to fit a trained equation to the data points and then you can obtain the forecasted value of the trend component for that time period. But if you do not have sufficient number of time periods, then you can use the moving average method. Once, the train value is obtained you have to adjust the data value using the cyclical or seasonal index values.

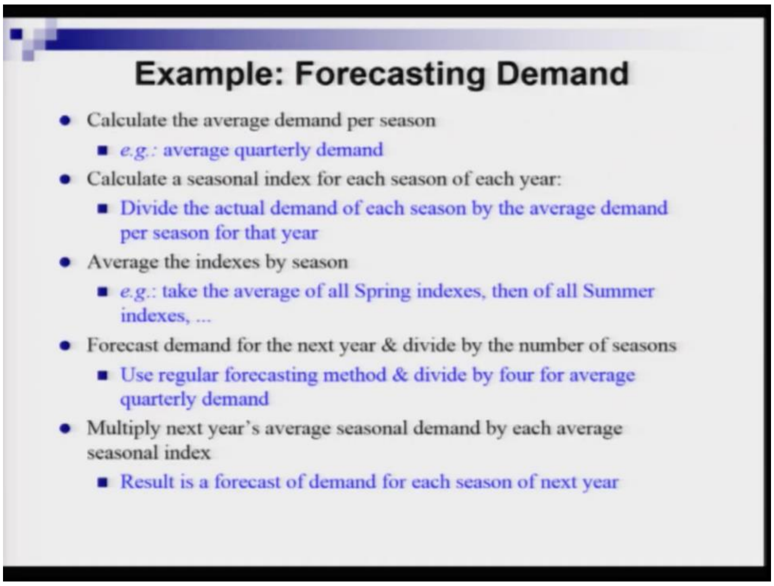
Now, for simplicity’s sake, we can assume that there is no cyclical component. If there is a cyclical component, of course, the task is even harder. And that is why for this simple discussion I am excluding the cyclical component. So, let us assume that there is only a seasonal component present in the time series data, so if there is seasonality then you know what to do, that is what we are going to study the next.

So, basically, what we have to do assume the fitted trend value for a particular quarter. Say, we are dealing with quarterly data for the sake of simplicity, because we have to take a particular value of k if we are interested to conduct an MA analysis, so I am assuming here that the data set given is a quarterly one so you can apply moving average method by adopting the value of k equal to 4.

And I must say here that if you have a time series data, quarterly times series data where you have many, many number of years. Say suppose you have say 10, 12 years of data where you observe the values of a variable over different quarters then also you can go for a least squares regression because 40 is a large enough number so that you can fit a linear trend equation.

So, here, let us come back to the example. So here, let us assume that for a particular quarter, there is a fitted train where from some method is 1044. And if the seasonal index for that quarter is 102.35 then the forecast with seasonal fluctuation is 1044 times 102.35 divided by 100, and that results into 1068.5. So, in the next slide, I am going to show you a step-by-step method in terms of simple language, how a company manager or a farm manager can forecast demand for an item that the farm is manufacturing.

(Refer Slide Time: 33:29)

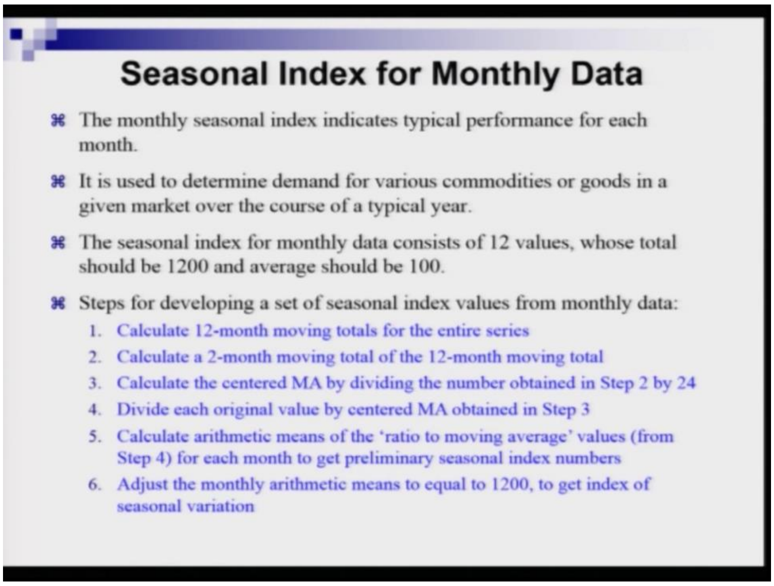


So, you start with by calculation of the average demand per season. And here, of course the season you can say that is represented by quarters in the calendared year. So, here we are talking about average quarterly demand for the item or the commodity. So, then you have to calculate the seasonal index for each season of each year. So, what to do? So then for that you need to divide the actual demand of each season by the average demand part season for that year.

So, this is the way you can compute a proxy seasonal index. Then in that hard step you have to average the indexes by season. So, in this particular context take the average of all spring indexes then all the summer indexes and so on, so forth. And then you finally forecast demand for the next year, and divide by the number of seasons. So, in that case what to do.

So, you first use the regular forecasting method, and divide by 4 for average quarterly demand. And finally, you multiply next year's average seasonal demand by each average seasonal index. So, now the result is a forecast of demand for each season of next year. So, now we are going to talk about an illustration with some hypothetical data.

(Refer Slide Time: 35:04)



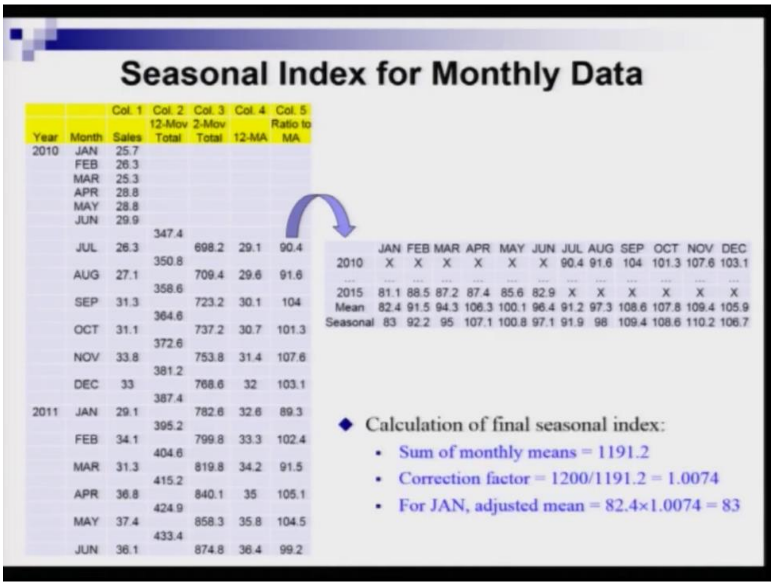
So now, in this slide, I am going to focus on the seasonal index calculation, if we come across monthly data. So, monthly seasonal index indicates typical performance for each month. So, what do we mean by that? So generally, it is used to determine demand for various commodities or goods in a given market situation over the course of a typical year. So that is why I am saying that this indicates typical performance for each month. And, of course, a seasonal index for monthly data consists of 12 values and the total should be 1200 and average should be a 100.

So, now this is somewhat on the monthly seasonal index concept. And now we are going to discuss briefly about the steps for calculation of seasonal index values from monthly data. So first you have to calculate a 12-month moving average total for the entire series, and then in the second step you calculate a two-month moving average total of that 12-month moving total. In step three you calculate the centered moving average by dividing the number obtained in step 2 by 24.

Now, why are you have to compute the centered moving average because if you remember here, the order of moving average k is even, so we have to center the calculated moving average and otherwise there will be a mismatching between the time period and the calculated moving average values.

So, in step 4, we divide each original value by the centered moving over to obtained in step 3 and then in step 5, we have to calculate the arithmetic means of the ratio of the. In step 5, we have to calculate arithmetic means of the ratio to moving average values that we have obtained from step 3, for each to get the preliminary seasonal index number. And then in the last step, we have to adjust the monthly seasonal index numbers such that the total becomes 1200 and the average becomes 100, so that we get the final index of seasonal variation. So here, we have given you the steps, but in words, but these steps are probably a bit complicated at first sight. So, this will be clear, if we show you an example and that is what we are going to do the next.

(Refer Slide Time: 37:43)



So, here in this slide, I am going to talk about an illustration using hypothetical data, so that you see how different steps that we spoke about in the last slide is executed. So, suppose I have data, I have monthly data on some variable say sales of a commodity for 6 years, starting from 2010 January to 2015 December. And now, I want to calculate the seasonal index for this monthly data.

What shall I do? So, you see the originally observed variable values are given under column one with heading sales. And then, as I told you in the previous slide, the first step is to calculate the 12-month moving total. So, that is given for column number 2. So, you see here I have to pick first 12 numbers, which is basically, which are basically the monthly sales figures for the year 2010, and then I have to calculate the total.

And where can I place this total note that there is a problem of even k here. So, the total will not before a month, June or not before month July, so where will it go. So, you have to basically place the total in the middle of the year, so in between the months of June and July. So, that is why I have created an extra row between the months, June and July to place this sum total. And that is where I placed my total 347.4. So, it corresponds to midpoint between June and July.

So, maybe 15th June, you can say, hypothetically. And then you have to compute the next entry for the 12-month moving total, what to do. So, you have to now exclude the value for January, 2010, which is 25.7. And now you have to include the value of 2011, January sales figure, which is 29.1, and then you calculate the sum.

And you see the difference between 25.7 and 29.1 is roughly a difference of 3.8 or something. So that is what you are observing here. So, the new total 350.8 is only three point something different from the previous total 347.4. And this is the way you continue. Now, you need to center this data, so that you can get the mapping between the calculated numbers and the time series points back. So, in column three, we are now calculating the two-point moving total. So, what to do for that? So, you have to first figure out the first two numbers, and then actually you can take the sum and then you can place it in the middle of these two sums.

So, if you take first two numbers 347 and 350.8, then the sum total is 698.2. And now you place that against the month of July. So, the mapping between the calculated number and the original time point is restored. And then what you have to do, you have to now calculate the 12-point to moving average and that is given under column 4.

So, from 698.2, you calculate the moving average value 29.1. And in the last column, column 5, I am showing you the ratio two moving average value calculation. So basically, what you have to do, you have to divide the original number 26.3 by this newly created moving average number 29.1 for the month of July, and then you have to multiply this with 100 to get the number 90.4.

And this is the way you can compute the other ratio to moving average numbers for other months. Now, you see this calculated ratio to moving average numbers are taken to a different table, table lumber two say. And here, in the rows I have six different years from 2010, 2015. To save space I am not showing the years in between 2010 and 2015. And in the columns, I have 12 months.

So, you see for first six months in 2010, I will not get this issue to moving average values. And I have shown you why I have also explained to you why we will not get the initial periods numbers in the last slide also. So here you can see that the ratio to moving average number will start from the month of July and it is 90.4, so that way you can see it corresponds to your table number one calculations.

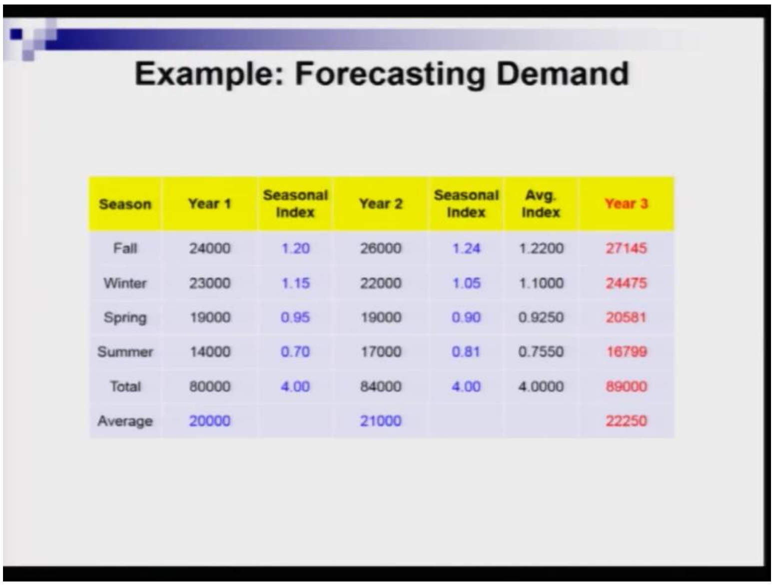
And you then then have other numbers for other months. But you note down that by following the same logic you will not have the ratio to moving average numbers for last six months of the final year, 2015. So, you have ratio to moving out of the numbers up to month, June. And again, these are just some hypothetical numbers because I am unable to show you the entire data set here.

In the next step, you have to calculate the arithmetic mean unadjusted, mean for this monthly ratio to moving average numbers. So, for the month of January you have five data points because you were losing out for the 2010 year so you calculate the mean, and hypothetically say that is 82.4. Similarly, you continue to get the monthly means for all months, and then you have to finally calculate the seasonal index. How to calculate the seasonal index? You have to first see, whether there is a need for correction factor or not.

So, why do we require correction factor? I told you, theoretically, the sum total of this monthly means should be 1200 and the arithmetic mean should be 100. So, if that is not case then we know we have some inconsistency in our calculation. So first, we need to check the sum of our monthly means calculated. So, if you take the sum of these 12-monthly means, the sum will become 1191.2, so it is not equal to 1200, so we need a correction factor. And the correction factor will be I have already given you the theoretical formula before. So here, the correction factor would be k times 100 so k here is 12, so it is 1200 divided by the sum of the monthly means and then you get this correction factor is 1.0074.

So now, you have to calculate the adjusted mean, and this will give you the final seasonal variation factor. So, for example, for the month of January, I am showing the calculation, so you have this unadjusted arithmetic mean for the month of January being 82.4. So, you multiply that number with the correction factor and you get 83. So, 83 is their finances and our variation factor for your dataset.

(Refer Slide Time: 45:43)



So, let us know talk about forecasting demand through a simple example. Suppose there is a farm and the farm has some data on the demand for an item for two years, year and two, and the demand can be further broken down in different seasons. So, you see for your 1, the total demand for these item a is 80,000. And for year 2 the total demand for the item is 84,000. Now the farm wants to know what will be the demand for year 3, and the farm also wants to see the seasonal breakdown of these demand figures for year 3. So how do we proceed?

So, suppose there is a statistician who has given some proxy measure for the demand annual demand for year three, and that is say 89,000. Now, the farm wants to know what will be the seasonal breakup of this demand figure annual demand figure. So, then what shall we do? So, you see the total for year one is 80,000 and total for year two is 84,000, so the first step would be to divide this total by different quarters.

So, there are 4 quarters here, so you get an average demand per quarter being 20,000. So, based on that you can now calculate a very rough measure of seasonal index. And how to do it? So, you have to divide the actual observed demand for each season by this average season demand.

So, for example, if I concentrate on the fall season or the autumn season, then I see the actual observed number is 24,000. So, I divide 24,000 by 20,000, I get this number 1.20 and that is my seasonal index for autumn or fall season. Similarly, I can calculate the seasonal index numbers for other three seasons. And you note that we are lucky that the sum total is four year one and similarly we can find out the seasonal index values for your two also.

And again, also we are lucky the sum is 4. If the sum is not equal to 4, then, of course, we need to use correction factor, but here we are not using correction factor because the sum total is as part theory. So now, given this what to do next for year three. We only know the annual demand for the item in year three, and that is 89,000. So, what we have to do, we have to first divide this estimated annual demand by the number of seasons, which is 4, so we get an average demand per season being 22,250 items. And now you have to apply some seasonal index number to this average season demand.

Now, which seasonal index to use? We have to take the arithmetic mean of the seasonal index values coming from two years, year one and year two. So, if we again, concentrate on the case of fall or autumn season, then we see from year one, the seasonal index values is 1.20 and from year two, the seasonal index value is 1.24. So, we take simple mean arithmetic mean and then we get an average index value of 1.22. So that is the mean seasonal index for season fall or autumn.

So, we can follow the same technique and calculate the average index values for other three seasons. And again, we sum them up, and we are again, lucky to see the sum being exactly equal to 4, so do not need a correction factor here.

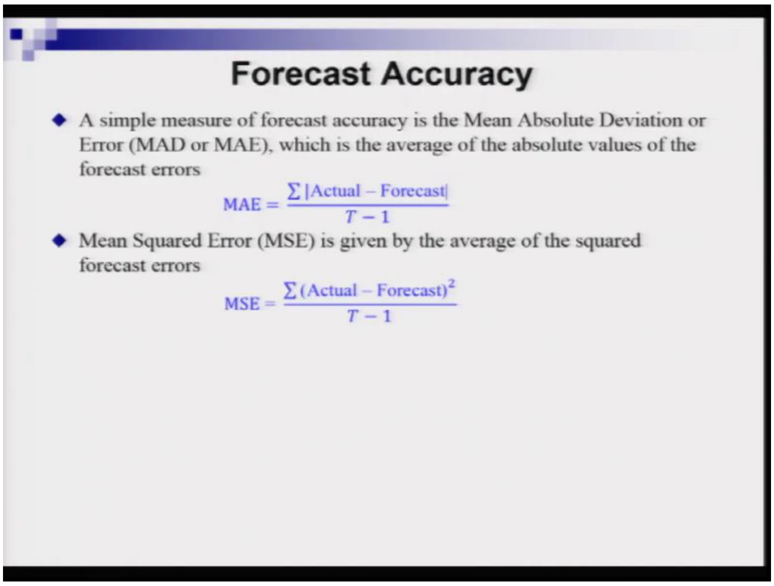
So next, you have to multiply this average index with this average part season demand for year three, so you have to multiply 1.22 with 22,250 to get the number 27,145 for the fall or autumn semester. So, this is the way you can calculate the expected demand or estimated demand for all other seasons for the year three. And as year three is totally being forecasted I have colored these numbers here in red, so that I can differentiate it from other two years. So, in the last slide, we are going to briefly discuss forecast accuracy.

So, if you remember our discussion on linear regression analyses, we said that when we fit a model then we are interested in the accuracy of the model fit or goodness of the model fit. And in the case of the linear regression, we had this measure of coefficient of determination or art square.

Now, in the case of classical time series analysis, when we are making projections, we may be interested to know how good or bad our projections are already or how good or bad in our model is. Because if you have come up with a bad model, then all the projections will be bad in the sense that they will be far off from the actual observed value of the times series variable. So, there is a need to have some measure for the goodness of fit of your time series analysis or modeling.

And there are several measures available, but here in this course, we are going to only talk about two.

(Refer Slide Time: 51:41)



So possibly the simplest measure of forecast accuracy could be the mean absolute deviation or error, so the abbreviation is MAD or MAE in textbooks and that is basically the average of the absolute values of the forecast error. So, forecast error is basically the difference between the actual value of the time series variable and the forecasted value of the times series variable then you take the absolute value of this difference and then you sum overall observations and then finally, you divide by capillarity T minus 1, the number of observations minus 1.

And then the second one is known as mean squared error MSE, and that is given by the average of the squared forecast errors. So, these squared forecast is basically very similar to the sum of squared of residuals that we have seen in the context of linear degression.

So, here MSE is divided by the sum of squared deviations, and the deviation should be by capital T minus 1. So, capital T is basically the number of times series observations, in the data set. Thank you.